Estimating Peer Effects on Career Choice: A Spatial Multinomial Logit Approach

Bolun Li[*], Robin Sickles[†], and Jenny Williams[‡§]

Abstract

Peers and friends are among the most influential social forces affecting adolescent behavior. In this paper we investigate peer effects on post-high school career decisions and on school choice. We define peers as students who are in the same classes and social clubs and measure peer effects as spatial dependence among them. Utilizing recent development in spatial econometrics, we formalize a spatial multinomial choice model in which individuals are spatially dependent in their preferences. We estimate the model via Pseudo Maximum Likelihood using data from the Texas Higher Education Opportunity Project. We do find that individuals are positively correlated in their career and college preferences and examine how such dependencies impact decisions directly and indirectly as peer effects are allowed to reverberate through the social network in which students reside.

JEL Classification: C31, C35
Keywords: spatial models, peer effect, random utility, multinomial logit models

[*]Department of Economics, Rice University. Email: bolun.li@rice.edu
[†]Department of Economics, Rice University. Email: rsickles@rice.edu
[‡]Department of Economics, University of Melbourne, Email: jenny.williams@unimelb.edu.au
[§]This research uses data from THEOP, a project directed by Marta Tienda, at Princeton University, in collaboration with Teresa A. Sullivan, formerly at the University of Texas at Austin. This research was supported by grants from the Ford, Mellon and Hewlett Foundations and NSF (GRANT # SES-0350990). We gratefully acknowledge institutional support from Princeton University's Office of Population Research (NICHD Grant # R24 H0047879). Special acknowledgment is due to Dawn Koffman, for preparing and documenting data for public use, and to Sunny X. Niu,who has been a stalwart collaborator and co-investigator.

1 Introduction

Peers and friends are among the most influential social forces affecting adolescent behavior. In this paper, we investigate peer effects on post-high school decisions. There is a substantial body of literature that studies peer influences on educational achievement (Ding & Lehrer, 2007; Lin, 2010; Fletcher & Tienda, 2009; Neidell & Waldfogel, 2010; Zimmerman, 2003). Yet little is known about how peers influence post school decisions, such as whether to attend college or not, and career choices. The few studies that investigate peer effects on college attendance typically measure peer effects as the proportion of high-school friends who intend to go to college (Alvarado & Turley, 2012; Arcidiacono & Nicholson, 2005; Fletcher, 2010; Lyle, 2007). However, these studies are silent on how the preferences of friends are formally linked and what mechanism best explains these links.

Utilizing developments in spatial econometrics (Anselin, 2002; Calabrese & Elkink, 2014; Chakir & Parent, 2009; Smirnov, 2010; Smirnov & Egan, 2012), we define peer effects as spatial dependence among individuals. In particular, we define peers as students who are in the same classes and social clubs. We develop and estimate a spatial multinomial logit choice model in which individuals' latent utilities are linked through a spatial weight matrix. In doing so, we are able to account for interdependency among individuals, and hence their related preferences and choices. Our model is estimated using detailed information on high school seniors and their post-school choices from the Texas Higher Education Opportunity Project (THEOP).

The remainder of the paper is organized as follows. In Section 2, we introduce a spatial multinomial choice model that allows interdependence among individual preference. Section 3 discusses details regarding our estimation strategy, including the calculation of indirect and indirect effects. In Section 4, we describe the data that we apply to our model. The estimation results are presented in Section 5, leading to our conclusions in Section 6.

2 Model

In this section, we outline our approach to measuring peer effects. We first define and construct the peer effects among high school seniors. Next, we discuss the spatial multinomial logit choice model which we use to model the interdependencies in peer choices.

2.1 Measuring Peer Effects

Peer effects are understood as common classes and social clubs which each high school senior i shares with her peers in high school s. We are interested in whether individuals decide to attend college or select a profession given that they are linked in space. We estimate these different sets of decisions in two separate analyses.[1]

To measure peer effects, we adapt a spatial econometrics approach. Let N^s denote the set of high school seniors in high school s and $|N^s| \equiv n^s$ be the total number of seniors in high school s. We define a symmetric square matrix A^s whose entity a_{ij}^s is the total number of classes and social clubs student i and j have in common in high school s. The dimension of A^s is the $n^s \times n^s$. We construct the spatial weight matrix W^s that measures the peer effect in high school s in the following way: w_{ij}^s is an element of W^s defined as:

$$w_{ij}^s = \begin{cases} a_{ij}^s, & \text{if} \quad i \neq j \\ 0, & \text{if} \quad i = j. \end{cases}$$

Notice that the matrix W^s has zero diagonal elements, as we do not allow for an individual to spatially affect herself. In addition, W^s is symmetric, i.e., $w_{ij}^s = w_{ji}^s$, implying the peer effect between any two individuals is symmetric.[2] The cardinal value of a_{ij}^s implies the ordinal closeness among friends.

[1] We assume that W is exogenous to the model and chosen by the researcher. However, W is possible to be endogenous if the friendship formation among individuals is taken into consideration (Kelejian & Piras, 2014; Qu & Lee, 2015). We do not model network formation in this paper.

[2] In what follows, the high school index superscript s is suppressed to ease notation.

2.2 Spatial Random Utility

Let C_i denote the choice set of individual i. We assume that everyone faces the same choice set, i.e., $C_i = C$. We assume that spatial dependence is specified as a relationship between individual preferences linked in space via a linear utility specification. The latent utility of individual i choosing alternative j is thus given by:

$$u_{ij} = \rho \sum_{k=1}^{n} w_{ik} u_{kj} + v_j(\beta) + \varepsilon_{ij}. \tag{1}$$

The deterministic utility component of each alternative is denoted by $v_j(\beta)$. The parameters of interest are ρ and β. Idiosyncratic shocks ε_{ij} are introduced to form the random utility structure. The model includes a spatial lag vector $\rho \sum_{k=1}^{n} w_{ik} u_{kj}$ which represents the linear combination of values of the latent dependent variable vector from neighboring observations. Because we allow for endogenous interaction and feedback effects, in the language of of LeSage (2014), the model considered in this paper falls into the category of a global spillover specification.

In addition, we assume ε_{ij} follows a Type I Extreme Value (TIEV) distribution. Let y_{ij} be an observed decision. The decision rule for individual i is:

$$y_{ij} = \begin{cases} 1, & \text{if } u_{ij} \geq u_{il}, \forall j, l \in C \\ 0, & \text{otherwise }, \end{cases} \tag{2}$$

and thus we allow for one and only one alternative to be chosen. By stacking the equations, we can write latent preferences as:

$$u_j = \rho W u_j + v_j(\beta) + \varepsilon_j, \tag{3}$$

where $u_j = (u_{1j}, u_{2j}, \ldots, u_{nj})'$ and $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{nj})'$. Equations (2) and (3) together with the distributional assumption on the error terms fully describe the basic spatial discrete choice model.[3]

[3] A richer model could also include the spatial impact δW on deterministic utility components $v_j(\beta)$. However, we do not consider it in this paper given the substantial nonlinearity of the current model.

3 Estimation

This section describes the statistical model derived from the utility structures in Section 2.2. We proceed in the following steps. First, we collect terms involving u in Equation (3) and premultiply by$Z \equiv (I - \rho V)^{-1}$, where V is the row and column normalized matrix of W in order to preserve symmetry. Row standardization is required because for larger schools with more clubs and classes, the possibility the same level of engagement with peers could be distorted among different schools simply because the students had more options. One could also impose column standardization to achieve the same goal. Reduced form preferences thus become:

$$u_j = Z(\rho)v_j(\beta) + Z(\rho)\varepsilon_j. \tag{4}$$

Recalling that $Z \equiv (I - \rho V)^{-1}$ and assuming convergence,[4] we have

$$\begin{aligned} Z(\rho) &= \sum_{k=0}^{\infty} (\rho V)^k \\ &= D(\rho) + (Z(\rho) - D(\rho)), \end{aligned} \tag{5}$$

where $D(\rho)$ is the $n \times n$ matrix with the diagonal elements of $Z(\rho)$, which we refer to as the "private" or "direct" effect, while $Z(\rho) - D(\rho)$ is the so-called "social" or indirect effect. Equation (4) then can be written as:

$$u_j = Z(\rho)v_j(\beta) + D(\rho)\varepsilon_j + (Z(\rho) - D(\rho))\varepsilon_j. \tag{6}$$

We next make a behavioral assumption that individuals ignore social shocks $(Z(\rho) - D(\rho))\varepsilon$ when making decisions.[5] That is,

$$u_j = Z(\rho)v_j(\beta) + D(\rho)\varepsilon_j. \tag{7}$$

[4]Convergence ensured by the identification conditions of the spatial discrete choice model. See the detailed discussion in Smirnov (2010).

[5]This assumption is violated however, when a student has strong preference over a certain school because this student doesn't have any peers.

Under the distributional assumption that ε follows the Type I Extreme Value distribution, the conditional probability of individual i choosing alternative j has the closed form expression:

$$p_{ij} = \frac{\exp(\sum_{k=1}^{n} z_{ik}(\rho) v_{kj}(\beta)/d_{ii}(\rho))}{\sum_{j \in J} \exp(\sum_{k=1}^{n} z_{ik}(\rho) v_{kl}(\beta)/d_{ii}(\rho))}. \tag{8}$$

Together with Equation (2), the log-likelihood is:

$$LL(\beta, \rho \mid y, X, W) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \ln(p_{ij}). \tag{9}$$

The pseudo maximum likelihood estimates (PMLE) are the values of parameters (ρ, β) that maximize LL in equation (9). Several remarks are worth mentioning. Equation (2) and (7) comprise an auxiliary form of the true model comprised by equations (3) and (2). Both models have the same observed deterministic components of individual random utilities. However, error terms in the auxiliary model are assumed to be independent. Since some of the information about effects of individual interdependence is not present in the auxiliary model, parameter estimates are not necessarily asymptotically efficient. Smirnov (2010) and Smirnov & Egan (2012), which our model is based on, provide Monte Carlo simulation results regarding to the bias, MSE and inference of this approach. They show that estimates and effects are consistently estimated.[6]

Another way to estimate spatial multinomial choice models is to integrate out the errors' dependency structure. Baltagi et al. (2016) provides a detailed discussion of the methodology and computation feasibility of this method. Although conceptually straightforward, numerical integration introduces a computational burden. Moreover, the parameters of interest in this paper, spatial

[6] We have also conducted a check on our methods by conducting a small Monte Carlo simulation. We proceeded in two steps. We allowed the total number of possible choices to be four and the number of different individual characteristics associated with each choice to be 5. We drew 5000 identically and independently distributed idiosyncratic shocks from the TEIV distribution and specified a symmetric and zero diagonal spatial weight matrix. After substituting the true parameter values, covariates, spatial weights, and idiosyncratic shocks into the utility function, we solved for the optimal decision for each observation and, given optimal decisions, individual characteristics and the spatial weight matrix, we used the estimation procedure just outlined to estimate model primitives with the data generated in the previous step. Simulation results based on 500 replications are available on request. Estimated parameters are close to the true parameter values and we are confident that our estimation algorithms are correct. We did not report our Monte Carlo simulation because the purpose of our simulation is simply to verify whether our code work properly.

effect ρ and utility parameters β, can be easily recovered without numerical integration.

Because we are interested in how peers influence choices, we define choice sets $C_{college} \equiv$ {attending college, not attending college}, if she faces a binary choice, and $C_{career} \equiv$ {attending college, working, serving military, staying at home}, if she faces multiple choices.[7] Let P denote the total number of individual characteristics. The deterministic utility component is specified as:

$$v_j(\beta_j) = x_i'\beta_j \tag{10}$$

where $x_i = (x_{i1}, x_{i2}, ..., x_{iP})$ denotes individual characteristics and $\beta_j \equiv (\beta_{j1}, \beta_{j2}, ..., \beta_{jP})$ are parameter of interest. The marginal utility of each characteristic is alternative specific and thus is indexed by j. Substituting Equation (10) into Equation (7) completes the model.

3.1 Direct & Indirect Impacts

Because the impact of changes in an explanatory variable differs over all observations, this section provides a measure of these varying impacts. The average total impact for coefficient β_{jp} is the average of all derivatives of the latent utility u_{ij} with respect to x_{ip} for any i. The average direct impact for coefficient β_{jp} is the average of all own derivatives. Naturally, the average indirect impact is defined as the difference between average total impact and average direct impact. Formally, the average total effect (TE), average direct effect (DE) and average indirect effect (IE) are defined as:

$$\begin{aligned} \text{TE}(\beta_{jp}) &= \frac{1}{N} l' \left[(I - \rho V)^{-1} \beta_{jp} \right] l \\ \text{DE}(\beta_{jp}) &= \frac{1}{N} tr\big((I - \rho V)^{-1} \beta_{jp} \big) \\ \text{IE}(\beta_{jp}) &= \text{TE}(\beta_{jp}) - \text{DE}(\beta_{jp}) \end{aligned} \tag{11}$$

where l is a $N \times 1$ vector of ones. [8]

[7] In the case where an individual is working and attending school, her choice is defined as school if she is working part time. Staying at home is served as base and hence parameters β_1 is normalized to 0 for the sake of identification.

[8] It is worth noting that DE (β_{jp}) equals β_{jp} because $\frac{1}{N} tr\big((I - \rho V)^{-1} \beta_{jp}\big) = \frac{1}{N} tr\big((I - \rho V)^{-1}\big) \beta_{jp} \approx \beta_{jp}$. As pointed out by the referee, the last equality of previous equation should be approximated given large N and small ρ. In

4 Data

We estimate our model using data from the Texas Higher Education Opportunity Project (THEOP).[9] The THEOP is accessible via the Princeton University Office of Population Research (OPR) data archive and requires confidentiality protocols that are rather extensive. It is a multi-year research evaluation study initiated by the Ford Foundation and undertaken by Princeton University's Office of Population Research. The centerpiece of THEOP is a two-cohort longitudinal survey of sophomores and seniors who were enrolled in Texas public high schools as of spring, 2002. Our estimation sample is drawn from the senior cohort only, which is a sample of 13,803 high school seniors attending 96 Texas public high schools. Students were randomly selected and surveyed during their last semester in high school and data were collected through a self-administered survey. Our estimation sample comprises 3125 seniors who responded to the self-administered survey.

Of relevance to this study, the THEOP provides a unique measure of peer effects. The survey asks each student what classes and social clubs she belongs to. We define peers as students who share common classes and social clubs. The THEOP also collects data on students' post high school activities in the second wave, which is used to construct the outcome variables for our analysis. We construct two outcomes. The first is a binary indicator for attending college. This variable is coded as one for all respondents who report attending college, either part or full time, in the follow-up survey. The second outcome is a multinomial measure for post high school career choice. This variable is coded as 1 if the respondent's primary activity is attending college, 2 if the respondent's primary activity is working, 3 if the respondent joins the military, and 4 if they are not attending college, not working and not in the military. Table 1 shows the sample distributions of the outcomes we study. 81% of our sample attend college either part or full time, while 19% report that they do not attend college at all. In terms of their primary activity, 70% of respondents report their primary activity is attending college, 15% report that their primary activity is working, 5% report being int he military, and 10% are not attending college, are not working, and are not in

calculating the different and indirect effect below, we use the definition in Equation (11) proposed in LeSage & Pace (2009).

[9] Papers based on the public and restricted use data are posted on the THEOP Web site at http://theop.princeton.edu/publications/

the military (are at home).

In addition, THEOP contains high school characteristics and individual demographics. Table 2 presents the sample descriptive statistics used in estimation. Figures 1 and 2 present undirected networks of peers in for two different Texas High Schools. It is clear from these two examples, which are representative of the other Texas hight school peer patterns, that such network linkages are pronounced and that there is variation in these networks across Texas high schools.

5 Estimation Results

5.1 Binary Models for College Attendance

Table 3 and Table 4 present maximum likelihood estimation of the coefficients of the binary choice logit model, and the spatial binary logit model for attending college. The impact of changes in explanatory variable for the logit model (both average partial effects (APE) and partial effects evaluated at sample averages (PEA)) are also presented in Table 3, and direct and indirect effects of explanatory variables for the spatial logit model are presented in Table 5.

Overall, the first point to note is that the spatial model finds significant evidence of peer effects, as evidenced by the positive and significant coefficient estimate on the peer effect term, ρ. Nonetheless, the coefficient estimates from the binary logit and the spatial binary logit model show general agreement in terms of the characteristics of the individual and their parents that are significant determinants of seniors' decision to go to college. In particular, the probability of attending college is lower for females, for seniors who are older, and who perform worse academically as measured by their percentile class rank, and for those whose family rents rather then owns their home. As expected, the probability of attending college is increasing in parental education, although only the education of the father (and not the mother) is significant in the spatial model that accounts for peer effects.[10]

School characteristics are also important determinants of student choices, however, the esti-

[10]In estimation, we treat parental education as a continuous variable instead of categorical variable because education is a monotonic ascending order in terms level and year and is discretized as a fine level.

mated effects differ when peer effects are accounted for. For example, after accounting for peer effects, the likelihood that an individual attends college is increasing in the college attendance rate of the high school they attend, and decreasing in the dropout rate of the high school they attend whereas in the binary logit model that does not account for peer effects, the percent attending college or percent dropping out of high school do not impact on individuals decision to attend college. Similarly, high school level variables related to the availability of advanced placement (AP) courses, the percent of students taking AP courses and the percent passing AP course are significantly and positively related to the individual attending college after accounting for peer effects, whereas only whether AP courses are offered is significantly related to attending college in the simple logit model. Finally, after accounting for peer effects, the probability of attending college is decreasing in the distance to a 4 year college and and the distance to a private college, whereas the estimates from the simple logit imply the opposite – that the probability of attending college is increasing in distance to four year and private colleges.

In terms of magnitudes of impacts of individual, family and school characteristics, the spatial model that accounts for peer effects implies much larger effects than the simple logit model, as a comparison of the APE (or PEA) in Table 3 with the direct and indirect effects (the sum of which produce the total partial effect) reported in Table 5 reveals. For example, the simple logit model estimates imply a partial effect of father's education on college attendance of 2 percentage points whereas the spatial model estimates imply an total effect of 14 percentage points, comprised of a 10 percentage point direct effect and a 4 percentage point indirect effect. Similarly, the logit model estimates imply a 4 percentage point reduction in the probability of attending college for seniors whose family live in a rented accommodation compared to those whose family own their home, whereas the coefficient estimates from the spatial model imply a total reduction in the probability of attending college of 60 percentage points, comprised of a 43 percentage point reduction due to the direct effect and a 17 percentage point reduction due to the indirect effect. Characteristics of the high school seniors attend typically have economically small impacts on the decision to attend college in both the spatial and non-spatial models.

In order to provide a more concrete insight into the modeling results, we use the coefficient estimates from the non-spatial and spatial models to predict the probability of attending college as a function of individual (and school characteristics). These are reported in Table 6.

Table 6 shows that for females, the predicted probability of attending college is lower when peer effects are taken into account compared to when they are not.

For example, using coefficient estimates from the spatial model that accounts for peer effects, the probability of attending college for a female from a racially diversified high school whose parents both have less than a high school education is 0.526 compared to 0.578 based on coefficients from the non-spatial model.[11] For a female from a less racially diversified high school with parents who both have less than a high school education, the probability of attending college is predicted to be 0.541 based on the spatial model, compared to 0.609 using the non-spatial model estimates.[12] The impact of increasing both parents education from less than a high school education to at least a high school education is, however, estimated to be slightly greater for females using the spatial model estimates compared to the non-spatial model estimates. Specifically, for females from a racially diversified high school, increasing both parents education raises the probability of attending college by 0.237 using the spatial model estimates compared to a 0.2216 increase using the non-spatial model estimates. For females from a less racially diversified high school the increase in the probability of attending college is 0.26 based on the spatial model and 0.236 based on the non-spatial model.

For males, the predicted probability of attending college is slightly greater based on the spatial model (compared to the non-spatial model) for those whose parents both have a low level of education, and is slightly greater based on the non-spatial model for those whose parents both have at least a high school education. In contrast to the results for females, the impact of increasing the education of both parents from less than high school to at least a high school graduate is larger for males when peer effects are not accounted for. Using the coefficient estimates from the non-spatial model, increasing both parents education leads to 0.194 improvement in the probability that a male from a racially diversified high school attends college and a 0.18 improvement in the probability that a male from a less racially diversified high school attends college. Using the spatial model,

[11] We define racially diversified high school as one with at least 50% of its student body is non-white.

[12] We define a less racially diversified high school as one with at least 50% of its student body is white.

the corresponding effects are 0.129 and 0.165.

Table 6 also shows that for a given race and parental education level, irrespective of whether or not peer effects are accounted for, males are more likely to attend college than females. The predicted probabilities based on the spatial (non-spatial) model for females and males indicated that females from a racially diversified high school with parents that have less than a high school education have a 0.526 (0.578) probability of attending college compared to 0.692 (0.666) for a male from a racially diversified high school with parents who have less than a high school education.

5.2 Mulinomial Models for Post High School Choice

Table 7 and Table 10 provide maximum likelihood coefficient estimates of the model of post high school choice, where senior choose to either attend college, to work, to join the military, or to stay home (where stay at home is the base category). Table 8 and Table 9 report partial effects of the explanatory variables for the Multinomial Logit Model and Table 11 reports direct and indirect partial effects from the spatial multinomial model.

As with the binary spatial model for college attendance, the multinomial spatial model provides evidence of the importance of high school peer effects in post school decisions. Specifically, the coefficient on the peer effect term, ρ, is statistically significant and positive. In terms of the impact of individual, family and school characteristics, the point estimates are quite similar for the simple multinomial and spatial multinomial models, although there are differences in the precision with which the coefficients are estimated. For example, in the spatial multinomial model, all else being equal, females are statistically significantly more likely than males to attend college, to work, and to join the military, whereas the coefficients estimates are significant for working and joining the military only in the simple multinomial model.

Irrespective of the model estimated, however, very few of the individual, family, or school characteristics predict joining the military after high school. And while individual academic performance as measured by percentile rank in class, parents education, and having a family who rents their home are statistically important determinants of attending college in both multinomial and spatial multinomial models, school characteristics have differing effects. In the spatial multinomial model, for example, the percentage of students from the high school attending college increases the likelihood of the individual attending college, while greater distances to 4 and 2 year colleges reduce it. In contrast, estimates from the simple multinomial model indicate that the probability of attending college is higher for students attending high schools a greater distance from a four year college, and lower for students from high schools with a higher proportion of low income students. Similarly, the estimates from the spatial multinomial logit indicate that the probability of choosing to work following high school is statistically significantly lower amongst individuals who attended high schools further away from 4 year colleges and with fewer nonwhites enrollments, whereas the simple multinomial logit estimates indicate a statistically significantly *higher* probability of working for students from high schools further away from 4 year colleges and with fewer nonwhites enrollments.

Overall, however, it is characteristics of the high school seniors and their family, not school characteristics, that are economically significant determinants of post-high school college and career decisions. This is demonstrated in Tables 8 and 9, which report average partial effects and partial effects at the average for the simple multinomial logit model, and in Table 11, which reports direct and indirect effects for the spatial model that accounts for peer effects. It is interesting to note that for the spatial multinomial logit model, the direct effects are typically around twice the magnitude of the indirect effects.

As with the binary logit models, we provide a more concrete comparison of the non-spatial and spatial multinomail logit model estimates by way of model predictions regarding the probability of attending college, working, and joining the military (leisure is the base category. These are reported for males and females from racially diversified high schools and less racially diversified high school who have parents with a low level of education (both parents have less than a high school education) and a high level of education (both parents are high school graduates) in Table 12.

As Table 12 shows, regardless of gender, race and parental education, and whether or not the impact of peers is accounted for, the high school students in our sample are more likely to attend college than work after completing high school, and more likely to work than join the military (males from a less racially diversified high school with parents who have each graduated high school are the exception). For example, a female from a racially diversified high school with parents with a low level of education attends college with probability 0.467, works with probability 0.239 and joins the military with probability 0.124 based on the non-spatial model. However, ignoring peer effects appears to lead to an upward bias in the estimated probabilities. Once peer effects are accounted for, the corresponding probabilities are estimated to be 0.386 for attending college, 0.193 for working and 0.112 for joining the military. Similarly for males, by failing to account for peer effects, the non-spatial model tends to lead to upward biased predictions probability of attending college, although the predictions for working and joining the military are similar. For a males from a racially diversified high school with parents who have a low level of education for example, the non-spatial model predicts the probability of attending college is 0.581, the probability of working is 0.164, and the probability of joining the military is 0.025. The corresponding predictions based on the spatial model that accounts for peer effects are 0.573, 0.165 and 0.03.

It is also the case that gender and parental education are more important determinants of sample members post-high school choices than race in both the spatial and non-spatial model. In the spatial model for example, the probability of a female from a less diversified high school with parents who have both graduated high school attending college is 0.654 compared to 0.787 for a male from a racially diversified high school with 2 parents who have graduated high school (the probabilities are 0.678 and 0.796, respectively for the non-spatial model). In terms of parental education, the probability that a female from a racially diversified high school with parents who have both not graduated high school attending college is 0.386 compared to 0.666 if both parents have graduated high school according the spatial model (the probabilities are 0.467 and 0.678, respectively for the non-spatial model)

Finally the impact of increasing parents education on attending college is sensitive to accounting for peer effects for females from a racially diversified high school and males from a less racially diversified high school (but not other categories). For example, the spatial model estimates the impact of increasing both parents education (from low to high) on attending college is for females from a racially diversified high school is 0.268 compared to 0.211 from the non-spatial model. For males from a less racially diversified high school, the corresponding predictions are 0.323 for the spatial model and 0.219 for the non-spatial model.

6 Conclusion

Our study has developed a spatial model of to study the impact of peer effects on post high school decision making based on the random utility paradigm and has applied this econometric model to a unique data set of High School Seniors in Texas. We have developed both a binary and multinomial logit specifications of the random utility model utilizing parametric specifications based on standard treatments used in the random utility literature and have considered a spatially autoregressive generalization of the classical random utility model to account for peer effects within Texas High Schools. Allowing interdependence among individual preference requires us to develop a new estimation strategy and interpretation of marginal effects that include the calculation of indirect and indirect effects. Our estimation results clearly point to the importance of peer effects in this age group of young people and that such peer effects, if ignored, may distort our understanding of the determinants of such decisions, both in terms of magnitudes and in terms of directions of effects.

References

Alvarado, S. E., & Turley, R. N. L. (2012). College-bound friends and college application choices: Heterogeneous effects for Latino and white students. *Social Science Research*, *41*(6), 1451–1468.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, *27*(3), 247–267.

Arcidiacono, P., & Nicholson, S. (2005). Peer effects in medical school. *Journal of Public Economics*, *89*(2-3), 327–350.

Baltagi, B. H., LeSage, J. P., & Pace, R. K. (2016). *Spatial econometrics: qualitative and limited dependent variables*. Emerald Group Publishing.

Calabrese, R., & Elkink, J. A. (2014). Estimators of binary spatial autoregressive models: A monte carlo study. *Journal of Regional Science*, *54*(4), 664–687.

Chakir, R., & Parent, O. (2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, *88*(2), 327–344.

Ding, W., & Lehrer, S. F. (2007). Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics*, *89*(2), 300–312.

Fletcher, J. M. (2010). Peer influences on college choices: New evidence from Texas.

Fletcher, J. M., & Tienda, M. (2009). High school classmates and college success. *Sociology of Education*, *82*(4), 287–314.

Kelejian, H. H., & Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, *46*, 140–149.

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *Available at SSRN 2420725*.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, *28*(4), 825–860.

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at west point. *Review of Economics and Statistics*, *89*(2), 289–299.

Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *Review of Economics and Statistics*, *92*(3), 562–576.

Qu, X., & Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, *184*(2), 209–232.

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional Science and Urban Economics*, *40*(5), 292–298.

Smirnov, O. A., & Egan, K. J. (2012). Spatial random utility model with an application to recreation demand. *Economic Modelling*, *29*(1), 72–78.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, *85*(1), 9–23.

Table 1: Occupational Choice

	College	Work	Military	Home
Number of Observation	2189	463	146	327
(Proportion)	(70%)	(15%)	(5%)	(10%)

	College	Not College
Number of Observation	2523	602
(Proportion)	(81%)	(19%)

[a] Total number of observations is 3125

Table 2: Descriptive Statistics

	Mean	Std. Dev.
Gender	0.45	0.49
Age	18.42	0.59
House renter	0.14	0.35
Father education	5.33	1.98
Mother education	5.18	1.85
Rank	39.29	48.98
HS College percentage	75.83	19.29
HS dropout rate	1.55	1.05
Low income	31.43	22.78
AP course offered	13.56	5.53
AP taking percentage	13.19	6.66
AP passing percentage	52.02	23.60
Total Enrollment	2393.42	1176.66
Distance 4yr college	10.13	8.89
Distance 2yr college	12.94	15.06
Distance private college	51.63	67.61
HS mean algebra score	1411.54	272.44
Nonwhite	1386.26	1192.10

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The higher the rank, the lower the percentage.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.
[i] Total number of observation is 3125

Figure 1: Example of An Undirected Network of Peers-School A

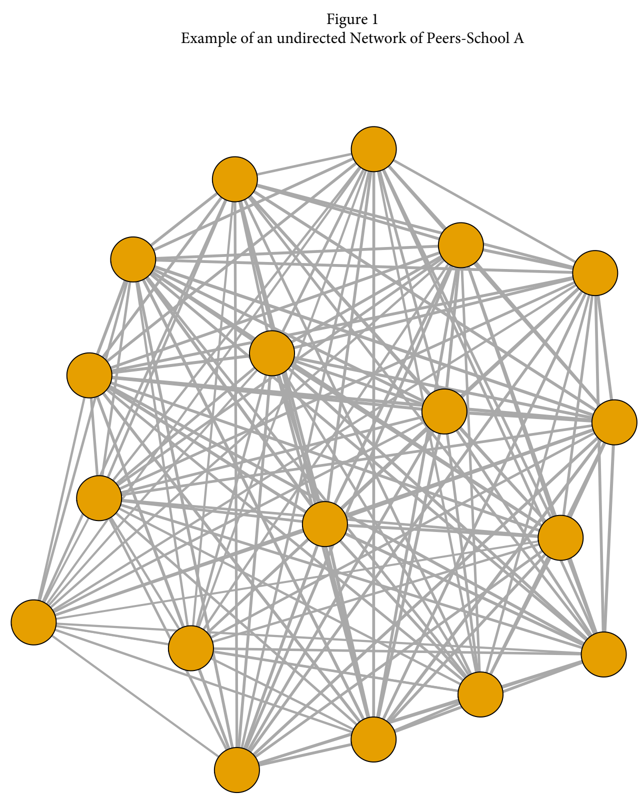

Figure 2: Example of An Undirected Network of Peers-School B

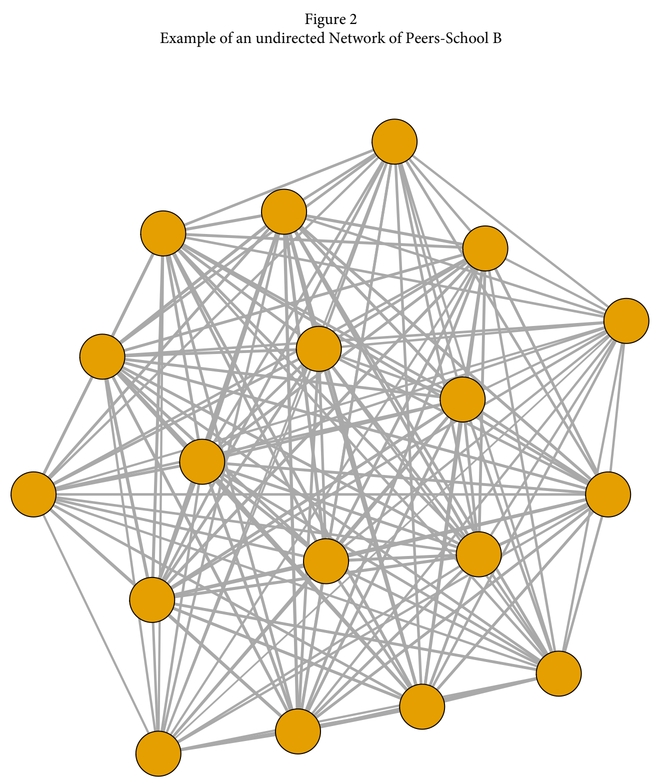

Table 3: Binary Logit Model

y = 1 if attending college

	β	APE	PEA
Gender	-0.37***	-0.05***	-0.05***
Age	-0.14*	-0.02**	-0.02**
House owner	-0.31**	-0.04***	-0.04***
Father education	0.14***	0.02***	0.02***
Mother education	0.23***	0.03***	0.03***
Rank	-0.01***	-1.36*e*-3***	-1.33*e*-3***
HS college percentage	0.01	2.63*e*-4	2.58*e*-4
HS dropout rate	-0.03	-4.62*e*-3	-4.53*e*-3
Low income	-0.02***	-2.01*e*-3***	-1.97*e*-3***
AP course offered	0.03*	3.81*e*-3	3.74*e*-3
AP taking percentage	-1.05*e*-3	-1.42*e*-4	-1.38*e*-4
AP passing percentage	6.39*e*-4	8.60*e*-5	8.44*e*-5
Total Enrollment	-8.12*e*-5	-1.09*e*-5	-1.07*e*-5
Distance to 4yr college	0.02***	2.56*e*-3***	2.51*e*-3***
Distance to 2yr college	-0.01***	-1.75*e*-3***	-1.72*e*-3***
Distance to private college	2.62*e*-3**	3.53*e*-4***	3.46*e*-4***
HS mean algebra	1.51*e*-4	2.03*e*-5	1.99*e*-5
Nonwhite	1.19*e*-4	1.61*e*-5	1.58*e*-5

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 4: Spatial Binary Logit Model: Point Estimates

y = 1 if attending college	
Peer effect ρ	0.29**
	β
Gender	-0.19***
Age	-0.14*
House renter	-0.43**
Father education	0.10***
Mother education	-0.13
Rank	-0.07***
HS college percentage	0.01**
HS dropout rate	-0.04**
Low income	-0.03*
AP course offered	0.02*
AP taking percentage	0.02*
AP passing percentage	3.30e-3*
Total Enrollment	2.80e-4
Distance to 4yr college	-0.03***
Distance to 2yr college	9.80e-3
Distance to private college	-4.21e-3**
HS mean algebra	1.92e-4
Nonwhite	-3.00e-4

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 5: Spatial Binary Logit Model: Direct & Indirect Effect

y = 1 if attending college

	DE	IE
Gender	-0.19***	-0.08***
Age	-0.14*	-0.06**
House renter	-0.43**	-0.17**
Father education	0.10***	0.04**
Mother education	-0.13	-0.05
Rank	-0.07***	-0.03***
HS college percentage	0.01**	4.05e-3**
HS dropout rate	-0.04**	-0.02*
Low income	-0.03*	-0.01*
AP course offered	0.02*	8.11e-3*
AP taking percentage	0.02*	8.01e-3
AP passing percentage	3.30e-3*	1.34e-3*
Total Enrollment	2.80e-4	1.13e-4
Distance to 4yr college	-0.03***	-0.01
Distance to 2yr college	9.80e-3	3.97e-3*
Distance to private college	-4.21e-3**	-1.71e-3**
HS mean algebra	1.92e-4	7.71e-5
Nonwhite	-3.00e-4	1.21e-4*

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 6: Predicted Probability: Binary Models

Predicted Probability		Variables		
Non-spatial	Spatial	Gender	Parental education	Racial diversity
0.578	0.526	Female	Low	Yes
0.609	0.541	Female	Low	No
0.794	0.763	Female	High	Yes
0.845	0.801	Female	High	No
0.666	0.692	Male	Low	Yes
0.705	0.712	Male	Low	No
0.860	0.821	Male	High	Yes
0.885	0.877	Male	High	No

[a] Predicted probability is the probability of attending college. Probability of all 2 choices sum up to 1 and therefore probability of not attending college is the residual.
[b] Predicted probability for spatial model is calculated by averaging spatial weight matrix as seen in Equation (11).
[c] Parental education is defined as high if both parents have at least high school degree.
[d] Racial diversity is yes if a high school with at least 50% of its student body is non-white.

Table 7: Multinomial Logit Model

y = 1 if attending college
y = 2 if working
y = 3 if serving in military
y = 4 if staying at home (base)

	College β_1	Work β_2	Military β_3
Gender	0.07	0.69***	1.86***
Age	-0.03	0.24***	-0.17
House renter	-0.31**	-0.32**	0.03
Father education	0.14***	-0.05	0.09
Mother education	0.20***	0.06	0.04
Rank	-1.35*e*-3***	-1.49*e*-3	-1.90*e*-3*
HS college percentage	4.22*e*-3	7.94*e*-3**	2.37*e*-3
HS dropout rate	-0.02	-0.16**	-0.03
Low income	-0.01*	-5.94*e*-3	-7.40*e*-4
AP course offered	0.02	3.26*e*-4	-0.04
AP taking percentage	-8.51*e*-3	-5.55*e*-3	-0.02
AP passing percentage	2.56*e*-3	-5.58*e*-3	3.51*e*-4
Total Enrollment	-5.50*e*-5	1.99*e*-4	7.65*e*-5
Distance to 4yr college	0.02***	0.03**	0.02
Distance to 2yr college	-0.02***	-3.09*e*-3	-0.02*
Distance to private college	1.21*e*-3	-3.32*e*-3**	-1.76*e*-3
HS mean algebra	-4.98*e*-4	-5.15*e*-4	-5.72*e*-4
Nonwhite	1.14*e*-4	-1.46*e*-4	1.52*e*-4
Constant	2.35	-3.24*	2.13

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 8: Multinomial Logit Model: Average Partial Effect

Variables	Gender	Age	House owner	Father education	Mother education	Rank
y = 1 if College	-0.102***	-0.024*	-0.028	0.028***	0.031***	-0.002***
y = 2 if Work	0.059***	0.033***	-0.010	-0.020***	-0.012***	0.001***
y = 3 if Military	0.071***	-0.008	0.011	4.081*e*-4	-0.005*	3.061*e*-4***
y = 4 if Home	-0.028***	-0.001	0.026*	-0.008**	-0.014***	8.891*e*-4***

Variables	HS college percentage	HS dropout rate	Low income	AP course offered	AP taking percentage	AP Passing percentage
y = 1 if College	-2.156*e*-05	0.011	-0.002**	0.004	-4.170*e*-4	0.001
y = 2 if Work	5.341*e*-4	-0.017**	4.731*e*-4	-0.001	3.222*e*-4	-0.001**
y = 3 if Military	-8.714*e*-05	7.391*e*-4	3.790*e*-4	-0.002	-6.735*e*-4	1.291*e*-4
y = 4 if Home	-4.261*e*-4	0.004	9.227*e*-4	-0.001	0.001	-7.481*e*-05

Variables	Total enrollment	Distance to 4 yr college	Distance 2yr college	Distance to private college	HS mean algebra	Nonwhite
y = 1 if College	-2.661*e*-05	8.939*e*-4	-0.003***	5.754*e*-4***	-2.801*e*-05	3.001*e*-05
y = 2 if Work	2.983*e*-05*	0.001	0.002**	-4.900*e*-4***	-1.041*e*-05	-2.914*e*-05*
y = 3 if Military	-3.462*e*-06	-3.849*e*-05	-1.547*e*-4	-7.973*e*-05	-5.841*e*-06	4.531*e*-06
y = 4 if Home	2.611*e*-07	-0.002**	0.001***	-6.028*e*-06	4.421*e*-05	-5.400*e*-06

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 9: Multinomial Logit Model: Partial Effect at Average

Variables	Gender	Age	House owner	Father education	Mother education	Rank
y = 1 if College	-0.102***	-0.026**	-0.027	0.030***	0.032***	-0.002***
y = 2 if Work	0.067***	0.033***	-0.006	-0.021***	-0.014***	0.001***
y = 3 if Military	0.054***	-0.005	0.009	-2.629*e*-4	-0.004**	0.001***
y = 4 if Home	-0.019*	3.551*e*-05	0.025*	-0.009**	-0.014***	0.001***

Variables	HS college percentage	HS dropout rate	Low income	AP course offered	AP taking percentage	AP Passing percentage
y = 1 if College	-5.681*e*-5	0.013	-0.002**	0.004	-5.530*e*-4	0.001
y = 2 if Work	5.040*e*-4	-0.017**	6.039*e*-4	-0.001	3.146*e*-4	-9.790*e*-4**
y = 3 if Military	-6.260*e*-5	2.951*e*-4	3.110*e*-4	-0.002*	-4.811*e*-4	7.536*e*-5
y = 4 if Home	-3.842*e*-4	0.003	9.222*e*-4*	-9.580*e*-4	6.961*e*-4	-1.073*e*-4

Variables	Total enrollment	Distance to 4 yr college	Distance 2yr college	Distance to private college	HS mean algebra	Nonwhite
y = 1 if College	-2.914*e*-05	8.870*e*-4	-0.003***	6.166*e*-4***	-2.951*e*-05	3.326*e*-05
y = 2 if Work	2.997*e*-05*	8.941*e*-4	0.0012***	-5.122*e*-4***	-7.771*e*-06	-2.978*e*-05*
y = 3 if Military	-1.988*e*-06	-4.519*e*-05	-6.701*e*-05	-6.973*e*-05	-3.703*e*-06	2.689*e*-06
y = 4 if Home	1.391*e*-06	-0.002**	0.001***	-3.411*e*-05	4.107*e*-05	-6.175*e*-06

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 10: Spatial Multinomial Logit Model: Point Estimates

y = 1 if attending college
y = 2 if working
y = 3 if serving in military
y = 4 if staying at home (base)

Peer effect ρ		0.36^{***}	
	College β_1	Work β_2	Military β_3
Gender	0.08^{*}	0.64^{*}	2.98^{*}
Age	-0.03	0.27^{**}	-0.16^{*}
House renter	-0.31^{**}	-0.31	0.04
Father education	0.15^{*}	-0.20^{*}	0.10
Mother education	0.24^{***}	0.07	0.03
Rank	-0.01^{*}	$-1.40e$-3	$-2.80e$-3
HS college percentage	$3.81e\text{-}4^{**}$	$9.51e$-4	$3.51e$-3
HS dropout rate	-0.02	0.14^{**}	-0.02
Low income	-0.07	$-6.72e$-4	$9.97e\text{-}3^{*}$
AP course offered	0.03	$-7.50e$-4	-0.04
AP taking percentage	$-3.51e$-3	$-3.61e$-3	-0.02
AP passing percentage	$1.00e$-3	$-3.53e$-4	$3.11e$-4
Total Enrollment	$3.24e$-3	$7.89e$-4	$8.75e$-4
Distance to 4yr college	-0.01^{***}	-0.04^{*}	$4.62e$-3
Distance to 2yr college	-0.015^{***}	$1.00e$-3	-0.01
Distance to private college	$7.11e$-4	$-4.11e\text{-}4^{**}$	$-3.17e$-3
HS mean algebra	$-5.24e$-4	$-3.20e\text{-}5^{**}$	$-5.01e$-4
Nonwhite	$-3.96e$-5	$1.16e\text{-}3^{*}$	$-8.70e$-5

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 11: Spatial Multinomial Logit Model: Direct & Indirect Effect

y = 1 if attending college
y = 2 if working
y = 3 if serving in military
y = 4 if staying at home (base)

	College		Work		Military	
	DE	IE	DE	IE	DE	IE
Gender	0.08*	0.04*	0.64*	0.36**	2.98*	1.66*
Age	-0.03	-0.02*	0.27**	0.15**	-0.16*	-0.08*
House renter	-0.31**	-0.17*	-0.31	-0.17	0.04	0.02
Father education	0.15*	0.08**	-0.20*	-0.11*	0.10	0.06
Mother education	0.24***	0.13***	0.07	0.04	0.03	0.02
Rank	-0.01*	-5.57*e*-3*	-1.40*e*-3	-7.88*e*-4	-2.83*e*-3	-1.56*e*-3
HS college percentage	3.84*e*-4**	2.23*e*-4*	9.51*e*-4	5.31*e*-4	3.55*e*-3	1.95*e*-3
HS dropout rate	-0.02	-0.01	0.14**	0.08*	-0.02	-0.01
Low income	-0.07	-0.04*	-6.72*e*-4	-3.73*e*-4*	9.94*e*-3*	5.52*e*-3*
AP course offered	0.03	0.02	-7.50*e*-4	-4.18*e*-4	-0.04	-0.02
AP taking percentage	-3.51*e*-3	1.67*e*-3	-3.61*e*-3	-2.01*e*-3	-0.02	-0.01
AP passing percentage	1.00*e*-3	5.58*e*-4*	-3.55*e*-4	-1.95*e*-4	3.11*e*-4	1.73*e*-4
Total Enrollment	3.22*e*-3	1.79*e*-4	7.89*e*-4	4.40*e*-4	8.74*e*-4	4.86*e*-4
Distance to 4yr college	-0.01***	-5.58*e*-4**	-0.04*	-0.02**	4.62*e*-3	2.57*e*-3
Distance to 2yr college	-0.02***	-8.37*e*-3**	1.00*e*-3	5.58*e*-4	-0.01	-5.58*e*-3
Distance to private college	7.11*e*-4	3.96*e*-4	-4.11*e*-4**	-2.29*e*-4*	-3.09*e*-3	-1.73*e*-4
HS mean algebra	-5.24*e*-4	-2.92*e*-4	-3.20*e*-5**	1.78*e*-5	-5.01*e*-4	-2.79*e*-4
Nonwhite	-3.94*e*-5	-2.18*e*-5	1.81*e*-3*	-6.13*e*-4*	-8.69*e*-5	-4.85*e*-5

[a] HS stands for high school.
[b] Gender equals to 1 if female.
[c] The variable House renter is an indicator that equals to 1 if a household rents.
[d] The variable Father/Mother education: 1-Illiterate/Semi-literate, 2-Elementary school, 3-Middle school, 4-High school, 5-some college, 6-2 year college, 7-4 year college, 8-Master degree, 9-professional grad school
[e] The variable Rank is class ranking of high school students indicated in percentile. The lower the percentage, the higher this person is ranked and therefore the coefficient is negative.
[f] The variable HS college percentage is the percentage of students plan to attend college in each high school.
[g] The variable Low income is the percentage of low income families in each high school.
[h] Nonwhite is the total number of nonwhite students in each high school.

Table 12: Predicted Probability: Multinomial Models

Predicted Probability		Variables		
Non-spatial	Spatial	Gender	Parental education	Racial diversity
y = 1 if attending college				
0.467	0.386	Female	Low	Yes
0.442	0.414	Female	Low	No
0.678	0.654	Female	High	Yes
0.698	0.666	Female	High	No
0.581	0.573	Male	Low	Yes
0.586	0.488	Male	Low	No
0.796	0.787	Male	High	Yes
0.805	0.811	Male	High	No
y = 2 if working				
0.239	0.193	Female	Low	Yes
0.309	0.279	Female	Low	No
0.143	0.144	Female	High	Yes
0.170	0.171	Female	High	No
0.164	0.165	Male	Low	Yes
0.203	0.202	Male	Low	No
0.083	0.069	Male	High	Yes
0.103	0.099	Male	High	No
y = 3 if serving in military				
0.124	0.112	Female	Low	Yes
0.108	0.117	Female	Low	No
0.085	0.078	Female	High	Yes
0.068	0.061	Female	High	No
0.025	0.030	Male	Low	Yes
0.023	0.021	Male	Low	No
0.158	0.152	Male	High	Yes
0.013	0.012	Male	High	No

[a] Probability of all 4 choices sum up to 1 and therefore probability of staying at home is the residual.
[b] Predicted probability for spatial model is calculated by averaging spatial weight matrix as seen in Equation (11).
[c] Parental education is defined as high if both parents have at least high school degree.
[d] Racial diversity is yes if a high school with at least 50% of its student body is non-white.