Regional Science and Urban Economics 67 (2017) 135–147

Contents lists available at ScienceDirect

Regional Science and Urban Economics

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

Gender and racial peer effects with endogenous network formation☆

Chih-Sheng Hsieh[a], Xu Lin[b,*]

[a] Department of Economics, The Chinese University of Hong Kong, CUHK, 9/F, Esther Lee Building, Shatin, China
[b] Department of Economics, Virginia Tech, 880 West Campus Drive, Blacksburg, VA 24061, USA

ARTICLE INFO

JEL:
C21
I21
J13

Keywords:
Bayesian estimation
Endogenous network
Gender peer effect
Racial peer effect
Spatial autoregressive

ABSTRACT

We apply a high order spatial autoregressive (SAR) model to simultaneously capture heterogeneous peer effects from multiple gender and racial groups, as well as endogenous network formation. In students' GPA and smoking behaviors, we find that within-gender endogenous effects are stronger than cross-gender effects. Females and whites are more sensitive to peer influences and more influential than other students. Intra-race spillover effects are stronger than inter-race effects for whites, but not for non-whites. Homophily on observed and unobserved characteristics are important for friendship formation. However, the formation of friendship is not necessary motivated by common interest in outcomes such as smoking. Our findings suggest that coeducational or desegregated schooling may help increase academic achievement, but not reduce smoking frequency.

1. Introduction

Peer influences have been investigated intensively for a variety of behaviors and outcomes. However, many existing studies focus only on the average magnitude of peer influences, without paying attention to the possible heterogeneous nature of peer effects along the dimensions of race and gender. As a result, informative policy implications are difficult to draw from those studies. For instance, to shed light on the debate on single-sex versus coeducational schooling, we need to know the mechanism of peer effects along the gender dimension, i.e., how an individual affects and is affected by his/her male and female peers respectively. Another example is the debate on school segregation versus desegregation, which calls for the analysis on the mechanism of peer interactions within and across racial groups.

Recently, a few studies have started to explore the heterogeneity of peer effects along the lines of gender and race.[1] However, most of these studies focus on the effects of gender or racial composition, in particular, female or black proportion, in a group on various outcomes.[2] These effects are termed "contextual effects" or "exogenous effects" by Manski (1993), which represent the effects of peers' characteristics, gender or race in this case, on an individual's outcome. Another type of peer effects are "endogenous effects", which capture the effects of peers' outcomes on an individual's outcome, and have only been examined in a few recent papers.[3] A typical assumption made in these studies is the absence of contextual peer effects.[4] As pointed out by Bramoullé et al. (2009); Fruehwirth (2010, 2013); Lee et al. (2010); Lin (2010), among others, separately identifying contextual and endogenous effects is important as endogenous effects generate a social multiplier whereas contextual effects do not. Therefore, in order to provide sound policy implications regarding single-sex versus coeducational schooling, as well as school segregation versus desegregation, it is necessary to identify not only the contextual effects along

☆ This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

* Corresponding author.
E-mail addresses: cshsieh@cuhk.edu.hk (C.-S. Hsieh), xulin88@vt.edu (X. Lin).

[1] Some papers, such as Patacchini et al. (2017), do not focus on racial or gender peer effects but they also study heterogeneous peer effects in school.

[2] For instance, Lavy and Schlosser (2011); Whitmore (2005) study gender proportion effects, and Angrist and Lang (2004); Hanushek et al. (2009) focus on racial composition effects, while Hoxby (2000) study both effects.

[3] Clark and Loheac (2007); Kooreman (2007); Soetevent and Kooreman (2007) and Trogdon et al. (2008) study endogenous peer effects along the dimension of gender, Nakajima (2007) investigates both gender and racial endogenous effects.

[4] A notable exception is Fruehwirth (2013), who examines the racial composition effects, as well as the spillover effects of different racial groups. But her model specification and identification are different from ours, as will be discussed in Section 3.

http://dx.doi.org/10.1016/j.regsciurbeco.2017.09.001
Received 21 October 2016; Received in revised form 28 August 2017; Accepted 3 September 2017
Available online 08 September 2017
0166-0462/

the dimensions of gender and race, but also the endogenous gender and racial effects.

Focusing on only one type of peer effects in most existing studies is due in great part to the identification challenge, i.e., the "reflection problem" (Manski, 1993), which refers to the impossibility to separately identify endogenous effects from contextual effects in the conventional linear-in-means model without imposing restrictions on the parameters. Fortunately, recent developments in spatial autoregressive (SAR) models demonstrate that the "reflection problem" is not an issue for SAR models as the nonlinearity introduced in the peer measurement breaks down the linearity between endogenous and contextual effects (e.g., Bramoullé et al. (2009); Calvó-Armengol et al. (2009); Lee et al. (2010); Lin (2010)). In this study, we employ a generalized SAR model to identify not only the effects of gender and racial composition on students' academic achievement (GPA) and smoking behaviors, but also the contemporaneous spillover effects from the outcomes of peers from different gender and racial groups.

Another identification difficulty in peer effects estimation is the omitted variable bias problem, caused by either endogenous group formation or unmeasured variables, e.g., school resources, teacher quality, etc. (Moffitt, 2001). To partially address this issue, existing studies employ either instrumental variable (IV) strategy (e.g., Evans et al. (1992); Rivkin (2001)); or group fixed effect strategy (e.g., Bramoullé et al. (2009); Calvó-Armengol et al. (2009); Lin (2010)); or rely on the experiment nature of the data (e.g., Sacerdote (2001); Zimmerman (2003)). However, as pointed out in Lin (2010) and Nathan (2008), among others, a more sophisticated strategy to deal with the omitted variable bias problem is to develop an equation system to simultaneously model endogenous network formation and peer influences, which has recently been investigated by Goldsmith-Pinkham and Imbens (2013); Hsieh and Lee (2016), among others.[5] In particular, Hsieh and Lee (2016) propose a network formation model where distances of unobserved variables are used to capture the influence of homophily in terms of unobserved characteristics on friendship formation. These unobserved variables are then applied to the SAR model as the control function to correct for the bias caused by omitted variables. However, Hsieh and Lee (2016) only consider homogenous peer effect.[6]

In this study, we generalize the model in Hsieh and Lee (2016) to a high order SAR model to incorporate heterogeneous peer effects from multiple gender and racial peer groups while accounting for endogenous network formation. For gender, we include four spatial weighting matrices to capture the spillover effects from female to female, male to female, male to male, female to male, respectively. For cross-race analysis, we consider three racial groups, i.e. white, black and other race, and specify nine weighting matrices to capture the spillover effects within and between groups, e.g., from white to white, black to white, and the like. Estimating the high order SAR model is more difficult than the standard SAR model as we need to perform a constrained optimization for multidimensional endogenous effect parameters which satisfy the stability condition. Moreover, the Jacobian determinant in the log likelihood function of the high order SAR model cannot be easily calculated (Lee and Liu, 2010). However, LeSage and Pace (2009) show that it is convenient to employ the Bayesian approach with Markov Chain Monte Carlo (MCMC) sampling to estimate the high order SAR model because the stability restriction can be easily imposed during the rejection step of the Metropolis-Hastings algorithm. Therefore we estimate our structural higher order SAR model by the Bayesian approach.

Data are from the National Longitudinal Study of Adolescent Health (Add Health). Our results show that failing to include both endogenous and contextual social effects, or to control for the endogenous group formation could seriously bias the estimated peer effects, especially for GPA. For endogenous gender spillover effects, we find that within-gender interaction effects are stronger than cross-gender effects, and females are subject to larger spillover effects and are more influential across gender. In terms of racial spillover effects, we find that whites are most sensitive to peer influences and most influential. Furthermore, for whites, intra-race spillover effects are stronger than inter-race effects, which is not true for blacks or students of other racial groups. And blacks appear to be the least influential group across race, as they do not generate any significant endogenous effect on their non-black friends. The parameters from the network formation model show that homophily in terms of both observed factors such as grade, race and gender, and unobserved characteristics are important for friendship formation. Interestingly, some unobserved factors are relevant for both network formation and outcomes (e.g., GPA), while the other latent characteristics affect network formation but not outcomes (e.g., smoking). It implies that friendship formation is not necessary motivated by common interest in outcomes such as smoking, even though an established friendship network does have significant endogenous effects on these outcomes. An important implication of our findings is that policy intervention targeting at female and/or white students will be most effective due to the influential status of female and/or white students as well as their responsiveness to peer influences. And as females and whites tend to perform relatively better in school, their influential status suggests that mixed schooling would be beneficial for improving academic performance. However, as females do not smoke less than males, and whites have a relative high smoking frequency, the mixing of genders and/or races do not help reduce smoking.[7]

The remainder of the paper is organized as follows. Section 3 reviews the related literature. Section 4 discusses our model in details. Section 5 describes the data and empirical results are presented in Section 6. Section 6 briefly concludes. Some additional results and technical details are provided in the Appendix.

2. Literature review

This section reviews some papers that examine social interactions along the lines of gender and race. As mentioned above, most existing papers focus on the contextual effects of female or black proportion in a group. Studies regarding endogenous gender or racial peer effects are rather limited. Hoxby (2000) estimates gender and racial peer effects in Texas elementary schools by exploiting exogenous idiosyncratic variations in gender and race composition in a school-grade in adjacent years, and finds that both male and female students achieve higher test scores in reading and math when there are more female students in the class. Using data from an introductory undergraduate management course, Hansen et al. (2006) find that male-dominated groups tend to perform worse than female-dominated and equally mixed gender groups in several outcomes. Relying on within school variations in the proportion of female students across adjacent cohorts in Israeli schools, Lavy and Schlosser (2011) show that a higher proportion of female peers has a positive and significant effect on academic achievements, with similar magnitude for both genders. They also find a positive and significant effect of female share on high school math and science classes enrollment rates for boys, while a smaller and less significant effect for girls. Han and Li (2009) use a randomized experiment data from a Chinese college and find that only female students respond to their roommates' influences. On the other hand, Whitmore (2005) exploits a randomized experiment data from Tennessee's Project STAR, and finds mixed results for the effects of female proportion. Christofides et al. (2011) show that

[5] Similar ideas of using Heckman type conditions to address network selection have been discussed in several papers, including Blume and Brock (2011), and Brock and Durlauf (2001, 2006).

[6] Arduini et al. (2014) propose a model of heterogeneous spillovers within and between groups similar to our high order SAR setting.

[7] We should note that the Add Health Data is composed by US students, the results may not apply in other countries.