2018-17

Peer-Induced Beliefs Regarding College Participation

Vincent Boucher
Finagnon A. Dedewanou
Arnaud Dufays

Septembre / September 2018

Centre de recherche sur les risques les enjeux économiques et les politiques publiques

www.crrep.ca

Abstract

We study peer effects on the formation of beliefs regarding college participation. We present a structural model of learning in friendship networks. We show that the model is identified and we present a Bayesian estimation procedure. We estimate the model using data on teenagers' beliefs regarding college participation, controlling for preferences and academic achievement. We find that, on average, friends' beliefs account for about 12% of the updating process. We also find strong heterogeneity among schools and individuals. In particular, we find substantial unobserved individual heterogeneity, which casts doubt on the efficiency of network-targeted public policies.

JEL Codes: D83, I20, C31, C11
Keywords: Social networks, Beliefs updating, College participation, Heterogeneous peer effects

Vincent Boucher: Corresponding author. Department of Economics, Université Laval, CRREP and CREATE
vincent.boucher@ecn.ulaval.ca
Finagnon A. Dedewanou: Department of Economics, Université Laval, CRREP,
finagnon-antoine.dedewanou.1@ulaval.ca
Arnaud Dufays: Department of Economics, Université Namur, Associate professor, Université Laval,
Arnaud.Dufays@unamur.be

1 Introduction

It is not news that teenagers often have biased views about the world. Indeed students' subjective expectations affect many of their choices, ranging from career choices (e.g. Ashby and Schoon, 2010) to risky sexual behaviour (e.g. Sipsma, Ickovics, Lin, and Kershaw, 2015). Typical predictors of teenagers' subjective biases include gender and the characteristics of their parents, such as parental education level (e.g. Attanasio and Kauffman, 2010).

In this paper, we study the role of friendship in explaining teenagers' subjective beliefs regarding college participation. We present a structural model of social learning in networks that we estimate using data from middle- and high-school teenagers in the US. A large set of observable variables allows us to measure the impact of beliefs, controlling for (stated) preferences, academic achievement, and many other relevant socio-economic characteristics.

We find that the average belief of a student's peers accounts for 12% of the belief updating process, while the remaining percentage is due to individual characteristics. We also find substantial heterogeneity among schools and individuals. In particular, we find substantial *unobserved* individual heterogeneity: the importance of peers' beliefs on the updating process ranges from 8% to 73%. Importantly, differences in the magnitude of peer effect cannot be well predicted using observable characteristics; an observation that has important policy implications. In particular, without accounting for unobserved heterogeneity, network-targeted policies such as *key-players* analysis (Ballester, Calvó-Armengol, and Zenou, 2006) are likely ineffective.

We present a model where, at each period, individuals update their beliefs according to their friends' average beliefs (e.g. DeGroot, 1974) and also according to their own particular characteristics (e.g. age, gender, parental education, etc.). Accordingly, the steady-state distribution of beliefs is affected by the shape of the social network, but it is not consensual (e.g. Friedkin and Johnsen, 1990).

We estimate the model using data related to students' beliefs regarding college participation. Specifically, each student is asked: *"On a scale of 1 to 5, where 1 is low and 5 is high, how likely is it that you will go to college?"*.[1] Since beliefs are only measured using an ordinal scale, we develop a multivariate ordered probit model. We show that all of the model's parameters are identified and estimate the model using a Bayesian approach, which allows for greater flexibility regarding the heterogeneity of social learning.

Estimating the model over the entire sample (16 schools), we find an average peer effect on beliefs of 12%. We also separately focus on the two largest schools in our sample and find effects of 13% and 25%. In addition, we study this heterogeneity in greater depth by then estimating a latent class model. Specifically, students are sorted into two (unobserved) groups according to the strength of peer effect on their beliefs. Estimating the model for the largest school in our sample (the school having an average effect of 13%), we observe marked differences between the two latent classes; peer influence on beliefs are respectively 8% and 73%.

To study the source of the students' heterogeneity, we try to find predictors for the probability of belonging to either class. We find that females and students having lower beliefs are slightly more likely to belong to the low-influence group. Overall, however, the predictive power is very weak. This is of great importance for policy making.

Indeed, one of the key implications of the literature on peer effects in social networks is that it allows the policy maker to target the individuals that would generate the largest spillovers (i.e. the *key player(s)*, see for example Ballester et al. (2006) and Zenou (2016)). Importantly, however, this literature mostly assumes that peer effects are homogeneous.

Since we observe (1) a high variation in peer influence among individuals, and

[1]The data also includes the students' preferences for college. See Section 3 for additional details.

that (2) this heterogeneity cannot be predicted using observables, targeted policies will be likely ineffective. That is, the most central individual in the network is not necessarily the key player when accounting for unobserved heterogeneity. Since this heterogeneity cannot be approximated by observable characteristics, and since uniform policies are relatively cheap (Hoxby and Turner, 2013), policies targeted at the level of the school are perhaps more sensible.

1.1 Related Literature

We contribute to the literature on peer effects on students' outcomes (e.g. Sacerdote et al., 2011; Black, Devereux, and Salvanes, 2013; Burke and Sass, 2013; Carrell, Fullerton, and West, 2009; Hatami, Kazemi, and Mehrabi, 2015; Zabel, 2008). As discussed by Manski (2000), interactions can affect constraints, preferences, or expectations. We estimate the impact of expectation interactions, and we show that they have a significant and heterogeneous impact on the steady-state beliefs regarding college participation.

We also contribute to the literature on peer effects in networks (see Boucher and Fortin, 2016; de Paula, 2017, for recent reviews). Our structural model is closely linked to the spatial autoregressive model. Although identification conditions are well known (e.g. Bramoullé, Djebbari, and Fortin, 2009; Lee, Liu, and Lin, 2010), few papers study the implications of such models when the outcome variable is ordered.[2] We present a Bayesian estimation procedure and show that the identification of the model is not impaired by the ordered nature of the outcome variable.

Finally, we also contribute to recent literature discussing the heterogeneity of peer effects. While some papers have assessed heterogeneous peer effects based on observable characteristics (e.g. Dieye, Fortin et al., 2017; Arduini, Patacchini, and Rainone, 2016), few papers have studied unobserved heterogeneity on peer

[2] An exception being Liu and Zhou (2017).

effects. One important exception is Peng (2016) who presented a model where the weight of the network is estimated using a LASSO technique. In our paper, the individuals belong to latent classes, grouped according to the magnitude of their peer effects. We find substantial differences among individuals' reliance on peers for belief formation.

The remainder of the paper is organized as follows. In Section 2, we introduce the model of social learning in networks, and we discuss the identification conditions and the estimation procedure. In Section 3, we present the data, while in the following Section 4, we present our results. Section 5 provides our conclusions.

2 Model

We consider a school of $n \geq 1$ students. Each student $i \in N$ is characterized by a set of observable socio-economic variables $\mathbf{x}_i$. Each student i also forms beliefs $p_i^* \in \mathbb{R}$ regarding the likelihood that they will go to college.

We assume that beliefs are formed as follows:

$$p_i^* = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where $\varepsilon_i \sim N(0,1)$. A similar specification is used for instance in Bellemare, Kröger, and Van Soest (2008). The variables in $\mathbf{x}_i$ are likely to influence the relative optimism (or pessimism) of the student regarding the likelihood that they will go to college, e.g. parents' education, college preferences, racial group, gender, etc.

Given that students at the same school are likely to discuss their beliefs, we allow the students' beliefs to evolve as a function of their peers' beliefs. Students who have no peers are called *isolated* and their beliefs are simply given by (1). For non-isolated students, we assume that the beliefs are updated according a modified DeGroot learning process (DeGroot, 1974):

$$p_i^{*(t)} = \frac{\alpha}{n_i} \sum_{j \in N_i} p_j^{*(t-1)} + (1-\alpha)\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{2}$$

where $N_i \subset N$ represent the set i's peers, i.e. the set of students from which i learns. We set $i \notin N_i$, $n_i = |N_i|$, and we assume that $\alpha \in (0, 1)$.

At any period t, a non-isolated student's belief is given by a convex combination between their type $\mathbf{x}_i\boldsymbol{\beta}$ and the average beliefs of their peers. Note that here, the fixed term $(1-\alpha)\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$ influences beliefs at every period t, which preserves the heterogeneity of beliefs at the steady state (see Equation (3) below). A similar approach is followed by Friedkin and Johnsen (1990).

Although naive, DeGroot learning processes are widely used in the context of belief updating in social networks (e.g. Golub and Jackson, 2010, 2012).[3] In particular, Chandrasekhar, Larreguy, and Xandri (2015) show, in a field experiment setting, that individuals' beliefs are better described by DeGroot learning relative to Bayesian learning.

Note that we can write the learning process in matrix form for all students at time t as:

$$\mathbf{p}^{*(t)} = \alpha\mathbf{T}\mathbf{p}^{*(t-1)} + \mathbf{B}_\alpha\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{B}_\alpha$ is a diagonal matrix taking the value $\mathbf{B}_{ii} = (1-\alpha)$ if the student is not isolated, and $\mathbf{B}_{ii} = 1$ otherwise. The matrix $\mathbf{T}$ summarizes the learning network and is such that $\mathbf{T}_{ij} = 1/n_i$ if i learns from j's beliefs (i.e. $j \in N_i$), and $\mathbf{T}_{ij} = 0$ otherwise.

Importantly, note that $\mathbf{T}$ is generally not block-diagonal; while isolated students do not learn from other students' beliefs, a non-isolated student may learn from an isolated students' beliefs.

For the remainder of this paper, we will assume that the data is generated

[3] See Golub and Sadler (2016) for a review of the learning approaches in a social network.

from the steady state of this learning process. (See the Appendix for a proof of convergence and of the invertibility of $[\mathbf{I} - \alpha\mathbf{T}]$.):

$$\mathbf{p}^* = [\mathbf{I} - \alpha\mathbf{T}]^{-1}\mathbf{B}_\alpha\mathbf{X}\boldsymbol{\beta} + [\mathbf{I} - \alpha\mathbf{T}]^{-1}\boldsymbol{\varepsilon} \tag{3}$$

A particularity of the model is that p_i^* is not observed. We assume that each student is asked to assess the likelihood that they will go to college using an ordinal scale. For student i, we therefore observe:

$$p_i = \begin{cases} 1 & \text{if } p_i^* \leq \gamma_1 \\ 2 & \text{if } \gamma_1 < p_i^* \leq \gamma_2 \\ \vdots & \\ K-1 & \text{if } \gamma_{K-2} < p_i^* \leq \gamma_{K-1} \\ K & \text{if } p_i^* > \gamma_{K-1} \end{cases}$$

where we normalize $\gamma_1 = 0$, and where $\boldsymbol{\gamma} = [\gamma_2, ..., \gamma_{K-1}]'$ must be estimated.

In the next section, we discuss the identification of α and $\boldsymbol{\beta}$. We discuss the estimation procedure in Section 2.2.

2.1 Identification

If we were to observe $\mathbf{p}^*$, identification of α and $\boldsymbol{\beta}$ would follow directly from the literature (e.g. Bramoullé et al. (2009)).[4] Here, however, we only observe the ordinal scale $\mathbf{p}$. The likelihood of $\mathbf{p}$ is therefore given by a multivariate ordered probit model.

We show that under the typical restrictions where the variances of ε_i are normalized to 1, and γ_1 is normalized to 0, the partial observability of $\mathbf{p}^*$ does not affect the identification of α and $\boldsymbol{\beta}$. To understand this reasoning, one can assume that there are no isolated individuals and that individuals interact in pairs

[4] Note that in that case, the normalization of the variance of ε_i would not be required.

(A general treatment is provided in the Appendix).

Under these simplifying assumptions, the model reduces to a simple bivariate ordered probit model:

$$\begin{aligned} p_i^* &= \alpha p_j^* + (1-\alpha)\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \\ p_j^* &= \alpha p_i^* + (1-\alpha)\mathbf{x}_j\boldsymbol{\beta} + \varepsilon_j \end{aligned}$$

It can also be rewritten as:

$$\begin{aligned} p_i^* &= \frac{\sqrt{1+\alpha^2}}{1-\alpha^2}\left[\frac{(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_i\boldsymbol{\beta} + \frac{\alpha(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_j\boldsymbol{\beta} + \nu_i\right] \\ p_j^* &= \frac{\sqrt{1+\alpha^2}}{1-\alpha^2}\left[\frac{(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_j\boldsymbol{\beta} + \frac{\alpha(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_i\boldsymbol{\beta} + \nu_j\right] \end{aligned}$$

where $\begin{bmatrix} \nu_i \\ \nu_j \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 2\alpha \\ 2\alpha & 1 \end{bmatrix}\right)$

Since we do not observe the latent variables p_i^* and p_j^*, this is observationally equivalent to:[5].

$$\begin{aligned} p_i^* &= \frac{(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_i\boldsymbol{\beta} + \frac{\alpha(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_j\boldsymbol{\beta} + \nu_i \\ p_j^* &= \frac{(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_j\boldsymbol{\beta} + \frac{\alpha(1-\alpha)}{\sqrt{1+\alpha^2}}\mathbf{x}_i\boldsymbol{\beta} + \nu_j \end{aligned}$$

Thus, α and $\boldsymbol{\beta}$ are identified. In particular, α is directly identified from the variance of $\boldsymbol{\nu}$ (e.g. Graham (2008), Rose (2017)) and also from the ratio of the coefficients for $\mathbf{x}_i$ and $\mathbf{x}_j$ (e.g. Moffitt et al. (2001), Bramoullé et al. (2009)). Notably, this implies that the identification of α *does not* follow from the normalization of the variance of ν_i.[6]

[5] This is due to the fact that we can rescale $\boldsymbol{\gamma}$

[6] As usual, $\boldsymbol{\beta}$ is only scale-identified, i.e. its identification depends on the normalization of the variance of ν_i.

2.2 Estimation

The estimation of a multinomial ordered probit model is challenging as, among other things, the maximum likelihood estimator cannot be written in close form. In this paper, we adopt a Bayesian approach, although a classical approach could also be used.[7]

One advantage of the Bayesian approach is that the estimation can be performed separately for different schools, allowing for the study of heterogeneity among schools (see Section 4). In a classical setting, the dependence between observations often requires that the estimation be performed for many independent populations (e.g. schools) to obtain consistent estimates.[8] This is important since we find large heterogenous effects between schools (see Section 4.2) and within schools (see Section 4.3).

The model parameters are estimated via an Markov chain Monte Carlo (MCMC) algorithm. Specifically, we follow Albert and Chib (1993) and augment the targeted posterior distributions with two random variables, p^* and γ (see Tanner and Wong (1987) for details). We use standard prior distributions that are set as $\tilde{\alpha} \equiv \ln \frac{\alpha}{1-\alpha} \sim N(0,4)$ and $\beta \sim N(\beta_0, \Sigma_0)$ with $\beta_0 = (0 \quad 0 \quad \ldots \quad 0)'$ and $\Sigma_0 = I_K$. The MCMC scheme is documented in Algorithm 1.

Steps [2] and [3] in Algorithm 1 are simple Gibbs steps where parameters are drawn from their conditional posterior distribution. Step [1] is a Metropolis–Hastings step where parameters α and $\boldsymbol{\beta}$ are both drawn before the accept/reject phase, which improves convergence.[9]

[7] For instance, using a GHK (see Geweke, Keane, and Runkle (1994)) algorithm.

[8] This is only a sufficient condition, see assumptions 2 and 3, and the discussion on page 1903 in Lee (2004). An alternative is to endogenize the interaction matrix $\mathbf{T}$ and invoke limited spatial dependence arguments, as in Qu and Lee (2015). In either case, the study of heterogeneity is not straightforward.

[9] An alternative would be to update $\boldsymbol{\beta}$ through a Gibbs step and then draw α using a Metropolis–Hastings step.

Algorithm 1 MCMC algorithm

Set initial values to $\boldsymbol{\beta}_0, \alpha_0, p_0^*, \gamma_0$

for $j = 1$ to M, where M denotes the number of MCMC iterations **do**

[1] - Sample $\boldsymbol{\beta}, \alpha \sim \boldsymbol{\beta}, \alpha | p^*, p, \gamma$:

- Draw $\tilde{\alpha}^* \sim N\left(\tilde{\alpha}_{j-1}, \bar{\sigma}_{j-1}^2\right)$ in which the scale $\bar{\sigma}_{j-1}^2$ is adapted over the MCMC iterations using the approach of Atchadé and Rosenthal (2005) and set $\alpha^* = e^{\tilde{\alpha}^*}/(1 + e^{\tilde{\alpha}^*})$.
- Sample $\beta^* | \alpha^* \sim N(\bar{\boldsymbol{\beta}}_{\alpha^*}, \boldsymbol{\Sigma}_{\alpha^*})$, where $\bar{\boldsymbol{\beta}}_{\alpha^*} = \boldsymbol{\Sigma}^{-1}[\mathbf{X}'\mathbf{B}_{\alpha^*}[\mathbf{I} - \alpha^*\mathbf{T}]\mathbf{p}^* + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0]$, and $\boldsymbol{\Sigma}_{\alpha^*} = [\boldsymbol{\Sigma}_0^{-1} + \mathbf{X}'\mathbf{B}_{\alpha^*}^2\mathbf{X}]^{-1}$.
- Accept or reject the draw $\boldsymbol{\beta}^*, \alpha^*$ according to the Metropolis-Hastings ratio,

$$\chi_{MH}(\boldsymbol{\beta}^*, \alpha^* | \bar{\boldsymbol{\beta}}_{j-1}, \alpha_{j-1}) = \text{Max}\{\frac{f(p^*|\alpha^*)f(\alpha^*)}{f(p^*|\alpha_{j-1})f(\alpha_{j-1})}, 1\},$$

  where $f(\alpha)$ is the prior density function of α, and $f(p^*|\alpha) = \frac{f(p^*|\alpha, \bar{\boldsymbol{\beta}}_\alpha)f(\bar{\boldsymbol{\beta}}_\alpha)}{f(\bar{\boldsymbol{\beta}}_\alpha|p^*, \alpha)}$.

[2] - Sample $\gamma \sim \gamma | p^*, p, \boldsymbol{\beta}, \alpha$:

- For $k = 2, ..., K - 1$, draw $\gamma_k | \gamma_{k-1}, \gamma_{k+1} \sim U[\underline{u}, \bar{u}]$, where $\underline{u} = \max\{\max_i \{p_i^* : p_i = k\}, \gamma_{k-1}\}$, and $\bar{u} = \min\{\min_i \{p_i^* : p_i = k+1\}, \gamma_{k+1}\}$.

[3] - Sample $p^* | p, \gamma, \boldsymbol{\beta}, \alpha$:

- For all i, draw p_i^* from a truncated normal distribution having a unit variance and a mean equal to $\mathbf{x}_i\boldsymbol{\beta}$ if i is isolated, and a mean $\alpha\mathbf{T}_i\mathbf{p}^* + (1-\alpha)\mathbf{x}_i\boldsymbol{\beta}$ if i is non-isolated (recall that $\mathbf{T}_{ii} = 0$). For $p_i = k$, the left truncation is γ_{k-1} and the right truncation is γ_k.

end for

3 Data

We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). The study began in 1994 (Wave 1) with a representative sample of students from grades 7 to 12. For our main purpose, we use data from the first wave to evaluate peer effects on the students' beliefs regarding the likelihood that they will go to college.[10] We also use the data from Wave 2 for our robustness analysis in sections 4.4 and 6.3.

In addition to the information about the students' beliefs, the survey contains information vis-à-vis the students' preferences regarding college, as well as many relevant socio-economic and demographic characteristics. In particular, students are asked how much they *want* to go to college as well as how *likely* they think this will happen. Although the two questions are clearly correlated, there is substantial variation, which allows us to capture the role of beliefs, while controlling for preferences.[11]

For a subset of 16 schools, the survey contains a full description of the students' friendship network. In the context of this paper, we assume that student i learns from student j if, and only if, i nominated j as a friend. Table 1 presents the summary statistics.

4 Results

In this section, we present the results of the estimation for the entire sample as well as for a subset of two relatively large schools. We begin by discussing the interpretation of the parameter α: essentially, the identification of social learning using model (3).

[10] Estimations are robust for the use of data of the second wave (1995) for which a similar set of variables is available.

[11] See Table 8 of the Appendix.

Table 1: Summary statistics

	Mean	Std. Dev.	Min	Max
Educational expectations				
Belief (ordinal)	4.05	1.14	1	5
College preference (ordinal)	4.37	1.05	1	5
Individual characteristics				
Age	15.01	1.41	11	19
Female	0.49	0.50	0	1
Grades 7–8	0.22	0.41	0	1
Grades 9–10	0.48	0.50	0	1
Grades 11–12	0.30	0.46	0	1
White	0.61	0.48	0	1
Black or African–American	0.15	0.36	0	1
Hispanic	0.18	0.38	0	1
Mother education				
Less than high school	0.07	0.26	0	1
High school	0.24	0.43	0	1
More than high school	0.29	0.45	0	1
Father education				
Less than high school	0.07	0.25	0	1
High school	0.18	0.38	0	1
More than high school	0.25	0.43	0	1
Household characteristics				
Household size	4.52	1.15	1	6
Network statistics				
Average number of friends	3.33	2.75	0	10
Number of female friends	1.68	1.59	0	5
Number of male friends	1.65	1.60	0	5
Isolated students		595		
Number of students		2,255		
Number of schools		16		

Notes: The variables "Belief" and "Preference" are both on an ordinal scale. The specific questions read: "On a scale of 1 to 5, where 1 is low and 5 is high, how likely is it that you will go to college?" and "On a scale of 1 to 5, where 1 is low and 5 is high, how much do you want to go to college?".

4.1 Interpretation of α

Although α is identified from the structural econometric model, an important concern is that it may not necessarily capture social learning (e.g. Angrist (2014) or Kline and Tamer (2014)). In fact, as for most of the existing literature, social interaction effects (here α) are picked up as residual correlations between outcomes (here, beliefs) after controlling for other potential sources of correlation.

Perhaps the most important source of correlation is the presence of common unobserved shocks.[12] The beliefs of students i and j may be correlated because they share a common (unobserved) source of information; for example, whether

[12] Also called "correlated effects" (Manski, 1993).

or not a career counselor is present in the school. In this paper, we control for these unobserved shocks using school-level fixed effects.

Another common source of correlation is introduced by the endogeneity of the network structure. In particular, α can not only capture social learning, but it can also capture the correlation between beliefs due to self-selection into the friendship network. For example, students i and j may become friends *because* they share common beliefs. If this point is conceptually valid, we are not worried about this issue in our context.

Indeed, multiple studies have shown repeatedly that the impact of self-selection is virtually nonexistent in the context of the particular database that we use (e.g. Goldsmith-Pinkham and Imbens, 2013; Boucher, 2016; Hsieh and Lee, 2016; Boucher and Fortin, 2016). A likely explanation is the extremely weak predictability of friendship relations. As presented in detail in Appendix 6.3, the marginal effects of most identified explanatory variables is very weak (on the order of 10^{-5} for the probability of creating a friendship relation). This is expected as friendship relations are rare events: for any two randomly selected individuals, the probability of friendship is very small.

The only variable that has some meaningful predictive power is the preexistence of a friendship relation. For the entire sample of 16 schools, the existence of a friendship relation in Wave 1 only increases the probability of friendship in Wave 2 by less than 2% (see Table 6 of Appendix 6.3). However, it may have some predictive power when we restrict the sample to specific schools. For example, for School 28, having a friendship relation in Wave 1 increases the probability of a friendship relation in Wave 2 by 42% (see Table 7 of Appendix 6.3). Regardless, this analysis suggests that any unobserved predictor of friendship relations is likely to have a very weak bias on the estimated value of α (see below for additional verifications of robustness).

A final source of correlation is due to omitted individual variables. Here, an important comment should be made: we make no causal claim for the estimated parameters $\boldsymbol{\beta}$. Indeed, many variables included in $\mathbf{X}$ are likely endogeneous. For example, unobserved effort is likely to be correlated with the students' beliefs, academic achievements, or college preferences. That being said, including academic achievement and college preference variables is nonetheless crucial to the interpretation of α as a social learning effect.

Many studies have shown the presence of peer effects on academic achievement (e.g. Sacerdote et al., 2011) or preferences for college (e.g. Giorgi, Pellizzari, and Redaelli, 2010). Thus, if we were to exclude achievement and preference variables, the within-network correlation of beliefs, picked up by α, would not only capture social learning, but it would also capture peer effects for academic achievement and college preferences.

Regardless, we also present, as a robustness check in Section 4.4, a version of the model estimated in deviations, which account for unobserved individual fixed effects. The fact that we obtain very similar results therefore reduces the likelihood that the measured value for α is due to unobserved common shocks, endogenous friendship selection, or missing individual variables.

A final concern for the interpretation of α comes from a potential mis-specification of the theoretical model. Although this is always a possibility, previous studies have preferred models similar to (3) to the Bayesian alternative (Chandrasekhar et al., 2015). In addition, a particular feature of social learning is that beliefs can be thought as being directly transmitted from a student to their peers (e.g. by talking), as opposed to other outcomes, such as body mass index (BMI). Indeed, since a student's BMI does not directly affect their peers' BMI, the social transmission necessarily goes through the students' (unobserved) effort, which also has to be modeled to interpret α as capturing social interactions (e.g. Boucher and Fortin, 2015).

4.2 Results and Heterogeneity Across Schools

We first perform the estimation for the entire sample (i.e. for the 16 schools). Results are presented in Table 2 and the histograms for all posterior distributions are presented in Figure 3 of Appendix 6.5. We find that, on average, the beliefs of a student's peers account for about 12% of the updating of their beliefs. Female students have a more positive attitude toward college, as do students having educated parents.

Table 2: Estimation of homogeneous peer effects

	Mean	Std. Dev.
Endogenous effect α	0.121***	0.026
Child's characteristics		
Female	0.292***	0.055
Grades 9–10	0.189	0.129
Grades 11–12	0.328**	0.165
Age	-0.052	0.038
White	-0.090	0.095
Black or African–American	-0.028	0.106
Hispanic	-0.079	0.089
Mother education		
Less than high school	-0.100	0.121
High school	0.073	0.089
More than high school	0.210***	0.092
Father education		
Less than high school	-0.018	0.120
High school	-0.040	0.088
More than high school	0.276***	0.085
Household characteristics		
Household size	-0.017	0.016
Urban	0.287	1.350
Constant	0.260	1.854
γ_2	0.855***	0.053
γ_3	2.047***	0.047
γ_4	2.961***	0.044
College preferences	yes	
Grade in mathematics, English, history, and science	yes	
School fixed effect	yes	
Observations	2,255	

Note: We let our chain run for 20,000 iterations, discarding the first 9,999 iterations; $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

To explore the heterogeneity of social learning, we also ran the model, separately, for the two largest schools in our sample. Figure 1 presents a kernel density estimate of the posterior distribution of α for both schools. The full results are presented in Table 9 in Appendix 6.4.

We see that social learning is more important in School 58. This is interesting

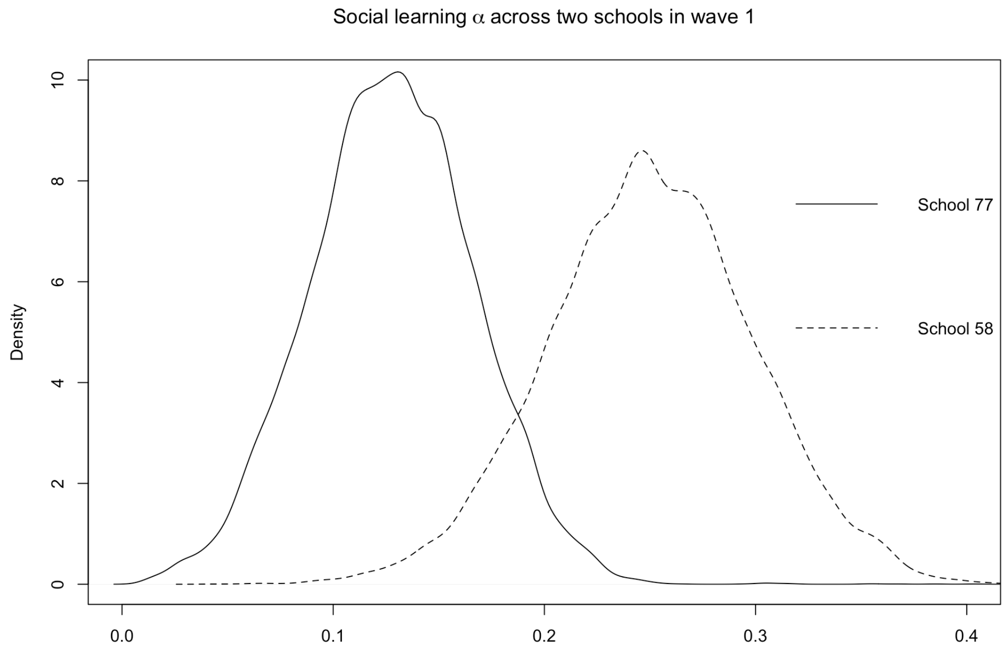


Figure 1: Heterogeneity among schools

as the distribution of beliefs is similar for both schools (see Figure 2 of Appendix 6.4). Thus, from a policy point of view, information campaigns are likely to be more effective in School 58 since beliefs are more dependent on social learning and less on factors that are difficult to influence through policies (e.g. parents' education).

4.3 Individual Heterogeneity

In this section, we explore in greater detail the role of heterogeneity in social learning. In the previous section, we found that the weight associated to friends' beliefs can vary among schools. Here, we assume that this weight can vary among students. Specifically, we assume that the learning process of student i is as follows:

$$p_i^{*(t)} = \frac{\alpha^{(i)}}{n_i} \sum_{j \in N_i} p_j^{*(t-1)} + (1 - \alpha^{(i)})\mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$$

where $\alpha^{(i)} = 0$ if i is isolated, and where $\alpha^{(i)} \in \boldsymbol{\alpha} \equiv \{\alpha_1, \alpha_2\}$ if i is not isolated. We therefore assume that non-isolated students can be classified into two (unobserved) groups. Using the same argument as before, we can show that the steady-state beliefs are given by:

$$\mathbf{p}^* = [\mathbf{I} - \mathbf{A}_{\boldsymbol{\alpha}}\mathbf{T}]^{-1} [\mathbf{I} - \mathbf{A}_{\boldsymbol{\alpha}}] \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} - \mathbf{A}_{\boldsymbol{\alpha}}\mathbf{T}]^{-1} \boldsymbol{\varepsilon}, \tag{4}$$

where $\mathbf{A}_{\boldsymbol{\alpha}}$ is a diagonal matrix taking $A_{ii} = \alpha^{(i)}$ for all i. One can easily verify that (3) is a special case of (4).[13]

To estimate the parameters (and to divide the students into two groups), we augment the model with a latent component $S = \{s_1, s_2, \ldots, s_N\}$. The variable $s_i \in \{1, 2\}$ represents the group to which the student i belongs and exhibits a prior distribution given by $P[s_i = 1] = q$. Given this new state variable, the model for non-isolated students can be restated as:

$$p_i^* | \mathbf{p}_{-i}^*, s_i, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim N(\alpha_{s_i}\mathbf{T}\mathbf{p}^* + (1 - \alpha_{s_i})\mathbf{x}_i\boldsymbol{\beta}, 1). \tag{5}$$

Equation (5) highlights that sampling the state is straightforward since, for $k \in \{1, 2\}$, the conditional probability function of the variable is given by $P(s_i = k | \mathbf{p}^*, \boldsymbol{\beta}, \boldsymbol{\alpha}) \propto f(p_i^* | \mathbf{p}_{-i}^*, s_i = k, \boldsymbol{\beta}, \boldsymbol{\alpha}) P[s_i = k]$. The MCMC steps used to estimate the model parameters are presented in Algorithm 2.[14]

Due to the computational burden of Algorithm 2, we only present results for School 77 (the largest school in our sample). Our results can be found in Table 3.

The difference between the two groups is striking. For some students (Group 1), social learning is small (around 8%). However, for other students (Group 2), social learning is very important (close to 80%). The composition of those two groups, as well as the strength of social learning within each group, is determined

[13]In a classical setting, this model would be called a finite mixture (or latent class) model and estimated using an EM (expectation-maximization) algorithm.

[14]Alternatively, see Algorithm 3 in Appendix 6.4

Algorithm 2 MCMC algorithm

Set initial values to $\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, p_0^*, \gamma_0, S_0, q_0$

for $j = 1$ to M, where M denotes the number of MCMC iterations **do**

[1] - Same as in Algorithm 1 but also conditional on S.

[2] - Same as in Algorithm 1.

[3] - Same as in Algorithm 1.

[4] - Sample $S \sim S|p^*, p, \gamma, \boldsymbol{\beta}, \boldsymbol{\alpha}$:

for $i = 1$ to N **do**

- Draw $u \sim U[0,1]$ and fix $s_i = 1$ if $u \leq P(s_i = 1|\mathbf{p}^*, \boldsymbol{\beta}, \boldsymbol{\alpha})$. Set $s_i = 2$ otherwise. Note that $P(s_i = 1|\mathbf{p}^*, \boldsymbol{\beta}, \boldsymbol{\alpha}) \propto f_N(p_i^*|\alpha_1 \mathbf{T}\mathbf{p}^* + (1-\alpha_1)\mathbf{x}_i\boldsymbol{\beta}, 1)q$, where $f_N(x|\mu, \sigma^2)$ stands for the normal density function with expectation μ and variance σ^2 evaluated at x.

end for

[5] - Sample $q|S \sim \text{Beta}(1 + n_1, 1 + n_2)$, where $n_k = \sum_{i=1}^{N} \mathbb{1}_{\{s_i = k\}}$.

end for

entirely by the data. Therefore, this method allows us to capture the unobserved heterogeneity in social learning.

The estimation technique also provides, for each student, an estimate of the probability of belonging to one of the two groups.[15] We can therefore search for predictors of this probability. In Table 4, we present the results of a (censored) regression for the probability of belonging to Group 1.

First, note that there are few significant predictors. This is important as it means that the strength of social learning cannot be directly predicted by observable variables in our sample. The sorting of the students between the two groups therefore truly captures unobserved heterogeneity.

That being said, one of the variables that appears to predict group membership is the students' beliefs. All else being equal, students having higher beliefs are less likely to belong to Group 1 and are therefore more likely to rely on social learning when forming their beliefs. This is somewhat intuitive. More pessimistic students (those having low beliefs, everything else being equal) are less likely to discuss their beliefs with their friends and therefore to update their beliefs as a

[15] Using the average value of s_i among draws from the posterior distribution.

Table 3: Heterogeneous peer effects in Wave 1 (School 77)

	Mean	**Std. Dev.**
Endogenous peer effects		
α_1	0.079**	0.038
α_2	0.735***	0.185
Child's characteristics		
Female	0.340***	0.084
Grades 9–10	-0.331	0.465
Grades 11–12	-0.439	0.491
Age	-0.016	0.053
Born in the USA	0.064	0.117
White	-0.153	0.106
Black or African–American	0.042	0.122
Hispanic	-0.109	0.111
Mother education		
Less than high school	0.087	0.179
High school	0.041	0.141
More than high school	0.199	0.139
Father education		
Less than high school	-0.140	0.170
High school	0.020	0.162
More than high school	0.139	0.130
Household characteristics		
Household size	-0.016	0.022
Constant	0.318	0.762
γ_2	0.813***	0.075
γ_3	2.136***	0.057
γ_4	3.004***	0.060
College preferences	yes	
Grade in mathematics, English, history and science	yes	
Observations	920	

Note: We let our chain run for $20,000$ iterations, discarding the first $11,999$ iterations; $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

result of social interactions.

Table 4: Predictors of Group 1

	Est.	Std. Dev.
Individual characteristics		
Intercept	0.126	0.354
Female	**0.048***	0.023
Grades 11–12	-0.043	0.032
Born in the USA	0.057	0.031
Age	0.036	0.022
White	-0.042	0.030
Black or African–American	0.002	0.035
Hispanic	-0.002	0.029
College preferences 3	0.010	0.067
College preferences 4–5	-0.064	0.065
Belief 3	**−0.118***	0.055
Belief 4–5	-0.095	0.054
Network structure		
Betweenness centrality	2.739	3.030
Mother education		
Less than high school	0.021	0.048
High school	-0.061	0.040
More than high school	-0.063	0.036
Father education		
Less than high school	-0.057	0.047
High school	0.004	0.044
More than high school	-0.001	0.034
Household characteristics		
Household size	0.010	0.006
Observations	630	

The importance of unobserved heterogeneity has powerful policy implications. The results of Table 4 imply that the heterogeneity of peer effects captured using our structural model (which is computationally and data intensive) cannot be easily predicted using observable characteristics.

This is important as the literature widely promotes the key player principle: policies should target the most central individual(s) in the network (often using Bonacich's centrality, see Ballester et al. (2006) or Zenou (2016) for details). However, when peer effects are heterogeneous, those centrality measures must be adjusted using a weighted network analysis. If the policy maker cannot easily capture the (strong) heterogeneity, their policies will be targeted toward the wrong individuals.

Thus, in the context of strong unobserved heterogeneity, and where large-base information policies are relatively cheap (Hoxby and Turner, 2013), we argue that policies not targeted toward networks should be privileged.[16] Moreover, the fact that we capture strong social learning is also an indication that those policies are likely to have a significant impact as they may act as a substitute to friends' subjective beliefs.

4.4 Robustness: Unobserved Fixed Effect

As described in Section 3, we observed students' beliefs for two consecutive years (waves 1 and 2). In this section, we use this feature of the data to account for the possibility of unobserved individual fixed effects. Specifically, we extend (3) as follows:

$$\mathbf{p}^{w*} = [\mathbf{I} - \alpha\mathbf{T}^w]^{-1}\mathbf{B}_\alpha[\mathbf{X}\boldsymbol{\beta}^w + \boldsymbol{\mu}\delta^w] + [\mathbf{I} - \alpha\mathbf{T}^w]^{-1}\boldsymbol{\varepsilon}^w \tag{6}$$

where $\boldsymbol{\mu}$ is unobserved and where the parameters $\boldsymbol{\beta}^w$ and δ^w are allowed to vary for each wave, $w = 1, 2$. We also normalize $\delta^1 = 1$.

Isolating $\boldsymbol{\mu}$ for $w = 1$ and then substituting, we easily find:

$$\mathbf{p}^{2*} = [\mathbf{I} - \alpha\mathbf{T}^2]^{-1}[\mathbf{B}_\alpha\mathbf{X}(\boldsymbol{\beta}^2 - \delta^2\boldsymbol{\beta}^1) + \delta^2\mathbf{p}^{1*} - \delta^2\alpha\mathbf{T}^1\mathbf{p}^{1*}] + [\mathbf{I} - \alpha\mathbf{T}^2]^{-1}\tilde{\boldsymbol{\varepsilon}}$$

where $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}^2 - \delta^2\boldsymbol{\varepsilon}^1$. Here, as the latent variable $\mathbf{p}^{1*}$ is not observed, we use the mean of the posterior distribution from Section 4.2 as a proxy.[17]

The estimated values are presented in Table 5. Results for α are remarkably close to the baseline estimates to add further credibility to the interpretation of α as social learning. Also, the small estimated effect for δ^2 implies that the contribution of the previous year's beliefs is small, which increases the probability that the learning process effectively converged in the data.

[16] Note that this argument holds even if the "true" number of latent classes is greater than two. Indeed, a sufficient condition for our argument is that we find some strong unobserved heterogeneity, not that we measure it perfectly.

[17] An alternative would be to estimate the models jointly. However, the computational burden makes this approach impractical.

Table 5: Contribution of unobserved fixed effects for School 77 in Wave 1.

	Mean	Std. Dev.
Endogenous effect α	0.113***	0.033
Unobserved fixed effect δ^2	0.068***	0.026
Child's characteristics		
Female	0.406	0.087
Grades 9–10	-0.336	0.354
Grades 11–12	-0.127	0.303
Age	0.021	0.040
Born in the USA	-0.127	0.114
White	-0.030	0.110
Black or African–American	0.229*	0.122
Hispanic	0.121	0.108
Mother education		
Less than high school	-0.087	0.186
High school	-0.223	0.145
More than high school	0.127	0.137
Father education		
Less than high school	0.103	0.169
High school	0.327**	0.162
More than high school	0.298**	0.133
Household characteristics		
Household size	-0.021	0.021
Constant	-0.902	0.751
γ_2	0.798***	0.109
γ_3	2.088***	0.156
γ_4	2.883***	0.156
College preferences	yes	
Grade in mathematics, English, history, and science	yes	
School fixed effect	yes	
Observations	920	

Note: We let our chain run for 20,000 iterations, discarding the first 9,999 iterations; $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

5 Conclusion

In this paper, we study peer effects on students' beliefs regarding college participation. To do so, we specify a structural model in which the students' beliefs are updated with respect to their friends' beliefs and other socio-economic characteristics. The steady state represents a multivariate ordered probit that is estimated using a Bayesian approach. Given that few papers take the discreetness of the outcome into account, we discuss identification issues and show that our model parameters can be identified. In addition, we also discuss the interpretations of parameters, and we estimate multiple models to ensure the robustness of our results.

When the 16 schools are taken together, we find a peer effect on beliefs that amounts to 12% on average. However, we highlight that the average effect may be meaningless since we find much network heterogeneity. If we estimate, separately, the same model for the two largest schools, we observe peer effects that reach 13% and 25%, respectively. While this result emphasizes some heterogeneity among the different schools, we also explore the unobserved heterogeneity *within* schools. Importantly, we find two groups of students. On one hand, we have students that have strong beliefs and that are hardly influenced by others. On the other hand, we also uncover a group that exhibits a very strong social learning effect (up to 73%). This finding, coupled with individual socio-economic characteristics being unable to explain group affiliation, leads us to believe that policies targeting *key player(s)* are likely ineffective. In particular, before finding the *key player(s)*, one must first understand the heterogeneity of the network and then target those who are sensitive to peers' beliefs so as to adapt the key player's argument.

As a final note, we see this work as a first step toward a better understanding of network heterogeneity. One important research avenue lies in capturing the variable that can predict the level of students' social learning.

6 Appendix

6.1 Convergence

Define the mapping:

$$F(\mathbf{p}^*) = \alpha \mathbf{T}\mathbf{p}^* + \mathbf{B}_\alpha \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

We have:

$$\|F(\mathbf{p}^*) - F(\tilde{\mathbf{p}}^*)\| \leq \alpha \|\mathbf{T}\| \|\mathbf{p}^* - \tilde{\mathbf{p}}^*\|$$

for any submultiplicative norm $\| \cdot \|$. Using the sup norm for instance, we have $\|\mathbf{T}\| \leq 1$ (it equals 1 whenever there is at least one non-isolated student). This implies that F is a contraction mapping. Therefore, there exists a unique $\mathbf{p}^*$, such that $\mathbf{p}^* = F(\mathbf{p}^*)$, which implies that the matrix $[\mathbf{I} - \alpha \mathbf{T}]$ solving the linear system is invertible.

6.2 Identification

In general, without isolated individuals (isolated individuals can only help identification), we then have:

$$\mathbf{p}^* = (1-\alpha)[\mathbf{I} - \alpha \mathbf{T}]^{-1} \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} - \alpha \mathbf{T}]^{-1} \boldsymbol{\varepsilon}$$

Letting $\mathbf{A}_\alpha$ be a diagonal matrix with the diagonal elements of $[\mathbf{I} - \alpha \mathbf{T}]^{-1}$, we then have:

$$\mathbf{p}^* = \mathbf{A}_\alpha [(1-\alpha)\mathbf{A}_\alpha^{-1} [\mathbf{I} - \alpha \mathbf{T}]^{-1} \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}]$$

where $\boldsymbol{\nu}$ is normally distributed with a mean of 0 and a variance-covariance matrix in correlation form (normalized variances): $\mathbf{A}_\alpha^{-1} [\mathbf{I} - \alpha \mathbf{T}]^{-1} [\mathbf{I} - \alpha \mathbf{T}']^{-1} \mathbf{A}_\alpha^{-1}$. The model is therefore identified whenever $\mathbf{A}_\alpha^{-1} [\mathbf{I} - \alpha \mathbf{T}]^{-1} \neq a\mathbf{I}$ for some $a > 0$, which is equivalent to saying that $\mathbf{T}$ is not a diagonal matrix; this is true by assumption.

6.3 Predictability of Friendship

We now present a simple probit model for the formation of friendship relations. The second column of Table 6 presents the marginal effects for the first wave. The third column presents the marginal effects for the second wave, adding the network structure in Wave 1 as an explanatory variable. We see that the explanatory power is extremely weak. Indeed, no variable has the potential of meaningfully influencing the probability of linking. In particular, even the status of the network in Wave 1 has little influence on the status of the network in Wave 2. Similar results are also found for more sophisticated models of network formation (e.g. Boucher (2016)).

The fact that most estimates are statistically significant is due to the extremely large number of observations. We therefore also present results for School 28 in Table 7. There, the loss in statistical significance is substantial, even if the number of observations is still large (close to 12,000). However, previous links have a significant explanatory power: being a friend in Wave 1 increases the probability of being a friend in Wave 2 by 42%.

Table 6: Network formation probit-marginal effects for all schools ($\times 10^{-4}$).

	$g_{ij,1}$		$g_{ij,2}$	
Variables	(1)	(2)	(3)	(4)
Link at wave I ($g_{ij,1}$)	-	-	137.67***	70.54***
Δ Female	−0.2745***	−0.1040***	−0.3018***	−0.1365***
Δ Grade	−1.0820***	−0.2113***	−0.7093***	−0.1285***
Δ Age	−0.1650***	−0.0221	−0.1032***	−0.0028
Δ Belief	−0.2356***	−0.0982***	−0.1159***	−0.0601***
Δ Math	−0.0005	0.0021	−0.0183	−0.0168**
Δ English	0.0068	0.0050	−0.0106	−0.0040
Δ History	−0.0259***	−0.0071*	−0.0052	0.0017
Δ Science	−0.0077	−0.0055*	0.0116*	0.0097***
Δ College preferences	−0.0325	−0.0250**	−0.0468	−0.0289*
Δ Father education	−0.0022	−0.0105	0.1098	0.0418
Δ Mother education	−0.1861	−0.1302**	−0.1684	−0.1259**
Δ Household size	−0.0759***	−0.0292***	−0.0876***	−0.0398***
Δ White	−1.2701***	−0.2993***	−1.0550***	−0.2784***
Δ Black	−1.3903***	−0.5386***	−1.0403***	−0.4715***
Δ Hispanic	−1.1540***	−0.4904***	−1.0920***	−0.5305***
Δ Urban	−7.3520***	−0.0488	−5.1057***	−0.0692
School fixed effect	no	yes	no	yes
Log likelihood	−33,349.402	−30,921.413	−25,305.575	−23,770.405
Pseudo R^2	0.145	0.207	0.354	0.394
Observations		**5,082,770**		

Note: *** p<0.01, ** p<0.05, * p<0.1. For any variable z, $\Delta z_{ij} = |z_i - z_j|$. Columns (1) and (3) investigate the network formation without a fixed effect. In both columns (2) and (4), we control for the network (school) fixed effect.

Table 7: Network formation probit-marginal effects in School 28 ($\times 10^{-4}$).

Variables	$T_{ij,1}$	$T_{ij,2}$
Link at wave I ($T_{ij,1}$)	-	$4,253.356^{***}$
Δ Belief	-2.682	13.993
Δ Female	-40.262^{*}	-16.394
Δ Grade	-52.629^{***}	-38.945^{**}
Δ Age	-9.267	-7.821
Δ White	-83.974^{*}	-90.938^{*}
Δ Black	-27.764	-29.234
Δ Hispanic	-86.913^{***}	-33.217
Δ College preferences	-25.881	17.338
Δ Father education	-0.213	-0.738^{**}
Δ Mother education	0.944^{**}	-0.560
Δ Household size	-5.785	2.083
Δ Math	5.472	3.370
Δ English	-16.523	-1.989
Δ History	-0.266	-12.382
Δ Science	1.059	-11.193
Log likelihood	$-$ 999.642	$-$ 992.153
Pseudo R^2	0.054	0.242
Observations	**11,990**	

6.4 Additional Material

Table 8: Students' beliefs and preferences (Wave 1)

	Preferences					
Beliefs	1	2	3	4	5	Total
1	32.87	12.48	22.18	9.70	22.77	100
2	12.28	14.04	29.61	19.96	24.12	100
3	5.85	6.50	23.34	23.13	41.18	100
4	2.64	2.59	12.28	20.01	62.48	100
5	1.43	1.20	3.61	8.25	85.50	100
Total	4.48	3.46	10.59	13.67	67.80	100

The detailed MCMC scheme to estimate model (5) is given in Algorithm 3.

Algorithm 3 MCMC algorithm

Set initial values to $\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, p_0^*, \gamma_0, S_0, q_0$

for $j = 1$ to M, where M denotes the number of MCMC iterations **do**

[1] - Sample $\boldsymbol{\beta}, \boldsymbol{\alpha} \sim \boldsymbol{\beta}, \boldsymbol{\alpha}|p^*, p, \gamma, S$:

- Draw $\tilde{\boldsymbol{\alpha}}^* \sim N\left(\tilde{\boldsymbol{\alpha}}_{j-1}, c_{j-1}\boldsymbol{\Sigma}_{j-1}\right)$ in which the scale c_{j-1} is adapted over the MCMC iterations using the approach of Atchadé and Rosenthal (2005). The covariance matrix $\boldsymbol{\Sigma}_{j-1}$ represents the empirical covariance matrix of $\tilde{\boldsymbol{\alpha}}$ realizations up to iteration $j-1$ (updated only up to $j = \lfloor 0.3M \rfloor$). For $i = \{1, 2\}$, set $\alpha_i^* = e^{\tilde{\alpha}_i^*}/(1 + e^{\tilde{\alpha}_i^*})$.
- Set $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}^*}^{-1} = [I - A_{\boldsymbol{\alpha}^*}T]'[I - A_{\boldsymbol{\alpha}^*}T]$, and $\tilde{\mathbf{X}}_{\boldsymbol{\alpha}^*} = [I - A_{\boldsymbol{\alpha}^*}T]^{-1}[I - A_{\boldsymbol{\alpha}^*}]\mathbf{X}$. Sample $\beta^*|\boldsymbol{\alpha}^* \sim N(\bar{\boldsymbol{\beta}}_{\boldsymbol{\alpha}^*}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}^*})$, where $\bar{\boldsymbol{\beta}}_{\boldsymbol{\alpha}^*} = \boldsymbol{\Sigma}^{-1}[\tilde{\mathbf{X}}_{\boldsymbol{\alpha}^*}' \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}^*}^{-1}\mathbf{p}^* + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0]$, and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}^*} = [\boldsymbol{\Sigma}_0^{-1} + \tilde{\mathbf{X}}_{\boldsymbol{\alpha}^*}' \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}^*}^{-1}\mathbf{X}_{\boldsymbol{\alpha}^*}]^{-1}$.
- Accept or reject the draw $\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*$ according to the Metropolis–Hastings ratio,

$$\chi_{MH}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*|\boldsymbol{\beta}_{j-1}, \boldsymbol{\alpha}_{j-1}) \quad = \quad \text{Max}\{\frac{f(p^*|\boldsymbol{\alpha}^*, S_{j-1})f(\boldsymbol{\alpha}^*)}{f(p^*|\boldsymbol{\alpha}_{j-1}, S_{j-1})f(\boldsymbol{\alpha}_{j-1})}, 1\},$$

  in which the function $f(\boldsymbol{\alpha}^*)$ represents the prior density function of α, and $f(p^*|\boldsymbol{\alpha}, S) = \frac{f(p^*|\boldsymbol{\alpha}, \bar{\boldsymbol{\beta}}_{\boldsymbol{\alpha}}, S)f(\bar{\boldsymbol{\beta}}_{\boldsymbol{\alpha}})}{f(\bar{\boldsymbol{\beta}}_{\boldsymbol{\alpha}}|p^*, \alpha, S)}$.

[2] - Sample $S \sim S|p^*, p, \gamma, \boldsymbol{\beta}, \boldsymbol{\alpha}$:

for $i = 1$ to N **do**

- Draw $u \sim U[0, 1]$ and fix $s_i = 1$, if $u \le P(s_i = 1|\mathbf{p}^*, \boldsymbol{\beta}, \boldsymbol{\alpha})$. Otherwise, set $s_i = 2$. Note that $P(s_i = 1|\mathbf{p}^*, \boldsymbol{\beta}, \boldsymbol{\alpha}) \propto f_N(p_i^*|\alpha_{s_i}\mathbf{T}\mathbf{p}^* + (1 - \alpha_{s_i})\mathbf{x}_i\boldsymbol{\beta}, 1)q$, where $f_N(x|\mu, \sigma^2)$ represents the normal density function with expectation μ and variance σ^2 evaluated at x.

end for

[3] - Sample $q|S \sim \text{Beta}(1 + n_1, 1 + n_2)$, where $n_k = \sum_{i=1}^N \mathbb{1}_{\{s_i = k\}}$.

[4] - Sample $\gamma \sim \gamma|p^*, p, \boldsymbol{\beta}, \alpha$:

for $k = 1$ to $K - 1$ **do**

- Draw $\gamma_k|\gamma_{k-1}, \gamma_{k+1} \sim U[\underline{u}, \bar{u}]$, where $\underline{u} = \max\left\{\max_i \left\{p_i^* : p_i = k\right\}, \gamma_{k-1}\right\}$, and $\bar{u} = \min\left\{\min_i \left\{p_i^* : p_i = k + 1\right\}, \gamma_{k+1}\right\}$.

end for

[5] - Sample $p^*|p, \gamma, \boldsymbol{\beta}, \alpha$:

for $i = 1$ to N **do**

- Draw p_i^* from a truncated normal distribution having a unit variance and a mean equal to $\mathbf{x}_i\boldsymbol{\beta}$ if i is isolated, and a mean $\alpha_{s_i}\mathbf{T}_i\mathbf{p}^* + (1 - \alpha_{s_i})\mathbf{x}_i\boldsymbol{\beta}$ if i is non-isolated. Recall that $\mathbf{T}_{ii} = 0$. For $p_i = k$, the left truncation is γ_{k-1} and the right truncation is γ_k.

end for

end for

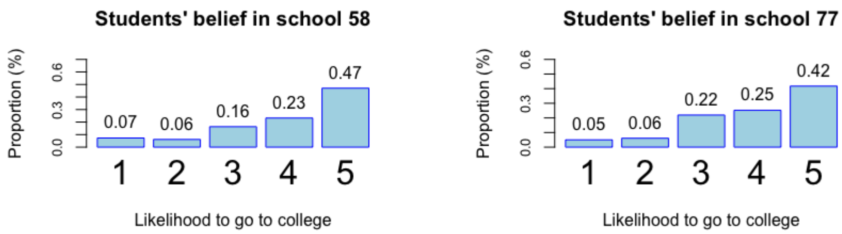


Figure 2: Distribution of beliefs for schools 58 and 77.

Table 9: Heterogeneity between schools

	School 77		School 58	
Endogenous effect α	0.129***	0.040	0.248***	0.050
Child's characteristics				
Female	0.315***	0.085	0.344**	0.137
Grades 9–10	-0.749	0.651	-0.203	0.601
Grades 11–12	-0.841	0.671	0.221	0.657
Age	-0.023	0.074	-0.063	0.093
White	-0.147	0.107	-0.397	0.686
Black or African–American	0.012	0.121	-0.047	3.203
Hispanic	-0.114	0.107	0.624	0.743
Mother education				
Less than high school	0.107	0.178	-0.027	0.319
High school	0.062	0.140	0.137	0.223
More than high school	0.173	0.134	0.366	0.237
Father education				
Less than high school	-0.166	0.172	-0.409	0.332
High school	-0.005	0.159	-0.017	0.206
More than high school	0.108	0.126	0.503	0.217
Household characteristics				
Household size	-0.015	0.021	-0.027	0.050
Constant	0.290	2.637	0.586	2.851
γ_2	0.803***	0.095	0.811***	0.111
γ_3	2.104***	0.096	1.960***	0.133
γ_4	2.959***	0.096	2.887***	0.135
College preferences	yes		yes	
Grade in mathematics, English, history, and science	yes		yes	
Observations	920		521	

Note: We let our chain run for 20,000 iterations, discarding the first 9,999 iterations; $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

6.5 Posterior Distributions

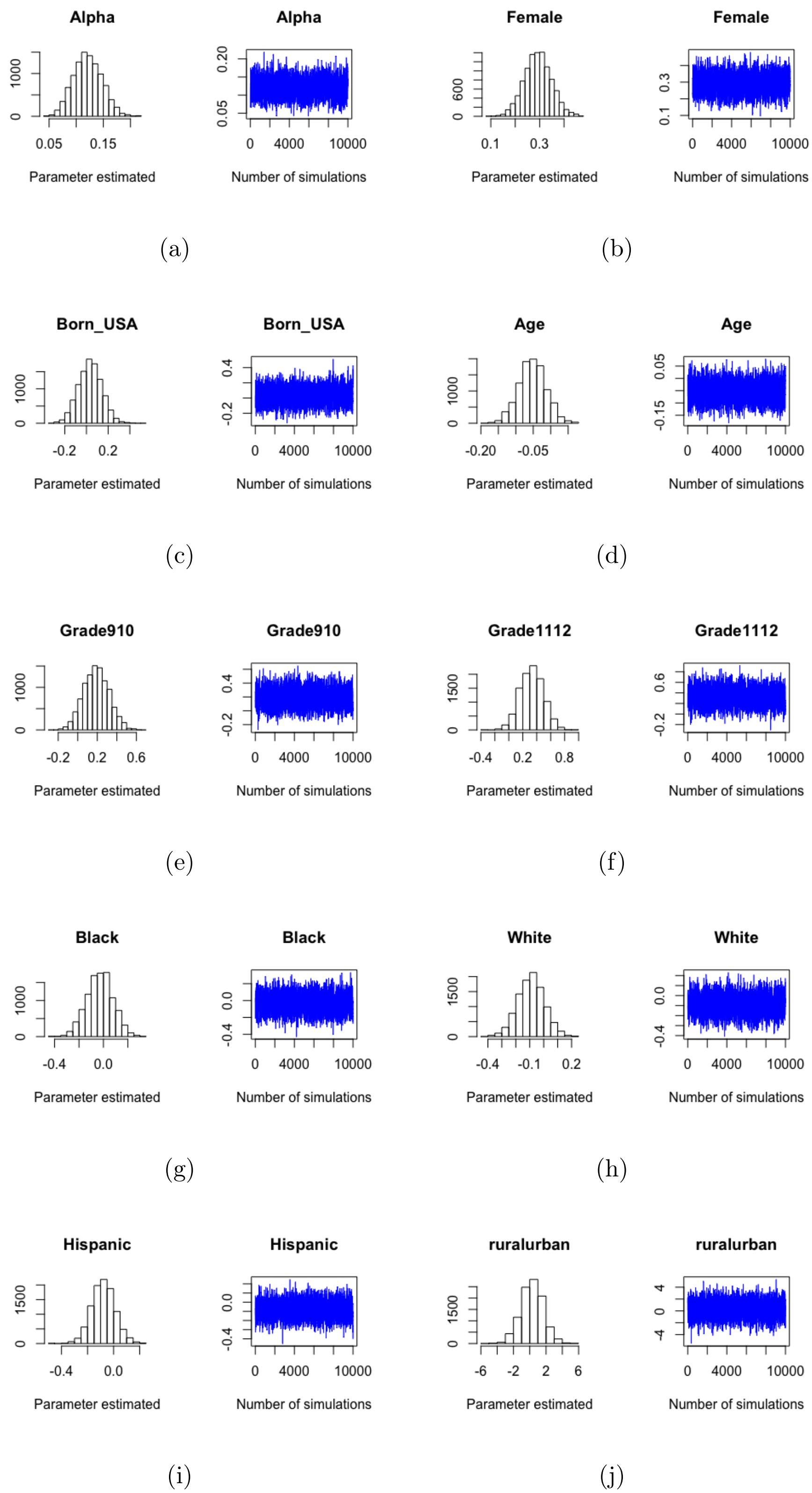
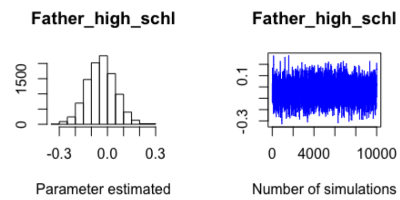
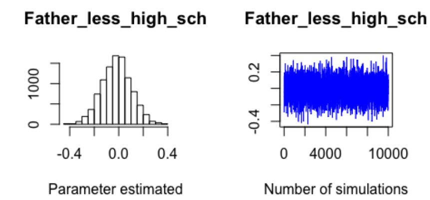


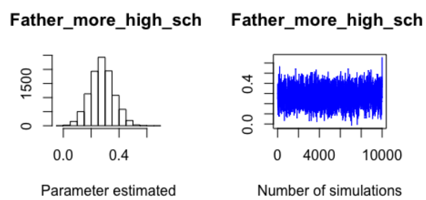
Figure 3: Posterior distributions for the estimation over the entire sample (continued).



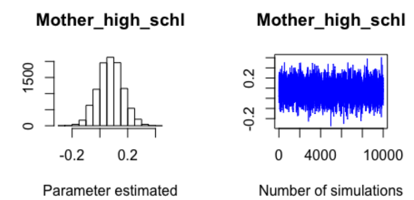
(a)



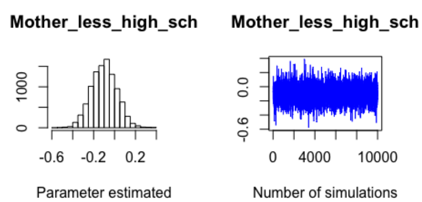
(b)



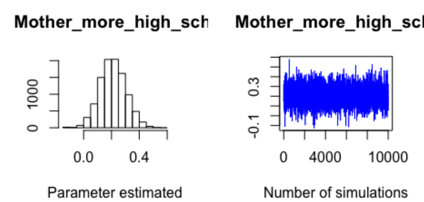
(c)



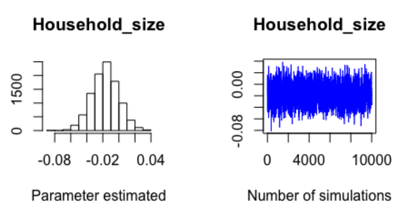
(d)



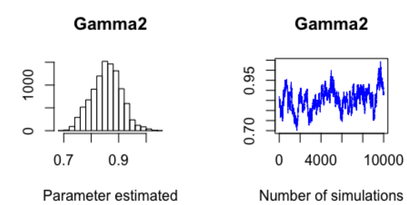
(e)



(f)



(g)



(h)

Figure 4: Posterior distributions for the estimation over the entire sample (continued).

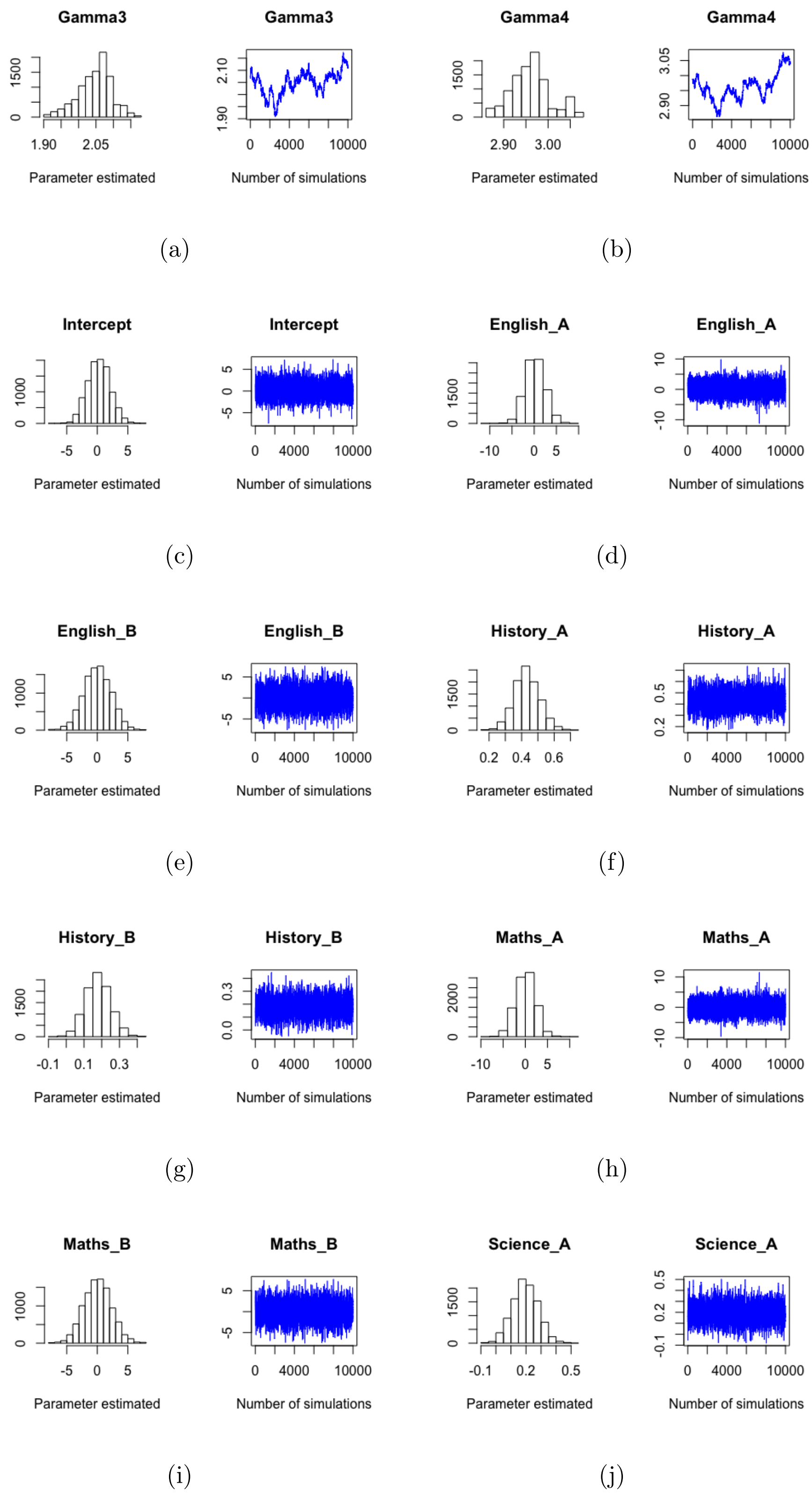
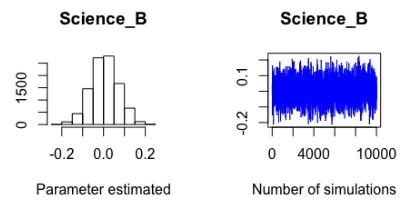
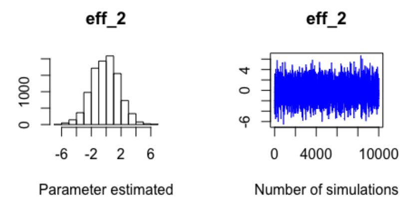


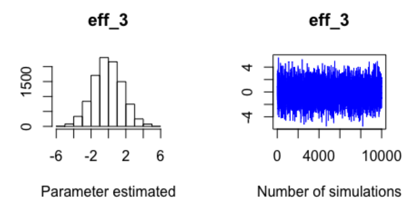
Figure 5: Posterior distributions for the estimation over the entire sample (continued).


Science_B
Science_B
Parameter estimated
Number of simulations

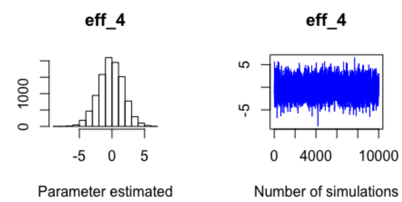
(a)


eff_2
eff_2
Parameter estimated
Number of simulations

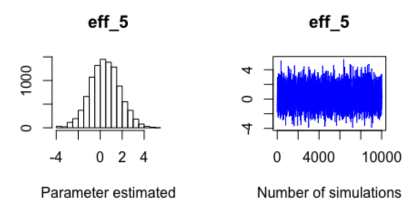
(b)


eff_3
eff_3
Parameter estimated
Number of simulations

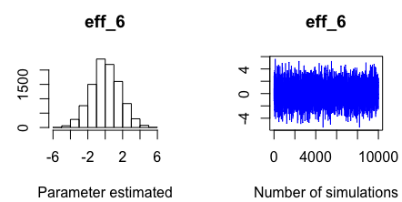
(c)


eff_4
eff_4
Parameter estimated
Number of simulations

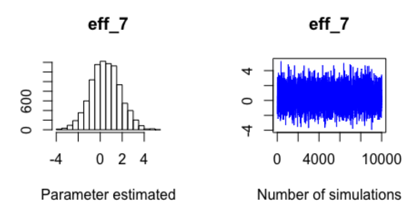
(d)


eff_5
eff_5
Parameter estimated
Number of simulations

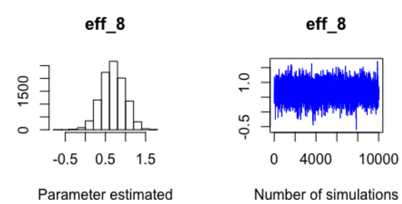
(e)


eff_6
eff_6
Parameter estimated
Number of simulations

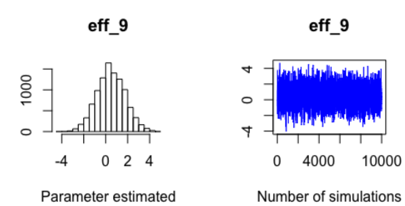
(f)


eff_7
eff_7
Parameter estimated
Number of simulations

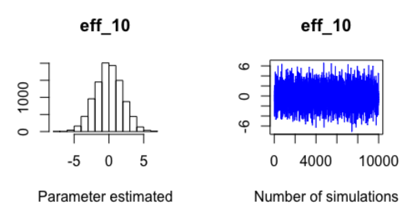
(g)


eff_8
eff_8
Parameter estimated
Number of simulations

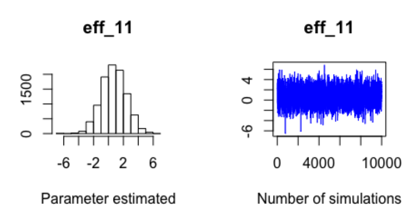
(h)

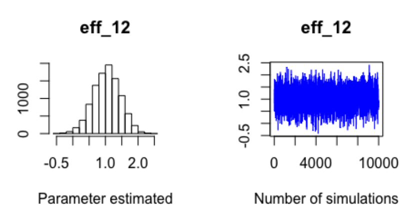

eff_9
eff_9
Parameter estimated
Number of simulations

(i)


eff_10
eff_10
Parameter estimated
Number of simulations

(j)


eff_11
eff_11
Parameter estimated
Number of simulations


eff_12
eff_12
Parameter estimated
Number of simulations

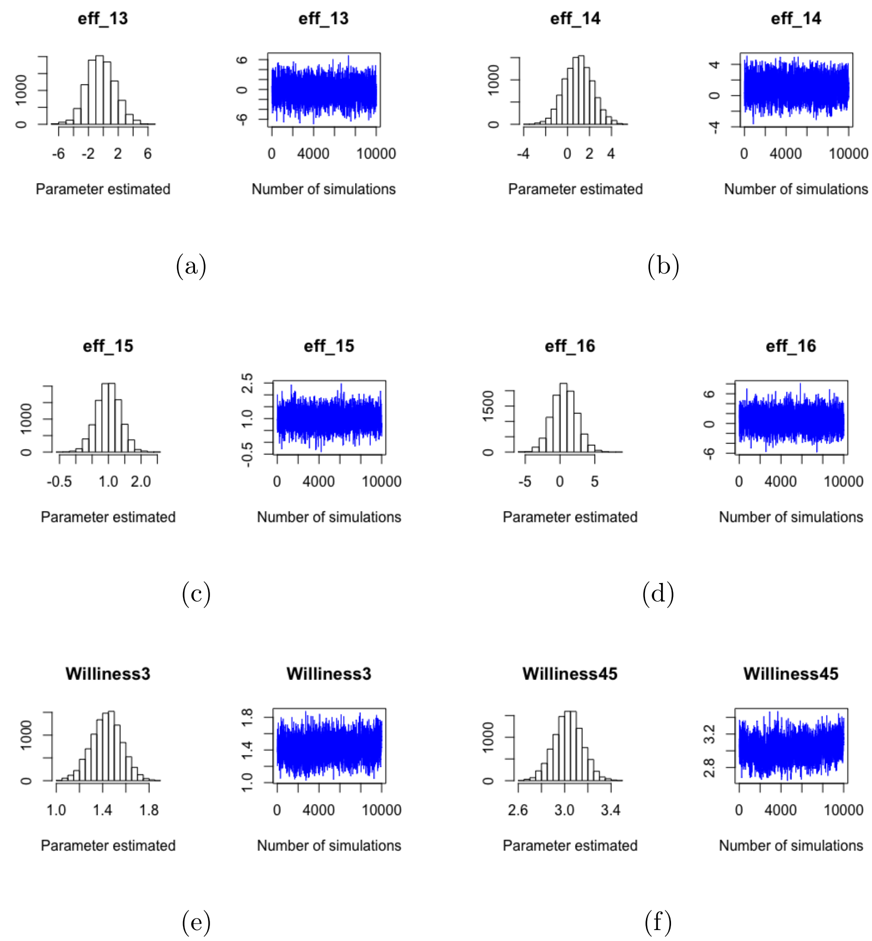


Figure 7: Posterior distributions for the estimation over the entire sample.

References

Albert, J. H. and S. Chib (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

Angrist, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

Arduini, T., E. Patacchini, and E. Rainone (2016): "Identification and estimation of network models with heterogeneous externalities," *Working paper*.

Ashby, J. S. and I. Schoon (2010): "Career success: The role of teenage career aspirations, ambition value and gender in predicting adult social status and earnings," *Journal of Vocational Behavior*, 77, 350–360.

Atchadé, Y. F. and J. S. Rosenthal (2005): "On adaptive markov chain monte carlo algorithms," *Bernoulli*, 11, 815–828.

Attanasio, O. P. and K. Kauffman (2010): "Educational Choices and Subjective Expectations of Returns: Evidence on Intra-Household Decision Making and Gender Differences," *NBER working paper*, 15087.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: The key player," *Econometrica*, 74, 1403–1417.

Bellemare, C., S. Kröger, and A. Van Soest (2008): "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities," *Econometrica*, 76, 815–839.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2013): "Under pressure? The effect of peers on outcomes of young adults," *Journal of Labor Economics*, 31, 119–153.

Boucher, V. (2016): "Conformism and self-selection in social networks," *Journal of Public Economics*, 136, 30–44.

Boucher, V. and B. Fortin (2015): "Some Challenges in the Empirics of the Effects of Networks," .

——— (2016): "Some Challenges in the Empirics of the Effects of Networks," *The Oxford Handbook of the Economics of Networks*.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

Burke, M. A. and T. R. Sass (2013): "Classroom peer effects and student achievement," *Journal of Labor Economics*, 31, 51–82.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009): "Does your cohort matter? Measuring peer effects in college achievement," *Journal of Labor Economics*, 27, 439–464.

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2015): "Testing models of social learning on networks: Evidence from a lab experiment in the field," Tech. rep., National Bureau of Economic Research.

de Paula, A. (2017): "Econometrics of Network Models," in *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress (Econometric Society Monographs)*, ed. by M. P. B. Honore, A. Pakes and L. Samuelson, Cambridge University Press, 268–323.

DeGroot, M. H. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69, 118–121.

Dieye, R., B. Fortin, et al. (2017): "Gender peer effects heterogeneity in obesity," *Working paper*.

Friedkin, N. E. and E. C. Johnsen (1990): "Social influence and opinions," *Journal of Mathematical Sociology*, 15, 193–206.

Geweke, J., M. Keane, and D. Runkle (1994): "Alternative computational approaches to inference in the multinomial probit model," *The review of economics and statistics*, 609–632.

Giorgi, G. D., M. Pellizzari, and S. Redaelli (2010): "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2, 241–275.

Goldsmith-Pinkham, P. and G. W. Imbens (2013): "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 31, 253–264.

Golub, B. and M. O. Jackson (2010): "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2, 112–149.

——— (2012): "How homophily affects the speed of learning and best-response dynamics," *The Quarterly Journal of Economics*, 127, 1287–1338.

Golub, B. and E. Sadler (2016): "Learning in Social Networks," *in The Oxford Handbook on the Economics of Networks, Editors: Yann Bramoullé, Andrea Galeotti, Brian W. Rogers.*

Graham, B. S. (2008): "Identifying social interactions through conditional variance restrictions," *Econometrica*, 76, 643–660.

Hatami, M., A. Kazemi, and T. Mehrabi (2015): "Effect of peer education in school on sexual health knowledge and attitude in girl adolescents," *Journal of Education and Health Promotion*, 4.

Hoxby, C. M. and S. Turner (2013): "Informing students about their college options: A proposal for broadening the expanding college opportunities project," *The Hamilton Project, Discussion Paper*, 3.

Hsieh, C.-S. and L. F. Lee (2016): "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 31, 301–319.

Kline, B. and E. Tamer (2014): "Some interpretation of the linear-in-means model of social interactions," *Working paper*.

Lee, L.-F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

Lee, L.-f., X. Liu, and X. Lin (2010): "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 13, 145–176.

Liu, X. and J. Zhou (2017): "A social interaction model with ordered choices," *Economics Letters*, 161, 86–89.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542.

——— (2000): "Economic analysis of social interactions," *Journal of Economic Perspectives*, 14, 115–136.

Moffitt, R. A. et al. (2001): "Policy interventions, low-level equilibria, and social interactions," *Social Dynamics*, 4, 6–17.

Peng, S. (2016): "Heterogeneous endogenous effects in networks," *Working paper*.

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

Rose, C. D. (2017): "Identification of peer effects through social networks using variance restrictions," *The Econometrics Journal*.

Sacerdote, B. et al. (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far," *Handbook of the Economics of Education*, 3, 249–277.

Sipsma, H. L., J. R. Ickovics, H. Lin, and T. S. Kershaw (2015): "The impact of future expectations on adolescent sexual risk behavior," *Journal of Youth and Adolescence*, 44, 170–183.

Tanner, M. A. and W. H. Wong (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Zabel, J. E. (2008): "The impact of peer effects on student outcomes in New York City public schools," *Education Finance and Policy*, 3, 197–249.

Zenou, Y. (2016): "Key Players," in *The Oxford Handbook of the Economics of Networks*, ed. by A. G. Yann Bramoullé and B. Rogers.