Identification and Estimation of Marginal Treatment Effect Through Social Networks

---

Timothy Hersey

THE UNIVERSITY
of ADELAIDE

School of Economics

University of Adelaide

October 2018

Contents

Abstract

This thesis develops a method for examining the role peer effects play in treatment effect models. We focus on the scenario where peer effects are important in determining the treatment decision. Identification of both the treatment decision and the associated marginal treatment effect is explored. In particular, exogenous and endogenous peer effects are used as instruments to identify marginal treatment effects. A Bayesian estimation procedure is presented, utilising a network formation model to adjust for unobserved peer effects. The performance of the model and the estimation procedure is analysed through a Monte Carlo experiment. The proposed method is then applied to estimate the effect of high school peers on the decision to attend college, and the return to education associated with such a decision using the *Add Health* data-set. Both the Monte Carlo experiment and the empirical application underscore the importance of accounting for the presence of peer effects in treatment effects models, and allows us to consider policy implications of peer effects.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

_	03/10/2018
Signature	**Date**

Acknowledgements

Firstly I would like to thank my supervisors Firmin Doko Tchatoka and Virginie Masson, who have supported and guided me along the whole journey. It has been a privilege to work with both of you over many years. Thank-you especially to my family and friends who have always encouraged me to persevere and kept me balanced. Thanks to the other postgraduate students and academics who took the time to check-in, have discussions and give advice. I would also like to acknowledge the assistance of the Carolina Population Center, University of North Carolina, in providing access to the Add-Health dataset. An extra thanks to Dr Chih-Sheng Hsieh, for providing Matlab codes and answering questions.

Chapter 1

Marginal Returns to Treatment and Social Networks

1 Introduction

We live and relate in an increasingly interconnected world. The social networks we associate in are ever expanding with the advent of globalisation and social media. Decisions and outcomes are not isolated to an individual, but are spread across networks and influenced by peers. Peer effects present themselves in a variety of contexts ranging from alcohol use (Kremer and Levy, 2008; Fletcher, 2012a) and delinquency (Patacchini and Zenou, 2012) to loan behaviour (Karlan et al., 2005) and obesity (Christakis and Fowler, 2007). Peer influence can act directly through conformism effects; as individual decision is swayed by the decisions of friends, or indirectly through the influence of personality and individual characteristics. Economics has begun to recognise the importance social ties play in the outcomes of individuals and the implications for policy. The influence of networks implies the existence of a social multiplier (Glaeser et al., 2003), such that the decision of one individual or the effect of policy on that individual is dispersed across their network. As a result, the net effect of policy becomes greater than the direct effect usually intended by policy makers, particularly if the individual is central in their network.

Peer effects are prominent in both the binary decisions made by individuals and observable continuous outcomes. In the education context, peer effects have been observed in the decision of an individual to attend college, their choice of major (Bifulco et al., 2011; Fletcher, 2013, 2012b; Wu, 2015), and resulting wages (Barbone and Dolton, 2015; Kramarz and Skans, 2014; Black et al., 2013). The structure of an econometric treatment effects model allows us to measure the returns to binary

treatment decisions. While peer influence can be expected to play a critical role in both binary decisions and observed outcomes, little research has focused on the incorporation of networks within a treatment framework. This thesis develops a model where peer effects are influential in shaping a binary treatment decision and, indirectly, observed outcomes and returns to treatment. Manski (2013) examines the identification of a treatment model where social interactions are important in the treatment response, i.e., when the treatment of peers affects individual's outcomes (e.g., immunisation). This thesis, in contrast, focuses on the role of social interactions on the treatment itself and the subsequent impact of the treatment on the observed outcome. We study the identification of such a model and quantify the role peer effects play in changes to policy.

The model is applied to analyse the critical decision of high school students to enter college. A common treatment model examines the effect of the binary decision to attend college on future wages (Carneiro et al., 2011). Returns to education should optimally be (and commonly are) considered in the decision to attend college. As a result, heterogeneous returns to education are expected, i.e. those who do attend college have higher returns, while those who do not attend college would not benefit greatly from doing so. The treatment effects model can be used to identify the effect of treatment on the outcome equation, accounting for heterogeneity in the returns to treatment.

Economic network literature has considered the effect of peers on both the decision to attend college and labour market outcomes. We bridge these two settings to analyse the interaction between peer effects and the returns to education.

This paper draws from both the treatment effects and peer effects literature to identify the effect of peers on the marginal returns to a binary decision. In particular, we consider the interaction between peer effect models, such as the one proposed by Goldsmith-Pinkham and Imbens (2013), and the treatment effects model. This is far from straightforward due to the numerous interrelated avenues through which peer effects operate.

In his 1993 pioneering research on the identification of peer effects, Manski ex-

plores the issue of identification in light of the "reflection problem". The "reflection problem" refers to the difficulty in isolating two types of peer effect; endogenous and exogenous effects, which are often strongly correlated.

Endogenous social effects refer to the direct influence of an individual's outcomes or decisions on their peers. The decision of one individual often directs the decisions made by the friends of that individual. Consider a high school student who decides to take up smoking. The friends of that student have a much higher probability of smoking than individuals who have no friends that smoke (Mercken et al., 2010; Christakis and Fowler, 2008).

The second effect, exogenous social effects (or contextual effects), describes the influence of an individual's characteristics on his peers' outcomes. For example, if an individual is proactive and hard working, he will likely have influence on the decisions of his friends, particularly in his college decision. Taking another example, Bifulco et al. (2014) show that having peers with college educated mothers leads to a higher probability of individuals attending college themselves.

A third non-peer effect, the correlated effect, often confounds identification of the exogenous and endogenous effects. Correlated effects refer to common factors that influence the outcomes of all individuals in the same peer group. Moffitt et al. (2001) segregates correlated effects into homophily and environmental effects. Individuals are prone to homophily, where friendships are chosen based on similarity in characteristics or personality. These characteristics, like social extroversion or athletic ability, can be difficult to observe, but lead to similar observed outcomes between peers. Furthermore, individuals in common environments (high schools, workplaces) face similar surroundings and influences. Students in the same high school will experience similarities in teaching style, school resources and socio-economic status, which lead to similarities in decisions and outcomes. The difficulty lies in determining the reason the decisions and outcomes of these peers look similar. Is it truly a peer effect, or is it simply because they face a common environment, or chose their friendship groups based on similarity in characteristics?

This paper serves as a preliminary investigation of the flexibility of the peer effect

model and its components and how they interact with existing econometric models. The model presented in this paper draws on the model in Goldsmith-Pinkham and Imbens (2013) in the network estimation literature and Carneiro et al. (2011) in the treatment effects literature. We consider the response of the marginal treatment effect (MTE) and the marginal policy relevant treatment effect (MPRTE), as presented in Carneiro et al. (2011), to the presence of peer effects. These measures allow us to control for heterogeneity in returns and are more robust than traditional measures (e.g. average treatment effect, treatment effect for the treated). The MTE describes the effect of treatment for individuals who are at the margin i.e. are indifferent to receiving treatment. The MPRTE, proposed by Carneiro, Heckman, and Vytlacil (2010), measures the returns to treatment for those induced into treatment as the result of a marginal change in policy. Estimation of the MPRTE in this setting enables us to consider the effect of a change in policy on the returns to a binary decision, when peer effects are present.

We start by outlining the theoretical model, before presenting identification results and proposing a Bayesian estimation method. The model is then tested using simulated data in a Monte Carlo experiment. An application studies the educational attainment model using the *Add-Health dataset*, which contains friendship network data for high school students. This application underscores the importance of controlling for peer effects when estimating the returns to education.

Throughout our analysis, the policy effects on the returns to treatment will also be considered. In particular, we investigate the implications of changes in *peer effects*, *network structures* and *policy initiatives* on the returns to treatment.

2 Relevant Literature

Comprehensive network data has only recently begun to emerge due to the complexities in accurately describing and measuring networks. Accompanying several key datasets, the research on networks and peer effects has expanded in recent years. Jackson (2013); Jackson et al. (2017) identify several of the challenges many re-

searchers encounter when measuring peer effects, namely, identification, endogenous networks and homophily, computation, measurement error and misspecification. In particular, the identification of peer effects has proven to be a difficult prospect.

Manski's (1993) infamous "reflection problem" describes the difficulty in separately identifying endogenous and exogenous peer effects. When an individual interacts with groups, he is both influenced by and influential in his peer's groups. As such the expected outcomes of a peer group (endogenous effects) and the mean characteristics of the group (exogenous effects) cannot be easily separated.

Traditionally, the peer effects literature has assumed that individuals interact in groups, with each member having an equal influence on all others in the group. Under such conditions identification is difficult (Manski, 1993; Moffitt et al., 2001). Lee (2007) established identification under the condition that the individual is excluded from their own peer group and peer groups have at least three distinct sizes. Relaxing the peer group assumption, the structure and non-linearity of social networks is a useful tool for identification. In the notable paper of Bramoullé, Djebbari, and Fortin (2009), generalising the models of Manski (1993), Moffitt et al. (2001) and Lee (2007), the structure of an individual's network is exploited to allow identification, using an instrumental variables approach. Consistent estimation, however, can only be achieved under the exogeneity of the network. This exogeneity assumption does not hold in the presence of correlated effects; characteristics which are correlated with both the choice of peers and the outcome variable.

Lee (2007), as in Lin (2010), Bramoullé et al. (2009), addresses the correlated effect empirically by introducing an unobserved network fixed effect variable into the model. This approach, while an improvement on existing measures, does not account for within-network variation which may influence the outcome. As an example, an unobserved personality characteristic such as social extroversion may influence not only who an individual becomes friends with, but also whether they attend college, their place of employment and resulting wage. If this effect is present and not accounted for, the network will be endogenous, causing bias in the peer effect estimates. Goldsmith-Pinkham and Imbens (2013) propose a network formation model

to control for this network endogeneity. The network formation model incorporates unobserved individual characteristics which influence the likelihood of friendship formation and observed outcomes. The model relies on asymptotic network theory to achieve identification. Hseih and Lee (2016) apply the model of Goldsmith-Pinkham and Imbens, relaxing some of their identifying assumptions.

The network formation literature holds promise for greater accuracy in the analysis of peer effects. Jackson et al. (2017) critiques the network formation model of Goldsmith-Pinkham and Imbens, while confirming the general theoretic approach. The link-by-link model of Goldsmith-Pinkham and Imbens is, in general, too simplistic to accurately reflect the interconnected network formation process. We use the original network formation model proposed by Goldsmith-Pinkham and Imbens, though alternate network formation approaches could also be integrated. Other simpler approaches have been taken, particularly in the applied literature (Patacchini et al., 2011; Kremer and Levy, 2008; Patacchini and Zenou, 2012), but generally fail to adequately control for the endogeneity issue. De Paula (2016) and Blume et al. (2011) offer a detailed overview of the literature concerning peer effect identification.

The linear in means model used in this paper can be derived from the utility theory setup of Blume et al. (2011). However, other theory approaches can be taken, specifically integrating the formation of friendships. Badev (2013) proposes one such network approach, deriving the Nash equilibrium of a friendship network to identify and decompose peer effects, while Boucher (2015) focuses heavily on the homophily aspects of friendships. Calvó-Armengol et al. (2009) combine the existing network formation model framework with network centrality concepts to achieve identification of peer and network effects.

Introducing non-linearities tends to complicate estimation in the peer effect model (Blume et al., 2011). In this paper we will examine the role of peers in a binary decision. This nonlinearity is explored in Soetevent and Kooreman (2007) using the standard utility theory set-up to derive the reduced form equations and equilibrium conditions. Similarly Brock and Durlauf (2001), and Blume et al. (2011) present alternate game theoretic approaches in a group interaction context. Lee, Li, and

Lin (2014) extend the approach of Brock and Durlauf (2001) from peer groups to networks. The binary model is applied in several empirical papers using a simple linear probability model setup (Patacchini and Arduini, 2016; Fletcher, 2012a). In most binary settings, Blume et al. (2011) and Blume et al. (2015) assert that the reflection problem is no longer pertinent.

Apart from these, limited research has been conducted on the estimation of an econometric binary peer model. The likelihood function of such a model requires an n-dimensional probit likelihood, which is computationally difficult to evaluate. The problem has, however, been considered in the spatial literature using the spatial autoregressive model (SAR). The estimation in this paper makes use of the spatial literature surrounding the estimation of a probit/logit SAR model. A comparison of the techniques to estimate such a model can be found in Calabrese and Elkink (2014). The method used in this paper follows that presented in LeSage (2000) and updated in LeSage and Pace (2009), who use a Bayesian estimation approach. This formalises and extends the original method of Chib (1992) and Albert and Chib (1993). The Bayesian approach tends to outperform other estimation methods, particularly under low spatial autocorrelation. The Bayesian approach can also be adapted to intersect with the Bayesian estimation of the Goldsmith-Pinkham and Imbens (2013) model.

Therefore, we work to combine the estimation methods of Goldsmith-Pinkham and Imbens, and Lesage to estimate a binary peer effects model in the first stage. In the second stage we follow Carneiro et al. (2011) in estimating the marginal returns to a binary decision.

The treatment model presented by Carneiro et al. (2011) estimates the heterogeneous returns to treatment, in the context of education, using the marginal treatment effect (MTE) measure. The MTE represents the effect of treatment for those who are indifferent to treatment (i.e. individuals who are at the margin). Björklund and Moffitt (1987) first proposed the MTE as an alternate measure to traditional estimators such as the average treatment effect (ATE) and average treatment effect of the treated (ATET) which do not account for heterogeneous treatment effects. Carneiro et al. (2011) use the MTE to consider the true effect of policy changes, deriving

the marginal policy relevant treatment effect (MPRTE) first developed by Carneiro et al. (2010). The MPRTE measures the returns to treatment for those induced into treatment by a marginal change in policy.

Carneiro et al. (2011) propose both a parametric and non-parametric estimation of these marginal effects, while Carneiro et al. (2010) examine the properties of these estimates. The instrumental variables method used to examine policy effects was first discussed in Heckman and Vytlacil (2007b, 1999, 2005).

3 Model

Let $\left\{ (y_i, X_i, Z_i, S_i) : 1 \leq i \leq N \right\}$ be an i.i.d. sample of N observations, where y_i is the i-th observation on an outcome variable, X_i is the i-th observation on K_1 covariates (possibly endogenous), and Z_i is a K_2-dimensional vector of the i-th observations on exogenous covariates (instruments). Let S be a treatment variable, i.e., $S_i = 1$ if individual i is included in the treatment and $S_i = 0$ otherwise. We assume that the treatment decision is latent, i.e., we observed $S_i = 1$ if $S_i^* > 0$ and $S_i = 0$ if $S_i^* \leq 0$, where S^* represents the net individual benefit of receiving treatment, which depends on observed exogenous covariates Z and unobserved variables v.

Our goal is to measure the marginal returns to treatment. To achieve this, we consider the framework of Carneiro et al. (2011) that generalises Roy (1958) and Quandt (1958, 1972). The model consists of potential outcome equations represented by

$$\boldsymbol{y}_j = \mu_j(X) + u_j, \ j = 0, 1, \tag{3.1}$$

where $\mu_j(x) = \mathbb{E}[\boldsymbol{y}_j | X = x]$. $\boldsymbol{y}_1$ describes the outcomes for individuals in the treatment group and $\boldsymbol{y}_0$ the outcomes for those not in the treatment group. The observed

outcome is then given by:

$$\begin{aligned} \boldsymbol{y} &=: S\boldsymbol{y}_1 + (1-S)\boldsymbol{y}_0 \\ &= \mu_0(X) + [\mu_1(X) - \mu_0(X)]S + (u_1 - u_0)S + u_0. \end{aligned} \tag{3.2}$$

Under this model, the return to treatment is defined as $\boldsymbol{y}_1 - \boldsymbol{y}_0 = \mu_1(X) - \mu_0(X) + u_1 - u_0$, $\mu_1(x) - \mu_0(x)$ represents the average treatment effect conditional on $X = x$, and $\mu_1(x) - \mu_0(x) + \mathbb{E}[u_1 - u_0|S = 1, X = x]$ is the average treatment effect for the treated. In the special case of a linear model [i.e., when $\mu_j(X) = X\beta_j$], (3.2) can be written as

$$\boldsymbol{y} = X\beta_0 + X[\beta_1 - \beta_0]S + [(u_1 - u_0)S + u_0] \tag{3.3}$$

so that $\delta \equiv \delta(x) =: x(\beta_1 - \beta_0)$ measures the average treatment effect conditional on $X = x$.

We suppose that the treatment decision is latent and represents the net individual benefit of receiving treatment. The standard treatment model[1] usually specifies S^* as the difference between observable variables Z and unobservable factors v, i.e.

$$S^* = \mu_S(Z) - v, \quad \mu_S(z) = \mathbb{E}[S^*|Z = z], \tag{3.4}$$

where Z may include exogenous variables in X in addition to other exogenous instruments. In this study, we emphasise the possibility that the specification of S^* in (3.4) may also include the characteristics of one's peers.

Let F denote the common distribution[2] of the sample $(y_i, X_i, Z_i)_{i=1}^N$. We assume $\mathbb{E}_F[(u_{ji}, v_i)|Z_i] = 0$ for all $i = 1, \ldots, N$ and all $j = 0, 1$, i.e., Z_i is uncorrelated with u_{ji} and v_i (orthogonality condition). We also assume that the v_i's are strictly increasing and continuous r.v. (random variables) with common distribution F_v. The latter assumption is also used in Carneiro et al. (2011) and implies that the

[1] e.g., see Carneiro et al. (2011, Eq.(3)).

[2] Note that F may depend on the sample size N, say F_N, but we drop the indexation by N for convenience.

probability of undertaking treatment ($S^* > 0$ or $S = 1$) conditional on $Z = z$ can then be expressed as

$$P(z) \equiv \mathbb{P}[S = 1|Z = z] = F_v(\mu_S(z)). \tag{3.5}$$

The quantity $P(z)$ in (3.5) is usually referred to as propensity score or probability of selection. Let $U_S = F_v(v)$, i.e., U_S is a uniform r.v. whose values correspond to the quantiles of v. Therefore, (3.4) holds if and only if $P(Z) \geq U_S$, so $P(Z)$ is interpreted as the mean scale utility function; see McFadden (1974).

3.1 Social networks and treatment decision

In this section, we explore the inclusion of peer effects in the treatment decision model (3.4). As in the linear-in-means model of Manski (1993), we assume that the decision of individuals to undertake treatment is affected by the mean characteristics of their peers.[3]

Suppose that individuals belong to m pre-specified groups and let $\boldsymbol{G}_g : N_g \times N_g$ represent a network associated with each group $g = 1, \ldots, m$. We consider the following model for the latent treatment variable S_g^* :

$$\begin{aligned} S_g^* &= \gamma_1 \boldsymbol{G}_g S_g^* + Z_g\gamma_2 + \boldsymbol{G}_g Z_g \gamma_3 + \alpha_g \iota_g + \gamma_4 \xi_g - v_g \\ &= \gamma_1 \boldsymbol{G}_g S_g^* + \sum_{k=1}^{K_2} (\gamma_{2,k} I_{N_g} + \gamma_{3,k} \boldsymbol{G}_g) Z_{gk} + \alpha_g \iota_g + \gamma_4 \xi_g - v_g, \end{aligned} \tag{3.6}$$

where $S_g^* : N_g \times 1$, $Z_g = [Z_{g1}, \ldots, Z_{gK_2}] : N_g \times K_2$ with $Z_{gk} : N_g \times 1$ for all $k = 1, \ldots, K_2$, $\iota_g : N_g \times 1$ is a vector of ones, $\xi_g : N_g \times 1$ contains unobserved within group characteristics affecting both the treatment decision and the formation of the *network*, γ_4 is the effect of these unobserved characteristics, $v_g : N_g \times 1$ is a vector of disturbances, $\gamma_1 : 1 \times 1$ represents the endogenous peer effect (the average benefits to treatment for an individual's peers in the network G_g), $\gamma_3 = (\gamma_{3,1}, \ldots, \gamma_{3,K_2})' : K_2 \times 1$ describes the exogenous peer effect (the average characteristics of an individual's

[3]We apply this to a binary model as done in the spatial literature– e.g., see Anselin (1988).

peers in the network $\boldsymbol{G}_g$), $\gamma_2 = (\gamma_{2,1}, \ldots, \gamma_{2,K_2})' : K_2 \times 1$ measures the direct impact of Z_g on the treatment decision, and $\alpha_g : 1 \times 1$ represents the group fixed effect that controls for correlated effects. For the remainder of the paper, we define $\gamma = (\gamma_1, \gamma_2', \gamma_3', \gamma_4)'$. The main difference between model (3.6) and the one considered in Bramoullé et al. (2009, Eq.(1)) and Goldsmith-Pinkham and Imbens (2013, Eq.(4.2)) is the latent nature of S_g^*, which introduces additional complexity in estimation of the parameters as we see later on. Following Goldsmith-Pinkham and Imbens (2013), the network $\boldsymbol{G}_g$ is constructed as:

$$\boldsymbol{G}_g \;=\; \text{diag}(\boldsymbol{M}_g)^{-1}\boldsymbol{D}_g \equiv \left[G_{g,ij}\right]_{1\leq i,j\leq N_g} :\; G_{g,ij} = D_{g,ij}/M_{g,i}, \tag{3.7}$$

where $\boldsymbol{D}_g = \left[D_{g,ij}\right]_{1\leq i,j\leq N_g}$ is a symmetric adjacency matrix such that $D_{g,ii} = 0$ for all i, $D_{g,ij} = 1$ if individuals i and j are friends and zero otherwise for all $i \neq j$; $\boldsymbol{M}_g = \left(M_{g,i}\right)_{1\leq i\leq N_g}$ is an $N_g \times 1$ vector with elements $M_{g,i} = \sum_{j=1}^{N_g} D_{g,ij}$ representing the number of friends of individual i.[4] Clearly, the network $\boldsymbol{G}_g$ is a row-normalised adjacency matrix, and can be used to determine the average friend of an individual. The network may be observed at multiple time periods but for simplicity, we drop the dependence over time. Hereinafter, $\boldsymbol{D}_{g\_}$ and $\boldsymbol{G}_{g\_}$ denote the non-normalised and row-normalised adjacency matrices at the previous period, respectively.

By noting that the determinant of the matrix $I_{N_g} - \gamma_1\boldsymbol{G}_g$ has the form written as $det(I_{N_g} - \gamma_1\boldsymbol{G}_g) = \prod_{j=1}(1 - \gamma_1\lambda_{jg})$, where $\lambda_{jg}, j = 1, \ldots, N_g$, are the eigenvalues of $\boldsymbol{G}_g$ satisfying $-1 < \lambda_{jg} \leq 1$ [e.g., see Case (1991, footnote 5)], it is straightforward to see that $I_{N_g} - \gamma_1\boldsymbol{G}_g$ is invertible as long as $|\gamma_1| < 1$. Under this condition, the reduced-form model for S_g^* can be expressed using the second equality in (3.6) as:

$$S_g^* = (I_{N_g} - \gamma_1\boldsymbol{G}_g)^{-1}\alpha_g\iota_g + (I_{N_g} - \gamma_1\boldsymbol{G}_g)^{-1}\boldsymbol{B}_q vec(Z_g) + (I_{N_g} - \gamma_1\boldsymbol{G}_g)^{-1}\gamma_4\xi_g - \eta_g, \tag{3.8}$$

where $vec(Z_g)$ is the $N_gK_2 \times 1$ dimensional column vectorization of Z_g, $\eta_g = (I_{N_g} -$

[4]Individuals with no friends are discounted from the model so that $M_{g,i} > 0$.

$\gamma_1 \boldsymbol{G}_g)^{-1} v_g \equiv (\eta_{g,1}, \ldots, \eta_{g,N_g})'$, and $\boldsymbol{B}_q$ is given by

$$\boldsymbol{B}_q =: \left[\gamma_{2,1} I_{N_g} + \gamma_{3,1} \boldsymbol{G}_g, \ldots, \gamma_{2,K_2} I_{N_g} + \gamma_{3,K_2} \boldsymbol{G}_g\right] : N_g \times N_g K_2.$$

In particular if $v_g \sim N(0, \sigma^2_{v_g} I_{N_g})$, then we have

$$\eta_g \sim N\big(0, \Sigma_\eta\big) \text{ with } \Sigma_\eta = \sigma^2_{v_g} \left[(I_{N_g} - \gamma_1 \boldsymbol{G}_g)'(I_{N_g} - \gamma_1 \boldsymbol{G}_g)\right]^{-1}, \tag{3.9}$$

i.e., $\eta_{g,i} \sim N(0, \sigma^2_{\eta_{g,i}})$ where $\sigma^2_{\eta_{g,i}}$ is the (i,i)th element of Σ_η. Clearly, the errors $\eta_{g,i}$ of the reduced-form regression (3.8) are heteroskedastic by construction. Let $\boldsymbol{b}_{g,i\cdot}$, $i = 1, \ldots, N_g$, denote the ith row of the matrix $\boldsymbol{b}_g$:

$$\boldsymbol{b}_g =: (I_{N_g} - \gamma_1 \boldsymbol{G}_g)^{-1} \left[\alpha_g I_{N_g} \vdots \boldsymbol{B}_q \vdots \gamma_4 I_{N_g}\right] : N_g \times N_g(K_2 + 2),$$

and define $\tilde{Z}_g =: [\iota_g' \vdots vec(Z_g)' \vdots \xi_g']' : N_g(K_2 + 2) \times 1$. Under the assumption that $v_g \sim N(0, \sigma^2_{v_g} I_{N_g})$, the propensity score (probability of treatment decision) of individual i in network $\boldsymbol{G}_g$, conditional on $\tilde{Z}_g = \tilde{z}_g$, is given by

$$\begin{aligned} P_i(\tilde{z}_g) &=: \mathbb{P}[S_{g,i} = 1 | \tilde{Z}_g = \tilde{z}_g] \\ &= \mathbb{P}[\eta_{g,i} \leq \boldsymbol{b}_{g,i\cdot} \tilde{z}_g] = \Phi\big(\boldsymbol{\gamma}_{g,i\cdot} \tilde{z}_g\big),\ i = 1, \ldots, N_g, \end{aligned} \tag{3.10}$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$ and $\boldsymbol{\gamma}_{g,i\cdot} = \sigma^{-1}_{\eta_{g,i}} \boldsymbol{b}_{g,i\cdot} : 1 \times N_g(K_2 + 2)$.

The marginal treatment effect (MTE)[5] is defined as the effect of treatment (attending college for example) on those indifferent to undertaking treatment, given the characteristics X_g and the propensity score $P(\tilde{Z}_g)$, i.e.

$$MTE(x_g, u_{S_g}) = \mathbb{E}(\boldsymbol{y}_{1g} - \boldsymbol{y}_{0g} | X_g = x_g, U_{S_g} = u_{S_g}),\ g = 1, \ldots, m. \tag{3.11}$$

Carneiro et al. (2011, Eqs.(5)-(6)) show that (3.10) along with the definition of the

[5]This was originally developed by Björklund and Moffitt (1987) and extended in Heckman and Vytlacil (2005, 1999, 2007b)

uniform r.v. U_{S_g} implies $MTE(x_g, u_{S_g})$ can be expressed as:

$$\frac{\partial \mathbb{E}(\boldsymbol{y}_g | X_g = x_g, P(\tilde{Z}_g) = p_g)}{\partial p_g} = MTE(x_g, p_g),\ g = 1, \ldots, m, \tag{3.12}$$

where $\boldsymbol{y}_g$ is the observed outcome in (3.2). As U_{S_g} has been normalized to be unit uniform, tracing $MTE(x_g, u_{S_g})$ over u_{S_g} values shows how the returns to treatment vary with different quantiles of the unobserved component of the index of the desire to undertake the treatment. Alternatively, it is the mean return to treatment for persons indifferent between undertaking treatment or not who have mean scale utility value $P(\tilde{Z}_g) = u_{S_g}$. From (3.2), it is easy to see that

$$\begin{aligned} \mathbb{E}[\boldsymbol{y}_g | X_g = X_g, P(\tilde{Z}_g) = p_g] \ &= \ \mathbb{E}[\boldsymbol{y}_{0g} | X_g = x_g, P(\tilde{Z}_g) = p_g] \\ &\quad + \mathbb{E}[(\boldsymbol{y}_{1g} - \boldsymbol{y}_{0g}) | X_g = x_g, P(\tilde{Z}_g) = p_g] p_g, g = 1, \ldots, m. \end{aligned} \tag{3.13}$$

In particular, if Model (3.1) is linear, i.e. if $\mu_j(x_g) = x_g \beta_j$ for $j \in \{0, 1\}$, we have:

$$\mathbb{E}[\boldsymbol{y}_g | X_g = x_g, P(\tilde{Z}_g) = p_g] = x_g \beta_0 + p_g x_g (\beta_1 - \beta_0) + K(p_g) \tag{3.14}$$

where $K(p_g) = \mathbb{E}[(u_{1g} - u_{0g} | S_g = 1, P(\tilde{Z}_g) = p_g]$, which can be estimated non-parametrically, for example, using local polynomial regressions; see Fan and Gijbels (1996). As outlined in Carneiro et al. (2011), aggregating the instruments $\tilde{Z}_g$ into the scalar index $P(\tilde{Z}_g)$ enlarges the range of values over which we can identify the MTE in comparison to using each instrument one at a time.

3.2 Network Formation Model

The inclusion of the *unobserved within group characteristics*, ξ_g, affecting both the treatment decision and the formation of the *network* in (3.6) raises additional difficulties in estimating the marginal treatment effect in (3.12). Studies such as Patacchini and Zenou (2012) often assume that the inclusion of network and school correlated effect dummies can help to identify the linear model. In the current context, this

approach implies dropping ξ_g out of model (3.6). Patacchini and Zenou (2012) argue that once the group fixed effect is controlled for, peer group formation is random conditional on the network. Fletcher (2012a), Kremer and Levy (2008) and Patacchini et al. (2011) also use a similar approach.

Similarly, Bramoullé et al. (2009) partially account for the problematic environmental effects by using network-level unobservable characteristics. They show that these unobserved characteristics can be partialled out in the estimation process using a within network transformation of the model. However, this approach does not account for the within network individual level variations. Controlling for these variations is important, especially in networks with a high diameter, where linked individuals share common characteristics, but those further away in the network are dissimilar in these characteristics. In our model, within-network variations are represented by ξ_g. In a network of high school students, these unobserved characteristics may represent social or communication skills that affect friendships that are made during high school, but also are important determinants of whether an individual decides to attend college. Goldsmith-Pinkham and Imbens (2013) allows the inclusion of unobserved terms by utilising a network formation process to determine the network adjacency matrix. As such, we follow this method in controlling for within network individual level variations [also, see Hseih and Lee (2016)].

For simplicity, we use the single unobserved factor setting of Goldsmith-Pinkham and Imbens (2013) for the friendship formation model (i.e., $\xi_{g,i}$ is a scalar for each individual i), but the setup can be generalised to a multiple factor setting as in Hseih and Lee (2016). Formally, the friendship formation model can be written as:

$$\begin{gathered} U_{g,ij} = \theta_1' c_{g,i} + \theta_2' c_{g,j} + \theta_3' c_{g,ij} + \theta_4 D_{g\_,ij} + \theta_5 F_{g\_,ij} + \delta|\xi_{g,i} - \xi_{g,j}| + \varepsilon_{g,ij} \\ D_{g,ij} = \mathbb{1}[U_{g,ij} > 0] \cdot \mathbb{1}[U_{g,ji} > 0] \end{gathered} \tag{3.15}$$

where $U_{g,ij}$ represents the utility of individual i from a friendship link with individual j in network $\boldsymbol{D}_g$, $c_{g,i}$ and $c_{g,j}$ represent the observed individual specific variables which may affect friendship formation, $c_{g,ij}$ represents dyad-specific variables which

may be either dummy variables indicating the same characteristic between individuals i and j (e.g. race or sex) or the difference between two continuous individual characteristics (e.g. difference in age, difference in household income),[6] and the θ's are unknown parameters. Differences in the unobserved characteristics towards every potential friendship (i, j)(i.e., $|\xi_{g,i} - \xi_{g,j}|$) are considered in (3.15) as key factors of network formation. A low value of $|\xi_{g,i} - \xi_{g,j}|$ will result in a likely friendship pair (i, j), while a higher value indicates that i and j will not be friends. As such, the parameter δ can be viewed as a measure of friendship intensity in this model. $\boldsymbol{D}_{g\_}$ and $\boldsymbol{F}_{g\_}$ characterize the network in the previous period. In particular, $D_{g\_,ij}$ is a dummy variable indicating whether i and j were friends in the previous period, while $F_{g\_,ij}$ is a dummy variable representing whether i and j had friends in common in the previous period. Clearly, we see that at least *two observations (data points)* on the network are required to fully identify the friendship relations from (3.15). Because of this dynamic aspect, model (3.15) is more general than the one considered, for example, in Hseih and Lee (2016). Finally, $\varepsilon_{g,ij}$ represent the error terms which are uncorrelated with the unobserved characteristics $\xi_{g,i}$ and the errors $v_{g,i}$ of the treatment decision model (3.6).

Specifying a model for network formation is complex for many reasons. First, misspecification is usually an issue in these types of models. Second, endogeneity issues may introduce bias, so that estimates of these models should be interpreted with caution (Jackson, 2013). Many factors must be accounted for in order to minimise these biases. For example, model (3.15) controls for the effect of homophily through the inclusion of unobserved characteristics ($\xi_{g,i}$) and transitivity in link formation through the inclusion of $\boldsymbol{F}_{g\_}$. Graham (2013) points out that a network formation model should also account for degree heterogeneity, such that some individuals are naturally "good friends" and thus give greater utility to friendship. Model (3.15) accounts for this degree heterogeneity by controlling for the observed characteristics in the network formation (i.e., $c_{g,i}, c_{g,j}$).

For the remainder of the paper, we define $\boldsymbol{C}_g = \left\{c_{g,i}, c_{g,j}, c_{g,ij} | (i, j) \in \boldsymbol{G}_g\right\}$ and

[6]These individual characteristics may overlap with those in $Z_{g,i}$ in equation (3.6).

consider the following assumptions on the model variables and parameters.

Assumption 3.1 *Given* $\boldsymbol{D}_{g\_}$, C_g, *and* ξ_g, *each link of network* $\boldsymbol{D}_g$ *is* (*conditionally*) *independent of other links.*

Assumption 3.2 *The errors* $\varepsilon_{g,ij}$ *of the regression* (3.15) *are i.i.d. and follow a standard logistic distribution given* $\boldsymbol{D}_{g\_}$, C_g, *and* ξ_g.

Assumption 3.3 $(v_{g,i}, \xi_{g,i})' \mid \boldsymbol{D}_{g\_}, \boldsymbol{C}_g \overset{i.i.d.}{\sim} N\left(0, \begin{bmatrix} \sigma^2_{v_g} & 0 \\ 0 & 1 \end{bmatrix}\right)$ *for all* $g = 1, \ldots, m$.

Assumptions 3.1–3.3 are commonly used in the literature on social networks– e.g., see Goldsmith-Pinkham and Imbens (2013) and Hseih and Lee (2016). Three features often observed from network data are homophily, transitivity of relations and clustering (Wasserman and Faust, 1994; Jackson, 2008), and all suggest that link decisions in networks tend to be dependent. Assumption 3.1 states that link decisions in networks are independent after controlling for homophily, transitivity of relations, and clustering, i.e., the links of $\boldsymbol{D}_g$ can be dependent unconditionally.

If we define $\boldsymbol{W}_{g,ij} = \{c_{g,i}, c_{g,j}, c_{g,ij}, c_{g,ji}, D_{g\_,ij}, D_{g\_,ji}, F_{g\_,ij}, F_{g\_,ji}, \xi_{g,i}, \xi_{g,j}\}$, the probability of friendship given $\boldsymbol{W}_{g,ij}$ (i.e., given the exogenous covariates and the network of the previous period) can be expressed under Assumption 3.2, [see Goldsmith-Pinkham and Imbens (2013)] as:

$$\begin{aligned} q_g\big(D_{g,ij} | \boldsymbol{W}_{g,ij}; \theta, \delta\big) &=: \mathbb{P}[D_{g,ij} = 1 | \boldsymbol{D}_{g\_}, \tilde{Z}_g] = \mathbb{P}[U_{g,ij} > 0, U_{g,ji} > 0 | \boldsymbol{D}_{g\_}, \tilde{Z}_g] \\ &= \boldsymbol{\Lambda}\Big(\psi_{g,ij}(\theta, \delta)\Big)\boldsymbol{\Lambda}\Big(\psi_{g,ji}(\theta, \delta)\Big), \end{aligned} \tag{3.16}$$

where $\theta = (\theta_1', \theta_2', \theta_3', \theta_4, \theta_5)'$, $\boldsymbol{\Lambda}(\cdot)$ is the cdf of the standard logistic random variable, and

$$\psi_{g,ij}(\theta, \delta) = \theta_1' c_{g,i} + \theta_2' c_{g,j} + \theta_3' c_{g,ij} + \theta_4 D_{g\_,ij} + \theta_5 F_{g\_,ij} + \delta |\xi_{g,i} - \xi_{g,j}|.$$

Assumption 3.3, along with the definition of η_g in (3.8) implies (3.9). Note that ξ_g may be correlated with X_g in the potential outcome equation (3.1) under this

assumption, although exploring cases where ξ_g is independent of X_g is a reasonable case to consider. Therefore, Assumption 3.3 allows for a possible dependence between ξ_g and the errors u_j $(j = 0, 1)$ in (3.1), in which case we refer to X_g as an endogenous regressor in (3.1). The variance of $\xi_{g,i}$ is normalised to 1 in Assumption 3.3 because it is not identified in either the treatment model (3.6) nor in the friendship formation model (3.15); see Hseih and Lee (2016). In particular, this normalisation implies that θ and γ can only be identified up to σ_ξ^2 in (3.15).

Under Assumptions 3.1 & 3.2, the likelihood function of network $\boldsymbol{D}_g$, conditional on $\boldsymbol{W}_g$, is given by

$$\mathcal{L}_{net}^{\xi_g}(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta) = \prod_{i\neq j}\Big[q_g\big(D_{g,ij}|\boldsymbol{W}_{g,ij};\theta,\delta\big)\Big]^{D_{g,ij}}\Big[1-q_g\big(D_{g,ij}|\boldsymbol{W}_{g,ij};\theta,\delta\big)\Big]^{1-D_{g,ij}} \tag{3.17}$$

where $q_g\big(D_{g,ij}|\boldsymbol{W}_{g,ij};\theta,\delta\big)$ is defined by (3.16). Therefore, the likelihood function of network $\boldsymbol{G}_g$, conditional on $\boldsymbol{W}_g$, is obtained by integrating over ξ_g, i.e.

$$\mathcal{L}_{net}\big(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta\big) = \int_{\xi_g}\mathcal{L}_{net}^{\xi_g}\big(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta\big)\phi(\xi_g)d\xi_g, \tag{3.18}$$

where $\mathcal{L}_{net}^{\xi_g}\big(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta\big)$ is the conditional likelihood function in (3.17) and $\phi(\xi_g)$ is the density of $\xi_g \sim N(0, I_{N_g})$ under Assumption 3.1.

Similarly, we can use Bayes rule to express the joint likelihood function of the latent treatment decision S_g^* and the network $\boldsymbol{G}_g$, conditional on Z_g and $\boldsymbol{W}_g$, as:

$$\mathcal{L}^{\xi_g}\big(S_g^*,\boldsymbol{G}_g|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2,\theta,\delta\big) = \mathcal{L}_{treat}^{\xi_g}\big(S_g^*|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2\big)\times\mathcal{L}_{net}^{\xi_g}\big(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta\big) \tag{3.19}$$

where $\mathcal{L}_{treat}^{\xi_g}\big(S_g^*|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2\big)$ is the joint likelihood function of S_g^* conditional on Z_g and $\boldsymbol{W}_g$. Therefore, the joint likelihood function of S_g^* and $\boldsymbol{G}_g$, conditional on observables Z_g $\boldsymbol{C}_g$, and $\boldsymbol{D}_{g\_}$, is obtained by integrating (3.19) over ξ_g, i.e.

$$\begin{aligned}\mathcal{L}\big(S_g^*,\boldsymbol{G}_g|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2,\theta,\delta\big) &= \int_{\xi_g}\mathcal{L}^{\xi_g}\big(S_g^*,\boldsymbol{G}_g|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2,\theta,\delta\big)\phi(\xi_g)d\xi_g \qquad (3.20)\\ &= \int_{\xi_g}\mathcal{L}_{treat}^{\xi_g}\big(S_g^*|Z_g,\boldsymbol{W}_g;\gamma,\alpha_g,\sigma_{v_g}^2\big)\times\mathcal{L}_{net}^{\xi_g}\big(\boldsymbol{D}_g|\boldsymbol{W}_g;\theta,\delta\big)\phi(\xi_g)d\xi_g.\end{aligned}$$

As S_g^* is not observed, but we rather observe $S_g = 1$ when $S_g^* > 0$ and $S_g = 0$ otherwise, the usual probit method could be used to compute $\mathcal{L}_{treat}^{\xi_g}\big(S_g^*|Z_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma_{v_g}^2\big)$ from the reduced-form equation (3.8), along with Assumption 3.3. However, identification in standard probit models requires normalising the variances of the errors $\eta_{g,i}$ (i.e., $\sigma_{\eta_{g,i}}^2$) to 1, as it cannot be jointly identified with the remaining parameters in (3.8). This means that the standard probit estimation can only identify the parameters of the treatment equation (3.6) up to $\sigma_{\eta_{g,i}}^2$ $(i = 1, \ldots, N_g)$ at best. Although this may be a reasonable assumption to consider, and usually is the case in binary outcome models, we employ an alternative approach that consists of simulating the latent outcome variable S_g^* from a multivariate truncated normal (TMVN) distribution using the Gibbs sampling technique proposed by Geweke (1991), and then applying the truncated regression methods to derive $\mathcal{L}_{treat}^{\xi_g}\big(S_g^*|Z_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma_{v_g}^2\big)$. This also allows us to estimate the model within the Bayesian setting, controlling for the ξ_g terms without having to numerically integrate them out.

Indeed, it is well known that sampling from $S_g^* \sim TMVN(\mu, \Sigma_\eta)$[7] subject to the inequality constraints $a \leq S_g^* \leq b$ is equivalent to sampling from $\tau_g \sim N(0, \Sigma_\eta)$ under the linear constraints $\underline{b} \leq \tau_g \leq \bar{b}$, where $\underline{b} = a - \mu$, $\bar{b} = b - \mu$, and then constructing the sample for S_g^* as $S_g^* = \mu + \tau_g$. Following Geweke (1991), we build up the sample for τ_g from the conditional distribution of $\tau_{g,i}$ given $\tau_{g,-i}$ for all $i = 1, \ldots, N_g$, where $\tau_{g,-i} = \tau_g \setminus \tau_{g,i}$ denotes the vector formed with the components of τ_g other than $\tau_{g,i}$. Geweke (1991) shows that

$$\mathbb{E}(\tau_{g,i}|\tau_{g,-i}) = \gamma_{g,-i}\tau_{g,-i}, \tag{3.21}$$

where $\gamma_{g,-i} = -\omega_{g,ii}^{-1}\omega_{g,-i}$, $\omega_{g,ij}$ is the (i,j)th element of Σ_η^{-1} and $\omega_{g,-i}$ is the ith row of Σ_η^{-1} excluding the ith element. Therefore, we can model $\tau_{g,i}$ as:

$$\tau_{g,i} = \gamma_{g,-i}\tau_{g,-i} + h_{g,i}\nu_{g,i}, \tag{3.22}$$

where $h_{g,i} = \omega_{g,ii}^{-1/2}$ and $\nu_{g,i} \sim N(0,1)$ for all i and g. As $\underline{b} \leq \tau_g \leq \bar{b}$, it follows from

[7]Note that $\Sigma_\eta = \sigma_{v_g}^2\big[(I_{N_g} - \gamma_1 \boldsymbol{G}_g)'(I_{N_g} - \gamma_1 \boldsymbol{G}_g)\big]^{-1}$ from (3.8).

(3.22) that $\nu_{g,i}$ satisfies the constraints:

$$h_{g,i}^{-1}(\underline{b_i} - \gamma_{g,-i}\tau_{g,-i}) < \quad \nu_{g,i} < h_{g,i}^{-1}(\bar{b}_i - \gamma_{g,-i}\tau_{g,-i}), \tag{3.23}$$

where $\underline{b_i} = -\infty$, $\bar{b}_i = -\mu_i$ if $S_{g,i} = 0$ (i.e., if $S_{g,i}^* \leq 0$) and $\underline{b_i} = -\mu_i$, $\bar{b}_i = +\infty$ if $S_{g,i} = 1$ (i.e., if $S_{g,i}^* > 0$). $\nu_{g,i} \sim N(0,1)$ can be simulated with the restrictions in (3.23) and $\tau_{g,i}$ can be generated following (3.22). Thus the sample for $S_g^* = \mu + \tau_g$ can be built up using this method.

Applying the above results to model (3.8) gives the following conditional truncated normal distribution of the latent variable S_g^* given $\tilde{Z}_g = \tilde{z}_g$ [similar to Geweke (1991)]:

$$S_{g,i}^* | \tilde{Z}_g, \boldsymbol{W}_g \sim \begin{cases} N\big(\boldsymbol{b}_{g,i\cdot}\tilde{z}_g, \sigma_{\eta_g,i}^2\big), & \text{truncated at the left by 0 if } S_{g,i} = 1 \\ \\ N\big(\boldsymbol{b}_{g,i\cdot}\tilde{z}_g, \sigma_{\eta_g,i}^2\big), & \text{truncated at the right by 0 if } S_{g,i} = 0, \end{cases} \tag{3.24}$$

for all $i = 1, \ldots, N_g$ and all $g = 1, \ldots, m$, where $\boldsymbol{b}_{g,i\cdot}$ is the ith row of $\boldsymbol{b}_g$ defined in (3.8) and $\sigma_{\eta_g,i}^2$ is the (i,i)th element of Σ_η. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of the standard normal random variable, respectively. The conditional density of $S_{g,i}^*$, given $\tilde{Z}_g$ and $\boldsymbol{W}_g$, can be expressed from (3.24) as:

$$f(s_{g,i}^* | \tilde{Z}_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma_{v_g}^2) = \begin{cases} \frac{\phi\left(\tilde{s}_{g,i}^* - \boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g\right)}{\sigma_{\eta_g,i}\Phi(\boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g)} \text{ truncated from above at 0 if } S_{g,i} = 1 \\ \\ \frac{\phi\left(\tilde{s}_{g,i}^* - \boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g\right)}{\sigma_{\eta_g,i}\left(1 - \Phi(\boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g)\right)} \text{ truncated from below at 0 if } S_{g,i} = 0, \end{cases} \tag{3.25}$$

where $\sigma_{\eta_g,i}^{-1}\boldsymbol{\gamma}_{g,i\cdot} = \boldsymbol{b}_{g,i\cdot}$, $\tilde{s}_{g,i}^* = \sigma_{\eta_g,i}^{-1} s_{g,i}^*$. We can therefore express the likelihood function

$\mathcal{L}^{\xi_g}_{treat}\big(S^*_g|Z_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma^2_{v_g}\big)$ as:

$$\begin{aligned}\mathcal{L}^{\xi_g}_{treat}\big(S^*_g|Z_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma^2_{v_g}\big) &= \prod_{i=1}^{N_g} f(s^*_{g,i}|\tilde{Z}_g, \boldsymbol{W}_g; \gamma, \alpha_g, \sigma^2_{v_g}) \qquad (3.26)\\ &= \prod_{\{i:\, s_{g,i}=1\}} \frac{\phi\big(\tilde{s}^*_{g,i} - \boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g\big)}{\sigma_{\eta_g,i}\Phi(\boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g)} \prod_{\{i:\, s_{g,i}=0\}} \frac{\phi\big(\tilde{s}^*_{g,i} - \boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g\big)}{\sigma_{\eta_g,i}\big(1-\Phi(\boldsymbol{\gamma}_{g,i\cdot}\tilde{z}_g)\big)}\,.\end{aligned}$$

4 Model identification

The marginal treatment effect (MTE) given in (3.14) can only be identified if both the treatment decision equation (3.6) and the outcome equation (3.1) are identified. In this section, we provide the conditions under which both equations are identified. For simplicity, we shall focus on the case where the outcome (3.1) is linear, i.e. $\mu_j(x_g) = x_g\beta_{jg}$ for $j \in \{0, 1\}$. Since the identification of the MTEs depends on that of the treatment decision model (3.6), it will be useful to investigate the latter first.

Various sources are threats to identifying the treatment model (3.6). First, the inclusion of direct endogenous peer effects (i.e., $\boldsymbol{G}_g S^*_g$) as explanatory variables is a threat to identification because these endogenous effects are subject to Manski's (1993) reflection problem. Bramoullé et al. (2009) propose to employ an instrumental variable (IV) method, where friends' decision to undertake treatment (i.e., $\boldsymbol{G}_g S^*_g$) is instrumented with friends of friends' characteristics (i.e., $\boldsymbol{G}^2_g Z_g$). So, as long as $\boldsymbol{G}^2_g Z_g$ is a valid and strong matrix of IVs, the reflection problem can be solved by using, for example, a two-stage least squares (2SLS) estimation. Second, the inclusion of the friendship network unobserved characteristic ξ_g in (3.6) is vital for achieving identification. If ξ_g is not controlled for, the error of this regression will incorporate the unobserved network effects and will thus be correlated with $G_g Z_g$, $G_g S^*_g$ and possibly Z_g. We identify the treatment model (3.6) following a two-step approach. Firstly, the network $\boldsymbol{G}_g$ is formed using the model (3.15). And secondly, the parameters of the treatment model (3.6) are identified, given the constructed network $\boldsymbol{G}_g$, using 2SLS estimation. We summarize in Proposition 4.1 the conditions under which equations (3.6) and (3.15) are identified.

Proposition 4.1 *Suppose that Assumptions* 3.1–3.3 *are satisfied. Then the following two statements are true.*

(*a*) *Conditional on* $\boldsymbol{W}_g$, θ *and* δ *are identified in* (3.15).

(*b*) *If* (*a*) *is satisfied and* $|\gamma_1| < 1$, $\gamma_2\gamma_1 + \gamma_3 \neq 0$, $\boldsymbol{G}_g^2 \neq 0$, *then* $\gamma = (\gamma_1, \gamma_2', \gamma_3', \gamma_4)'$ *and* α_g *are identified in* (*3.6*) *for all* $g = 1, \ldots, m$.

Proof of Proposition 4.1:

The proof of (a) follows straightforwardly from the logistic distributional assumption of $\varepsilon_{g,ij}$ in Assumption 3.2 along with Assumption 3.1, therefore it is omitted. The proof of (b) follows identical steps as that of Proposition 3 of Bramoullé et al. (2009), hence is also omitted to simplify the exposition.

While the inclusion of a network formation process allows us to estimate the effect of unobservables in the treatment decision equation, the nonlinearity of the outcome equation does not allow an equivalent approach using the proposed estimation method. We can include observed variables in the outcome equation in an attempt to control for any network correlated characteristics which also influence the outcome. While an imperfect approach, if network effects are minimal as in our empirical setting, endogenous effects are not a problem. We assume that the outcome of one individual is independent from their peers, controlling for the effect of peers directly on the treatment, i.e. there are no factors outside of the treatment through which an individual may influence a peer.

5 Estimation

We follow the networks and spatial literature (e.g., see Goldsmith-Pinkham and Imbens (2013) and LeSage and Pace (2009)) to estimate the networks and treatment models simultaneously using a Bayesian method. To do this, we define the following prior distributions for the parameters θ, δ, $\bar{\gamma}$, the unobserved characteristics ξ, and

the latent variable S_g^*:

$$\begin{aligned} (\theta', \delta)' &\sim N(\phi_0, \Phi_0),\ \xi_{i,g} \sim N(0,1), \\ \bar{\gamma} = (\gamma_2' \ \gamma_3' \ \gamma_4 \ \alpha_g')' &\sim N(\gamma_0, \Gamma_0),\ \gamma_1 \sim U[-1,1], \end{aligned} \tag{5.1}$$

where $\phi_0 : (q+3) \times 1$, $\Phi_0 : (q+3) \times (q+3)$, $\gamma_0 : (2k+m+1) \times 1$, and $\Gamma_0 : (2k+m+1) \times (2k+m+1)$ are fixed. These prior distributions in (5.1) are commonly used in Bayesian literature (Hseih and Lee, 2016; LeSage and Pace, 2009). The prior distribution for γ_1 is restricted to the interval $[-1,1]$ to ensure the matrix $(I_{N_g} - \gamma_1 \boldsymbol{G}_g)$ is non-singular, and we use the normalization $\sigma_v = 1$. The posterior distributions needed to estimate the model are constructed as follows.

1. First, we construct the posterior distribution of $\xi_{i,g}$ as:

$$P(\xi_{i,g}|S_g^*, \boldsymbol{G}_g, \xi_{-i,g}, \bar{\gamma}, \theta, \delta, \sigma_v, \alpha_g) \sim \pi(\xi_{i,g}) \cdot P(S_g^*, \boldsymbol{G}_g|\xi_g, \bar{\gamma}, \theta, \delta, \sigma_v, \alpha_g), \tag{5.2}$$

   where $P(S_g^*, \boldsymbol{G}_g|\xi_g, \bar{\gamma}, \theta, \delta, \sigma_v, \alpha_g)$ is given by the likelihood in (3.19).

2. The conditional posterior for $(\theta', \delta)'$ can be simplified to

$$P(\theta, \delta|\boldsymbol{G}_g, \xi_g) \propto \pi(\theta, \delta) \cdot \prod_{g=1}^{m} P(\boldsymbol{G}_g|\xi_g, \theta, \delta). \tag{5.3}$$

3. The posterior for γ_1 is constructed as

$$P(\gamma_1|S_g^*, \boldsymbol{G}_g, \xi_g, \bar{\gamma}, \sigma_v, \alpha_g) \sim \prod_{g=1}^{m} P(S_g^*|\boldsymbol{G}_g, \xi_g, \gamma_1, \bar{\gamma}, \sigma_v, \alpha_g). \tag{5.4}$$

4. Following Albert and Chib (1993), the posterior of $\bar{\gamma} = (\gamma_2' \ \gamma_3' \ \gamma_4 \ \alpha_g')'$ is con-

structed as:

$$P(\bar{\gamma}|S_g^*, \boldsymbol{G}_g, \xi_g, \gamma_1, \sigma_v) \propto N(\bar{\gamma}; \gamma_0, \Gamma_0) \cdot \prod_{g=1}^{m} P(S_g^*|\boldsymbol{G}_g, \xi_g, \gamma_1, \bar{\gamma}, \sigma_v, \alpha_g).$$

$$\therefore P(\bar{\gamma}|S_g^*, \boldsymbol{G}_g, \xi_g, \gamma_1, \sigma_v) \propto N(\bar{\gamma}; \gamma_0^*, \Gamma_0^*),$$

$$\begin{aligned}
\gamma_0^* &= (\frac{1}{\sigma_v^2}\sum_{g=1}^{m} \tilde{Z}_g'\tilde{Z}_g + \Gamma_0^{-1})^{-1}(\frac{1}{\sigma_v^2}\sum_{g=1}^{m} \tilde{Z}_g' A_g S_g^* + \Gamma_0^{-1}\gamma_0), \\
\Gamma_0^* &= (\frac{1}{\sigma_v^2}\sum_{g=1}^{m} \tilde{Z}_g'\tilde{Z}_g + \Gamma_0^{-1})^{-1}, \\
A_g &= I_{N_g} - \gamma_1 \boldsymbol{G}_g.
\end{aligned} \tag{5.5}$$

Initial values for all parameters and unknowns are chosen by the user. We set all initial values to zero for simplicity. Parameters and unknowns are updated sequentially using an MCMC approach. Given the likelihood in (3.26), we draw samples from the posterior distributions given above at each stage of the MCMC. We run the algorithm for $T = 30000$ iterations, where the first 20000 iterations are discarded. At the tth iteration, the following steps are taken:

1. The Metropolis-Hastings algorithm is used to draw samples $\xi_{i,g}^{(t)}$ from the posterior distribution $P(\xi_{i,g}|S_g^{*(t-1)}, \boldsymbol{G}_g, \xi_{-i,g}^{(t-1)}, \bar{\gamma}^{(t-1)}, \theta^{(t-1)}, \delta^{(t-1)}, \sigma_v, \alpha_g^{(t-1)})$ given in (5.2), where $\xi_{-i,g}^{(t-1)} = (\xi_{1,g}^{(t)}, \ldots, \xi_{i-1,g}^{(t)}, \xi_{i+1,g}^{(t-1)}, \ldots, \xi_{N_g,g}^{(t-1)})$. This occurs for every individual $i = 1, \ldots, N_g$ and network $g = 1, \ldots, m$. The M-H procedure is as follows:

   (1) Propose $\tilde{\xi}_{i,g} \sim N(\xi_{i,g}^{(t-1)}, \kappa_\xi^2)$, where κ_ξ^2 is chosen by the user, and let $\tilde{\xi}_g = (\xi_{1,g}^{(t-1)}, \ldots, \xi_{i-1,g}^{(t-1)}, \tilde{\xi}_{i,g}, \xi_{i+1,g}^{(t-1)}, \ldots, \xi_{N_g,g}^{(t-1)})$. The value of κ_ξ^2 is adjusted to achieve an acceptance rate between 20% and 40%.

   (2) With probability equal to $a(\xi_{i,g}^{(t-1)}; \tilde{\xi}_{i,g}) =$

$$\min\left\{ \frac{P(S_g^*, \boldsymbol{G}_g|\tilde{\xi}_g, \bar{\gamma}^{(t-1)}, \theta^{(t-1)}, \delta^{(t-1)}, \sigma_v, \alpha_g^{(t-1)})}{P(S_g^*, \boldsymbol{G}_g|\xi_g^{(t-1)}, \bar{\gamma}^{(t-1)}, \theta^{(t-1)}, \delta^{(t-1)}, \sigma_v, \alpha_g^{(t-1)})} \cdot \frac{N(\tilde{\xi}_{i,g}; 0, 1)}{N(\xi_{i,g}^{(t-1)}; 0, 1)}, 1 \right\},$$

set $\xi_{i,g}^{(t)}$ equal to $\tilde{\xi}_{i,g}$, otherwise, set it to $\xi_{i,g}^{(t-1)}$.

2. The M-H procedure is used to sample $(\theta^{(t)'}, \delta^{(t)})'$ from $P(\theta, \delta|\boldsymbol{G}_g, \xi_g^{(t)})$ given in (5.3):

   (1) Propose $(\tilde{\theta}', \tilde{\delta})' \sim N_{q+3}\big((\theta^{(t-1)'}, \delta^{(t-1)'}), \kappa_{\theta,\delta}^2 I_{q+3}\big)$, where $\kappa_{\theta,\delta}^2$ is chosen by the user.

   (2) With probability equal to $a(\theta^{(t-1)}, \delta^{(t-1)}; \tilde{\theta}, \tilde{\delta}) =$

$$\min\left\{\prod_{g=1}^{m}\frac{P(\boldsymbol{G}_g|\xi_g^{(t)}, \tilde{\theta}, \tilde{\delta})}{P(\boldsymbol{G}_g|\xi_g^{(t)}, \theta^{(t-1)}, \delta^{(t-1)})}\cdot\frac{N_{q+3}(\tilde{\theta}, \tilde{\delta}; \phi_0, \Phi_0)}{N_{q+3}(\theta^{(t-1)}, \delta^{(t-1)}; \phi_0, \Phi_0)}, 1\right\},$$

   set $(\theta^{(t)'}, \delta^{(t)})'$ equal to $(\tilde{\theta}', \tilde{\delta})'$, otherwise, set it to $(\theta^{(t-1)'}, \delta^{(t-1)})'$.

3. γ_1 is sampled from the distribution $P(\gamma_1|S_g^{*(t-1)}, \boldsymbol{G}_g, \xi_g^{(t)}, \bar{\gamma}^{(t-1)}, \sigma_v, \alpha_g^{(t-1)})$ in (5.4) using the M-H algorithm, proceeding as follows:

   (1) Propose $\tilde{\gamma}_1 \sim N(\gamma_1^{(t-1)}, \kappa_{\gamma_1}^2)$, where $\kappa_{\gamma_1}^2$ is chosen by the user.

   (2) Let $A = [-1, 1]$, with probability equal to $a(\gamma_1^{(t-1)}; \tilde{\gamma}_1) =$

$$\min\left\{\prod_{g=1}^{m}\left(\frac{P(S_g^*|\boldsymbol{G}_g, \xi_g^{(t)}, \bar{\gamma}^{(t-1)}, \tilde{\gamma}_1, \sigma_v, \alpha_g^{(t-1)})}{P(S_g^*|\boldsymbol{G}_g, \xi_g^{(t)}, \bar{\gamma}^{(t-1)}, \gamma_1^{(t-1)}, \sigma_v, \alpha_g^{(t-1)})}\right)\cdot\frac{I(\tilde{\gamma}_1 \in A)}{I(\gamma_1^{(t-1)} \in A)}, 1\right\},$$

   set $\gamma_1^{(t)}$ to $\tilde{\gamma}_1$. Otherwise, set it to $\gamma_1^{(t-1)}$.

4. The Gibbs sampling method is used to draw samples for $\bar{\gamma}^{(t)} = \big(\gamma_2^{(t)'},\ \gamma_3^{(t)'},\ \gamma_4^{(t)},\ \alpha_g^{(t)'}\big)'$ from the posterior distribution $P(\bar{\gamma}|S_g^{*(t-1)}, \boldsymbol{G}_g, \xi_g^{(t)}, \gamma_1^{(t)}, \sigma_v)$ in (5.5). The sign of γ_4 will not be determined, as $|\xi_{g,i} - \xi_{g,j}|$ is not affected by a change in the signs of $\xi_{g,i}, \xi_{g,j}$. To address this, we fix γ_4 to be positive using the acceptance-rejection algorithm.

5. Following LeSage and Pace (2009); Geweke (1991), $S_g^{*(t)}$ is sampled from the truncated multivariate normal distribution, $P(S_g^*|\boldsymbol{G}_g, \xi_g^{(t)}, \gamma_1^{(t)}, \bar{\gamma}^{(t)}, \sigma_v, \alpha_g^{(t)})$ given in (3.24).

6. The propensity scores are calculated using the definition in (3.10). For simplicity, we initially assume that (X_g, Z_g) are independent of (u_{0g}, u_{1g}, v_g). Then we obtain the propensity scores as follows:

   (a) The Peter M. Robinson (1998) method for estimating partially linear equations is used to obtain estimates of β_0 and $\beta_1 - \beta_0$:

      i. The difference between the outcome equation and its expected value is taken to remove the non-linear component in P:

$$\boldsymbol{y}_g - \mathbb{E}\big(\boldsymbol{y}_g | P(\tilde{Z}_g)\big) = \Big[x_g - \mathbb{E}\big(x_g | P(\tilde{Z}_g)\big)\Big]\beta_0 + p_g\Big[x_g - \mathbb{E}\big(x_g | P(\tilde{Z}_g)\big)\Big](\beta_1 - \beta_0). \tag{5.6}$$

      ii. Kernel regressions of the dependent variable and each of the regressors are run on P_g in order to estimate the expected values in equation (5.6).

      iii. The residuals from these kernel regressions are regressed on each other to determine β_0 and $\beta_1 - \beta_0$.

   (b) Following (3.14), a local polynomial regression of $\boldsymbol{y}_g - x_g\hat{\beta}_0 - P(\tilde{Z}_g)x_g(\hat{\beta}_1 - \hat{\beta}_0)$ is run on $P(\tilde{Z}_g)$ to estimate the function $K(P(\tilde{Z}_g))$ and its partial derivative with respect to $P(\tilde{Z}_g)$[8]. Adding the partial derivative to $x_g(\hat{\beta}_1 - \hat{\beta}_0)$ results in an estimate for the MTE:

$$MTE = x_g(\hat{\beta}_1 - \hat{\beta}_0) + \frac{dK\big(P(\tilde{Z}_g)\big)}{dP(\tilde{Z}_g)}.$$

   (c) To estimate the marginal policy relevant treatment effect (MPRTE), a weighted average of the MTE is taken across the support of $P(Z_g)$. The relevant weight is expressed conditional on the value of X_g and must be integrated over the distribution of X_g.[9] We can measure different forms

[8] Fan and Gijbels (1996) recommend using a local quadratic estimator for fitting a first order derivative. We therefore use a local quadratic estimator with a bandwidth that minimises the residual square criterion proposed by Fan and Gijbels.

[9] Since conditioning on X_g is computationally demanding due to the possible high dimension of X_g, as in Carneiro et al. (2011) we condition on the index $X_g(\beta_1 - \beta_0)$ as an approximation.

of the MPRTE using different weighting functions and definitions of the policy change. In particular we consider the following three scenarios: (1) a policy change that directly increases the probability of treatment equally for all individuals, i.e., $P_\alpha = P_g + \alpha$; (2) a policy that proportionally increases the probability of treatment, i.e., $P_\alpha = P_g(1 + \alpha)$; and (3) a policy that affects one of the instruments used in the treatment equation, i.e., $Z_\alpha = Z_g + \alpha$.

7. Relaxing the assumption of exogeneity, identification of the model is considered under the new assumption that Z_g is independent of (u_{0g}, u_{1g}, v_g) given X_g. The MTE is identified over the support of $P(\tilde{Z}_g)$ holding the values of X_g constant. The process of identification of the MTE is as follows:

   (a) The value of $(I_{N_g} - \gamma_1 G_g)^{-1} X_g$ is calculated at the $25th$ and $75th$ percentile of the distribution of $(I_{N_g} - \gamma_1 G_g)^{-1} X_g(\beta_1 - \beta_0)$.

   (b) Holding $(I_{N_g} - \gamma_1 G_g)^{-1} X_g$ constant at this point, the instruments are allowed to vary.

   (c) The MTE is calculated over the sections of $P(\tilde{Z}_g)$ that are supported as the instruments are allowed to vary.

   The MPRTE is a weighted average of the MTE, placing weights only on those sections of the MTE that are identified. This MPRTE is calculated conditional on the value of X_g so that the MPRTE is still identified even under the endogeneity of X_g.

6 Monte Carlo Simulations

We simulate data in order to study the performance of the model. The MCMC algorithm is iterated 30000 times, with the first 20000 samples discarded to obtain estimates of the parameters to be estimated. This is repeated for $R = 30$ different data samples to obtain estimates of the standard error. The data is simulated as

follows:

$$[u_{1g}, u_{0g}, v_g, X_g, Z^*_{g,1}, Z_{g,2}, \xi_g, c_{g,1}] \sim N(\mu, \Sigma),$$

where,

$$\mu = (0, 0, 0, 1, 0.8, 0, 0.5, 0.2)',$$

and Σ is an 8×8 matrix with ones along the diagonal. The correlations between variables are set to zero except for:

$$\mathbb{E}(X_g' Z^*_{1g}) = \mathbb{E}(X_g' Z_{2g}) = 0.2,\ \mathbb{E}(Z_{1g}'\xi) = 0.1,\ \mathbb{E}(Z_{2g}'\xi) = -0.2$$
$$\mathbb{E}(u_{0g}' u_{1g}) = 0.3,\ \mathbb{E}(u_{0g}' v_g) = 0.3,\ \mathbb{E}(u_{1g}' v_g) = -0.5.$$

$Z_{g,1i}$ is a binary variable, such that $Z_{g,1i} = \mathbb{1}_{[Z^*_{g,1i} \geq 0]}$. We assume that ξ_g is exogenous in this setting. ε_g is generated according to a logistic distribution with mean 0 and standard deviation 1. The network formation variables are constructed as:

$$c_{g,i} = c_{g,1i}, \quad c_{g,j} = \emptyset, \quad c_{g,ij} = \left[|Z_{g,2i} - Z_{g,2j}|,\ \mathbb{1}_{[Z_{g,1i}=Z_{g,1j}=1]},\ \mathbb{1}_{[Z_{g,1i}=Z_{g,1j}=0]}\right].$$

Two networks of size 250, $\boldsymbol{D}_1, \boldsymbol{D}_2$, are constructed using the network formation model in (3.15) at two time periods. Individuals with no friends are removed from the sample. The network formation process is calibrated to ensure the number of individuals with no friends is less than 20% of the original sample and the average number of friends is greater than 10. We set

$$\theta_0 = 0.2,\ \theta_1 = 0.2,\ \theta_3 = (-0.9, 0.4, 0.5)',\ \theta_4 = 0.2,\ \theta_5 = 0.1,\ \delta = -1.2.$$

On average, 7 individuals with no friends are removed from the network, and the remaining individuals have an average of 13.65 friendship connections. The average dynamics of the network are given in Table 1.2. If individuals are friends in period 0, they are likely to remain friends in period 1 (2.5% of possible friendships). Similarly, if individuals are not friends in period 0, they are unlikely to form a new friendship in period 1 (3.2% of possible friendships). Due to the large number of possible

friendships (23100), we see more new friendships in period 1 than existing ones carried from period 0.

Some network statistics are included in Table 1.1. An example of a generated network can be seen in Figure 1.1. We see that the generated network is relatively dense for the size of the network, due to the high number of average friendships. We use this high degree in order to establish stronger results for the peer effects. Average path length and diameter are low, with the average distance between any two individuals being 2.62. On average, one large component is generated by the network formation process. There is some evidence of clustering with 15% of possible triplets closed. Connections occur mostly on the same side of the network, with fewer cross network links. This aligns with the homophily constructed in the model such that an individual is likely to connect with their friend's friends who are also close in homophilic characteristics. This clustering would be more pronounced with greater correlation between these variables in the network formation. As the generated variables are relatively uncorrelated, homophily occurs independently across the characteristics.

One potential drawback of the proposed network model is the lack of dependence on the number of friends. As a result, we observe a high variance in the number of friends. While some individuals only have one network connection, others have very many (the maximum in the plotted network is 32).

The outcome variables S_g and $\boldsymbol{y}_g$ are constructed using equations (3.6) and (3.2) respectively, with the coefficients:

$$\gamma_1 = 0.05,\ \gamma_2 = (1, 1, 0.5)',\ \gamma_3 = (0.5, 0.5, 0.2)',\ \gamma_4 = 0.8,\ \alpha = (-2.8, -2.8)',$$
$$\beta_0 = (0, 0.8)',\ \beta_1 = (0.5, 1)'.$$

Under this construction an average of 55% of individuals undertake treatment and the average value of Y_g is 1.48, with a standard deviation of 2.01.

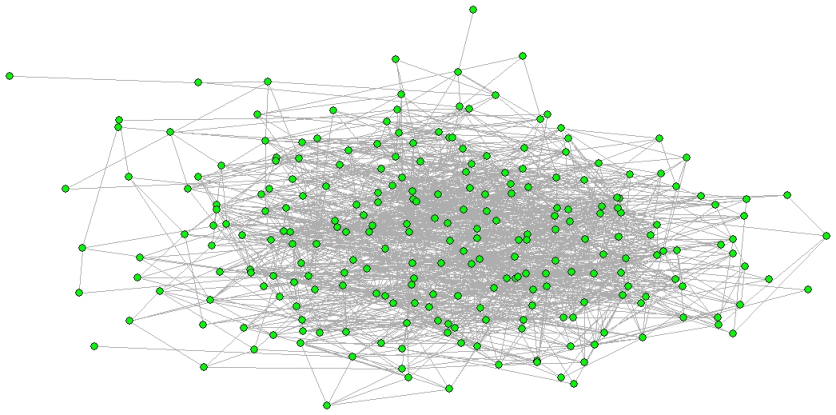


Figure 1.1: Generated network using (3.15)

Table 1.1: Networks Summary

Property	Mean	S.D.
Number of Nodes	242.74	2.97
Number of Friendships	1659	140.42
Link Density	0.06	0.005
Average Degree	13.67	1.16
Clustering Coefficient	0.15	0.01
Number of Components	1.08	0.27
Average Path Length (of largest component)	2.62	0.08
Diameter (of largest component)	6.56	0.79

Number of Networks = 2

6.1 Exogeniety of X_g

As described in the estimation section above, we begin with the assumption that X_g is strictly exogenous, i.e. $\mathbb{E}[X_g' u_{1g}] = \mathbb{E}[X_g' u_{0g}] = 0$ for all g. We present results for the full model (model I), the model with no unobserved ξ_g term (model II) and the model with no network effects (model III). Table 1.3 displays the estimates of the network formation process for the full model. All estimates except θ_4 in the network formation are highly significant with the correct signs. θ_1 and θ_3 appear to be overestimated, while θ_4 and θ_5 are underestimated. We would expect the network in the previous period $D_{g\_}$ to be highly correlated with the characteristics C_g, so this result is not surprising, and should not affect the estimation of the ξ_g parameters. We note that the coefficients in the network formation and treatment equation (δ, γ_4)

Table 1.2: Dynamic Friendships

		Period 1	
		Friends	Not Friends
Period 0	Friends	614 (2.48%)	249 (1.01%)
	Not Friends	790 (3.19%)	23100 (93.32%)

Table 1.3: Estimation in Network Formation

Variable	Estimate	Variable	Estimate
θ_1	0.2422***	θ_4	0.1416*
	(0.0180)		(0.0810)
$\theta_{3,1}$	−0.8344***	θ_5	0.0768***
	(0.0207)		(0.0228)
$\theta_{3,2}$	0.5470***	δ	−0.7922***
	(0.0417)		(0.0486)
$\theta_{3,3}$	0.6744***		
	(0.0370)		

* p < 0.1, ** p < 0.05, *** p < 0.01.

absorb σ_ξ (normalised to one in this construction).

Table 1.4 contains the estimates for the coefficients in the treatment equation. As we are assuming a probit type specification, the variance is not identified. The model with no network components, Model (III), performs poorly, especially when identifying $\gamma_{2,1}$. In general, the simplistic model underestimates the parameters. Standard errors are relatively small, and all estimates are significant.

The intermediate model that does not include the unobserved ξ_g (Model II) offers an improvement in the estimated coefficients, coupled with an increase in the standard errors. In this case, only the coefficients on individuals' characteristics and the fixed effects are significant. None of the peer effect terms are significant due to to high standard error estimates.

Moving to the full model (Model I) that now accounts for the presence of unobservable peer effects in the estimation, we see some improvements. Firstly, the endogenous effect estimate (γ_1) has doubled in size compared to Model II and is significant. The individual characteristic coefficient (α's) estimates are relatively

unchanged, while the peer effects estimates (γ's) have reduced in magnitude. The standard error estimates have also decreased both set of variables, but the accuracy of the estimates themselves seems close to Model II. In particular, the same individual characteristics are significant in both Model I and Model II. Endogenous effects are significant at the 5% level, though exogenous effects remain insignificant. We may be concerned with the coefficient of the unobserved terms, γ_4, which in this setup is positive, but not statistically significant to high standard error estimated, probably due to a relatively small number of simulations.

Computational limits impede the ability to run the estimation with more iterations or with a greater sample size. We run the 30 simulations in parallel on the University of Adelaide's Phoenix High Performance Computing service. The simulations with two networks of size 250 recorded an average running time of 1 day, 14 hours. We note that while the results are reasonable, larger sample size and number of iterations will improve the accuracy of the estimation.

The MTE for each model is plotted in Figure 1.2. The MTE curve in these figures is calculated at the mean value of X_g (note that X_g is exogenous in this subsection). The MTE is identified over the support of P_g which can be seen in Figure 1.3. We see that low values of X_g are associated with low values of P_g. The more complex models increasingly polarise the predicted propensities, so that propensities become more highly correlated with the observed treatment. We identify the MTE at areas where we have common support for P_g, i.e. when we observe values of P_g for both $S = 0$ and $S = 1$. On average, Model (III) has support on the interval **(0.0237, 0.8796)**, Model (II) has support on **(0.0443,0.9633)** and Model (I) has support over **(0.0536,0.9568)**. We see that the introduction of peer effects gives us a fairer spread of propensities, allowing us to identify a greater portion of the MTE. The regions of the MTE that are not identified are those areas of higher variance, corresponding to the extreme values of the propensity scores. Estimation of traditional treatment measures such as the average treatment effect (ATE) or treatment effect for the treated (TT), require support over the entire [0,1] interval. As such, these measures cannot be identified in the model (and rarely are: Carneiro et al. (2011)), making

Table 1.4: Estimation of the Treatment Paramaters, X_g exogenous

Variable	(I)	(II)	(III)
γ_1	0.3624**	0.1893	-
	(0.1810)	(0.2422)	
$\gamma_{2,1}$	0.6262***	0.6142***	0.3773***
	(0.0669)	(0.0666)	(0.0505)
$\gamma_{2,2}$	0.6630***	0.6642***	0.4382***
	(0.1809)	(0.2114)	(0.1251)
$\gamma_{2,3}$	0.3127**	0.6918***	0.3012***
	(0.1405)	(0.2145)	(0.0551)
$\gamma_{3,1}$	0.0553	0.2939	-
	(0.1769)	(0.2144)	
$\gamma_{3,2}$	−0.1306	0.2004	-
	(0.2650)	(0.5007)	
$\gamma_{3,3}$	−0.0090	0.1116	-
	(0.2234)	(0.3548)	
γ_4	0.1022	-	-
	(0.0810)		
α_1	−1.1579***	−1.7283***	−1.3130***
	(0.2669)	(0.3629)	(0.1755)
α_2	−1.1541***	−1.7146***	−1.3198***
	(0.2773)	(0.3781)	(0.1594)

* p < 0.1, ** p < 0.05, *** p < 0.01.

Note: Model (I) - full model with network formation process
Model (II) - excludes ξ and the network formation process
Model (III) - excludes all network terms.

the MTE and MPRTE more practical and relevant measures of the true treatment effect.

The MTE is traced across values of U_S, corresponding to the quantiles of v_g. A high value of U_S corresponds to a low probability of treatment, while a low value of U_S corresponds to a high probability of treatment. At a high value of U_S, individuals at the margin have a corresponding high propensity score P_g. The MTE at this point is defined as the expected increase in the outcome variable Y_g when P_g is varied so the individual is induced into treatment, starting at a high value of P_g. The converse is true for a low value of U_S. In this construction of the model, we clearly see that a high probability of treatment is associated with a high return to treatment (around 90% in each model), with the opposite observed for a low probability of treatment

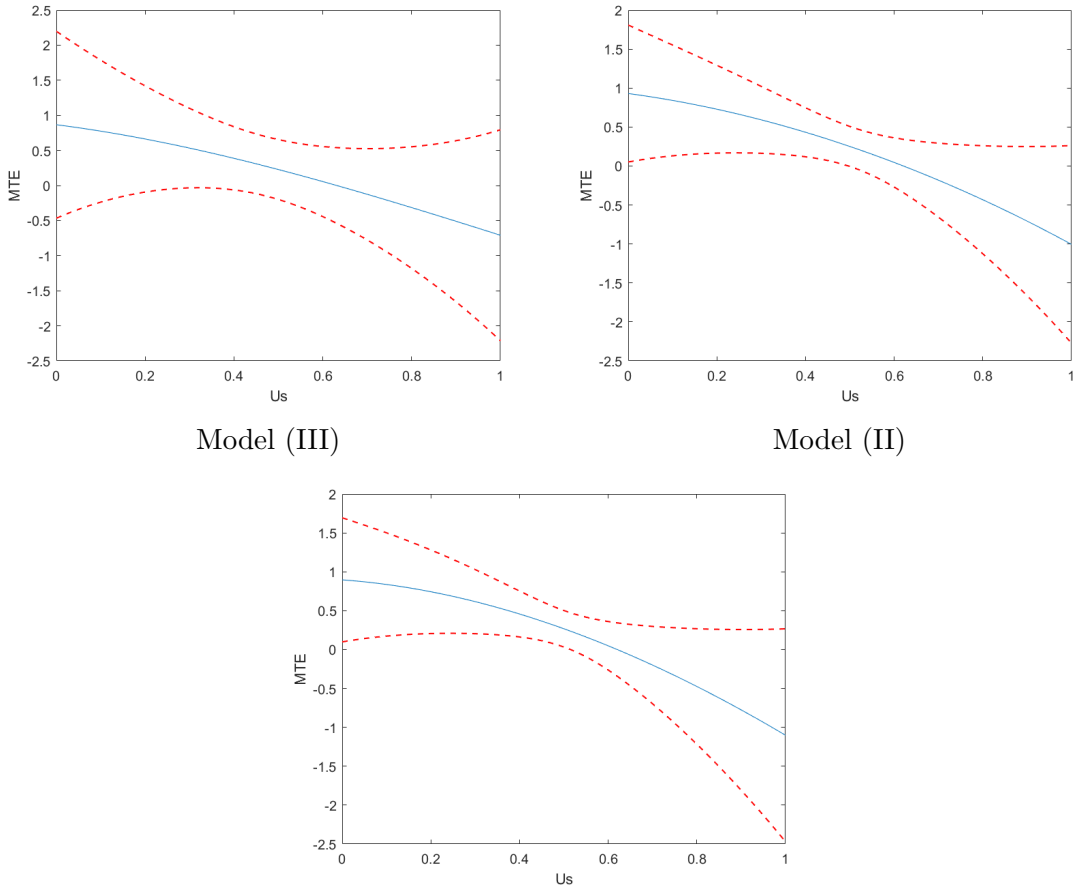
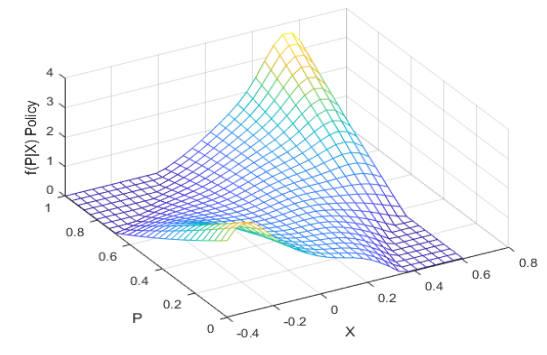


Figure 1.2: MTE curve with 90% confidence bands

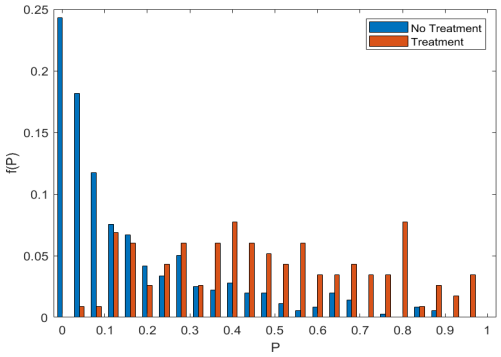
(around -90%).

MTE estimates in each model appear relatively similar. The model with the full estimation appears to estimate slightly more extreme results for high and low values of U_S, with the intermediate model splitting the results of the other two models. Under the full model, the marginal treatment effect ranges from 0.89 for low U_S to -1.1 for high U_S, compared to 0.8 and -0.7 for model (III). The shape of the curve also differs, with model (I) showing a steeper, more convex curve. Most notably, there is a progressive improvement in the width of the 90% confidence bands moving towards the more robust model. In this case, there is evidence that the full network model enables more precise identification of the true returns to treatment.

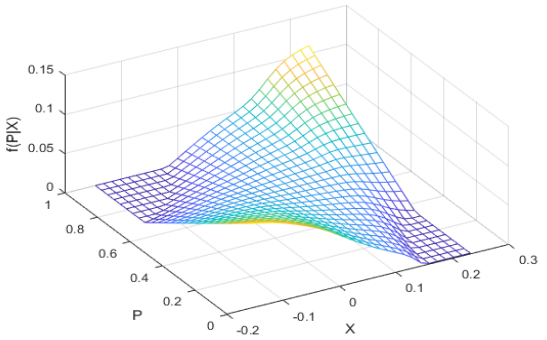
The mean line in the three plots in figure 1.2 are clearly downwards sloping so that



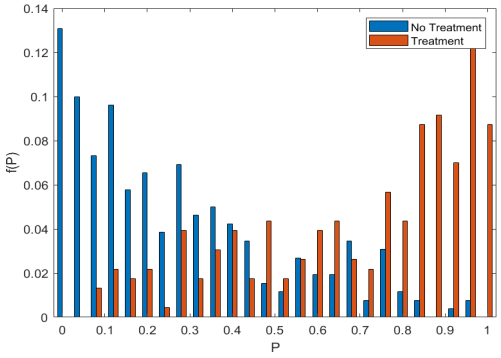
(a) Model (III)



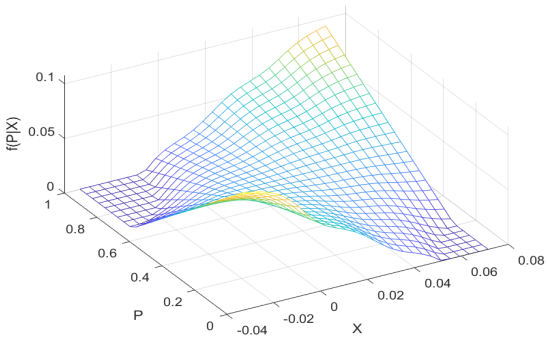
(b) Model (III)



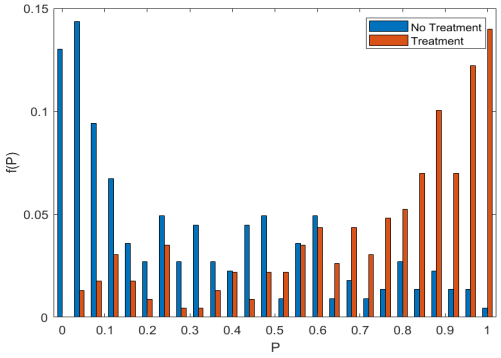
(c) Model (II)



(d) Model (II)



(e) Model (I)



(f) Model (I)

Figure 1.3: Support of P_g given X_g, support of P_g for $S = 0$ and $S = 1$

individuals self select into treatment. We can test the MTE curve for zero slope at each point, as well as running a joint p-test for an overall constant slope as described by Carneiro et al. (2011). The results of these tests are given in appendix table

2.1. The MTE curve is statistically downwards sloping for the middle sections of the MTE, from around 0.25 to 0.75. There is more variation at the ends of the MTE, and we cannot statistically determine a negative slope. The p-value for a constant slope across the whole MTE is 0.1333, so we cannot reject at the 90% level that the overall slope of the MTE curve is non-constant. For the more simplistic models, these p-values are clearly higher, and it is difficult to conclude a non-constant trend.

Table 1.5 reports the estimates of the MPRTE for the three different changes in policy. The models involving network effects appear to estimate a higher marginal policy effect for those induced into treatment. Using the more complex models we have greater significance, with models (I) and (II) displaying lower standard errors and achieving positive significance.

Table 1.5: MPRTE, X_g exogenous

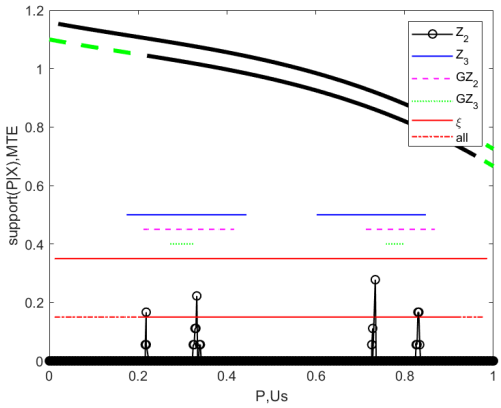
Policy Change	(I)	(II)	(III)
$Z^k_\alpha = Z^k + \alpha$	0.6797***	0.6293**	0.4880
	(0.2630)	(0.2848)	(0.3707)
$P_\alpha = P + \alpha$	0.6680**	0.6185**	0.4823
	(0.2637)	(0.2858)	(0.3373)
$P_\alpha = (1+\alpha)P$	0.5561***	0.5219**	0.4209
	(0.2083)	(0.2303)	(0.2940)

* p < 0.1, ** p < 0.05, *** p < 0.01.

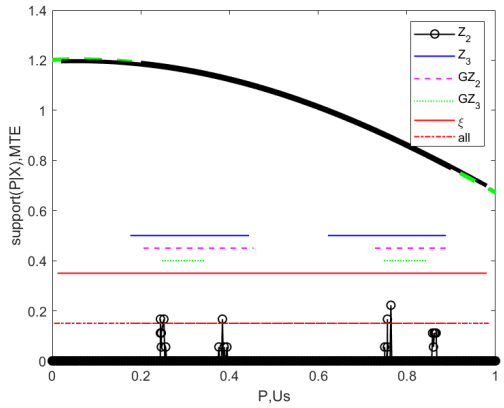
6.2 Endogeniety of X_g

The tables and figures discussed thus far rely on the strong assumption that X_g and Z_g are both exogenous of the error terms (u_{0g}, u_{1g}, v_g). This assumption would not hold, for example, in a wage equation, where X_g represents educational achievement or GPA. If unobserved ability is also important, X_g will not be exogenous.

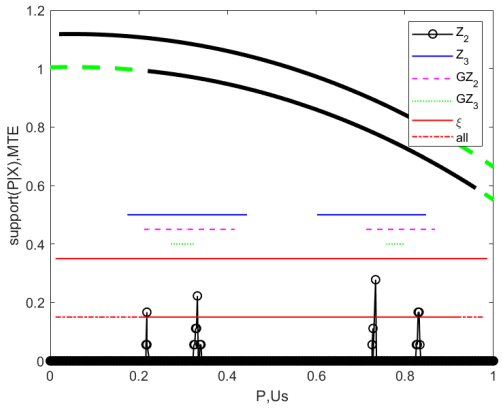
Relaxing the assumption of the exogeneity of X_g, we examine the model under weak endogeneity ($corr(X_g, u_{1g}) = corr(X_g, u_{0g}) = 0.3$) and strong endogeneity ($corr(X_g, u_{1g}) = corr(X_g, u_{0g}) = 0.75$). Results in both cases reflect those in the case where X_g is exogenous. In this case all coefficients are still exogenous of the treat-



$corr(X_g, U) = 0$



$corr(X_g, U) = 0.3$

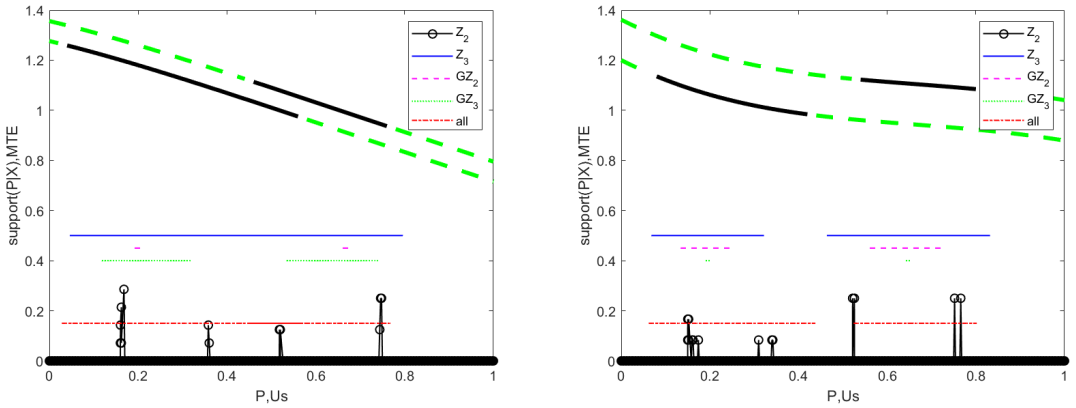


$corr(X_g, U) = 0.75$

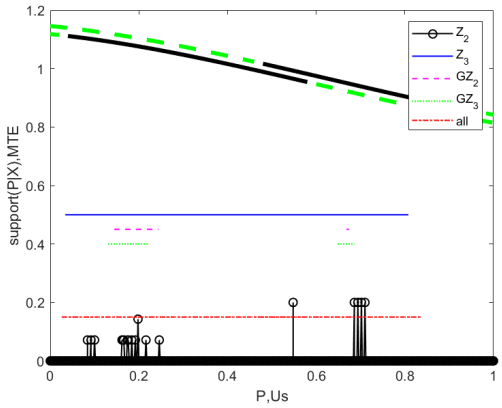
Figure 1.4: MTE support, Model (I)

ment error v_g, so endogeneity in the outcome will not have an effect. The estimation of the MTE, however, will be affected. The MTE is plotted in appendix figures 2.1 and 2.2. Under this model the MTE curves will not be identified, and clearly in 2.1 and 2.2, the plotted MTE is not stable. Stronger endogeneity has greater effects on the shape of the MTE, and completely removes the downwards slope we would like to see.

In figures 1.4-1.6 we present the identification of the MTE when X_g is endogenous. Identification is displayed in each of the models (I)-(III) respectively. When X_g is held constant, we rely on the variation of the instruments Z_g to achieve identification. $(I_{N_g} - \gamma_1 G_g)^{-1} X_g$ is held constant at its $25th$ and $75th$ percentile with each instrument varied collectively and individually to produce the figures below. As in the MTE,



$corr(X_g, U) = 0$ $corr(X_g, U) = 0.3$

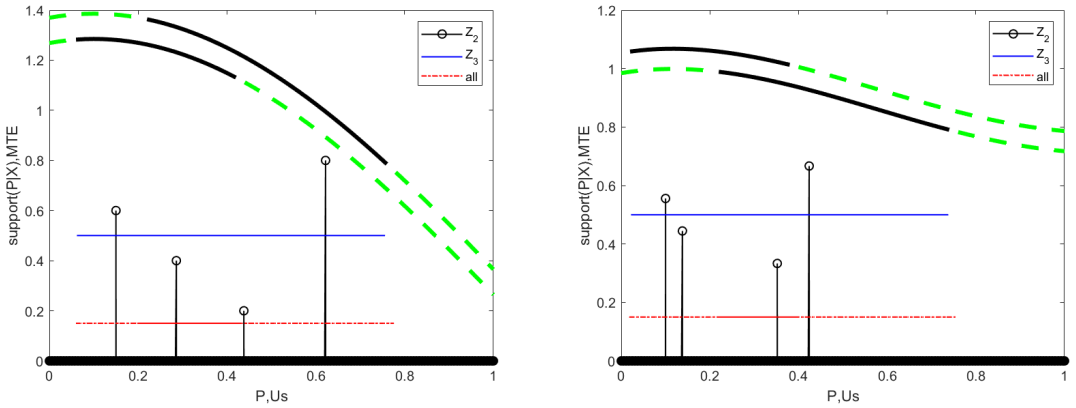


$corr(X_g, U) = 0.75$

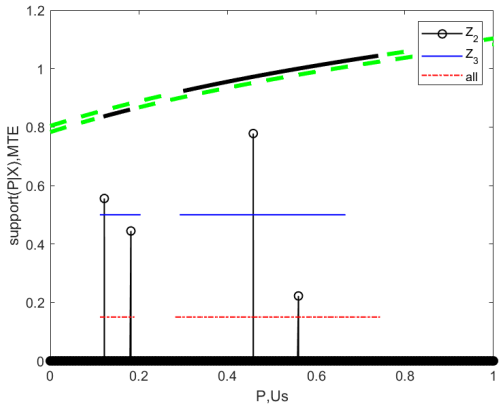
Figure 1.5: MTE support, Model (II)

we condition on $(I_{N_g} - \gamma_1 G_g)^{-1} X_g(\beta_1 - \beta_0)$ as an approximation. The black lines represent the portion of the MTE that is identified when X_g is endogenous and held constant at its 25th and 75th percentile.

The dashed green line represents the sections of the MTE we do not identify. The MTE is identified over the support of P_g when the instruments are varied. The red line corresponds to this interval of P_g that is identified when all instruments are allowed to vary. The lines above this correspond to the portion of the MTE identified when a single instrument is allowed to vary and all other instruments are held at their mean values. As Z_1 is a binary variable, the support associated with varying Z_1 are only single points in the black circles. The blue lines correspond to the portion of the MTE identified when Z_2 is allowed to vary.



$corr(X_g, U) = 0$ $corr(X_g, U) = 0.3$



$corr(X_g, U) = 0.75$

Figure 1.6: MTE support, Model (III)

The higher order models gift us additional instruments for identification. The exogenous network effects ($G_g Z_g$) and unobserved effects (ξ_g) now act as instruments, allowing greater variation in the collective instruments and achieving a greater region of identification for the MTE. Of course, this relies on the assumption that G_g and ξ_g are independent of (u_{0g}, u_{1g}). The shape of the MTE also appears to be affected. The slope of the MTE under model (III) increases as the endogeneity increases. This seems to indicate the inability of this simplistic model to identify the true MTE once endogeneity is introduced. In the full model, however, the MTE maintains its negative slope. The full model not only gives a more precise identification, but it extends the range of identification as more instruments are added.

7 Discussion

The results of the network formation process are encouraging. All terms, except θ_4 are highly significant. The model generally performs well in estimating most of these coefficients. We note that all coefficients attain the correct sign, particularly the coefficient of the unobserved characteristics.

Comparing the results of the treatment equation, we see that the simple model with no peer effects (model (III)), while having the lowest standard errors, is subject to bias in its estimates. The simplistic model is problematic for two reasons. Firstly, individual characteristics are correlated with the characteristics of their peers. By omitting peer effects, these individual characteristics become endogenous. Secondly, the simplistic model assumes standard normality of the error term. If the endogenous effect is truly significant, then in the reduced form the error term will have variance given by $[(I_{N_g} - \gamma_1 G_g)'(I_{N_g} - \gamma_1 G_g)]^{-1}$. This will result in misidentification in the simpler model, reflected in the underestimation of the coefficients in the treatment equation.

While the increase in standard error moving to the peer effect models (models (I) and (II)) may be initially concerning, we note that 1. standard errors are not identified in this treatment equation and, 2. variance is dependent on the estimation of the γ_1 term, which also varies between samples. Another source of of variation is generated in the addition of the N new parameters that must be estimated, ξ_g. Each individual ξ_g is randomly generated, introducing considerable variation in each sample, but reducing endogeneity bias. It is therefore surprising that moving from model (II) to model (I), we observe a reduction in the standard error. Clearly adequately controlling for endogeneity in unobserved characteristics is important in improving the precision of the treatment estimation.

The introduction of peer effects and the network formation process allows us to control for homophilic tendencies between individuals. The effect of homophily in the network model is to polarise the sample. There are both direct and indirect effects working here. Firstly, if individuals choose friendships based on a characteristic which

also influences the probability of treatment; then the decision of these individuals are likely to be correlated. Friends with shared characteristics will make similar decisions. Thus, independent of any peer effect, the decisions of these individuals will look similar. Exogenous and endogenous peer effects reinforce this. If an individual has similar characteristics to their friends, then the exogenous effect of a friend will reinforce the effect of the characteristic itself. Taking parental education as an example; if an individual has college educated parents, they may choose friends who also have college educated parents. Their friends will be more likely to enter college due to the college education of their parents, influencing the individual directly through the endogenous effect of similar decisions. The individual is also influenced by the average characteristics of their peers, i.e. by the college education of their friend's parents. This exogenous effect will work in parallel, reinforcing the decision of the individual to attend college. Thus the network model predicts more extreme propensities as the peer effects reinforce individual characteristics.

Individuals who were initially likely to enter treatment, now are further incentivised by the similarity of their friends, while those unlikely are further dissuaded. We see evidence of this polarisation in both the MTE plots 1.2, and in the support plots 1.3. We observe more extreme propensities in the support plots, as the predicted probabilities are emphasised by peer effects. As a result, the higher and lower values of U_S will be more extreme, and we would expect more extreme returns at the highest and lowest quantiles of v_g. The fuller model more accurately accounts for the reinforcement effects of peers, and predicts more polarised propensity scores for individuals entering treatment. The model that does not account for ξ_g, while an improvement on the basic model, does not adequately control for endogeneity in the network and still underestimates the true effect. Thus introducing network effects and adequately controlling for endogeneity in the network is critical when estimating such a model.

The clear downwards slope in each of the MTE curves implies that the individual selects into treatment based on their expected return, i.e. if you observe a low expected return to treatment, you will be unlikely to choose to enter into the treatment

group. The improvement in the confidence bands is most notable. Even with the greater standard errors in the treatment equation, the fuller models have narrower confidence bands when we estimate the MTE and MPRTE.

While the full model improves the confidence interval for the MTE, the width still may be of concern. This can partly be attributed to the difficulties in using non-parametric estimation for smaller datasets (we would expect these to narrow with sufficient sample size). The fuller model does, however, cause estimation of the MTE to become both more accurate and more extreme.

The positive MPRTE in the estimation implies that individuals induced into treatment would attain positive returns from doing so. This holds for all types of marginal policy changes, whether a direct increase in propensity or an increase in the instruments, although increasing propensities multiplicatively appears to result in the smallest return. The peer effect model estimates a higher return to these marginal policy changes. The model predicts that those at the margin are now more likely to attain higher returns. Homophily appears to work positively in this case, such that those who should obtain treatment have already done so, and those with negative returns are dissuaded from entering treatment. Those at the margin have more moderate characteristics and are therefore likely to receive higher returns from entering treatment than that predicted by a non-peer effect model.

We also note that as the peer effects are functioning as instruments in the model, these MPRTE values indicate that an increase in peer effects would also result in positive returns for those induced into treatment by this change. Peer effects can be used as a tool in this case by policy makers to encourage treatment, leading to positive returns.

Turning to the results of the endogenous estimation, we see that the proposed model performs well in comparison to models (II) and (III). In this setting (holding X_g fixed and creating identification through the variation in instruments), the network paramaters act as additional instruments. Exogenous peer effect terms are less useful than the non-peer instruments. Exogenous peer effects are simply averages of the original individual characteristics and therefore exhibit less variation than the

characteristics themselves (X_g has an average standard deviation of approximately 1.4, while $G_g X_g$ has an average standard deviation of of 0.6). The parameter which is of some use is the unobserved peer characteristics. In particular, these characteristics generally do not have a high correlation with the other variables, and vary considerably between individuals, independent of the outcome. As a result, these unobserved characteristics are particularly effective instruments.

We must acknowledge the limitations in the specification of the proposed network model. It is unlikely that the network model accurately reflects the network formation process, though it should be sufficient to identify most of the unobserved component and it certainly offers an improvement on neglecting these effects. The assumption that individuals are either close in observed or unobserved characteristics is a strong one. In reality, friendship formation is difficult to model and is reliant on many unobserved, intangible factors, including significant randomness which is unlikely to adhere to any normality assumptions.

In our model we assume that peer effects only influence the binary treatment decision, exogenous of the outcome. It is trivial to include exogenous peer effects in the outcome equation, as long as these terms are truly exogenous. Introducing endogenous and unobserved effects is a more complex process. The non-linearity of the outcome equation complicates the Bayesian estimation, and while theoretically possible, we do not include such an extension here.

In conclusion, the addition of peer effects is important in correcting for potential endogeneity in the model. The introduction of the network formation process allows us to correct for homophily in characteristics and as a result, we observe more polarised estimates in the MTE, and larger estimates for the returns to policy changes. Estimates of the MTE and MPRTE become more accurate and more precise. Furthermore, the peer effect components can act as additional instruments, which is particularly useful when the X_g terms are endogenous, and lead to a greater region of identification.

Chapter 2

Application to High-School Networks and College Attainment Decision

1 Introduction

High schools act as incubators of peer effects; creating an environment where strong friendship connections are established, and social norms are firmly enforced. Significant spillovers in achievement and other outcomes are easily observable between students. As a result, the classroom has become a popular area for peer effect analysis. Sacerdote et al. (2011) summarises the burgeoning literature on peer effects in education. The literature demonstrates strong evidence of the importance and strength of peer effects, especially regarding social outcomes such as delinquent behaviour. When considering peer effects in achievement, the exogenous effect of peer background is moderate. Estimates of the endogenous effect are mixed, but are generally moderate to large.

In this application, we are concerned with the peer effect of high school students on the decision to attend college, and the indirect effect this has on wages. The literature generally supports the hypothesis of positive peer effects on college attendance and future wages.

Even in high school, the networks an individual finds themselves in are likely to influence future wages. The common adage that getting a job is more about

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

who you know than what you know seems to hold elements of truth. Considering peer effects, Kramarz and Skans (2014) find that the parents of classmates are an important determinant of where an individual finds employment. Similarly, Black et al. (2013) find that the average paternal income of one's high school peers has a small exogenous peer effect on the future wage of the individual. In contrast, Bifulco et al. (2014) find little evidence of persistent peer effects on wages, but suggest that the exogenous effect of having peers with college educated mothers is positive and significant. Furthermore, they find that the delay in college education due to the decision of peers not to attend college has a temporary effect on wages. Taking networks as a proxy for social adeptness, Galeotti and Mueller (2005) and Barbone and Dolton (2015) find that the social skills associated with network position in high school lead to a significant increase in future earnings.

Evidence of the influence of peers on the decision to attend college is more widespread. Bifulco et al. (2011) find that peers with college educated mothers create a positive exogenous peer effect on the probability of attending college and reduce the likelihood of dropping out of high school. Fletcher (2012b) and Wu (2015) focus on the exogenous peer effect of college decisions, both finding strong social influence of friendship networks on the likelihood of college enrolment and college preferences. Fletcher (2013) finds that an increase in classmates attending college by 10% leads to an increase in the probability of an individual attending college by a significant 2-3%. De Giorgi et al. (2010) find a significant peer effect on the college major decided by students when at college, which influence both academic achievement outcomes and entry wages.

College educated individuals earn significantly more than those with only high school education. According to the OECD (2016), the increase in earnings in 25 OECD countries from a bachelors level degree is around 48%. As we have already explored, both endogenous and exogenous peer effects play an important role in determining the decision to attend college. Clearly, if peers are playing an important role in determining which students are attending college, there will be a corresponding influence on wage outcomes. The result mentioned of Bifulco et al. (2014), provides

positive evidence for this indirect college effect. Following a similar intuition, we utilise the model presented in chapter 1 to estimate the role of high school peers in the returns to education.

High school networks hold significant potential in the dissemination of knowledge and the effectiveness of policy. In addition to the obvious impacts of networks (such as social multiplier effects), other more subtle network-based phenomena may be pertinent to education policy changes. For example, changes in social norms have been shown to significantly affect student achievement through intra-school networks and the learning of these social norms. Del Bello et al. (2015) argue that policies that utilise the networks within a school, particularly focused on social norms, are highly effective compared to neighbourhood based policy changes. As evidence, Kremer et al. (2009) show that the introduction of scholarships for some high school students can affect all students in the school, even those ineligible or unlikely to attain the scholarship. Therefore, the effect of networks and peer groups should be considered when education policy is enacted. As such, we use the approach of Carneiro et al. (2011) to estimate the potential effect of networks on policy changes in our model and in the Add-Health data set.

In what follows, we apply the model presented in chapter 1 to a dataset containing the networks of high school students. We measure the marginal returns to education these individuals receive, when networks influence the decision of students to attend college. We examine the role these networks play in the treatment decision and on the marginal returns to education. We then demonstrate and discuss the influence these networks hold on changes to policy in the education setting.

2 Data Description

The Add Health Dataset is a longitudinal study across high schools in the US. Surveys were conducted for 90,118 individuals in school years 7-12 in representative high schools during the 1994-1995 school year. A core sample of 20,745 students were selected to take part in a detailed in-home survey across four waves; wave I: 1994-

1995, wave II: 1996, wave III: 2001-2002 and wave IV: 2008. A saturated sample comprised 16 schools, where all students in the school were selected were selected for this core sample. These 16 heterogeneous schools included two large schools (total enrolment in excess of 3100) and 14 smaller school (enrolment fewer than 300 each). One of the large schools is located in a mid-sized town with a predominantly white enrolment, while the other is located in a metropolitan area and is ethnically diverse. The smaller schools are a mix of public and private schools located in rural and urban areas. In waves I and II, students were asked to name up to 10 of their closest friends (5 male, 5 female). We use these friendship rosters to construct high school networks in waves I and II. Friendships named outside of the selected school are excluded, as these individuals are not in the sample.

Wave II responses are used to construct most variables for the treatment and outcome equations, while observed wages and some contextual variables are collected from wave IV. Individuals who did not take part in all of waves I, II and IV are removed from the sample. We also remove individuals with no friendship nominations (409 individuals) as they experience no measurable peer effect. After removing individuals with incomplete data or no friends we have a total sample of $N = 1696$ observations across $g = 15$ schools. Due to the high computational capacity required for estimation of large networks, we limit estimation to the 13 smaller schools with a total sample size of $N = 631$. Using the University of Adelaide's Phoenix High Performance Computing service, estimation of these 13 networks takes approximately 23 hours. The larger networks take over a week to estimate, so results are not reported here. It is difficult to obtain many iterations across such a time horizon, so standard errors cannot be estimated without greater computational capacity.

Summary statistics are given in table 2.3. Descriptions of the variables can be found in figure 2.4. Network properties are given in table 2.1. Figure 2.1 displays the networks of two schools in our sample. On average, individuals in the network have 3.85 friends. Only two networks are completely connected, but most have a large component dominating the network. The network statistics reflect the small size of the schools; link density is comparatively high while average path length and

diameter are smaller. Each individual is on average 3.68 links away from anyone in their network, with a maximum distance of 9.08 per network. Friends tend to cluster together, with few connections across the network. The clustering coefficient indicates that just over a third of possible triplets are closed, i.e. you are likely to be connected with your friend's friends. This pattern is clear in the second graph in figure 2.1, with evidence of three distinct groups. Table 2.2 gives a transition matrix of friendships from period 0 to period 1. Less than half of friendships in period 0 are still present in the next period. In fact, the number of new friendships is not far below the number of existing friendships transferred from period 0.

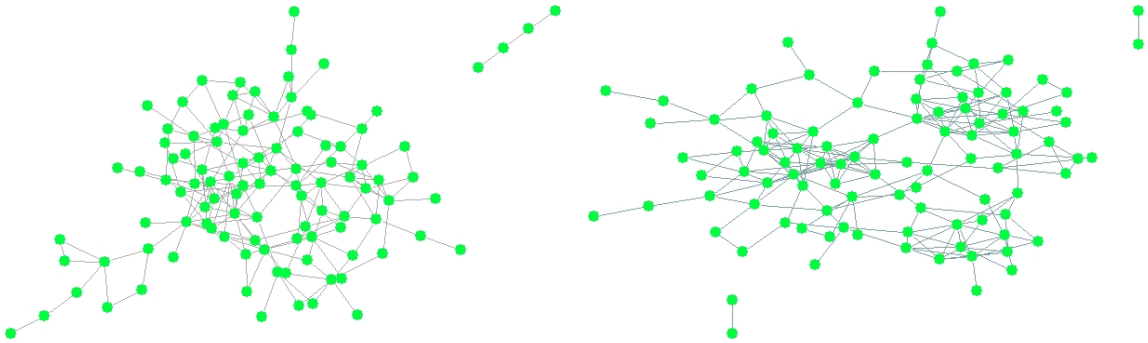


Figure 2.1: School networks

Table 2.1: Networks Summary

Property	Mean	S.D.
Number of Nodes	48.54	26.19
Number of Friendships	93.77	58.08
Link Density	0.11	0.08
Average Degree	3.85	1.02
Clustering Coefficient	0.35	0.14
Number of Components	2.30	1.55
Average Path Length (of largest component)	3.68	0.91
Diameter (of largest component)	9.08	2.63

Number of Networks = 13

As with most survey data, we must be careful when interpreting the results due to the potential effect of measurement error. In particular, self reported variables such as wage and subject grade are liable to misreporting or unconscious bias. However, most variables we are concerned with are easily verifiable (race, gender, college

Table 2.2: Transition of friendships from period 0 to period 1

		Period 1	
		Friends	Not Friends
Period 0	Friends	661 (3.52%)	726 (3.86%)
	Not Friends	558 (2.97%)	16860 (89.66%)

decision). We assume that the sample covers the majority of in-school friendships. While the survey restricts students to naming 10 close friends, only one student in our sample reaches this limit.

As in chapter 1, we treat friendships as a binary relation and normalise the network so that the average peer effect is equal for every individual regardless of the number of friends. Each of an individual's friends are given equal weight. These are necessary and reasonable simplifications of the true friendship network, but as a result, estimates may underestimate the true influence of peer effects.

The variables X_g, Y_g and Z_g are defined in table 2.4. We take the levels of all variables, except wage and local income, for which logs are taken. We use terms of order one, without interactions in each variable for simplicity. The instruments in the treatment equation are number of siblings, innovations in local labour market variables and the proportion of college education in the local area. As in Carneiro et al. (2011), we utilise innovations in labour market area statistics. Specifically we use local income in 1990 and local unemployment in 1993, observed around the time of the first wave. If unemployment is higher than normal and wages are lower, the individual will be more likely to pursue further education rather than enter the labour force directly. X_g contains local income and unemployment in 2008, when wage is observed, to control for long term labour market conditions in the area. The use of these instruments is discussed in Cameron and Heckman (1998), Cameron and Heckman (2001) and Cameron and Taber (2004), who follow a similar procedure in estimating the returns to education. Independent of local labour market conditions, the proportion of college education in the local area will clearly affect the decision to attend college, but should not impact on resulting wages. Similarly, we expect

number of siblings to influence the likelihood of attending college, but not observed wage in the long run.

Table 2.3: Data Summary

Variable Name	Min	Max	Mean	SD
Wage Y_g (\$)	0	300000	30386	27531
College Attendance S	0	1	0.71	0.45
Age (months)	130	222	162.1	18.92
Male	0	1	0.47	0.50
Female	0	1	0.53	0.50
Race:				
White	0	1	0.84	0.36
Black	0	1	0.14	0.35
Other race	0	1	0.05	0.22
Mother's education:				
High School	0	1	0.46	0.50
College	0	1	0.39	0.49
Less than High School	0	1	0.10	0.30
Father's education:				
High School	0	1	0.33	0.48
College	0	1	0.35	0.47
Less than High School	0	1	0.11	0.31
Grade in Maths	0	4	2.54	1.34
Grade in English	0	4	2.84	1.14
Appearance	1	5	3.65	0.78
Shy	1	5	3.31	1.26
Independent	1	5	1.80	0.76
Instruments:				
Number of siblings	0	14	2.59	1.99
Local income 1990	3817	28501	11155	4090
Local unem. 1993	0.02	0.15	0.07	0.02
College ed. 1990	0.03	0.48	0.18	0.11
Variables in X_g only:				
Local income 2008	8500	94950	23419	10230
Local unem. 2008	0	0.35	0.07	0.04
Married	0	1	0.55	0.50
Number of Children	0	5	0.90	1.08

Table 2.4: Description of Variables

Variable Name	Description
Outcome variable Y_g, wage	Earnings reported at wave IV (Individuals who do not report exact earnings are asked to report earnings within a range of values. We take the mid-value of this range.)
Treatment variable S_g, college attendance	Dummy variable indicating education level is at least some college
Variables in X_g and Z_g	
Age	Age of participant at the time of the first wave (1993)
Female	Dummy variable indicating female, male
Race	Dummy variables indicating white, black or other race
School	Dummy variable indicating the school of the respondent
Mother's Education	Dummy variables indicating the respondent's mother 1. graduated high school and 2. attended college
Father's Education	Dummy variables indicating the respondent's father 1. graduated high school and 2. attended college
GPA	Sum of reported grades in Mathematics and English; A=4, B=3, C=2, D=1, no grade=0
Appearance	Response from the surveyor on a Likert scale to the question "How physically attractive is the respondent?"
Personality variables	Variables on a Likert scale for Shyness and Independence (E.g. Response to the question "How much do you agree with the statement 'You are Shy.'"
Variables in Z_g, not X_g	
Number of Siblings	Variable of the number of reported siblings
Local income at wave II	Income per capita at the local tract area level in 1990
Local unemployment at wave II	Unemployment at the local tract area level in 1990
Local proportion with a college degree at wave II	Proportion of residents over 25 who hold a bachelor degree or higher, taken at the local tract area level
Variables in X_g, not Z_g	
Married	Dummy variable indicating if the respondent has ever been married
Number of Kids	Variable of the number of reported children
Local income at wave IV	Income per capita at the local tract area level in 2008
Local unemployment at wave IV	Unemployment at the local tract area level in 2008

3 Results

We follow the model and estimation method proposed in chapter 1 to estimate the treatment, outcome and returns to education (Model I). We run the MCMC procedure $t = 30000$ times, discarding the first 20000 iterations. To obtain standard errors, we repeat this process $R = 30$ times. We compare these results to those derived from the model with no network components (Model III). In this case we bootstrap the sample to obtain standard errors.

Table 2.5 presents the results of the network formation process for the full model. Most terms, except those in $C_{g,i}$ are significant. Characteristics such as shyness or independence do not statistically determine the number of friends an individual has. The homophily effects all prove to be important in the model. In particular, individuals select friendships based on similarity in age, race, gender, parent's education, GPA and even appearance. Having been friends in the previous period, or having friends in common in the previous period are particularly strong determinants of friendship formation. Similarity in unobserved characteristics has a negative effect on friendship choices. While this effect is reasonably large, it is not highly significant in this estimation.

Table 2.6 presents the results of the estimation of the treatment equation. Again most terms are significant. Individual effects are all highly significant, with race, gender and parent's education relatively important. Most of the corresponding exogenous effects are significant, particularly having college educated parents and having friends with a higher GPA. Although being black has a positive effect on treatment, having black friends appears to negatively impact the decision to attend college. The effect of unobserved characteristics, ξ_g is small but significant. Three of our instruments are significant; number of siblings and proportion of college education positively influence the college decision. Local income has a strong negative effect, as we would predict.

The results of the model with no peer effects are in table 2.7. Most estimates for the coefficients of X_g are comparable, with some appearing to absorb the peer effects

Table 2.5: High School Network Estimation

Variable	Estimate	Variable	Estimate
$C_{g,i}$		Father: College	0.1661***
Shy	−0.0578		(0.0330)
	(0.0732)	Male	0.1671***
Independent	−0.1373		(0.0390)
	(0.1089)	Female	0.0381***
No. of Siblings	−0.0189		(0.0142)
	(0.0224)	White	0.4123***
$C_{g,ij}$			(0.1179)
$\lvert Age_{g,i} - Age_{g,j}\rvert$	−0.4687***	Black	0.4494***
	(0.0076)		(0.0836)
$\lvert GPA_{g,i} - GPA_{g,j}\rvert$	−0.0641***	$G_{g\_}$	2.0982***
	(0.0038)		(0.1251)
$\lvert APER_{g,i} - APER_{g,j}\rvert$	−0.0665***	$F_{g\_}$	0.3973***
	(0.0159)		(0.0394)
Father: No School	−0.2923***	$\lvert\xi_{g,i} - \xi_{g,j}\rvert$	−1.1199*
	(0.1117)		(0.6533)
Father: High School	−0.0540***		
	(0.0197)		

* p < 0.1, ** p < 0.05, *** p < 0.01.

of the full model. However, estimates of the instruments are considerably reduced, while the fixed effects are much lower. The variances in the base model are high, particularly for the fixed effects.

As in Carneiro et al. (2011), we divide the MTE by four to obtain annualised estimates. The result in figure 2.2 is quite striking. Those with low values of U_S (and a high probability of attending college) have a high return (49.5%), while those with high values of U_S have a negative return (−72%). A 49.5% return indicates that these individuals would expect to have a wage 49.5% higher if they attended college than if they didn't attend college. The MTE is identified over the range (0.1063,0.9695), where it has common support, while the base model is identified over the region (0.1435,0.9729). The two models are comparable at low values of U_S, although neither are identified in this region. At high U_S, the full model predicts a less negative return. The confidence bands for the MTE are much wider for the basic model, as we saw in chapter 1. We test the difference in LATE's for a negative slope. The p-values in table 2.2 demonstrate an undeniably downwards slope at each

Table 2.6: High School Treatment Estimation

Variable	Estimate	Variable	Estimate
Endogenous effect	−0.0271**	GPA	0.0830***
	(0.0124)		(0.0028)
Female	0.3779***	ξ_g	0.0593***
	(0.0043)		(0.0136)
White	0.1695***	*Fixed Effects*	
	(0.0104)	School 1	−0.1640***
Black	0.3422***		(0.0069)
	(0.0124)	School 2	0.0129***
Parents: High School	0.0450***		(0.0058)
	(0.0046)	School 3	−0.4094***
Parents: College	0.7959***		(0.0077)
	(0.0041)	School 4	−0.4111***
GPA	0.1225***		(0.0090)
	(0.0009)	School 5	0.3157***
No. of Siblings	0.0027***		(0.0095)
	(0.0012)	School 6	−0.0263***
Income 1990	−0.2027***		(0.0086)
	(0.0033)	School 7	−0.4915***
Unemployment 1993	0.0044***		(0.0088)
	(0.0086)	School 8	−0.0773***
College Prop 1990	0.5045***		(0.0169)
	(0.0108)	School 9	0.0920***
Exogenous Effects			(0.0106)
Female	0.0496***	School 10	0.2663***
	(0.0094)		(0.0065)
White	0.1161***	School 11	0.2387***
	(0.0127)		(0.0115)
Black	−0.2448***	School 12	0.1207***
	(0.0220)		(0.0124)
Parents: High School	0.0908***	School 13	0.3465***
	(0.0100)		(0.0083)
Parents: College	0.5521***		
	(0.0142)		

* p < 0.1, ** p < 0.05, *** p < 0.01.

point of the MTE. Clearly individuals are self-selecting into college in this model, with those who are likely to attend college receiving much higher returns.

The support for the propensity scores, P_g, can be seen in figure 2.3. As expected, we see that at low values of X_g, we observe most support at low values of P_g. Conversely, high values of X_g are associated with high propensity scores. In figure 2.3b we see that, largely, the model correctly assigns high propensity scores to those who attend college. Those who don't attend college have a more diverse spread of

Table 2.7: High School Treatment Estimation: No Networks

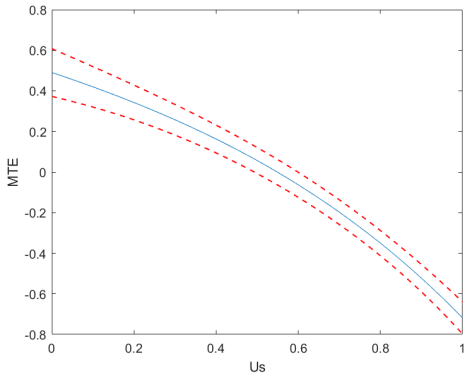
Variable	Estimate	Variable	Estimate
Female	0.3537**	School 4	−2.2977
	(0.1568)		(5.0353)
White	0.3126	School 5	−1.4049
	(0.5606)		(5.0171)
Black	0.2732	School 6	−1.7360
	(0.5572)		(5.0526)
Parents: High School	0.1084	School 7	−2.4233
	(0.1569)		(5.0636)
Parents: College	0.7982***	School 8	−1.7595
	(0.1685)		(4.7392)
GPA	0.1267***	School 9	−1.8119
	(0.0321)		(5.0316)
No. of Siblings	0.0090	School 10	−1.6826
	(0.0384)		(5.0952)
Income 1990	0.0979	School 11	−1.5257
	(0.5652)		(5.1571)
College Prop 1990	0.0486	School 12	−1.6394
	(1.5111)		(5.1949)
Fixed Effects		School 13	−1.3898
School 1	−1.8591		(5.1689)
	(4.9976)		
School 2	−1.8531		
	(5.0886)		
School 3	−2.3501		
	(5.0971)		

* p < 0.1, ** p < 0.05, *** p < 0.01.

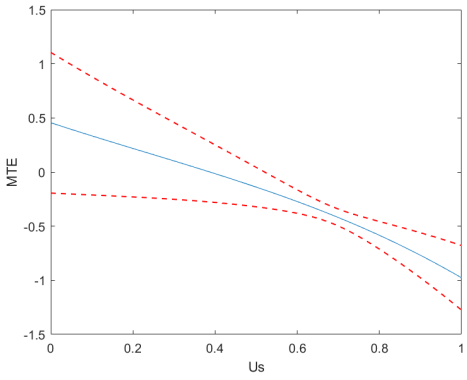
propensities, but most weight is given under 0.5.

We see similar results for the base model. The base model generally assigns low P_g values to low X_g values, but in figure 2.3c, we see a tendency to assign high P_g values across a range of X_g values. Figure 2.3d shows that the base model is slightly worse at separating those who attend college and those who do not, with fewer low propensities assigned.

Table 2.8 describes the results of the MPRTE for the high school networks. As in the MTE, we observe high returns in each case. Those induced into treatment by a small change in policy, receive a return of around 35%. This is similar for both the case when we increase the instruments or increase propensity additively. The return is slightly reduced if propensity is multiplicatively increased. As our exogenous effects are effectively used as instruments, we can conclude that a marginal

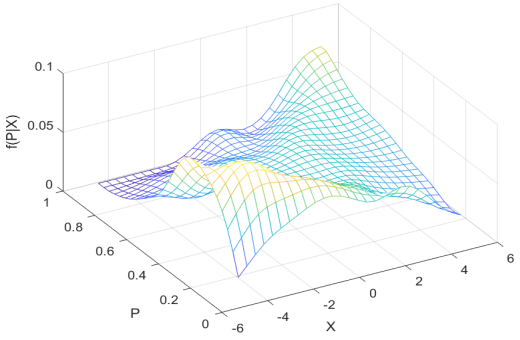


(a) Full Model

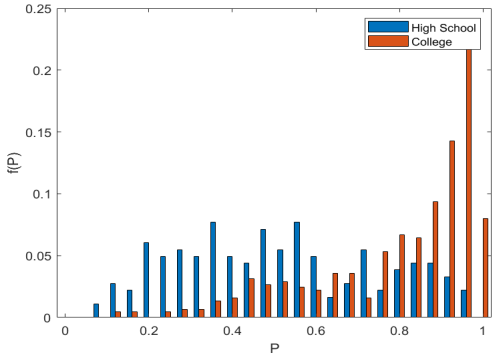


(b) Base Model

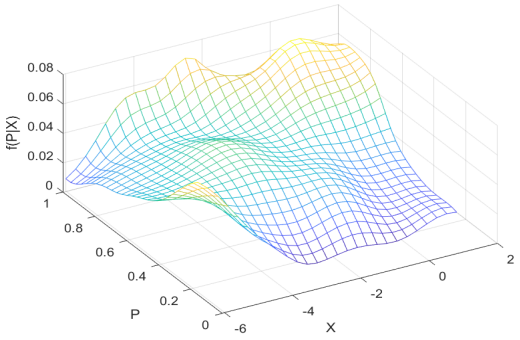
Figure 2.2: MTE for the returns to college education



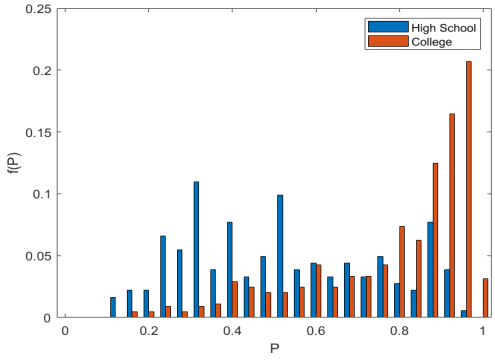
(a) Support of P_g given X_g



(b) Support of P_g given $S = 1, 0$



(c) Support of P_g given X_g - no networks



(d) Support of P_g given $S = 1, 0$ - no networks

Figure 2.3: Support for high school network data

increase in the exogenous peer effect will lead to a positive return to those induced into attending college. The base model policy effects are more muted. Those induced to attend college by a change in policy are predicted to receive a 20% increase in

wage, although the standard error of this estimate is reasonably high.

Table 2.8: MPRTE in High School Networks

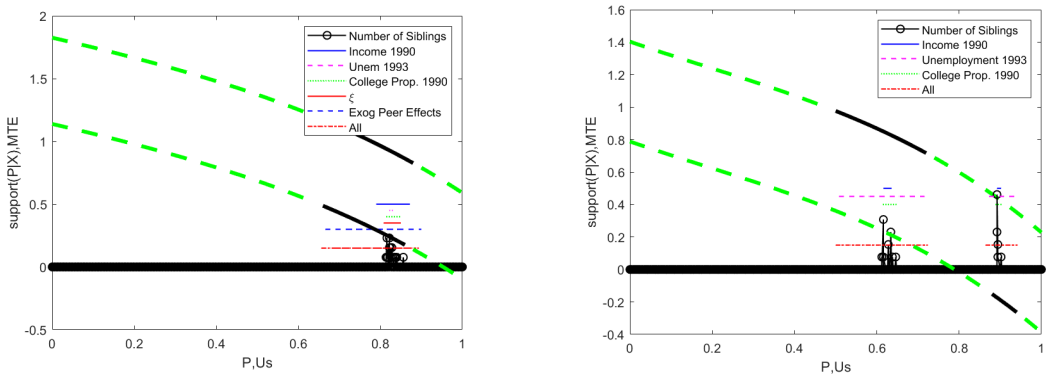
Policy Change	Model with Networks	Model without networks
$Z^k_\alpha = Z^k + \alpha$	0.3503***	0.2095
	(0.0606)	(0.2663)
$P_\alpha = P + \alpha$	0.3519***	0.1956
	(0.0610)	(0.2579)
$P_\alpha = (1+\alpha)P$	0.2764***	0.0828
	(0.0531)	(0.2057)

* p < 0.1, ** p < 0.05, *** p < 0.01.

We can also directly measure the effect of other changes in our model. In particular, we can measure the effect of an increase or decrease in particular characteristics. For example, if we increase the college education of every 5th student's parents by 10%, we observe a treatment effect of 34.64% with a standard error of 21.68%.

Relaxing the assumption of the exogeniety of X_g we obtain figure 2.4. The three labour market variables and ξ_g, functioning as instruments are allowed to vary as X_g is held constant. In this case, it appears the strength of the instruments in identifying the MTE is low. We identify the black portion of the MTE curve only. We see in figure 2.3 that most propensities are clustered around 0.8. This corresponds to the portion of the MTE identified in the model when X_g is endogenous. The MTE is still downwards sloping in these regions. The MTE has drastically different intercepts depending on the value we hold X_g at. This is to be expected, given the number of variables in X_g in comparison to the number of instruments included. Including exogenous peer effects as instruments gives us the result in 2.4a.

In comparison, the base model has fewer instruments and as a result identifies a slightly smaller region for identification in this endogenous setting. The ξ_g variable and the exogenous peer effects offer greater variation in instruments, though this improvement is not dramatic in this case.



(a) Full Model

(b) No Networks

Figure 2.4: College Network Under Endogeneity

4 Discussion

Comparing the two models presented above, we see, in general, that the peer effect model produces results with a smaller variance and greater identification. In the treatment equation, most individual effects have similar estimates. It appears that in the base model some of the exogenous effects are absorbed into the individual characteristics. The greatest difference in the treatment estimates lies in the fixed effects. The basic model predicts lower fixed effects than the full model. The standard errors in the full model are considerably smaller, particularly for the fixed effect terms.

Results in the MTE are similar between the models, with the full model displaying a less negative return for higher values of U_S. The confidence bands for the full model are much smaller, as in chapter 1. The most notable difference is in the MPRTE. The full model predicts a significantly higher return for those affected by changes in policy. This may be due to the prediction of propensities clustered largely above 0.5, corresponding to the region where the full model predicts higher returns to education. Thus most individuals at the margin will experience a higher predicted return under the peer effect model. In addition, we observe a lower MPRTE variance in the peer effect model.

In addition to the changes noted above, introducing peer effects allows us to

achieve a greater range of identification in both the case when X_g is exogenous and when X_g is endogenous. In the exogenous case, we identify the MTE at slightly lower values of U_S, according to figure 2.3. In the endogenous case, the addition of peer effects allows us additional instruments for identification. As in figure 2.4, we identify a greater range across the MTE, as the additional instruments allow more variation when X_g is held constant.

The proposed peer effect model is unique in its ability to control for homophily in observed and unobserved characteristics. In the network model, we note strong homophily in several dimensions including age, GPA, appearance, parental education, gender and race. High school students appear to select strongest across age and race. While the age affect is unsurprising (students are grouped into classes by age), the lack of racial crossover may be more concerning. We find modest but significant effects for GPA and appearance i.e. students select friends based on similarity in appearance and achievement. Males are more likely to nominate friendships with other males, while the homophily of females is somewhat weaker. Students with fathers who did not finish high school are less likely to become friends with similar students. This may, however, just be an indication that these students make less friends in general. Students with college educated fathers do tend to cluster together.

Homophily, by definition, groups individuals together so that we observe similar outcomes and similar decisions across segments of the population. While homophily often works against those on the negative side of the distribution, in our setting the effect of homophily is largely positive. In fact, it may be possible for policy makers to use homophily as an instrument to achieve positive outcomes. Many of the characteristics in which homophily is displayed; gender, race, parental education and GPA, strongly influence the probability that a student decides to enrol in college. The result of such homophily is to reinforce the decision of the student. As discussed in chapter 1, this reinforcement occurs through both endogenous and exogenous peer effects. As students select friends based on similarity in characteristics, characteristics encouraging college attendance will be reinforced by positive exogenous effects and the increased likelihood that their friend will also attend college (endogenous

effects).

Given the evidence that individuals self-select into college based on their returns, homophily assists in allocating students to the decision which results in the highest return. Those who should attend college (who would obtain a positive return) are now encouraged to do so, while those unlikely to attend college (who would experience negative returns) are further dissuaded. This does, however, leave some individuals somewhat entrenched. If a student is surrounded by other students unlikely to attend college, the probability they will separate from their friends in the college decision is low, even if positive returns are possible.

From a policy standpoint, homophily makes it easier to target certain groups of individuals. If policy makers want to assist one particular racial group, homophily and peer effects can work in tandem to magnify these policy effects, and spread them across the peer group. Changing the social norms in a large group, however, is usually a difficult prospect and requires a critical mass to shift the direction of the peer effect.

Turning to the results of the treatment equation, we note that most coefficients are significant. Race, gender and parental education all prove to be especially important in determining the likelihood of attending college. GPA is less important but still significant. Each of our instruments are significant, though number of siblings is a weak determinant of attendance in college. As expected, lower income and higher college attendance in the local area positively influence the likelihood of college enrolment. Most of our exogenous effects are also significant. Considering exogenous effects, having white friends and having female friends positively influence the probability of treatment. Similarly if friends have a higher GPA, or have college educated parents they will be more likely to attend college. Having black friends appears to reduce the probability of college attendance, though being black is a strong positive influence on the probability of treatment. This may be a product of multicollinearity between individual and peer characteristics, given the racial homophily we have already noted. The base model predicts a similar positive result of black students on college attendance. In this sample, black and white students dominate

the sample and are more likely to attend college than those of other races (mostly Hispanic and Asian in our sample). The much investigated disparity between black and white students is not prevalent in our sample. In fact, black students appear to have higher college attendance and higher average wage than white students (French et al. (2015) find and discuss a similar result in the Add-Health data). Being female significantly increases the probability of attending college, a result which has been well documented. Much research has discussed the factors underlying this reversal of the gender gap including the effects of childhood neighbourhood and environment (Chetty et al., 2016), teaching style (Dee, 2005) and school quality (Figlio et al., 2016). The effect of GPA on college attendance is only small, although the effects are accompanied by an associated small exogenous effect of peer GPA.

The presence of college educated parents is the strongest indicator of college education, accompanied by a strong exogenous peer effect of having college educated parents. A social norm effect can be asserted: students who are accustomed to interacting with those who have previously attended college, whether their parents, or friend's parents, are themselves likely to enter college. Being surrounded by highly educated individuals creates expectations for the individual of their college enrolment. Students who only have experience with college education conclude that further education is the required norm, while those who interact only with non-college educated individuals are less likely to anticipate a college future. We do not, however, find evidence of a positive endogenous effect. The direct effect of friend's intentions to attend college is irrelevant, having controlled for all other individual and peer factors. Thus having friend's with college educated parents is much more important than having friends who plan on attending college themselves. These findings may be particularly pertinent in a policy setting. Students must view college education as a viable option. Interaction with people who have attended college allows college education to appear more realistic, or more normal, with the benefits of college education more directly observable. Social norms are a strong mechanism to encourage certain individual decisions and behaviours. Policies that exploit social norms can be particularly effective (Del Bello et al., 2015). One other possible ex-

planation for the insignificance of peer effects is that the endogenous effect is simply absorbed by several of the exogenous effects, which influence the likelihood that both the individual and the peer will attend college.

Given the significance of the peer effect terms (particularly exogenous peer effects) in the treatment model, and the positive returns associated with policy changes, consideration of peer effects is likely to enhance education policy strategies. If policy works to increase any of the discussed characteristics within a school, then benefits may spread across the network through the social multiplier effect. Peer effects can be used as an instrument to implement change or to improve the effectiveness of policy.

The MTE curve demonstrates strong evidence of self selection into college. Those with high probability of attending college benefit greatly from doing so, while those unlikely to attend would achieve a very low return if they did. The data in this case may represent an extreme sample. Taking simple averages, those who attend college have a mean observed wage 93.29% higher than those who do not attend college. Thus, those along the extremes would expect to achieve a dramatic differential in wage depending on their choice to attend college. It is worth asking the question why returns are lower for individuals who are unlikely to attend college. Choice of college or college major may be one factor, so that even if these individuals do attend college, they are either unable or unlikely to enrol in college courses which would increase their returns. If this were indeed the key issue, then encouraging college attendance for all students would not be enough to increase their returns; selection of college and college major, and the probability of success within this degree would be an important consideration.

Endogeneity in the X_g variables proves difficult to manage when estimating the MTE. We only identify a small section of the MTE curve, when X_g is held constant. This problem would be mitigated under the presence of stronger instruments, uncorrelated with the regressors in the outcome equation. As noted, however, peer effects do afford additional variation, particularly in the unobserved ξ_g effects. The estimation in this case is limited by the small sample size. Only 13 smaller schools

are available here, and may not be representative of the true population of schools in the US. A larger sample size would allow more accurate estimation of the nonlinear outcome equation, and would provide increased variation in parameters for greater identification.

The MPRTE in each case is positive, such that those influenced by policy changes would receive positive returns from attending college. Thus any change in policy is generally constructive, including changes in the peer effect. Given the importance of peer effects in the college decision, peer effects provide a useful tool to encourage college attendance.

The estimation is potentially distorted by the structure of the data. Due to the limitations of the survey, we cannot map friendships outside of the school. This means that modelled friendships are only a subset of the true friendships and our estimates of peer effects are likely to underestimate the true peer effect. We have already discussed the important role of parents and friend's parents in the peer effect, so we would expect other individuals outside the school to also contribute to the peer effect. In addition, removing individuals with no named friends from the sample may also affect the results. It is difficult to know the net impact of this omission without further investigation. We would still expect friendless students to be subject to some peer effect, whether in relationships outside of the school, or with unlisted classmates. It is likely this peer effect would be weaker than the one measured within the school. Those deleted from the sample went to college less often and received a lower wage on average, although returns to college were higher for these individuals. Of those deleted from the sample, 62.6% went to college, receiving an average wage of \$31942 and a return to education of 1.58 (compared to 69.34%, \$34159 and 1.29 for those in sample). We may conclude, therefore, that removing these individuals from the sample causes returns to be underestimated.

We noted that friendships were generally transient in the data sample. Almost 50% of friendships from one period were not transferred to the next. We can either attribute this high number to measurement error, such that individuals were not complete in their recording of friendships from one period to the next, or conclude

that most friendships are generally short-lived at this stage of high school. This is surprising given the significance of the peer effects we have found in the decision to attend college. One explanation may be that friendships observed in the first period are observed earlier in high school, when students are adjusting to the new environment of high school and making many new friends. By the time we record period 1 friendships, these friendships may carry a greater weight and be more stable over time. This may also lead to the conclusion that the peer effect is not equally weighted over a student's friends. More stable, deeper friendships may hold a greater weight in the decision making processes of students. The peer effect we found may be distributed more heavily over a closer group of friends rather than being an average of the reported friendship roster of the individual. One could place weights on the order of the friendship roster to test this hypothesis.

5 Conclusion

This paper proposes a model to enable the incorporation of peer effects into a standard treatment effects framework. We explore identification within this model and present a Bayesian estimation method. The model is tested using a Monte Carlo experiment and applied to a data-set of high school students.

We find that the inclusion of peer effects allows wider identification and greater accuracy in estimation. Without controlling for peer effects in a model where peer effects are important, estimates will be biased. The inclusion of peer effect terms provides us with useful instruments for estimation and allows a greater range of identification when X_g is both exogenous and endogenous. Our model obtains more accurate measures of the marginal treatment effect and reduces the associated standard error.

We adjust for unobserved peer effect terms by including a network formation model. The proposed network formation model relies on homophily, where individuals select friendships based on similarity in characteristics for identification. Homophily plays a strong role in the estimation, increasing the size of propensity scores

in the model, increasing returns at each quantile and allowing stronger identification.

In our data section, we find that peer effects are significant in the treatment equation. There is evidence that peers play an important role in the decision to attend college. Homophily in friendship selection is strong and prevalent, and controlling for unobserved characteristics is important in preventing the problems associated with homophily. The inclusion of peer effects is important from a policy analysis perspective. Peer effects can either distort policy intentions or promote the diffusion of the benefits of a policy across the network. Peer effects can be directly used as a policy instrument. We find that when peer effects are increased in the college decision process, those students induced to attend college benefit from doing so. Furthermore, when we increase some characteristic in a school, such as the proportion of college educated parents, the student benefits not only from the increase in their own parent's education, but also from the increase in friend's parents education, and the increase in probability that their friends will now attend college.

The proposed model can be used in settings where peers affect the binary treatment decision of an individual, independent of a continuous outcome. We have explored the setting where peers influence the decision of an individual to attend college, independent of an observed future wage outcome. Other possible settings could explore binary decisions such as the decision to smoke, the decision to engage in delinquent behaviour or the decision to marry. Associated outcomes may be continuous health or economic related outcomes, independent of the original peer effect.

The peer effect treatment model presented here, allows us to adjust for and conduct measurement of important peer effects in econometric settings. The proposed method adequately controls for potential biases and addresses many of the complications in network estimation. Wider acceptance and use of peer effect models within econometrics will allow us to more widely contemplate and exploit the potential role of networks in enacting and dispersing economic and social policy in numerous settings, and may allow significant improvements in the effectiveness of community based initiatives.

Appendix

1 Identification in the College Decision Equation

Identification as in Goldsmith-Pinkham and Imbens (2013) is achieved in the network formation model through asymptotic theory, particularly in the case when few networks are present. Homophily in observed characteristics, such that individuals who are very dissimilar are unlikely to become friends, allows us to assume a level of independence in the network. A summary of the theory can be found in Chandrasekhar (2016) and is expounded in Leung (2014). This theory derives a law of large numbers and central limit theorem that can be applied to the network model. The argument is repeated below in the context of this paper. The link formation model can be expressed as

$$D_{g,ij} = \mathbb{1}[U_{g,ij}(D_g, \boldsymbol{W}_g; \theta, \delta) > 0] \cdot \mathbb{1}[U_{g,ji}(D_g, \boldsymbol{W}_g; \theta, \delta) > 0] \tag{1.1}$$

such that,

$$\boldsymbol{W}_g = \{c_{g,i}, c_{g,j}, c_{g,ij}, D_{g\_ij}, F_{g\_ij}, \xi_{g,i}, \xi_{g,j}, \varepsilon_{ij}; i, j \in \mathcal{N}_n\},$$

where $\mathcal{N}_n$ is the set of all nodes. Leung defines the node statistic of node i, $\psi_i(D_g)$, such that $\psi_i(D_g)$ is a function that depends only on the network D_g through the set of links between all nodes connected to node i (the component of node i). We would like to prove that a law of large numbers and central limit theorem exists for corresponding network moments, which can be written as $\frac{1}{N_g}\sum_{i=1}^{N_g}\psi_i(D_g)$. In our model we are interested in the network moments corresponding to the probability of linkage i.e. $P(D_{g,ij} = 1)$. Under certain conditions, $\{\psi_i(D_g); i \in \mathcal{N}_n\}$ forms an α-mixing field, such that the linking decisions of individuals in the network are sufficiently uncorrelated. This α-mixing field allows the establishment of a law of large numbers and central limit theorem for network moments.

As laid out by Leung (2014) and Chandrasekhar (2016) the assumptions required

for an α-mixing field to exist are as follows:

1. No coordination: Under θ, δ and for a given W_g, sets of nodes that aren't connected under any equilibrium make friendship linkage decisions independently. i.e. isolated friendship networks have no incentive to coordinate on their friendship decisions.

2. Homophily: Irrespective of endogenous network effects on utility, it is typically not worthwhile to link directly to someone who is very far away in homophilic characteristics.

Let $d(i,j)$ be a matrix representing the difference in characteristics in which individuals display homophily. For example, in our model $d(i,j)$ could contain the difference in age or observed GPA for individuals i and j: $d(i,j) = \big[|AGE_{g,i} - AGE_{g,j}|, |GPA_{g,i} - GPAg, j|\big]$. As long as individuals are typically unlikely to link with those far away in these characteristics, the homophily condition will hold. We can express

$$U_{g,ij}(D_g, w; \theta, \delta) = U_{ij}(\underbrace{d(i,j)}_{distance}, \underbrace{z(D,w)}_{endogenous}, \underbrace{f(w)}_{exogenous}; \theta, \delta) \tag{1.2}$$

so that z represents non-homophilic endogenous network effects (which may depend on the network D) and f represents non-homophilic exogenous network effects. As such the homophily assumption says:

$$\lim_{d(i,j)\to\infty} U_{ij}(d(i,j), \bar{z}, f(w); \bar{\theta}, \bar{\delta}) < 0 \tag{1.3}$$

for $f(w) \leq \bar{f}$, and the probability of $f(w) > \bar{f}$ is sufficiently low, such that the exogenous and endogenous effects are generally bounded.

3. Thin Tails: The distribution of ε_{ij} has thin tails such that $P(\varepsilon_{ij} > r) \leq Ce^{-\kappa r}$ for some C,$\kappa > 0$. This condition is satisfied if ε_{ij} is normally distributed as in our model.

4. Increasing Domain: $\sup_n \max_{i \in \mathcal{N}_n}(\#j \in \mathcal{N}_n : |d(i,j) < r_0|) < \infty$, for some $r_0 > 0$. This says that the largest number of individuals that are at most a distance r_0 away from any other individual is finite. Communities are small relative to the overall network.

5. Diversity: For any individual i and distance r, there exists a set of nodes S containing i such that any $k \in S$ satisfies $d(i,k) < r$ and $d(k,l) \geq 9\kappa^{-1} \log r$ for $l \notin S$ and κ^{-1} given in condition 3.

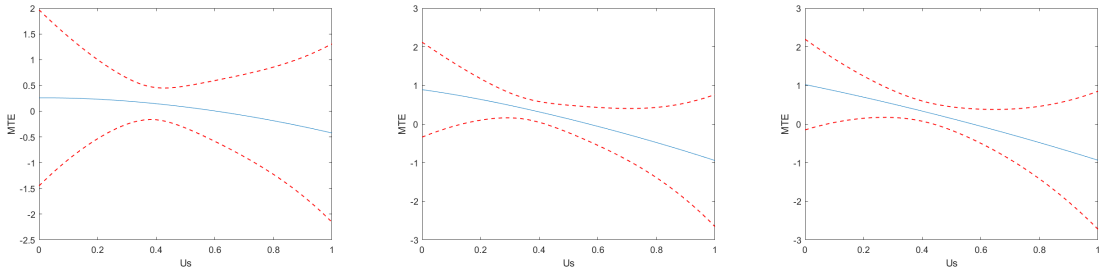
This condition ensures that there is sufficient diversity in the distance characteristics. Diversity ensures that $\alpha(i,j)$ decays at a sufficiently fast rate for a CLT. This is a sufficient condition to prove that $P(D_{i,j} = 1) \to 0$ as $d(i,j) \to \infty$. In our case, we use several homophilic indicators; parents education, personality variables, School achievement, age, race and gender to ensure sufficient diversity.

2 Additional tables and figures

2.1 Chapter 1

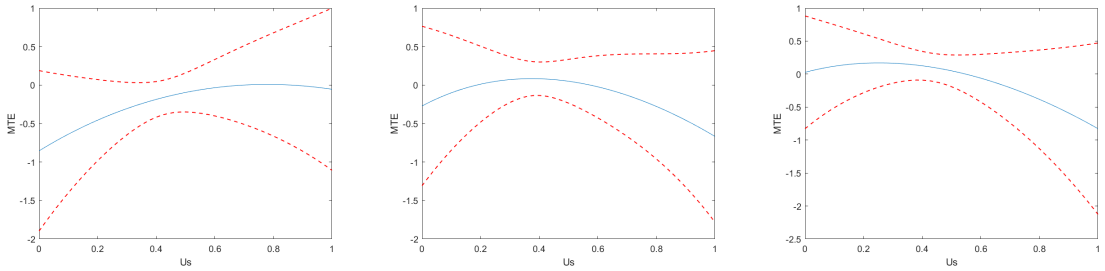
Table 2.1: Test of Equality of LATEs over different intervals

Ranges of U_S for $LATE_j$	(0,0.04)	(0.08,0.12)	(0.16,0.20)	(0.24,0.28)	(0.32,0.36)	(0.40,0.44)
Ranges of U_S for $LATE_{j+1}$	(0.08,0.12)	(0.16,0.20)	(0.24,0.28)	(0.32,0.36)	(0.40,0.44)	(0.48,0.52)
Difference in LATEs	0.0502	0.0723	0.0937	0.1145	0.1346	0.1542
p-value	0.5333	0.3000	0.2333	0.1333	0.0667	0.0667
Ranges of U_S for $LATE_j$	(0.48,0.52)	(0.56,0.60)	(0.64,0.68)	(0.72,0.76)	(0.80,0.84)	(0.88,0.92)
Ranges of U_S for $LATE_{j+1}$	(0.56,0.60)	(0.64,0.68)	(0.72,0.76)	(0.80,0.84)	(0.88,0.92)	(0.96,1)
Difference in LATEs	0.1731	0.1915	0.2093	0.2264	0.2430	0.2589
p-value	0.1333	0.1333	0.1667	0.1667	0.1333	0.1333
joint p-value	0.13333					



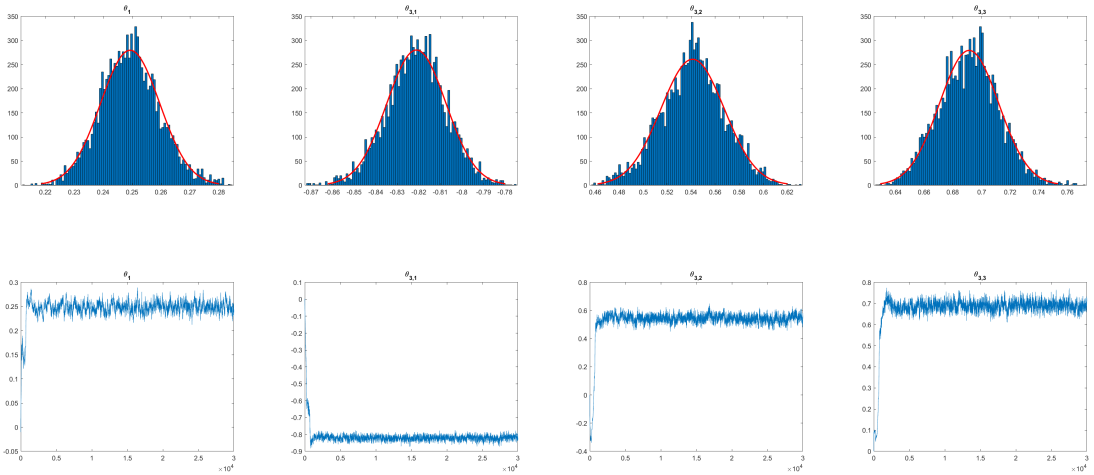
No Networks No ξ_g Full Model

Figure 2.1: MTE with $corr(X_g, U) = 0.3$

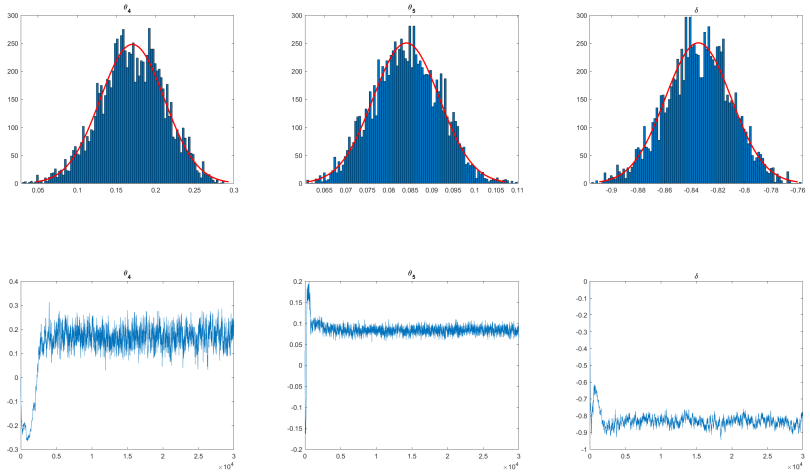


No Networks No ξ_g Full Model

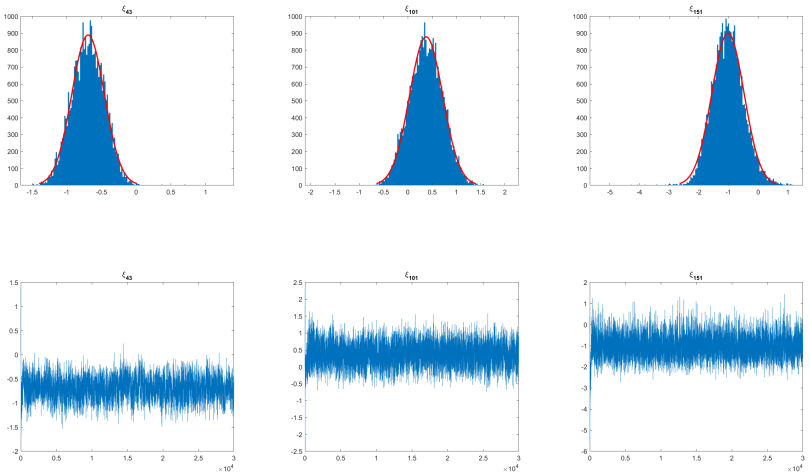
Figure 2.2: MTE with $corr(X_g, U) = 0.75$



(a) θ_1 (b) $\theta_{3,1}$ (c) $\theta_{3,2}$ (d) $\theta_{3,3}$

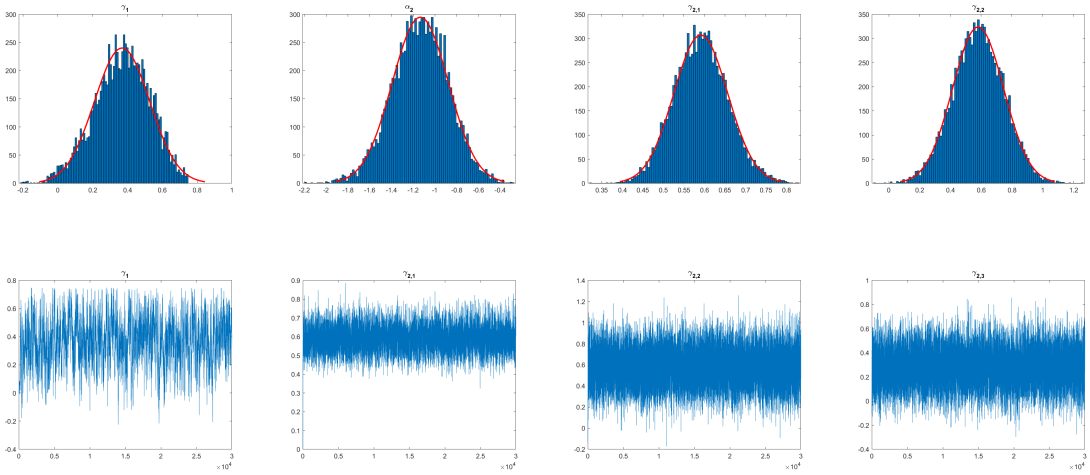


(e) θ_4 (f) θ_5 (g) δ

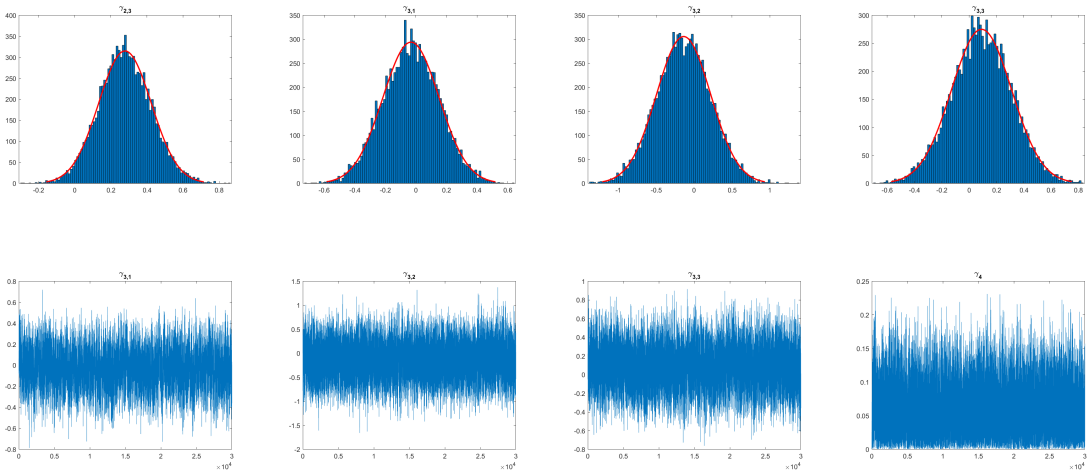


(h) ξ_{43} (i) ξ_{101} (j) ξ_{151}

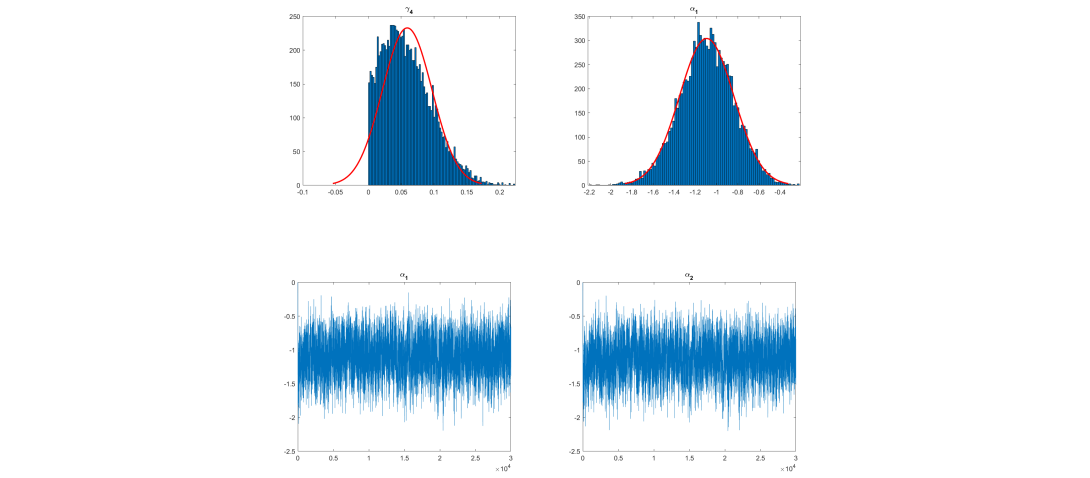
Figure 2.3: Network Formation MCMC plots



(a) γ_1 (b) γ_0 (c) $\gamma_{2,1}$ (d) $\gamma_{2,2}$



(e) $\gamma_{2,3}$ (f) $\gamma_{3,1}$ (g) $\gamma_{3,2}$ (h) $\gamma_{3,3}$



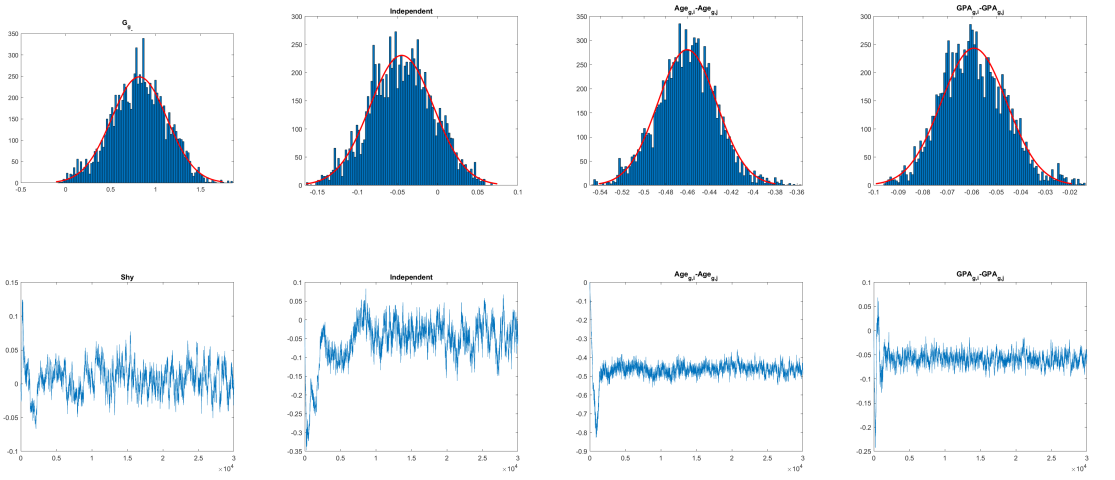
(i) α_1 (j) α_2

Figure 2.4: Treatment Equation MCMC plots

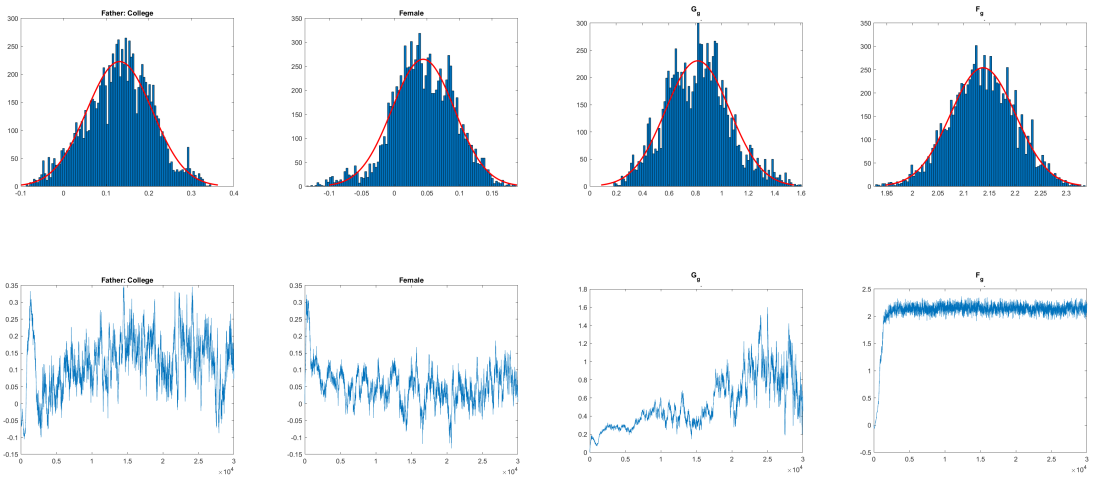
2.2 Chapter 2

Table 2.2: Test of Equality of LATEs over different intervals for high school networks

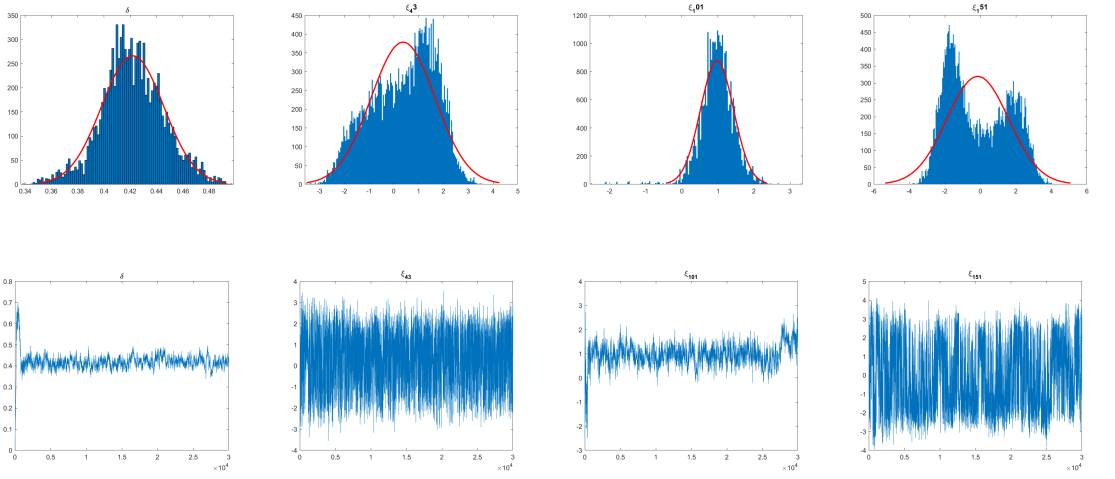
Ranges of U_S for $LATE_j$	(0,0.04)	(0.08,0.12)	(0.16,0.20)	(0.24,0.28)	(0.32,0.36)	(0.40,0.44)
Ranges of U_S for $LATE_{j+1}$	(0.08,0.12)	(0.16,0.20)	(0.24,0.28)	(0.32,0.36)	(0.40,0.44)	(0.48,0.52)
Difference in LATEs	0.2325	0.2497	0.2698	0.2929	0.3194	0.3496
p-value	0	0	0	0	0	0
Ranges of U_S for $LATE_j$	(0.48,0.52)	(0.56,0.60)	(0.64,0.68)	(0.72,0.76)	(0.80,0.84)	(0.88,0.92)
Ranges of U_S for $LATE_{j+1}$	(0.56,0.60)	(0.64,0.68)	(0.72,0.76)	(0.80,0.84)	(0.88,0.92)	(0.96,1)
Difference in LATEs	0.3838	0.4222	0.4650	0.5124	0.5646	0.6216
p-value	0	0	0	0	0	0
joint p-value			0			



(a) Shy (b) Independent (c) $Age_{g,i} - Age_{g,j}$ (d) $GPA_{g,i} - GPA_{g,j}$

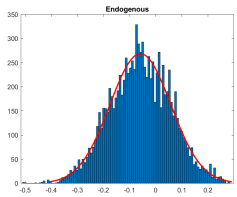


(e) Father College (f) Female (g) $G_{g\_}$ (h) $F_{g\_}$

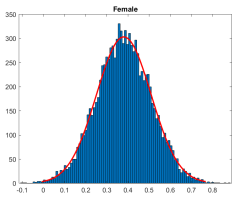


(i) δ (j) ξ_{43} (k) ξ_{101} (l) ξ_{151}

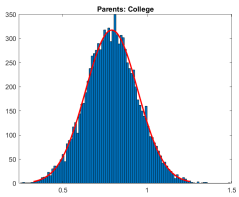
Figure 2.5: Network formation MCMC plots for high school networks



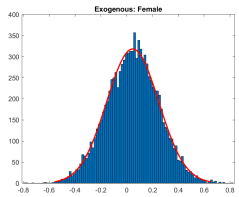
(a) Endogenous



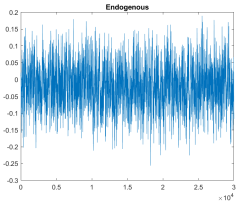
(b) Female

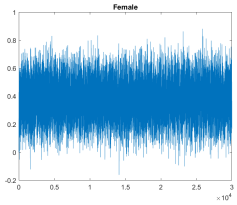


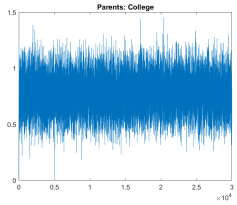
(c) Parents: College

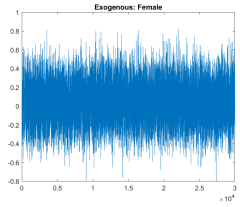


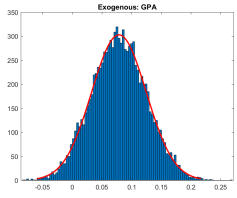
(d) Exogenous: Female



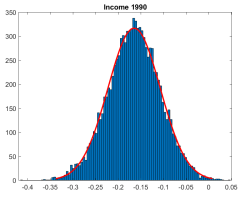




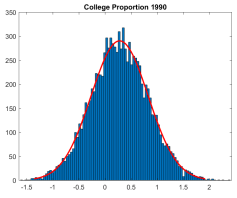




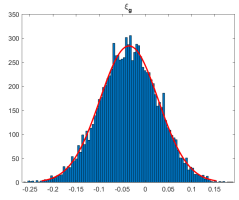
(e) Exogenous: GPA



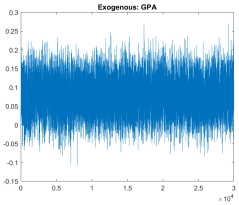
(f) Income 1990

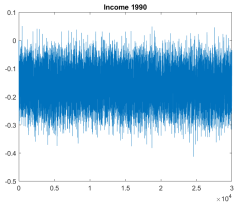


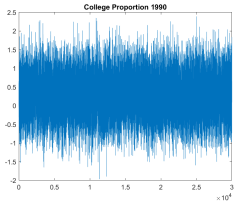
(g) College Proportion 1990

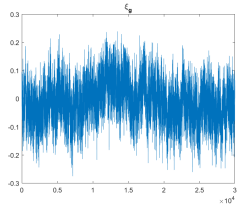


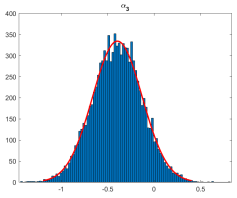
(h) ξ_g



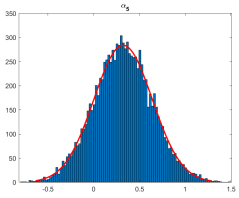








(i) α_3



(j) α_5

Figure 2.6: Treatment Equation MCMC plots for high school networks

Bibliography

James H Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422): 669–679, 1993.

L Anselin. *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media, 1988.

Anton Badev. Discrete Games in Endogenous Networks: Theory and Policy. PSC Working Paper Series 45., 2013. Available at `http://repository.upenn.edu/psc_working_papers/45`.

Lucia Barbone and Peter Dolton. Key Players? Identifying High School Networking Effects on Earnings. CESinfo Area Conference on the Economics of Education, 2015.

Robert Bifulco, Jason M Fletcher, and Stephen L Ross. The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1):25–53, 2011.

Robert Bifulco, Jason M Fletcher, Sun Jung Oh, and Stephen L Ross. Do high school peers have persistent effects on college attainment and other life outcomes? *Labour Economics*, 29:83–90, 2014.

Anders Björklund and Robert Moffitt. The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, pages 42–49, 1987.

Sandra E Black, Paul J Devereux, and Kjell G Salvanes. Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153, 2013.

Lawrence E Blume, William A Brock, Steven N. Durlauf, and Yannis M. Ioannides. Identification of Social Interactions. In Jess Benhabib, Alberto Bisin, and

Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, chapter 18, pages 853–964. Elsevier, 2011.

Lawrence E Blume, William A Brock, Steven N Durlauf, and Rajshri Jayaraman. Linear social interactions models. *Journal of Political Economy*, 123(2):444–496, 2015.

Vincent Boucher. Structural homophily. *International Economic Review*, 56(1):235–264, 2015.

Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55, 2009.

William A Brock and Steven N Durlauf. Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260, 2001.

Raffaella Calabrese and Johan A Elkink. Estimators of binary spatial autoregressive models: A Monte Carlo study. *Journal of Regional Science*, 54(4):664–687, 2014.

Antoni Calvó-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267, 2009.

Stephen V Cameron and James J Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political economy*, 106(2):262–333, 1998.

Stephen V Cameron and James J Heckman. The dynamics of educational attainment for black, hispanic, and white males. *Journal of political Economy*, 109(3):455–499, 2001.

Stephen V Cameron and Christopher Taber. Estimation of educational borrowing constraints using returns to schooling. *Journal of political Economy*, 112(1):132–182, 2004.

Pedro Carneiro, James J Heckman, and Edward Vytlacil. Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica*, 78(1):377–394, 2010.

Pedro Carneiro, James J Heckman, and Edward J Vytlacil. Estimating marginal returns to education. *The American Economic Review*, 101(6):2754–2781, 2011.

Anne C Case. Spatial patterns in household demand. *Econometrica*, pages 953–965, 1991.

Arun Chandrasekhar. Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, pages 303–357, 2016.

Raj Chetty, Nathaniel Hendren, Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi. Childhood environment and gender gaps in adulthood. *American Economic Review*, 106(5):282–88, 2016.

Siddhartha Chib. Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, 51(1-2):79–99, 1992.

Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

Nicholas A Christakis and James H Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

Giacomo De Giorgi, Michele Pellizzari, and Silvia Redaelli. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–275, 2010.

Aureo De Paula. Econometrics of network models. Technical report, cemmap working paper, Centre for Microdata Methods and Practice, 2016.

Thomas S Dee. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165, 2005.

Carlo L Del Bello, Eleonora Patacchini, and Yves Zenou. Neighborhood effects in education. IZA Discussion Paper No. 8956., 2015. Available at SSRN:`https://ssrn.com/abstract=2589818`.

Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

David Figlio, Krzysztof Karbownik, Jeffrey Roth, Melanie Wasserman, et al. School quality and the gender gap in educational achievement. *American Economic Review*, 106(5):289–95, 2016.

Jason Fletcher. Social interactions and college enrollment: evidence from the national education longitudinal study. *Contemporary Economic Policy*, 31(4):762–778, 2013.

Jason M Fletcher. Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25(4):1265–1286, 2012a.

Jason M Fletcher. Similarity in peer college preferences: New evidence from Texas. *Social science research*, 41(2):321–330, 2012b.

Michael T French, Jenny F Homer, Ioana Popovici, and Philip K Robins. What you do in high school matters: High school gpa, educational attainment, and labor market earnings as a young adult. *Eastern Economic Journal*, 41(3):370–386, 2015.

Andrea Galeotti and Gerrit Mueller. Friendship relations in the school class and adult economic attainment. IZA discussion paper, 2005.

John Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pages 571–578. Interface Foundation of North America, Inc., Fairfax Station, VA, 1991.

Edward L Glaeser, Bruce I Sacerdote, and Jose A Scheinkman. The social multiplier. *Journal of the European Economic Association*, 1(2-3):345–353, 2003.

Paul Goldsmith-Pinkham and Guido W Imbens. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.

Brian S Graham. Comment. *Journal of Business & Economic Statistics*, 31(3): 266–270, 2013.

James J Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005.

James J Heckman and Edward J Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, 96(8):4730–4734, 1999.

James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6:4875–5143, 2007b.

Chih-Sheng Hseih and Lung Fei Lee. A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, 31(2): 301–319, 2016.

Matthew O Jackson. *Social and economic networks*. Princeton university press, 2008.

Matthew O Jackson. Unraveling Peers and Peer Effects: Comments on Goldsmith-Pinkham and Imbens'"Social Networks and the Identification of Peer Effects". *Journal of Business & Economic Statistics*, 31:270–273, 2013.

Matthew O Jackson, Brian W Rogers, and Yves Zenou. The Economic Consequences of Social-Network Structure. *Journal of Economic Literature*, 55(1):49–95, 2017.

Dean Karlan, Markus M Mobius, Tanya S Rosenblat, and Adam Szeidl. Measuring trust in Peruvian shantytowns. *Unpublished manuscript*, 2005.

Francis Kramarz and Oskar Nordström Skans. When strong ties are strong: Networks and youth labour market entry. *The Review of Economic Studies*, 81(3):1164–1200, 2014.

Michael Kremer and Dan Levy. Peer effects and alcohol use among college students. *The Journal of Economic Perspectives*, 22(3):189–189, 2008.

Michael Kremer, Edward Miguel, and Rebecca Thornton. Incentives to learn. *The Review of Economics and Statistics*, 91(3):437–456, 2009.

Lung-fei Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2): 333–374, 2007.

Lung-fei Lee, Ji Li, and Xu Lin. Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics*, 96(3): 402–417, 2014.

James P LeSage. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1):19–35, 2000.

James P LeSage and R Kelley Pace. *Introduction to Spatial Econometrics (Statistics, textbooks and monographs)*. CRC Press, 2009.

Michael Leung. A random-field approach to inference in large models of network formation. *Available at SSRN*, 2014.

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

Charles F Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1), 2013.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Economics*, pages 105–142, 1974.

Liesbeth Mercken, Tom AB Snijders, Christian Steglich, Erkki Vartiainen, and Hein De Vries. Dynamics of adolescent friendship networks and smoking behavior. *Social Networks*, 32(1):72–81, 2010.

Robert A Moffitt et al. Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 4(45-82):6–17, 2001.

OECD. Education at a Glance 2016. 2016. doi: http://dx.doi.org/10.1787/eag-2016-en. DOI: http://dx.doi.org/10.1787/eag-2016-en, OECD Publishing.

Eleonora Patacchini and Tiziano Arduini. Residential choices of young americans. *Journal of Housing Economics*, 34:69–81, 2016.

Eleonora Patacchini and Yves Zenou. Juvenile delinquency and conformism. *Journal of Law, Economics, and Organization*, 28(1):1–31, 2012.

Eleonora Patacchini, Edoardo Rainone, and Yves Zenou. Dynamic aspects of teenage friendships and educational attainment. CEPR Discussion Paper No. DP8223, 2011.

Richard E. Quandt. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*, 53(284):873–880, 1958.

Richard E. Quandt. A New Approach to Estimating Switching Regressions. *Journal ofthe American Statistical Association*, 67(338):306–310, 1972.

Peter M Robinson. Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

A. D. Roy. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers New Series*, 3(2):135–146, 1951.

Bruce Sacerdote et al. 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(3):249–277, 2011.

Adriaan R Soetevent and Peter Kooreman. A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics*, 22(3):599–624, 2007.

Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994. URL `http://scholar.google.com/scholar.bib?q=info:gET6m8icitMJ:scholar.google.com/&output=citation&hl=en&as_sdt=0,5&as_vis=1&ct=citation&cd=0`.

Zebing Wu. *Adolescent friendship network and college enrollment: a longitudinal network analysis of selection and influence processes*. PhD thesis, University of Iowa, 2015.