

Identifying Marginal Returns to Education Through Social Networks

Timothy Hersey (B.Ec)

Supervisors: Dr Virginie Masson & Dr Firmin Doko
Tchatoka



THE UNIVERSITY
of ADELAIDE

School of Economics
University of Adelaide
November 4, 2016

Thesis submitted in partial completion of the requirements for the degree of Bachelor of
Economics (Honours)

Declaration

Except where appropriately acknowledged this thesis is my own work, has been expressed in my own words and has not previously been submitted for assessment.

Word count: approx. 12,000

Signature

_____ 04/11/2016

Date

Acknowledgements

I would firstly like to thank my supervisors Virginie Masson and Firmin Doko Tchato, whose passionate teaching and support has always inspired me. Thank you to Virginie for helping me keep perspective over the year, for the chats and for your honesty. Thank you also to Firmin for your well-timed encouragements and positivity, particularly when things didn't appear to be going to plan. I would also like to thank my friends and family, whose incredible support has allowed me to do this, putting up with my insanity along the way. Finally, thank you to the Honours cohort; each of you have added something unique and valuable to the group and made this year enjoyable amongst the madness.

Identifying Marginal Returns to Education Through Social Networks

Timothy Hersey

Abstract

This thesis explores the role of peers in influencing the decision of individuals to attend college and the resulting labour market outcomes. It proposes a model, combining the econometric methods of networks and treatment effects, to estimate the marginal treatment effect of education when peers have influence on the wage outcome and probability of treatment for an individual. Using Monte Carlo simulations, the effect of networks on the treatment effects model is investigated. We further explore the model by varying the network structure and conducting sensitivity analyses, considering the impact of networks on policy. Our results suggest that networks initially have a significant positive impact on the returns to education and the effects of policy. However, this effect is reduced once homophily in characteristics is introduced.

Contents

1	Introduction	1
2	Review of Literature	3
	2.1 Peer Effects	3
	2.2 Returns to Education	7
3	Model	9
	3.1 College Attendance Decision	11
	3.2 The Wage Model and Returns to Education	14
4	Estimation	17
	4.1 Estimating College Attendance	17
	4.2 Estimating Marginal Returns to Education	20
5	Monte Carlo	21
	5.1 Baseline Model	21
	5.2 Alternate Network Models	27
	5.3 Relating Network Formation to \mathbf{Z}	29
	5.4 Sensitivity Analysis	33
	5.5 Policy Effects	43
6	Conclusion	44
	Appendix A Additional Figures and Tables	52

List of Figures

1	Watts-Strogatz network graphs	23
2	Support of P conditional on \mathbf{X}	25
3	Baseline graphs of the marginal treatment effect (MTE)	26
4	MTE - Changes in network specification	28
5	MTE - Correlating network formation with \mathbf{Z}	31
6	MTE - Network size=[200 100 100 100]	34
7	MTE - Altering model coefficients	39
8	MTE - \mathbf{X}, \mathbf{Z} and α correlated	41
9	MTE - Correlated error terms	42
A.1	MTE - Bootstrapped errors	52
A.2	MTE - Networks dependent on \mathbf{X}	53
A.3	MTE - Correlated network	53
A.4	MTE - Network size=[300 150 40 10]	54
A.5	MTE - Alternate model specifications	55

List of Tables

1	Incorporating peer effects	27
2	Effect of networks on policy	29
3	Linking networks and \mathbf{Z}	30
4	Changing the unobserved network effect	35
5	Alternate model specifications	37
6	Coefficient specification	39
7	\mathbf{X}, \mathbf{Z} and α correlated	41
8	Correlated error terms	43
9	The effect of localised policies	43
A.1	Network size=[300 150 40 10]	54

1 Introduction

People live and relate in an increasingly interconnected world. The social networks in which we associate are ever expanding with the advent of globalisation and social media. Behaviours are generally not isolated to an individual, but are spread across networks and influenced by peers. Although the economic literature has begun to recognize the effects of social ties on the decisions and outcomes of an individual, the integration of peer effects into existing frameworks remain seldom. This is due to the complex interaction of the components defining peer effects and its subsequent implication for identification strategies.

In this thesis, we aim at estimating the returns to education in terms of increased future wage, recognising the fundamental role of peer effects in both education decisions and employment outcomes. Our work draws on the network model of Bramoullé, Djebbari & Fortin (2009) and the treatment effects model of Carneiro, Heckman & Vytlacil (2011). More precisely, we investigate the influence of high school student networks on both the college decision and future wages of individuals.

We use Monte-Carlo simulations to analyse the incorporation of the network model in the estimation of marginal returns, exploring the flexibility of networks and the role of each individual driver of peer influence on the returns to education. We then consider the effect of a change in policy on a system involving networks, determining if the influence of peers allows more effective policy. We find that while networks initially have a significant positive effect on the returns to education, the addition of homophily (the tendency of individuals to interact with similar peers) reduces this positive impact; polarising the college decision and reducing policy influence.

Naive estimation of the average returns to education does not capture heterogeneity in returns and can often be misleading when considering the effect of policy. As a consequence, the returns to college for individuals targeted by policy are likely to be lower than the returns for those already attending college. To account for these biases, we consider the model of Carneiro, Heckman & Vytlacil (2011). In their

model, Carneiro et al. estimate the returns to education by considering education as the treatment in the standard treatment effects model. The structure of the econometric treatment effects model allows for the heterogeneity that should be expected in the returns to education, facilitating the analysis of variables influencing both the inclusion of an individual into the treatment group and the resulting outcomes. As such, it is natural to consider this model in the light of peer influence.

It has been shown that peers have a considerable impact on whether an individual chooses to commence tertiary education (Bifulco, Fletcher & Ross 2011, Fletcher 2013, Fletcher 2012, Wu 2015) and where they are finally employed (Kramarz & Skans 2014, Black, Devereux & Salvanes 2013, Bifulco et al. 2014). Given the eminent effect of peers in these processes, we would also expect these networks to influence the returns an individual receives from attending college, particularly if they are the defining influence in the college decision. Bramoullé, Djebbari & Fortin (2009) exploit the structure of networks in order to identify each mode of the network peer effect. To analyse the integral role of peers in the returns to education, we utilise the methods of Bramoullé et al. in identifying the components of peer effect, incorporated into the treatment model.

More precisely, we consider the response of the marginal treatment effect (MTE) and the marginal policy relevant treatment effect (MPRTE), as employed by Carneiro et al. (2011), to changes in the model specification and network structure, demonstrating the robustness and intricacies of the model. The MTE describes the effect of treatment for individuals who are at the margin i.e. are indifferent to attending college, allowing us to consider the true returns to college attendance.

The MPRTE, proposed by Carneiro, Heckman & Vytlačil (2010), describes the returns to treatment for those induced to attend college as a result of policy. The presence of networks implies the existence of a social multiplier (Glaeser et al. 2002), such that the decision of one individual or the effect of policy on an 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 to their network. The structure of the MPRTE allows us to investigate the

role of this social multiplier in the returns to education.

Although navigating the identification and separation of the components of peer effect proves somewhat difficult in the non-linear treatment effects model, the construction of the network theoretical model allows us to decompose the interaction of networks with the treatment and outcome equations in the context of the returns to college education. The integration of these distinct areas of literature demonstrates the flexibility of the network model when brought to existing econometric models. The specification of the components of peer effects inside the treatment effects model proves to have significant implications on the performance of the model.

This thesis is organised as follows. Section two reviews the literature on peer effects and the returns to education. Section three presents the theoretical model. Section four describes the estimation technique. Section five details the Monte Carlo simulations and sensitivity analysis. Finally, section six makes concluding remarks.

2 Review of Literature

2.1 Peer Effects

Peer effects comprise a growing body of literature in economics, particularly of recent years. The identification of peer effects, however, has proven to be a difficult prospect. Jackson (2013) identifies several of the challenges many researchers encounter when considering peer effects, namely, identification, endogenous networks and homophily, computation, measurement error and misspecification. Comprehensive network data has only recently begun to emerge due to the complexities in measuring true networks. Accompanying several key datasets, the research on networks and peer effects has expanded greatly. Many of the challenges Jackson mentions are associated with the difficulty in identifying and disentangling the components of peer effects.

In his seminal paper on the identification of peer effects, Manski (1993) explores the issue of identification in the light of the “reflection problem”. The “reflection problem” refers to the difficulty in isolating endogenous and exogenous effects due

to the strong correlation between them. Endogenous and exogenous effects, as identified by Manski (along with correlated effects), comprise two of the major channels through which peer effects can occur.

Endogenous social effects refer to the influence of an individual's outcomes or decisions on their peers. The decision of an individual often has repercussions for decisions made by the friends of that individual. Consider the decision of a high school student in taking up smoking. The friends of that student have a much higher probability of smoking than individuals who have no friends smoking at all, whether by explicit or implicit peer pressure (Mercken et al. (2009) and Christakis & Fowler (2008)).

Exogenous social effects, also called contextual effects, describe the influence of an individual's characteristics on their peers' outcomes. For example, if an individual is a hard worker, they will likely encourage their friends to be hard workers as well. This peer effect may have numerous positive implications, with individuals more likely to attend college and obtain a higher wage.

Manski's (1993) "reflection problem" derives from the perfect multicollinearity between the expected outcomes of a peer group and the mean characteristics of the group. Simultaneity in behaviour, leads to a group of peers displaying similar characteristics and as a result making similar decisions.

Unravelling endogenous and exogenous effects has been the subject of several research papers. Traditionally the peer effects literature assumed that individuals interacted in defined groups, with each member having an equal influence on all others in the group. Lee (2007), however, draws upon the comparisons between network estimation models and spatial autoregressive models (SAR models); utilising the structure and non-linearity of a social network to enable identification of the exogenous and endogenous effects. Similarly in the notable paper by Bramoullé, Djebbari and Fortin (2009), the structure of an individual's peer network is exploited to allow identification, using an instrument variable approach.

A third effect must also be considered; the effect of homophily, or more broadly, correlated effects as outlined by Manski (1993). Moffitt (2001) segregates correlated

effects into two key components; individual preferences leading to the creation of friendship ties and unobserved common environmental factors. Expanding on the first component, individuals are prone to homophily, such that friendships are more common with people that display similar characteristics or personality. These effects, like social extraversion or athletic ability, can be difficult to observe. As an example of common environmental effects, students in the same high school face similarities in teaching styles, school resources and the average socio-economic standing of the school. This similarity in environment may lead to similar college preferences or employment outcomes.

In an attempt to distinguish the peer effect from correlated effects, Sacerdote (2000) and Zimmerman (2003) analyse natural experiments where individuals are allocated randomly into a peer group, removing the effect of homophily that may occur under individual choice. Once the noise of homophily is accounted for, the true peer effect can be estimated.

Lee (2007), as in Lin (2010), addresses the correlated effect empirically by introducing an unobserved network fixed effect variable into the model. This term accounts for all unobservables (correlated and homophily effects) common to the network. Bramoullé et al. (2009) employ the same method for dealing with correlated effects, generalising the models of Manski (1993), Moffitt (2001) and Lee (2007), achieving not just identification of endogenous and exogenous effects, but also controlling for correlated effects. Lin (2010) and Lee et al. (2010) build upon the methods of Lee (2007) and Bramoullé (2009), broadening the context and assumptions of the peer effect model to explore the applicability and versatility of the peer effect model in more complex real world settings. Lin (2010) generalises the context of the model by considering variations in group structures, while Lee et al. (2010) introduce disturbances to the SAR model to enable contextual and group unobservables to be considered.

However, the addition of a network term is an imperfect approach, not accounting for unobserved within-network and individual characteristics that may influence both outcomes and the decision to establish links. Taking the unobserved personality

characteristic social extraversion as an example, extraversion can influence not only who an individual befriends but also key outcomes and decisions made by the individual, such as their place of employment and observed wages. If social extraversion is not accounted for and varies across a network, the network will be endogenous, causing bias in the peer effect estimates. Goldsmith-Pinkham & Imbens (2013) propose a network formation model to control for this network endogeneity, incorporating unobserved individual characteristics which influence the likelihood of friendship formation and perceived outcomes. Goldsmith-Pinkham & Imbens examine the network formation process and peer effect estimation through Monte Carlo simulations and in the context of high school friendship networks, using the Add Health dataset (2009)¹. Hsieh & Lee (2014), apply the model of Goldsmith-Pinkham & Imbens, extending the model by allowing for non-reciprocal friendships and modelling unobserved individual characteristics as multidimensional continuous variables, rather than the binary variable adopted by Goldsmith-Pinkham & Imbens.

Several authors have followed the example of Goldsmith-Pinkham & Imbens in utilising a network formation model to consider network interaction and peer effects. The network formation literature holds promise for greater accuracy in the analysis of peer effects. As part of this literature, Badev (2013) uses a game theoretic approach in order to determine the Nash equilibrium of a friendship network to identify and decompose peer effects. Boucher (2015) develops a network formation model, focusing heavily on the homophily aspects of friendships. Calvó-Armengol et al. (2008) combine the existing network formation model framework with network centrality concepts to achieve identification of peer and network effects. In each of these cases, while the individual correlated effect may not have been strong enough to warrant a network formation model, the network formation process allows exploration of some aspect of peer effects and the relation to the initial building of the network.

Due to the limitations in network data, applied research tends to take a more practical approach. Rather than overcomplicating the issue of correlated effects,

¹The Add Health Dataset tracks the outcomes and friendships of students in a sample of US high schools over time, providing an opportunity for expansive peer effect research.

the authors include variables in their equations that effectively account for environmental effects. This approach is taken by Patacchini & Zenou (2012), by including network and school correlated effect dummies, and exploiting the depth of the Add Health dataset; including several individual characteristic and personality variables which typically remain unobserved. Once these effects are controlled for, the authors are able to show that peer group formation is random conditional on the network. Similar approaches are taken in the applied literature by Fletcher (2012), Kremer & Levey (2008) and Patacchini, Rainone & Zenou (2012) due to the simplicity of the method and robustness with the data. In general, these applied papers utilise the method proposed by Bramoullé et al. (2009) in accounting for unobserved variables, exploiting the network structure in order to achieve identification.

2.2 Returns to Education

As mentioned, the scope of peer influence is broad. These effects can be found in numerous decision making processes, such as the decision to attend college (Patacchini, Rainone & Zenou 2011, Fletcher 2013), engage in delinquent behaviour (Patacchini & Zenou 2009) and consume alcohol (Kremer & Levey 2003, Fletcher 2012). Peer effects can also be found when describing outcomes and behaviour, such as high school GPA (Goldsmith-Pinkham & Imbens 2013), obesity (Christakis & Fowler 2007) and loan behaviour (Karlan, Mobius, Rosenblat & Szeidl 2009). Two key channels of peer influence lie in the friendship networks of high schools, and the peer networks in entering the labour market. The common adage that finding a job is more about who you know than what you know, seems to hold elements of truth when taken to the literature. Considering peer effects, Kramarz & Skans (2014) find that the parents of classmates are an important determinant of where an individual finds employment. Similarly, Black, Devereux & Salvanes (2013) find that the average income of the fathers of high school peers has a small exogenous peer effect on the future wage of the individual. Bifulco et al. (2014), however, find little evidence of persistent peer effects of high school friends on wages, with the suggestion that the college educated mothers of one's friends may have a significant effect on future wages. Significantly,

they do find the delay in college education from the decision of peers not to attend college has a temporary negative effect on wages. Networks can also be used as a determinant of social skills. Using the networks of high school students, Galeotti & Mueller (2005) and Barbone & Dolton (2015) find that the social skills associated with network position lead to a significant increase in future earnings.

High school students have also been shown to have a large effect on the decision of their friends to attend college. Bifulco, Fletcher & Ross (2011) find a positive exogenous peer effect in that students who have classmates with college educated mothers are more likely to attend college and are less likely to drop out of high school. Fletcher (2012) and Wu (2015) focus on the endogenous 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%. We see, therefore, a peer effect on both wages and college enrolment, with a particularly strong peer effect emerging in the college enrolment decision.

College educated individuals earn significantly more than those with only a high school education level. According to the OECD (2016), the benefit to an individual (in the 25 OECD countries) of obtaining a bachelors level degree is around 48%. Peer effects on wages come from two sources; the direct endogenous peer effect of friends on wages and the indirect effect of peers on the college decision and therefore wages. The result mentioned of Bifulco et al. (2014), provides positive evidence for this indirect college effect of high school friends on wages. The paper of De Giorgi, Pellizzari & Redaelli (2010) takes a similar trajectory, finding a significant peer effect on the college major decided by students when at college, associated with adverse outcomes on academic achievement and entry wages.

Through the use of a treatment model, Carneiro, Heckman & Vytlacil (2011) present a model to analyse the heterogeneous returns to education. Specifically, they identify the marginal treatment effect (MTE); the effect of going to college on wages for those indifferent to attending college. Björklund & 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 instruments from the treatment model in order to calculate the MTE at each supported probability of attending college. This allows them to effectively consider the true effect of general policy changes, using the marginal policy relevant treatment effect (MPRTE), first developed by Carneiro, Heckman & Vytlacil (2010). Carneiro et al. (2011) propose both a parametric and non-parametric method for estimating these marginal effects, while Carneiro et al. (2010) examine the properties of these estimates.

The method of Carneiro, Heckman & Vytlacil (2011) has been applied across countries and contexts as evidence for the heterogeneous returns to education and to examine the true implications of policy on those who are affected. Kyui (2016) examines the returns to education in Russia, while Lavaggio & Verzillo (2016) consider the Italian context. Joensen & Nielson (2016) conduct a gender comparison on the importance of taking high school mathematics. They utilise the heterogeneous returns the model allows for, to consider the differing effects of policy on convincing boys and girls to take high school mathematics.

3 Model

Following Carneiro et al. (2011), we apply the generalised Roy model² to estimate the effect of peers on the decision to attend college, and from there the effect on wage outcomes and the marginal returns to college. The model comprises an outcome equation of wage and a treatment equation of the probability of attending college. The outcome can be represented by the equations

$$Y_1 = \mu_1(\mathbf{X}) + U_1 \quad \text{and} \quad Y_0 = \mu_0(\mathbf{X}) + U_0 \quad (1)$$

²The model was first developed by A.D. Roy (1951) and Richard E. Quandt (1958,1972).

where Y_1 is the potential log wage when an individual attends college and Y_0 is the potential log wage when an individual does not attend college. \mathbf{X} is a matrix of observed variables that influence wage outcomes. U_0 and U_1 are the unobserved components.

The decision to attend college is modelled by the following discrete choice latent variable model, where $S = 1$ if the individual chooses to attend college and $S = 0$ otherwise. This choice depends on observed variables, \mathbf{Z} and unobserved variables, V .

$$\begin{aligned} \mathbf{S}^* &= \mu_s(\mathbf{Z}) - V \\ S &= 1 \text{ if } \mathbf{S}^* \geq 0; \quad S = 0 \text{ otherwise} \end{aligned} \tag{2}$$

\mathbf{S}^* is a latent variable representing the net benefit of attending college. Under this setup, \mathbf{Z} contains some or all of the variables in \mathbf{X} . Additionally it must contain variables not included in \mathbf{X} to be used as instruments for college attendance in the outcome equations in (1). V is a strictly increasing continuous random variable with distribution function F_V . We assume that (U_0, U_1, V) are uncorrelated with \mathbf{Z} given \mathbf{X} . Throughout this thesis, $P(\mathbf{z}) = \Pr(S = 1 | \mathbf{Z} = \mathbf{z}) = F_V(\mu_s(\mathbf{z}))$ is used to denote the probability that an individual who has characteristics \mathbf{z} attends college.

Using equation (2), the observed outcome is expressed by

$$Y = (S)Y_1 + (1 - S)Y_0 \tag{3}$$

which can be rewritten, substituting in the expressions of (1), as

$$Y = \mu_0(\mathbf{X}) + [\mu_1(\mathbf{X}) - \mu_0(\mathbf{X}) + U_1 - U_0]S + U_0 \tag{4}$$

As in Carneiro et al. (2011) the conditional expectation of Y given $\mathbf{X} = \mathbf{x}$ and

$P(\mathbf{Z}) = p$ is then expressed as

$$\begin{aligned} \mathbb{E}(Y|\mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p) &= \mathbb{E}(Y_0|\mathbf{X} = \mathbf{x}, P(\mathbf{Z}) = p) \\ &+ \mathbb{E}(Y_1 - Y_0|\mathbf{X} = \mathbf{x}, S = 1, P(\mathbf{Z}) = p)p \end{aligned} \quad (5)$$

3.1 College Attendance Decision

College attendance can be viewed as the treatment in our model, such that attending college (or being in the treatment group), would be expected, on average, to result in a higher wage.

Considering the first stage, and following Carneiro et al. (2011) we take college attendance as a binary variable, indicating the completion of a college degree for an individual. We aim to incorporate peer effects in the model in (2). For simplicity, we assume linearity in parameters. This restriction can be relaxed without altering the findings of this thesis. To achieve this goal, we consider the following latent model:

$$S_{il}^* = \beta_z Z_{il} + \beta_{\bar{s}} \bar{S}_{(il)}^* + \beta_{\bar{z}} \bar{Z}_{(il)} + \eta_l - \varepsilon_{il}, \quad (6)$$

such that if $S_{il}^* > 0$, then $S = 1$ and individual i attends college, and if $S_{il}^* < 0$ then $S = 0$ and the individual i does not attend college. This basic model is based upon Manski's (1993) Linear in Means model with individual i belonging to the network l . In this equation S_{il}^* is an unobserved continuous variable indicating the benefits for individual i of attending college and Z_i describes all observable individual characteristics that influence the college decision process.

To account for the role that peers play in the decision to attend college, three additional terms are added to the college attendance equation (6), relating to the three avenues through which peer effects may occur. If S_{il} describes the decision to attend college for individual i in network l , then $\bar{S}_{(il)}$ describes the average decision for individual i 's friends (who also appear in network l) to attend college, corresponding to the endogenous peer effect. Similarly, if Z_{il} describes the individual characteristics affecting the college decision, then $\bar{Z}_{(il)}$ describes the average characteristics of an

individual's friends corresponding with the exogenous peer effect. In both cases, the effect is averaged over friends, such that every friend has an equal impact on an individual. Taking the average across peers additionally ensures that the expected total peer effect is equal for every individual, so that an individual with more friends is not subject to a greater peer effect than an individual with less friends.

The final peer effect that must be accounted for is the correlated or environmental effect. This is controlled for by the inclusion of the term, η_l that controls for all unobserved characteristics that are common to a network and may affect both the outcome variable and the network formation process. As discussed, this term does not account for unobserved individual characteristics that vary within a network that may influence the likelihood of forming friendships and simultaneously the decision to attend college. As in much of the literature, due to the difficulty in appropriately controlling for unobserved individual characteristics, we assume this individual effect is negligible and statistically insignificant after controlling for network effects.

The vector Z_i should contain economic, contextual and academic information as well as personality based variables, to account for as much within-network variation as possible. $\bar{S}_{(il)}^*$ represents the benefits of attending college for the average of individual i 's peers. If this value is higher, then individual i 's average friend will have a higher probability of attending college.

We next express the model in matrix form to demonstrate the relation between variables from the inclusion of network components. Following Goldsmith-Pinkham & Imbens (2013), the network is constructed by introducing the adjacency matrix \mathbf{D} such that an element of \mathbf{D} , D_{ij} is equal to 1 if individual i and j are friends and 0 otherwise. We begin under the assumption that friendship is a mutual relation, resulting in a symmetric adjacency matrix. This assumption could be relaxed so that friendship is modelled as a one-way relation, with unequal effects of friendships. By convention, we set the diagonal elements of \mathbf{D} to 0. The number of friends of individual i is denoted $M_i = \sum_{j=1}^N D_{ij}$. Individuals with no friends are discounted from the model so that M_i is positive and peer effects can be estimated for all individuals. \mathbf{M} is the corresponding $n \times 1$ vector so that the i th element describes

the number of friends of individual i . It is necessary to construct a row-normalised adjacency matrix, to describe the average friend of individual i . This matrix is defined as

$$G = \text{diag}(\mathbf{M})^{-1}\mathbf{D}, \quad \text{such that } G_{ij} = \begin{cases} D_{ij}/M_i & \text{if } D_{ij} = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Writing the model in equation (6) in matrix notation we obtain the expression

$$\mathbf{S}_i^* = \beta_z \mathbf{Z}_i + \beta_s G \mathbf{S}_i^* + \beta_z G \mathbf{Z}_i + \boldsymbol{\eta}_i \boldsymbol{\iota} - \boldsymbol{\varepsilon}_i, \quad \mathbb{E}[\boldsymbol{\varepsilon}_i | \mathbf{Z}] = 0 \quad (8)$$

where \mathbf{S}_i^* is an $n \times 1$ vector of the individual benefits of attending college, $\boldsymbol{\iota}$ is an $n \times 1$ vector of ones, \mathbf{Z}_i is an $n \times q$ matrix of individual characteristics, G is an $n \times n$ row normalised interaction matrix and $\boldsymbol{\eta}_i$ is an $n \times 1$ vector describing the unobserved network effect. $\boldsymbol{\varepsilon}_i$ is the unobserved component of the college decision. We require the matrix G to be exogenous conditional on $\boldsymbol{\eta}_i$ and \mathbf{Z}_i . This condition will only hold if there are no unobserved individual characteristics affecting both the network formation process and the probability of attending college. This model is estimated in order to recover the fitted values to be used in outcome equation (3), as in a two-stage least squares (2SLS) treatment effects model. \mathbf{Z} is a matrix made up of the regressors from the outcome equation (1), \mathbf{X} , and the instruments \mathbf{Z}_2 such that $\mathbf{Z} = [\mathbf{X} \ \mathbf{Z}_2]$.

We define the probability of attending college given characteristics $\tilde{\mathbf{Z}}$ to be $P(\tilde{\mathbf{Z}}) = Pr(S = 1 | \tilde{\mathbf{Z}})$ where $\tilde{\mathbf{Z}}$ is a matrix consisting of \mathbf{Z} , the network effect $\boldsymbol{\eta}_i$ and the endogenous and exogenous peer effects, GS and $G\mathbf{Z}$. We also define the variable $U_S = F_\varepsilon(\boldsymbol{\varepsilon})$, where F_ε is the cumulative distribution function of the unobserved component $\boldsymbol{\varepsilon}$. As such it has a uniform distribution between 0 and 1, and U_S corresponds to the different quantiles of $\boldsymbol{\varepsilon}$. U_S is related to $P(\tilde{\mathbf{Z}})$ via the relation $F_\varepsilon(\beta_z \mathbf{Z}_i + \beta_s G \mathbf{S}_i^* + \beta_z G \mathbf{Z}_i + \boldsymbol{\eta}_i \boldsymbol{\iota}) = P(\tilde{\mathbf{Z}})$ such that $S=1$ if $P(\tilde{\mathbf{Z}}) \geq U_S$. When considering changes in the probability of attending college, we see that at the margin when an individual is indifferent to attending college, $P(\tilde{\mathbf{Z}}) = U_S$. Therefore each

value of U_S describes the quantiles of the unobserved component for an individual wanting to attend college. This interpretation becomes important when considering the MTE.

3.2 The Wage Model and Returns to Education

Once the fitted values of the treatment equation are obtained, the wage equation and the marginal returns to a college education can be estimated. $\tilde{\mathbf{Z}}$ must include some or all elements of \mathbf{X} , with some added variables not correlated with the wage outcome, that are used as instruments for the treatment effect. We must therefore consider which peer effect and individual variables included in $\tilde{\mathbf{Z}}$ are determinants of future wages and which can be used as instruments. One difficulty of developing the proposed model is the presence of a time gap in the two estimation steps. The decision to attend college is observed earlier than wages by several years. We must therefore be cautious in the definition of variables to be included in the wage equation. However, this time discontinuity can also be beneficial, allowing some relaxation of the required assumptions. During the time between high school and full employment, an individual's network will evolve fundamentally until it is almost unrecognisable, with many friendships disappearing altogether. Consequentially, peer effects in the original network will diminish over time. The expected direct effect of high school friends on wage is negligible in terms of endogenous effects. The wages and college experience of high school peers would reasonably not be expected to influence the future wage of the individual.

The characteristics of friends, however, may have a longer lasting legacy on an individual and their own characteristics. While the wage or college decision of a high school friend may not have a direct impact on a person's wage, the study habits and social skills developed with friends may influence the performance in the labour market. Therefore, in this initial specification, we include the effect of friend's characteristics, $G\mathbf{X}$ on wages, but do not include the effect of friends' wages, or friends' college decisions. A similar argument can be made for the unobserved network components which are assumed to not be important determinants of wage

in the baseline model. As a result, the instruments we use in the treatment equation (8) are $[\mathbf{Z}_2, G\mathbf{Z}_2, G\mathbf{S}, \boldsymbol{\eta}]$.

Assuming the wage equation (3) is linear in parameters ($\mu_0(\mathbf{X}) = \delta_0\mathbf{X} + \gamma_0G\mathbf{X}$, $\mu_1(\mathbf{X}) = \delta_1\mathbf{X} + \gamma_1G\mathbf{X}$), we obtain the following expression describing the effect of college education on individual wages:

$$Y = \delta_0\mathbf{X} + [\delta_1\mathbf{X} - \delta_0\mathbf{X}]S + [(U_1 - U_0)S + U_0]. \quad (9)$$

Y is an $n \times 1$ matrix of wages and S is an $n \times 1$ matrix of the binary variable indicating college attendance. Adding in the influence of friends' characteristics gives us:

$$Y = \delta_0\mathbf{X} + \gamma_0G\mathbf{X} + [\delta_1\mathbf{X} + \gamma_1G\mathbf{X} - \delta_0\mathbf{X} - \gamma_0G\mathbf{X}]S + [(U_1 - U_0)S + U_0]. \quad (10)$$

Taking expectations conditional on the value of \mathbf{X} and the value of $P(\tilde{\mathbf{Z}})$, we obtain as in (5)

$$\begin{aligned} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p] &= \mathbb{E}[Y_0|\mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p] \\ &\quad + \mathbb{E}[(Y_1 - Y_0)|\mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p]p, \end{aligned} \quad (11)$$

noting that $\mathbb{E}(S|P(\tilde{\mathbf{Z}}) = p) = P(\tilde{\mathbf{Z}}) = p$ by the conditional expectation. From equation (10) with peer effects, we have

$$\mathbb{E}[Y|\mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p] = \mathbf{x}\delta_0 + G\mathbf{x}\gamma_0 + p\mathbf{x}[\delta_1 - \delta_0] + pG\mathbf{x}[\gamma_1 - \gamma_0] + K(p), \quad (12)$$

where $K(p) = \mathbb{E}[(U_1 - U_0)S = 1, P(\tilde{\mathbf{Z}}) = p]$. Letting $\tilde{\mathbf{X}} = [\mathbf{X} \ G\mathbf{X}]$, equation (12) can be expressed as

$$\mathbb{E}[Y|\mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p] = \tilde{\mathbf{x}}\boldsymbol{\psi}_0 + p\tilde{\mathbf{x}}\boldsymbol{\psi}_1 + K(p), \quad (13)$$

where $\boldsymbol{\psi}_0 = [\delta_0 \ \gamma_0]'$ represents a $2n \times k$ matrix with the first n elements equal to δ_0 and the second n elements γ_0 , k is the number of parameters in \mathbf{X} . $\boldsymbol{\psi}_1 = [\delta_1 - \delta_0 \ \gamma_1 - \gamma_0]'$

has a similar interpretation.

The marginal treatment effect (MTE) originally developed by Björklund & Moffitt (1987) and extended in Heckman & Vytlacil (1999,2005,2007b), is defined as the effect of treatment (attending college) on those indifferent to attending college, given characteristics \mathbf{x} and probability of attending college p i.e.

$$\text{MTE}(\mathbf{x}, u_s) = \mathbb{E}(Y_1 - Y_0 | \mathbf{X} = \mathbf{x}, U_s = u_s), \quad (14)$$

which Carneiro et al. (2011) show is equal to

$$\frac{\partial \mathbb{E}(Y | \mathbf{X} = \mathbf{x}, P(\tilde{\mathbf{Z}}) = p)}{\partial p} = \text{MTE}(\mathbf{x}, p). \quad (15)$$

We are concerned with the case when U_S and $P(\tilde{\mathbf{Z}})$ are equal, i.e. when an individual is indifferent to attending college. Individuals with a high value of U_S have a high unobserved cost to attending college, and require a similarly high $P(\tilde{\mathbf{Z}})$ to attend college. A marginal increase in $P(\tilde{\mathbf{Z}})$ starting from a high value of $P(\tilde{\mathbf{Z}})$ induces those who are less likely to attend college, to attend. Those who are already attending college will not be affected by this marginal increase in $P(\tilde{\mathbf{Z}})$. Therefore the MTE at a high value of U_S describes the effect of treatment on those who have a large unobserved characteristic component, making them unlikely to attend college. Similarly at low values of U_S , the probability of attending college only needs to be low for an individual to attend. This means that a marginal change in $P(\tilde{\mathbf{Z}})$ starting from a low value will not affect those who have a high U_S and are already unlikely to attend college. The change in $P(\tilde{\mathbf{Z}})$ will only affect those who are likely to attend college and have a low cost to attending. Therefore tracing the MTE across increasing values of U_S shows us the effect of treatment on those who are increasingly unlikely to attend college. Differentiating (13) with respect to p gives us the MTE.

From this point, we can evaluate the effect of a change in policy, or the effect of a change in one of our instruments on those induced to attend college. As in Carneiro et al. (2011) we use the marginal policy relevant treatment effect (MPRTE) to

evaluate the outcome of a policy on those who are actually affected.

Consider a policy which changes the probability of attending college by an amount epsilon ($P_\epsilon = P + \epsilon$). The policy relevant treatment effect (PRTE) is defined by Heckman and Vytlacil (2005,2007a) as the average effect of going from a baseline policy to an alternative policy for each person who changes treatment due to the policy. Explicitly this is

$$\frac{\mathbb{E}(Y|\text{Alternative Policy}) - \mathbb{E}(Y|\text{Baseline Policy})}{\mathbb{E}(S^*|\text{Alternative Policy}) - \mathbb{E}(S|\text{Baseline Policy})} \quad (16)$$

The MP RTE is simply the marginal version of this statistic and is found by taking the limit of the PRTE as ϵ goes to zero. This ensures that rather than measuring the number of individuals affected, the MP RTE simply describes the effect on those who do change treatment. For example consider an increase in the proportion of an individual's friends that attend college. This is equivalent to increasing the variable GS in the college equation, i.e., $GS_\epsilon = GS + \epsilon$. The MP RTE measures this effect as ϵ goes to 0. Carneiro et al. (2011) find that estimating the MP RTE is equivalent to taking a weighted average of the MTE across the support of the data. The appropriate weight for estimating the MP RTE in terms of the MTE in this example is

$$h_{MP RTE}(\mathbf{x}, u_s) = \frac{f_{P|\mathbf{x}}(u_s)f_{v|\mathbf{x}}(F_{V|\mathbf{x}}^{-1}(u_s))}{\mathbb{E}(f_{v|\mathbf{x}}(\mu_s(\mathbf{Z}))|\mathbf{X})} \quad (17)$$

such that the MP RTE is equal to

$$\int_0^1 MTE(\mathbf{X}, u_s)h_{MP RTE}(\mathbf{x}, u_s)du_s. \quad (18)$$

4 Estimation

4.1 Estimating College Attendance

The first step is to estimate the college attendance model in equation (8). The fitted probabilities must then be recovered in order to estimate the wage equation and the

marginal returns to education. The fitted wage will be a function of the decision to attend college. The college decision model closely follows the setup of Bramoullé et al. (2009), with the additional complication of a binary dependent variable. Therefore, we estimate this model, adapting the method of Bramoullé et al. (2009). As in the spatial econometrics literature, a local within transformation can be used to remove the unobserved network effect. This involves subtracting from individual i 's college equation the average of equation (20) over all i 's friends.

$$(\mathbf{I} - G)\mathbf{S}_l^* = \beta_z(\mathbf{I} - G)\mathbf{Z}_l + \beta_{\bar{s}}(\mathbf{I} - G)G\mathbf{S}_l^* + \beta_{\bar{z}}(\mathbf{I} - G)G\mathbf{Z}_l - (\mathbf{I} - G)\boldsymbol{\varepsilon}_l \quad (19)$$

Difficulties arise when attempting to recover the fitted values from this model. As the matrix G is row-normalised, $(\mathbf{I} - G)$ is non-invertible. Therefore, we can only recover the difference between an individual's outcomes and their average friend's outcomes. The raw fitted values are an integral component in estimating the wage equation and marginal treatment effect. Therefore, this method of simply first differencing is untenable. As an alternative, we include a dummy variable for each network. This will give us an equivalent estimation, with each dummy able to control for the unobserved components that influence the college decision common to the network. For this, we use the following equation:

$$\mathbf{S}_l^* = \beta_z\mathbf{Z}_l + \beta_{\bar{s}}G\mathbf{S}_l^* + \beta_{\bar{z}}G\mathbf{Z}_l + \boldsymbol{\eta}_l\boldsymbol{\alpha}_l - \boldsymbol{\varepsilon}_l, \quad \mathbb{E}[\boldsymbol{\varepsilon}_l|\mathbf{z}] = 0, \quad (20)$$

where $\boldsymbol{\eta}_l$ is a matrix of dummy variables describing which individual belongs to each network and $\boldsymbol{\alpha}_l$ is a vector of coefficients on the dummy variables.

Perfect multicollinearity stemming from the reflection problem (Manski, 1993) must also be addressed. Multicollinearity causes problems in estimation and may produce inconsistent estimates. Bramoullé et al. (2009) propose the following transformation to remove the \mathbf{S}^* matrix from the RHS of the equation:

$$\mathbf{S}_l^* = (\mathbf{I} - \beta_{\bar{s}}G)^{-1}(\beta_z\mathbf{I} + \beta_{\bar{z}}G)(\mathbf{I} - G)\mathbf{Z}_l - (\mathbf{I} - \beta_{\bar{s}}G)^{-1}(\mathbf{I} - G)\boldsymbol{\varepsilon}_l, \quad (21)$$

where the matrix $(\mathbf{I} - \beta_{\bar{s}}G)$ is invertible as $\beta_{\bar{s}}$ is less than one.

Bramoullé et al. (2009) indicate that under this transformation, the model is identifiable as long as \mathbf{I} , G , G^2 and G^3 are linearly independent. This will occur as long as the network diameter is greater than or equal to 3. The diameter is defined as the greatest distance between two individuals within a network. A diameter of three indicates the existence of at least two individuals in the same network that are neither friends nor friends of friends (have a friendship distance of at least three). In this case, $(G^2\mathbf{Z}_i, G^3\mathbf{Z}_i, \dots)$ can be used as valid instruments for $G\mathbf{S}_i^*$. This occurs indirectly, such that the characteristics of an individual i will influence the outcomes of their friends, which will in turn influence the outcomes of their friends of friends.

The resulting model in equation (10), with $S = 1$ if $\mathbf{S}_i^* > 0$ and $S = 0$ otherwise, can be estimated via a linear probability model using instrumental variables. While a probit or logit model with instrument variables may be desirable here, the presence of network effects makes the recovery of the non-linear fitted values difficult, such that the linear probability model provides more consistent estimates for the values of S . We may be concerned about fitted values that lie outside the range of $[0, 1]$, however these only represent a small proportion of the fitted values and can therefore be trimmed from the sample. This trimming may lead to inconsistent estimates for those at the extremes of the distribution and must be considered when interpreting the results.

Under the assumptions of the model, the error terms are i.i.d but need not be homoskedastic. The 2SLS linear probability model is estimated using $\mathbf{W} = [\mathbf{Z} \ G\mathbf{Z} \ G^2\mathbf{Z} \ \boldsymbol{\eta}]$ as instruments to obtain an estimate for $\hat{\theta}$. The regressors are $\tilde{\mathbf{Z}} = [\mathbf{Z} \ G\mathbf{Z} \ G\mathbf{S} \ \boldsymbol{\eta}]$ with endogenous variable $G\mathbf{S}$. This estimation results in consistent estimates for the model parameters.

We use the estimated θ 's to determine the predicted value of $P(\tilde{\mathbf{Z}})$, i.e. the probability of attending college for an individual given their individual attributes, the average individual attributes of their friends, the average college decision of their friends and the unobserved network effect.

4.2 Estimating Marginal Returns to Education

Estimation of the wage equation and marginal treatment effects can be achieved following one of the approaches proposed by Carneiro et al. (2011). The first is a parametric approach using maximum likelihood. This requires the assumption of joint normality of (U_0, U_1, ε) and independence of these error terms from (\mathbf{X}, \mathbf{Z}) . This assumption is strong and will not hold in most cases, indicating that a semiparametric approach may be necessary. In the discussion that follows, the semiparametric method will be estimated under the more flexible assumption that (\mathbf{X}, \mathbf{Z}) are independent of (ε, U_1, U_0) . Under this semiparametric model, the MTE is identified over the marginal support of $P(\tilde{\mathbf{Z}})$ i.e. the values of $P(\tilde{\mathbf{Z}})$ that are observed such that the distribution of P is greater than zero given observed variables $\tilde{\mathbf{X}} = [\mathbf{X} \ G\mathbf{X}]$ and college attendance S . Estimating conventional treatment parameters such as the average treatment effect (ATE), the treatment effect for the treated (TT) and the average treatment effect for the untreated (TUT), requires the full support of P across the interval. If we do not observe this full support of P , then the non-parametric estimation cannot identify these statistics. These measures do not take into account heterogeneity of returns. However, the non-parametric estimation can identify the MP RTE. Estimating the MP RTE occurs on the MTE over the support of the data, controlling for heterogeneity.

Estimation of the MTE from this point follows a two stage methodology. Firstly, the Peter M. Robinson (1998) method for estimating partially linear equations is used to obtain estimates of ψ_0 and ψ_1 . Subtracting the expected value of equation (13) from itself removes the non-linear component in P . We obtain the following equation:

$$Y - \mathbb{E}(Y|P(\tilde{\mathbf{Z}})) = [\tilde{\mathbf{x}} - \mathbb{E}(\tilde{\mathbf{x}}|P(\tilde{\mathbf{Z}}))]\psi_0 + p[\tilde{\mathbf{x}} - \mathbb{E}(\tilde{\mathbf{x}}|P(\tilde{\mathbf{Z}}))]\psi_1 \quad (22)$$

Estimation is achieved by running kernel regressions of the dependent variable and each of the regressors on P in order to estimate the expected values in equation (22). OLS is then run using the residuals from these kernel regressions to determine δ_0 ,

δ_1 , γ_0 and γ_1 .

Once we have these parameter estimates, we consider estimating $K(P(\tilde{\mathbf{Z}}))$. From equation (12) we see that

$$\mathbb{E}(Y - \tilde{\mathbf{X}}\hat{\psi}_0 - P(\tilde{\mathbf{Z}})\tilde{\mathbf{X}}\hat{\psi}_1|P(\tilde{\mathbf{Z}})) = K(P(\tilde{\mathbf{Z}})). \quad (23)$$

From here, a local polynomial regression of $Y - \tilde{\mathbf{X}}\hat{\psi}_0 - \hat{P}(\tilde{\mathbf{Z}})\tilde{\mathbf{X}}\hat{\psi}_1$ on $\hat{P}(\tilde{\mathbf{Z}})$ is run in order to estimate the function $K(P(\tilde{\mathbf{Z}}))$ and its partial derivative with respect to $P(\tilde{\mathbf{Z}})$.³ The partial derivative is then added to the linear component of equation (13) in order to estimate the MTE.

Recall that in order to estimate the MP RTE we take a weighted average of the MTE across the support of $P(\tilde{\mathbf{Z}})$. The formula for this weighted average is given in equation (17). However, the relevant weight is expressed conditional on the value of \mathbf{X} and therefore to determine the MP RTE, after applying the weight, the parameters must be integrated over the distribution of \mathbf{X} .⁴ We can measure different forms of the MP RTE using different weighting functions and definitions of policy changes. In particular we consider a policy change that directly increases the probability of attending college equally for all individuals ($P_\epsilon = P + \epsilon$), a policy which proportionally increases the probability of attending college ($P_\epsilon = P(1 + \epsilon)$) or a policy that affects one of the instruments used in the treatment equation ($Z_\epsilon = Z + \epsilon$). In each of these cases we will be able to ascertain the effect of peers on policy outcomes.

5 Monte Carlos

5.1 Baseline Model

From this point, we use Monte Carlo experiments to examine alternate network specifications and conduct sensitivity analyses. This process involves generating

³Fan 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.

⁴Since conditioning on \mathbf{X} is computationally demanding due to the possible high dimension of \mathbf{X} , as in Carneiro et al. (2011), we condition on the index $\mathbf{X}(\delta_1 - \delta_0)$ as an approximation.

data from predefined distributions and estimating the model, enabling us to not only examine the sensitivity of the model to changes in specification, but also determine the relative effect of a change in peer effects on the outcome variable. As such we can explore the role networks play in policy changes that impact the decision of individuals to receive treatment.

We begin using the specification given in the model section, running $N = 1000$ simulations with $n = 500$ observations per simulation. The simulation codes are developed and run in Matlab, using the codes of Carneiro et al. (2011) as reference.⁵

To simulate the data, the error terms (ε, U_1, U_0) are drawn from a joint normal distribution such that $[\varepsilon, U_1, U_0] \sim N(\mu_\varepsilon, \Sigma_\varepsilon)$ where,

$$\mu_\varepsilon = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & \rho_1 & \rho_0 \\ \rho_1 & 1 & \rho_2 \\ \rho_0 & \rho_2 & 1 \end{pmatrix}$$

with $\boldsymbol{\rho} = [\rho_0, \rho_1, \rho_2] = [0.2, 0.6, 0.5]$. We require the value of ρ_1 to be greater than ρ_0 , so that the unobserved component of the treatment equation has a greater relation with the outcome equation for those attending college and the wages are positively related to the probability of attending college. For simplicity we let $\mathbf{Z} = [\mathbf{X} \ \mathbf{Z}_2]$ be an $N \times 2$ joint normally distributed matrix, such that $Z_i \sim N(0, I_2)$, where the variable \mathbf{X} is the first column and the second column is the instrument for the selection model, \mathbf{Z}_2 . In this initial specification, we estimate the network graph randomly. We start by generating four complete networks. The size of the networks are drawn as a random partition of the sample size n . We construct complete graphs using the Watts-Strogatz (1998) random graph generation process. A ring lattice graph is generated with each individual connected to its four closest neighbours (two on either side). Links are rewired with a defined probability 0.5, such that with probability 0.5 a link is removed and the individual is connected randomly to another individual in the network. Changing the rewiring probability affects the randomness of the graph,

⁵The codes (Stata, Gauss, Matlab and R) of Carneiro et al. (2011) are available as part of the additional materials for their paper on the AEA journal website.

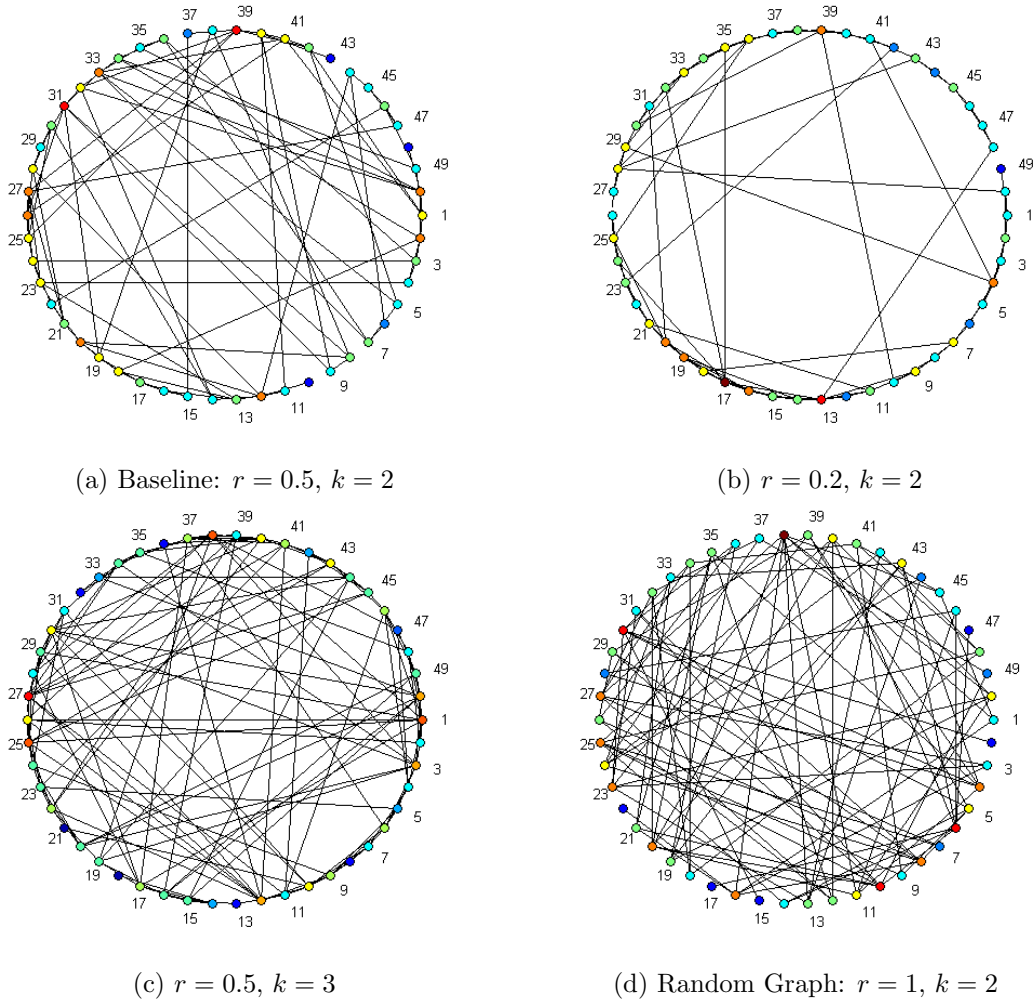


Figure 1: Watts-Strogatz network graphs

and as a result the diameter of the network. The density of the network can easily be modified by changing the number of neighbours each individual is initially linked to. Associated with each of the four networks is a dummy accounting for unobserved network effects. In Figure 1 we see network graphs across values of the rewiring probability r and the number of initial neighbours on each side k . Figure 1d displays a completely random graph corresponding to a rewiring probability of 1.

We define the initial values of the parameters to be estimated as

$$\begin{aligned} [\beta_X, \beta_{Z_2}, \beta_{\bar{X}}, \beta_{\bar{Z}_2}, \beta_{\bar{S}}] &= [1, 1, 0.5, 0.5, 0.5], \\ [\delta_0, \delta_1, \gamma_0, \gamma_1] &= [1, 1, 0.5, 0.5], \\ \boldsymbol{\alpha} &= (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \sim N(0, I_4). \end{aligned}$$

The outcome variables S^* and Y are constructed using these defined parameters and variables in the model equations. The binary variable S is determined using the indicator function such that $S = \mathbb{1}(S^*)$. This specification will act as our initial baseline model.

Note that under this initial specification, we assume that \mathbf{X} and $G\mathbf{X}$ have an equal effect on the wage regardless of the decision to attend college. This implies that, holding all else constant, the sole effect of increasing the value of \mathbf{X} and $G\mathbf{X}$ will be a negative proportional reduction in returns. The peer effect in both the treatment and wage equation is assumed to be exactly half of the individual effect.

Running this simulation and bootstrapping the errors, we obtain a plot of the MTE across values of U_S . The MTE is only defined over the support of $P(\tilde{\mathbf{Z}})$. Figure 2 displays the density function for the support of P , conditional on the observations of \mathbf{X} . Due to the linear relationship between \mathbf{X} and P , we see that at low values of \mathbf{X} , there is a high density of low P values and at high values of \mathbf{X} , there is a high density of high P values. In general, we observe that the support of P covers the majority of the interval.

We proceed conducting a comparison of the model containing no peer effects with the baseline model described, utilising the Watts-Strogatz network formation process. The MTE graphs are presented in Figure 3. Firstly, note that the graph is downwards sloping. Recall U_S represents the quantiles of the desire to attend college, or the unobserved components of the college decision.

Regarding Figure 3a, at low values of U_S (and high probabilities of attending college), the returns to attending college are positive, with a maximum return of 17.2% to receiving treatment. At high values of U_S (and low probabilities of attending

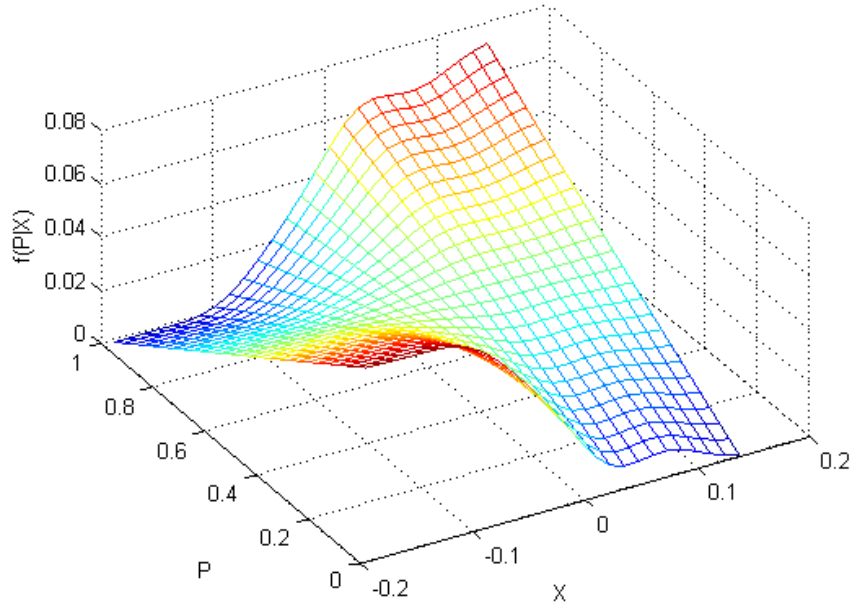


Figure 2: Support of P conditional on \mathbf{X}

college), the returns to attending college are negative, with a minimum return of -16.5%. The MTE graph therefore takes the desired shape and is roughly symmetric around zero, accompanying the symmetry of the initial assumptions. Comparing this with Figure 3b, we see that the initial introduction of networks has little effect on the scale or shape of the MTE. This similarity should be expected and can be explained by the specification of zero average network effects and completely random friendships. Both the size of the networks and the associated unobserved network components are randomly generated with mean zero, so that the effect of a network on the treatment and the outcome is, on average, zero. While the values of \mathbf{X} and $G\mathbf{X}$ have a positive effect on the probability of attending college, the peer effect $G\mathbf{X}$ is independent of the individual and completely random. Consider a student who has friends, on average, with high values of the characteristic \mathbf{X} , leading to a high value of $G\mathbf{X}$ and increasing the value of $P(\tilde{\mathbf{Z}})$ for the individual. The value of U_S for which they are at the margin will now increase. The effect of this high value in $G\mathbf{X}$ is to increase the wage additively by a factor of 0.5, proportionally reducing the return. As $G\mathbf{X}$ acts almost identically to \mathbf{X} and is independent, random and centred

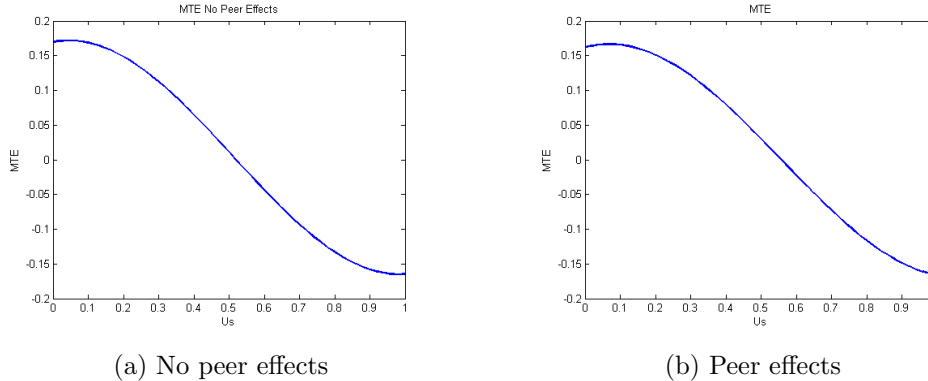


Figure 3: Baseline graphs of the marginal treatment effect (MTE)

around zero, this small proportional effect on the MTE is negligible. The inclusion of networks, however, is more noticeable when considering changes in policy.

We examine two types of policies; regular substantive policy changes and marginal policy changes to the probability of treatment. The marginal policies we consider are an increase in the instruments $\tilde{\mathbf{Z}}^6$ by the marginal amount ϵ and an increase in the probability of attending college for all individuals both additively (an equal increase in probability for each individual) and multiplicatively (an equal proportional probability increase for each individual). We also consider the PRTE for increases in peer related variables, namely, an increase in the proportion of peers that undergo treatment (GS) by 1% and an increase in the characteristic \mathbf{Z}_2 by 0.01.

A blanket policy change affecting all individuals will induce a greater number of people into treatment when networks are present due to the social multiplier effect. The PRTE and the MPRTTE, however, do not examine the number of individuals that transition into the treatment group, but the returns for those individuals that do so. Table 1 presents the comparison of the effect of networks on policy changes. We see that for marginal policy changes, the introduction of networks has a minimal effect on the MPRTTE values, which can be discounted by randomness. However, an increase in the value of \mathbf{Z}_2 by 0.01 has a significantly higher impact in the peer effects model. As the variable \mathbf{Z}_2 increases, more individuals are enticed into attending college in the networks model through the influence of peers, both exogenously (through the

⁶The variables used as instruments in this model are $\tilde{\mathbf{Z}} = [GS \ \mathbf{Z}_2 \ G\mathbf{Z}_2 \ \boldsymbol{\eta}]$.

increase in $G\mathbf{Z}_2$) and endogenously as the number of individuals, and therefore peers, attending college (GS) increases. Furthermore, those who are induced into college by an increase in \mathbf{Z}_2 , on average, gain more in the network model. Those enticed into treatment are likely to be those with more moderate \mathbf{X} and $G\mathbf{X}$ values, who receive higher returns from attending college. This increase in returns is partially due to the higher proportion of individuals who now receive treatment.

Confidence intervals are not included in this image but can be obtained via bootstrapping (see Figure A.1). The bootstrap shows us that the associated standard error is large, so that we cannot determine if the slope is significant for a single simulation. These results are likely to improve if we increase the value and size of our initial data.

We must be careful when attempting to draw intuition from these results. In what follows, analysis is conducted solely at the margin. These returns are not representative of individuals not at the margin, who may display significantly different returns to college. Nonetheless, analysis at the margin is desirable, as these individuals are the ones most likely to be affected by changes in policy.

Table 1: Incorporating peer effects

Policy Change (MPRTE)	No peer effects model	Peer effects model
$\mathbf{Z}_2 = \mathbf{Z}_2 + 0.01$	0.0925	0.1058
$Z = Z + \epsilon$	0.1140	0.1179
$P = P + \epsilon$	0.1081	0.1104
$P = (1 + \epsilon)P$	0.08550	0.0867

5.2 Alternate Network Models

Network structure can determine the nature and strength of the peer effect. Altering the density and randomness of the network changes the number of friends an individual has and how connected they are in the network. Figure 4a demonstrates the effect of increasing the density of the network on the MTE ($k=3$ so that each individual is initially connected to 6 neighbours). Figure 4b shows the effect of de-

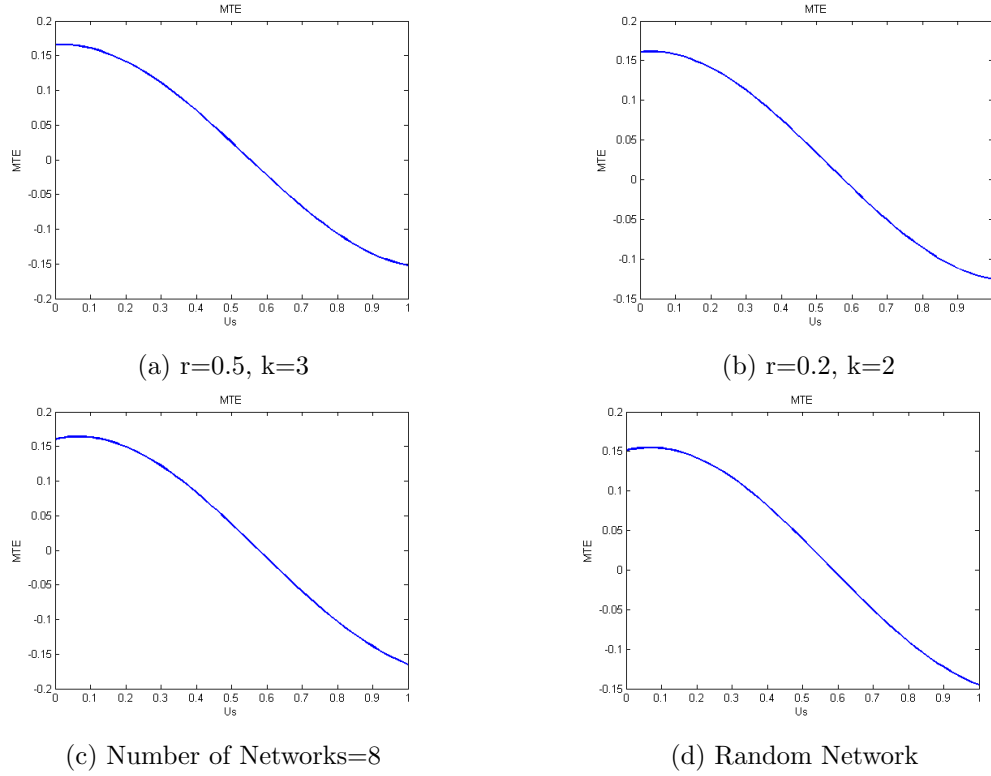


Figure 4: MTE - Changes in network specification

creasing the rewiring probability to $r = 0.2$ and Figure 4c the effect of increasing the number of networks to eight. Figure 4d displays the MTE for a completely random network, such that $r = 1$.

There appear to be no major differences in the resulting MTE graphs, such that the network structure does not have a significant impact on the MTE, given the initial assumptions. In each case most of the variation can be attributed to randomness. This is not surprising, given the introduction of networks has very little initial impact on the MTE.

The effect of changes in network structures on policy are recorded in Table 2. Again we see negligible differences in the MPRTE values. We see that all network models outperform the no network model when increasing the value of Z_2 , consistent with the established result. Some networks appear to have a stronger effect than others. The density of the network appears to have no effect on the policy returns. Increasing the density of the graph narrows the distribution of $G\mathbf{X}$, as the average

characteristics of friends are drawn closer to zero, reducing the already negligible effect of networks. Adjusting the rewiring probability appears to influence the returns to treatment, with the random network displaying the highest return. Increasing this randomness implies that individuals are more interconnected, increasing the dispersion and therefore effect of policy in the network.

Table 2: Effect of networks on policy

Policy Change (MPRTE)	r=0.5, k=2	r=0.5, k=3	r=0.2, k=2	8 Networks	Random Network	No Network
$Z_2 = Z_2 + 0.01$	0.1058	0.1024	0.1107	0.0978	0.1133	0.0925
$GS = GS + 0.01$	0.1106	0.1130	0.1091	0.1189	0.1108	-
$Z = Z + \epsilon$	0.1179	0.1116	0.1135	0.1142	0.1135	0.1140
$P = P + \epsilon$	0.1104	0.1036	0.1065	0.1062	0.1064	0.1081
$P = (1 + \epsilon)P$	0.0867	0.0793	0.0842	0.0861	0.0862	0.0855

5.3 Relating Network Formation to Z

Considering again the network structure, the assumption that the network formation process is unrelated to the values of Z seems unlikely. Furthermore, without correlations between friends, the characteristics of friends of friends cannot be used as valid instruments in the estimation. There are several ways in which we can construct networks based upon the observed variables. The simplest method is to compare all values of X and establish links between individuals if the difference between their characteristic X lies within a certain bound. If both variables in Z are used, we can analyse the distance between both variables to establish links, either by taking the sum of characteristics, or by establishing links with the variables separately. This process, however, leads to a very particular graph structure with a large diameter, so that those with high values of X are very far from those with low values of X . The process is improved by imperfectly defining these links, establishing friendships with a predefined probability. The number of links (and density of the graph) can be adjusted by varying the bound on the distance between characteristics. However, this still implies that all friendship connections occur between those with similar characteristics, with no randomness occurring in the graph.

To improve this network formation process, we can adapt the Watts-Strogatz method. We start by linking all individuals based upon the values of the variables in \mathbf{X} . Individuals are linked to their k neighbours with the closest values of \mathbf{X} . We then destroy links and create new random links with probability r . As in the original Watts-Strogatz method, increasing r increases the randomness of the graph and increasing k increases the density of the graph. In this case, however, r also determines to what extent individuals with similar characteristics befriend each other. We refer to this parameter from now on as homophily, such that a graph with $r = 0$, links individuals exactly with those closest to their characteristics \mathbf{X} . Such a network can be described as perfectly homophilic.

The MTE plots for the random and Watts-Strogatz graphs based upon both variables in \mathbf{Z} are found in Figure 5 with policy effects given in Table 3. The results are presented for networks where links are defined by the sum of the characteristics in \mathbf{Z} and for networks with links defined by either variable in \mathbf{Z} .⁷ In reality, we would expect a mixture of these two phenomena, such that individuals seek friends that are strongly similar in a few characteristics and not too opposingly different in other characteristics. Graphs based on the value of \mathbf{X} alone are in the appendix A.2.

Table 3: Linking networks and \mathbf{Z}

Policy Change (PRTE)	Basic Link, $\mathbf{X}+\mathbf{Z}_2$	Basic Link $\mathbf{X}\&\mathbf{Z}_2$	Watts-Strogatz, $\mathbf{X}+\mathbf{Z}_2$	Watts-Strogatz, $\mathbf{X}\&\mathbf{Z}_2$	Random Network
$\mathbf{Z}_2 = \mathbf{Z}_2 + 0.01$	0.0952	0.1136	0.0917	0.1111	0.1133
$GS = GS + 0.01$	0.1046	0.1184	0.0958	0.1236	0.1108
$Z = Z + \epsilon$	0.1001	0.1185	0.0949	0.1152	0.1132
$P = P + \epsilon$	0.0922	0.1099	0.0882	0.1078	0.1066
$P = (1 + \epsilon)P$	0.0813	0.0983	0.0725	.0875	0.0898

Comparing the Watts-Strogatz graphs and the basic link graphs, we see that while the scale of the MTE are similar, the shape of the graphs differ. The shape of the Watts-Strogatz graph bears a greater resemblance to the baseline graph with no network correlation, explained by the increase in randomness.

Considering the MTE for the network with Watts-Strogatz links in relation to the

⁷In this case, two friendships are created based on the value of \mathbf{X} and two are created based on \mathbf{Z}_2 .

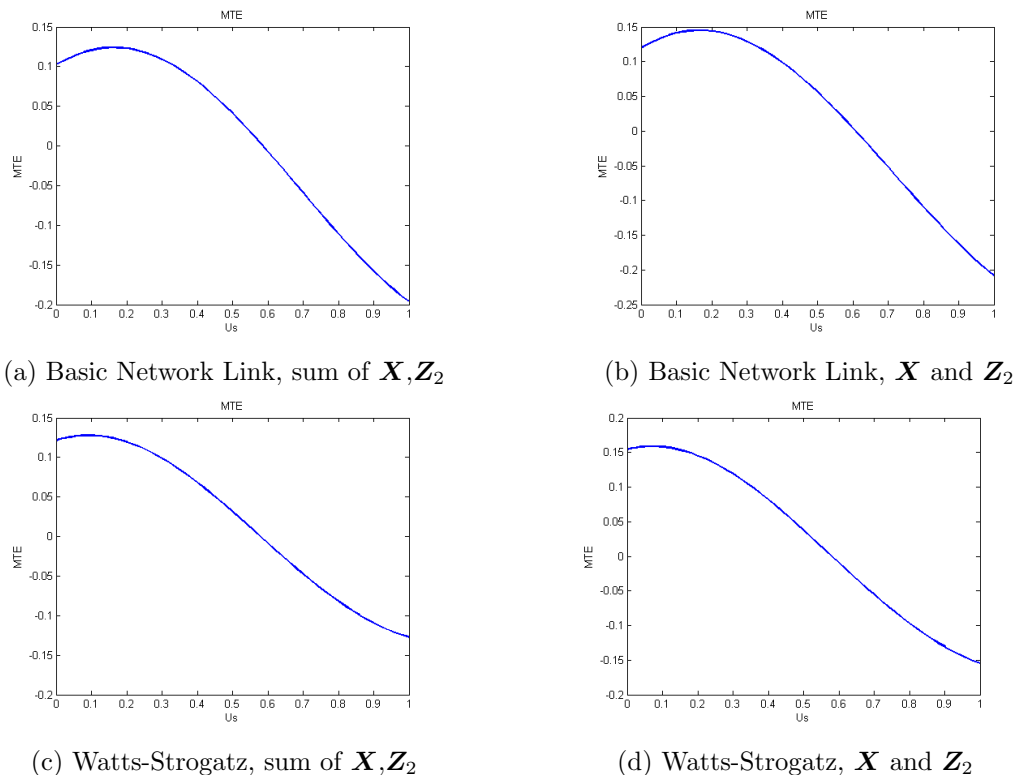


Figure 5: MTE - Correlating network formation with \mathbf{Z}

baseline model, we see that at high values of U_S , the negative returns are increased, while at low values of U_S , returns are reduced. An individual who has a high positive value of \mathbf{X} and \mathbf{Z}_2 is likely to be at the margin only for high values of U_S . With the introduction of correlated networks, peer characteristics are much more correlated with one's own characteristics, leading to high values of $G\mathbf{X}$ and $G\mathbf{Z}_2$. This increase in correlation is exaggerated further through the endogenous peer effect, as similar peers are likely to exhibit the same college decision, correlating GS with the other variables. With this higher correlation between regressors, the distribution of $P(\tilde{\mathbf{Z}})$ is widened so that there are more individuals with high or low values of $P(\tilde{\mathbf{Z}})$ and less in the centre. Accompanying this increased distribution, the average size of \mathbf{X} and $G\mathbf{X}$ at the margin is reduced (values are reduced at high U_S and increased at low U_S). The effect of this is to increase returns for those with a low probability of attending college (high U_S), and reduce returns for those with a high probability of attending college (low U_S).

Comparing this to the graphs where networks are determined perfectly by the values in \mathbf{Z} , we see that the size of the MTE is diminished further, with some distortion at high values of U_S . This distortion is likely to result from the almost perfect collinearity of \mathbf{Z} with $G\mathbf{Z}$. By introducing some randomness into the network formation, the contracting effects on the MTE are diminished, particularly at very high and low values of \mathbf{X} .

Regarding the networks where the components of \mathbf{Z} are used deterministically, we see that the range of the MTE expands. When friendships are determined by either \mathbf{X} or \mathbf{Z}_2 , the effect of homophily is reduced. In this case, the \mathbf{X} and \mathbf{Z}_2 values are treated separately and are therefore not correlated. An individual may be friends with someone with a high \mathbf{X} but low \mathbf{Z}_2 characteristics, reducing the total peer effect in this way. While there is a correlation between \mathbf{X} and $G\mathbf{X}$, these variables are independent of the rest of the regressors, such that the effect on the distribution of $P(\tilde{\mathbf{Z}})$ does not differ significantly from the distribution in the randomly linked network. Acknowledging this, we observe that the network performs similarly to that of a random network.

In contrast, the summation linked graphs reduce the policy effect in each case. Those induced into treatment will now, on average, be those with more moderate negative values of \mathbf{Z} , as those with extreme values in \mathbf{Z} experience similar peer effects, pulling them away from the margin. Those affected by policy now receive less benefit from attending college, decreasing the values of the PRTE and MPRTE.

We see, therefore, that the overall effect of homophily in the Watts-Strogatz graph is to reduce the spread of the MTE, so that the size of returns is diminished. Strikingly, homophily also works to reduce the effect of policy. Those with high probability of attending college will be even more likely to attend college, due to the similarity of their peers, while those with unfavourable characteristics will be even less likely to enrol. Those who are influenced to attend college by policy, now have less favourable characteristics and will therefore have lower returns to attending college.

Varying the network properties of the correlated graph (Figure A.3), we see that

as the probability of rewiring increases and the network becomes more random; the range of the MTE increases, reducing the effect of homophily. When modifying the density of the graph, the shape of the MTE curve changes. If the probability of rewiring is kept constant, the effect of lowering the density is to increase the effect of homophily. As the density decreases, the average number of friendship connections decreases and therefore the correlation with friend’s characteristics increases. In this case the scale of the graph is reduced as homophily is amplified.

From here we proceed making comparisons with the three base models discussed; the model with no networks, the model with peer effects and random friendships using Watts-Strogatz and the model with networks determined by the sum of the values in \mathbf{Z} , involving homophily. We denote the model with random friendships N_{rand} and the model with correlated friendships N_{cor} .

5.4 Sensitivity Analysis

Varying α

To extract the greatest understanding of the implications of incorporating networks into the model, we consider the case when the unobserved network correlated effects have a mean that differs from zero. We consider this in the context of fixed network sizes so that the number of individuals who experience the unobserved network effect is constant over time. The results are presented in Figure 6, with MPRTE values given in Table 4. We can also extract insight into the impact of network sizes by varying the distribution of people across the networks. Figure 6 plots the MTE when networks have sizes [200 100 100 100], while Figure A.4 and Table A.1 in the appendix contain the results for networks of size [300 150 40 10].

Firstly, comparing the MTE across the random and correlated networks provides clear evidence of homophily reducing the inequality in the MTE, particularly evident in the networks of size [300 150 40 10].

α is used as an instrument in the treatment equation and therefore has no effect on the wage outcome. A change in α influences only the probability of attending

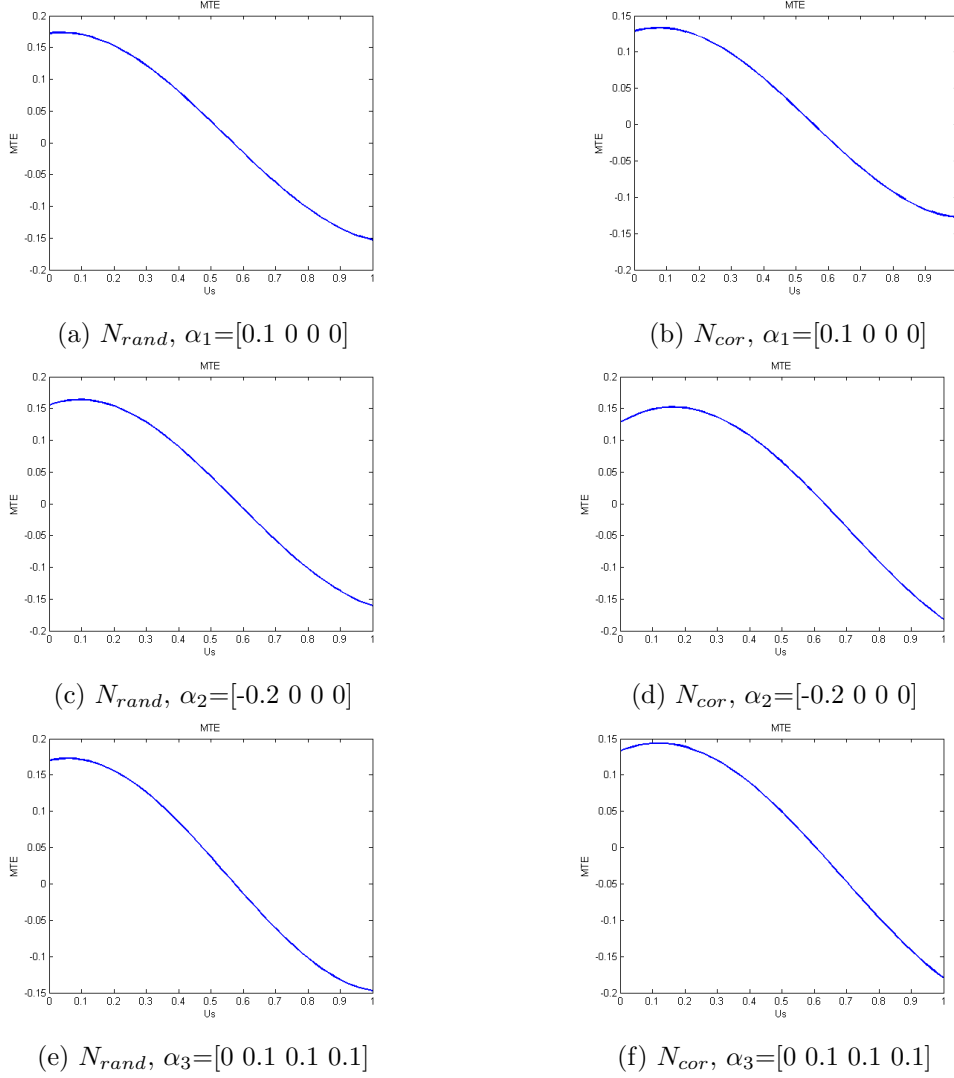


Figure 6: MTE - Network size=[200 100 100 100]

college, and therefore the distribution of $P(\mathbf{Z})$ compared to U_S . An increase in the α values appears to shift the MTE curve upwards at all U_S . α increases the probability of attending college for all individuals in that network. As a result, those at the same margin of U_S have lower average values of the other regressors, including \mathbf{X} and $G\mathbf{X}$, so that lower values of \mathbf{X} and $G\mathbf{X}$ are required for attendance at college. Recalling that reducing \mathbf{X} and $G\mathbf{X}$ will proportionally increase returns, we see a small shift in the MTE corresponding with the change in α . The opposite effect occurs for a negative α value. These effects are only small and localised to the network that is affected. With an increase in α for the first network, there will be less individuals in

network 1 at the margin for very low U_S values and more at high U_S values. This will distort the MTE at the edges of the curve.

The network sizes appear to have little effect, except in defining the number of individuals subject to the increase in the unobserved component.

Some of these graphs display a slight peak in the MTE. This is a product of the polynomial modelling and generally appears once the MTE has been shifted. It indicates a shift in the proportion of individuals undergoing treatment. As in the α case, this section may be subject to some distortionary effect representing individuals in networks not affected by the change in α . However, the results on this section of the MTE curve cannot be taken as deterministic due to the potential trimming of extreme values and the polynomial modelling.

Table 4 and A.1 describe the policy effects when the unobserved network effects are varied. In general increasing α leads to a small increase in the MP RTE in line with the MTE effect. Therefore we see that increasing the environmental effect induces a greater effect of policy and greater treatment effects for all individuals.

Table 4: Changing the unobserved network effect

Policy Change (MP RTE)	N_{rand}, α_1	N_{cor}, α_1	N_{rand}, α_2	N_{cor}, α_2	N_{rand}, α_3	N_{corr}, α_3	Baseline N_{cor}
$\mathbf{Z}_2 = \mathbf{Z}_2 + 0.01$	0.1324	0.08399	0.1345	0.0820	0.1448	0.1164	0.0917
$GS = GS + 0.01$	0.1071	0.0902	0.1071	0.0901	0.0879	0.1101	0.0958
$Z = Z + \epsilon$	0.1218	0.08801	0.1188	0.0924	0.1193	0.1154	0.0949
$P = P + \epsilon$	0.1148	0.0811	0.1111	0.0876	0.1114	0.1087	0.0882
$P = (1 + \epsilon)P$	0.0898	0.0651	0.0906	0.0829	0.0879	0.0932	0.0725

Examining the outcome equation

We next analyse alterations to the specification of the model to test for robustness. We consider the effect of removing $G\mathbf{X}$ from the outcome equation, such that the characteristics of peers do not affect the outcome. We also consider the inclusion of the unobserved network components and GS individually into the outcome equation. GS and $G\mathbf{X}$ cannot be incorporated simultaneously in the outcome equation due to the presence of the “reflection problem” which proves difficult to navigate when

estimating the nonlinear MTE.

Figure A.5 demonstrates that these specifications dramatically impact the results of the model. In the case where GS is included in the outcome equation, the MTE curve becomes solely negative, so that there are negative returns to attending college for those at the margin. Including the unobserved network components, we find the curve tends to be condensed. In this model, an increase in the unobserved component increases both the probability of attending college and the wage by an identical amount. Those within a network who have the same α value are more likely to have similar wages, increasing the effect of homophily and ensuring the change in returns are smaller across the MTE. Some of these specifications prove to make identification of the MTE difficult. Removing all peer effects from the outcome equation is especially problematic. We see, therefore, that the model is sensitive to the inclusion of the peer effect terms in the outcome and treatment equations. This is indicative of the contrasting roles of the endogenous and exogenous peer effects, and the interactions between the peer effects and the non-linear estimation.

The policy effects of such specifications are found in Table 5. Most specifications have little effect on the MP RTE measures despite the impacts on the MTE. The MP RTE policy changes consider only the effect of policy on a general increase in the probability of treatment and are therefore invariant to the inclusion of variables in the outcome equation. The specification does play a role in the effect of non-marginal policy changes (PRTE). For example, the policy effect of increasing the value of \mathbf{X} for every individual is reduced by introducing new terms into the outcome equation which increase the scale of wages. Each of these have an intuitive explanation. Removing $G\mathbf{X}$ from the outcome equation reduces the size of wages for all individuals, amplifying the proportional difference between the outcome equations and exaggerating the effect of homophily. This explanation is confirmed by comparing the peer effect for the randomly linked networks, which have a much higher policy effect in each case.

Table 5: Alternate model specifications

Policy Change (MPRTE)	N_{rand} , GX removed	N_{cor} , GX removed	N_{rand} , GS included	N_{cor} , GS included	N_{rand} , η_l included	N_{cor} , η_l included	Baseline N_{cor}
$Z_2 = Z_2 + 0.01$	0.0951	0.1077	0.1266	0.076132	0.1430	0.0852	0.0917
$GS = GS + 0.01$	0.0782	0.0981	0.1171	-	0.1292	0.1050	0.0958
$Z = Z + \epsilon$	0.1090	0.0953	0.1105	0.093023	0.1236	0.0904	0.0949
$P = P + \epsilon$	0.1000	0.0888	0.1021	0.084944	0.1159	0.0839	0.0882
$P = (1 + \epsilon)P$	0.0743	0.0677	0.0808	0.070242	0.0919	0.0654	0.0725

Effects of coefficients

Adjusting the coefficients in the specification enables us to further analyse the performance of the model. We consider changes in the coefficients of \mathbf{X} and \mathbf{Z} in both the treatment and outcome equations. The results in Figure 7 and Table 6 demonstrate that the basic structure of the MTE graphs are not affected, with the most significant change found in the scale of the graphs and changes in the MPRTE estimates.

Notably, increasing the effect of \mathbf{X} and $G\mathbf{X}$ in the outcome model changes the shape of the MTE curve. Considering this increase for the non-attendance equation Y_0 , the benefits of attending college reduce for those with \mathbf{X} and $G\mathbf{X}$ values greater than 0. These individuals are likely to be at the margin for high values of U_S and in this case the gap between the wages will decrease. For \mathbf{X} and $G\mathbf{X}$ values less than 0, returns will increase, as the gap between the wages increases. The MTE therefore spreads out, particularly given the correlation of \mathbf{X} and $G\mathbf{X}$.

Increasing the effect of \mathbf{X} and $G\mathbf{X}$ in the attendance outcome equation Y_1 , modifies the shape of the MTE considerably. The size of the MTE reduces substantially, and the shape appears to be, for the most part, concave. Following a similar thought process we see that at high U_S we expect to see a reduction in returns and at low U_S , increased returns.

Figure 7c further decomposes this effect by looking at an increase in the effect of peers' characteristics only for those attending college. Under this change, the MTE still reduces in size but not to the same extent. As in the MTE, the MPRTE values reduce significantly under these conditions.

In conclusion, altering the coefficients in the outcome equation has a direct effect on the wage returns. Alterations for the college attendance outcome increase the

variation in returns at the margin, while alterations for the non-attendance outcome has the opposite effect.

Turning to the treatment equation, an increase in the endogenous peer effect slightly reduces the MTE and MP RTE. The effect is identical to an increase in α , but is more pronounced due to the correlation of GS with \mathbf{Z} and $G\mathbf{Z}$ associated with homophily, diminishing the MTE over U_S . Decreasing the coefficient of \mathbf{X} or \mathbf{Z}_2 reduces the effect of homophily for the MTE and MP RTE. As before, those with positive values of \mathbf{X} will now have a smaller value for $P(\tilde{\mathbf{Z}})$. The value of \mathbf{X} at the margin is shifted, spreading out the values of the MTE .

When considering changes in the coefficients of the instruments in the treatment effects model, the effects on the MTE and MP RTE measures are smaller. The effect is on the distribution of $P(\tilde{\mathbf{Z}})$, rather than directly impacting wages. However, the effects on policy changes are more pronounced, determining the scale and characteristics of the individuals that are induced into treatment.

The MP RTE, in this case, is somewhat more informative. While the MTE describes the returns to college across all quantiles of the desire to attend college, the MP RTE weights the policy response for those induced into college attendance. We see that changing the coefficient of GS marginally reduces the effect of policy, while decreasing the coefficients of \mathbf{Z} and $G\mathbf{Z}$ has a positive effect on the MP RTE.

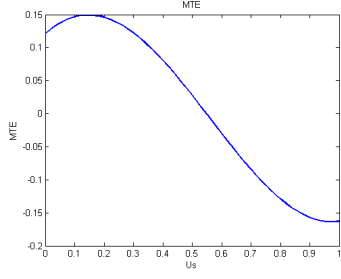
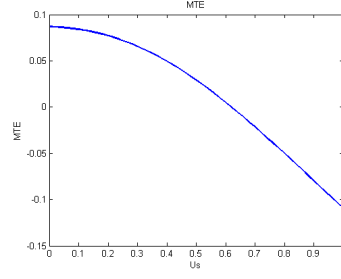
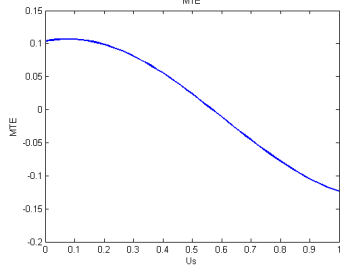
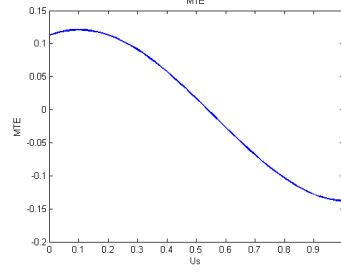
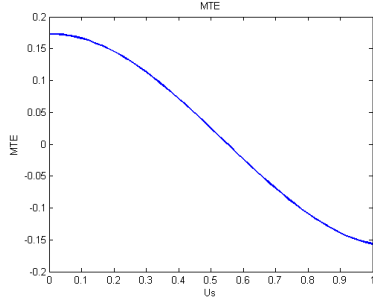
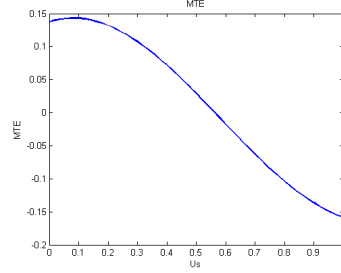
(a) $N_{cor}, \delta'_0 = [1.5 \ 0.8]$ (b) $N_{cor}, \delta'_1 = [1.5 \ 0.8]$ (c) $N_{cor}, \delta''_1 = [1 \ 0.8]$ (d) $N_{cor}, \beta_{\bar{s}} = 0.8$ (e) $N_{cor}, \beta'_Z = [0.5 \ 1], \beta''_Z = [0.2 \ 0.5]$ (f) $N_{cor}, \beta''_Z = [1 \ 0.5], \beta'_Z = [0.5 \ 0.2]$

Figure 7: MTE - Altering model coefficients

Table 6: Coefficient specification

Policy Change (MPRTE)	N_{rand}, δ'_0	N_{cor}, δ'_0	N_{rand}, δ'_1	N_{cor}, δ'_1	N_{rand}, δ''_1	N_{cor}, δ''_1	Baseline N_{cor}
$Z_2 = Z_2 + 0.01$	0.0809	0.0520	0.1183	0.0595	0.1152	0.0894	0.0917
$GS = GS + 0.01$	0.0911	0.0043	0.1209	0.0925	0.0896	0.0832	0.0958
$Z = Z + \epsilon$	0.0883	0.0574	0.1236	0.0612	0.0953	0.0774	0.0949
$P = P + \epsilon$	0.0853	0.0549	0.1144	0.0592	0.0890	0.0726	0.0882
$P = (1 + \epsilon)P$	0.0648	0.0452	0.0935	0.0483	0.0724	0.0582	0.0725
Policy Change (MPRTE)	$N_{rand}, \beta_{\bar{s}} = 0.8$	$N_{cor}, \beta_{\bar{s}} = 0.8$	N_{rand}, β'_Z	N_{cor}, β'_Z	N_{rand}, β''_Z	N_{cor}, β''_Z	Baseline N_{cor}
$X = X + 0.01$	0.0950	0.0822	0.1589	0.1147	0.0899	0.1030	0.0917
$GS = GS + 0.01$	0.0908	0.0727	0.1579	0.1163	0.1090	0.1116	0.0958
$Z = Z + \epsilon$	0.0986	0.0867	0.1606	0.1189	0.1097	0.1075	0.0949
$P = P + \epsilon$	0.0914	0.0804	0.1539	0.1128	0.1026	0.0995	0.0882
$P = (1 + \epsilon)P$	0.0731	0.0631	0.1163	0.0883	0.0842	0.078	0.0725

Effects of distributions and correlations

Turning to the distributions and correlations of the data, we note that the variables \mathbf{X} and \mathbf{Z}_2 in \mathbf{Z} are likely to have a high correlation due to the influence of homophily and conformity. As an extension of this, the unobserved network component may be correlated with \mathbf{X} and \mathbf{Z}_2 , such that the network effects are similar to the characteristics observed within a network. We see that increasing the correlation between \mathbf{X} and \mathbf{Z}_2 will reduce the size of the MTE, pulling it towards zero. An increase in this correlation will lead to a greater degree of homophily such that individuals with high \mathbf{X} and therefore \mathbf{Z}_2 are friends with others with high \mathbf{X} and \mathbf{Z}_2 . The distribution of $P(\tilde{\mathbf{Z}})$ will become much wider, so that there are less individuals in the centre, and more at the extremes of $P(\tilde{\mathbf{Z}})$. As such, the effect on the MTE builds on the effect of homophily, further compacting the MTE curve. All policy effects are diminished with this correlation. An increase in the instruments now has a smaller effect for those induced into treatment. On average, those with \mathbf{X} values that are slightly negative will be the ones transferring to the treatment group more often. Those with a high \mathbf{X} are likely to be in the treatment group already due to the correlation with \mathbf{Z}_2 .

When these variables are correlated with the unobserved network effects, the change in MTE becomes difficult to estimate. In fact, the model is not robust to a high correlation between \mathbf{X} , \mathbf{Z}_2 and α . These parameters almost perfectly predict the outcome S , making estimation difficult. When estimation is achieved we see that the policy effects are very close to zero for random networks, and meaningless for the correlated networks.

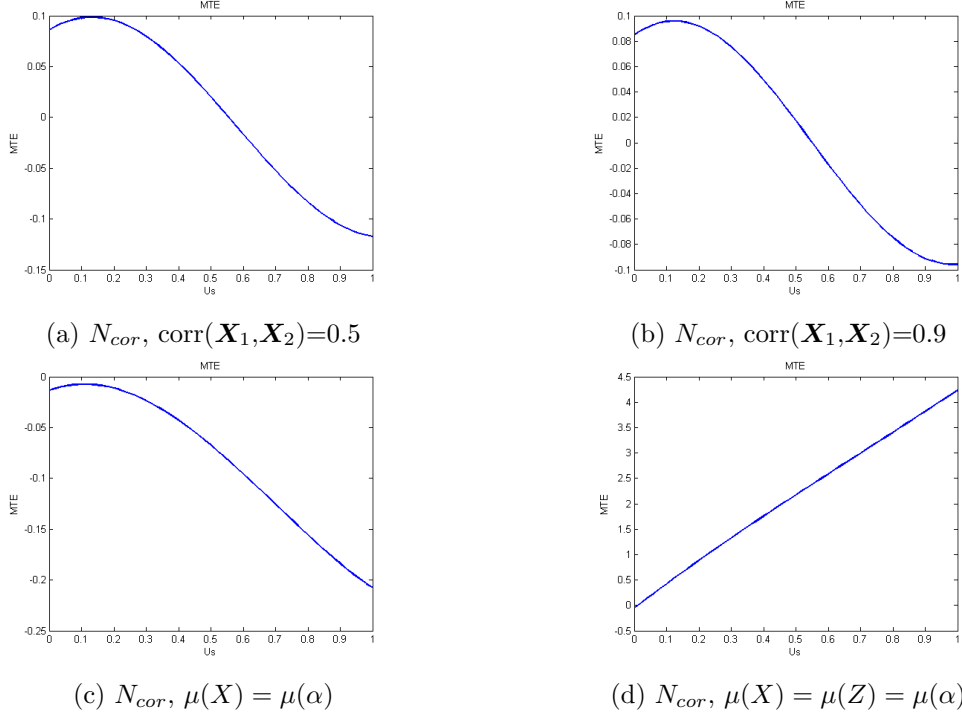


Figure 8: MTE - \mathbf{X}, \mathbf{Z} and α correlated

Table 7: \mathbf{X}, \mathbf{Z} and α correlated

Policy Change (MPRTE)	$N_{rand}, \text{corr}(\mathbf{X}_1, \mathbf{X}_2)=0.5$	$N_{cor}, \text{corr}(\mathbf{X}_1, \mathbf{X}_2)=0.5$	$N_{rand}, \text{corr}(\mathbf{X}_1, \mathbf{X}_2)=0.9$	$N_{cor}, \text{corr}(\mathbf{X}_1, \mathbf{X}_2)=0.9$	Baseline N_{cor}
$\mathbf{Z}_2 = \mathbf{Z}_2 + 0.01$	0.0869	0.0568	0.0943	0.0641	0.0917
$GS = GS + 0.01$	0.1237	0.0654	0.0936	0.0699	0.0958
$Z = Z + \epsilon$	0.1190	0.0723	0.0922	0.0672	0.0949
$P = P + \epsilon$	0.1099	0.0657	0.0856	0.0610	0.0882
$P = (1 + \epsilon)P$	0.0905	0.0545	0.0692	0.0493	0.0725
Policy Change (MPRTE)	$N_{rand}, \mu(X) = \mu(\alpha)$	$N_{cor}, \mu(X) = \mu(\alpha)$	$N_{rand}, \mu(X) = \mu(Z) = \mu(\alpha)$	$N_{cor}, \mu(X) = \mu(Z) = \mu(\alpha)$	Baseline N_{cor}
$\mathbf{Z}_2 = \mathbf{Z}_2 + 0.01$	0.0561	-0.0699	0.0218	0.962598	0.0917
$GS = GS + 0.01$	0.0094	-0.3186	0.0259	28.762337	0.0958
$Z = Z + \epsilon$	0.0425	-0.0515	0.0235	1.349076	0.0949
$P = P + \epsilon$	0.0388	-0.0490	0.0034	37.822344	0.0882
$P = (1 + \epsilon)P$	0.0543	-0.0574	0.0211	4.235802	0.0725

Correlating the error terms, the results are reported in Figure 9 and Table 8. These correlations play a pivotal role in determining the shape and results of the model. As the unobserved component U_S of the college decision determines the probability of an individual attending college, understanding the relation between ϵ , U_0 and U_1 is primordial. Altering the correlation between the error term in the treatment equation and the errors of the outcome equation (ρ_0, ρ_1), we see that when the correlation with U_0 increases, the MTE reduces and loses its shape. When the

correlation with U_1 decreases, the MTE becomes very small, retaining its downward slope. When ρ_0 and ρ_1 are similar, the change in returns to education fall away. Individuals receive a similar wage regardless of whether they go to college. Furthermore, the probability of attending college, reflected in U_S will not have differing effects on wages. The only variation in the MTE and MPRTE will occur due to the value of \mathbf{X} which plays a role in both outcome equations. This assumption does not give us the desired shape of the MTE graph. In both cases, the policy effects are accordingly greatly reduced. The model is therefore not robust to these changes in the error correlations. In both cases, making the correlations similar has adverse affects on the MTE. Due to the important role the unobserved components of the treatment equation play in determining the MTE and MPRTE, changing these correlations has a large impact. Altering the correlation between the error terms U_0 and U_1 is, however, of small concern, not having a substantial impact on our results.

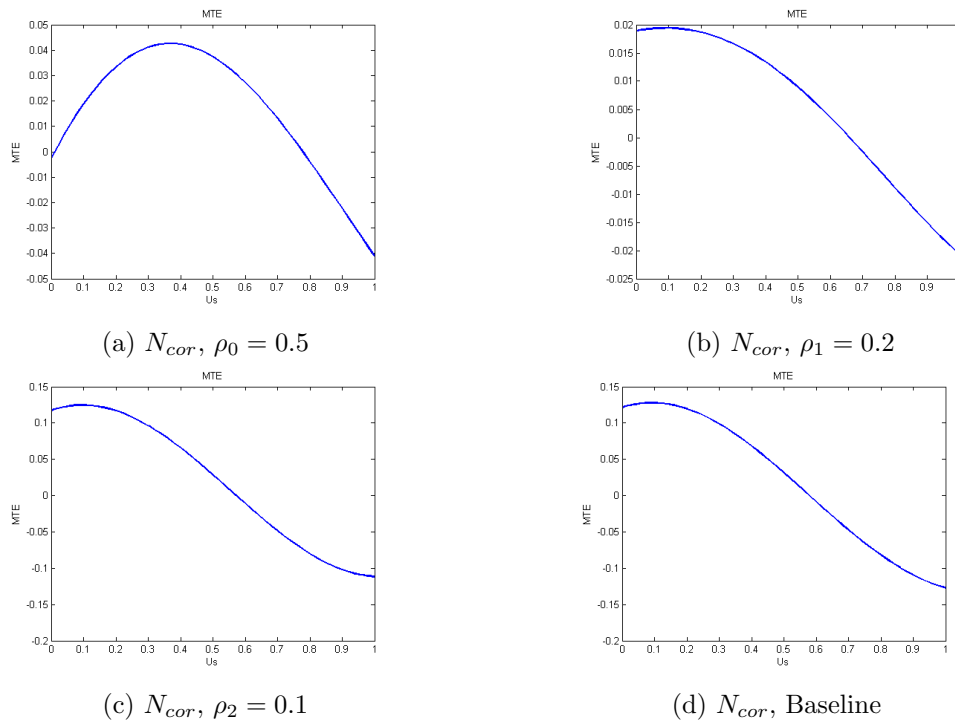


Figure 9: MTE - Correlated error terms

Table 8: Correlated error terms

Policy Change (MPRTE)	$N_{rand}, \rho_0 = 0.5$	$N_{cor}, \rho_0 = 0.5$	$N_{rand}, \rho_1 = 0.2$	$N_{cor}, \rho_1 = 0.2$	$N_{rand}, \rho_2 = 0.1$	$N_{cor}, \rho_2 = 0.1$	Baseline N_{cor}
$Z_2 = Z_2 + 0.01$	-0.0650	0.0315	0.0135	0.0118	0.0910	0.0897	0.0917
$GS = GS + 0.01$	-0.0174	0.0261	0.0249	0.0122	0.1207	0.0937	0.0958
$Z = Z + \epsilon$	0.0631	0.0282	0.0108	0.0102	0.1130	0.0915	0.0949
$P = P + \epsilon$	0.0593	0.0241	0.0089	0.0093	0.1056	0.0852	0.0882
$P = (1 + \epsilon)P$	0.0603	0.0309	0.0113	0.0079	0.0823	0.0694	0.0725

5.5 Policy Effects

Understanding the structure of peer effects allows us to now consider the important policy implications of networks. As seen, there is significant evidence of the presence of a social multiplier effect in the model setup. So far, we have only interpreted a policy change as affecting all individuals in our model. However, policy is commonly individual specific such that the policy only affects some people, or has a differing impact on individuals. To consider the effect of an individual specific policy, we set up the policy vector $pol = [p_1 \ p_2 \ \dots \ p_N]$ and consider the effect of a policy that has the effect $P = P + (pol + \mathbf{X})\beta_X + G(pol + \mathbf{X})\beta_{\bar{X}}$ (an increase in the characteristics \mathbf{X} for some individuals only). We can define a change in the number of peers receiving treatment similarly. The results are presented in Table 9. We see that the effect of a localised policy is significantly higher than a general one. Those that enter the treatment group as a result have much greater returns. The returns for those who receive treatment due to the increase in characteristics of a few friends are greater than the returns for those who receive treatment because all friends have more favourable characteristics. One factor driving this is that fewer individuals are affected and receive treatment, with these individuals being more likely to obtain greater benefits from the treatment. A policy affecting one network only has a reduced effect due to the level of homophily and is similar to the effect for all individuals.

Table 9: The effect of localised policies

Policy Change (MPRTE)	All individuals	Every 5th individual	First Network Only
$X = X + 0.01$	0.0917	0.1326	0.0917
$GS = GS + 0.01$	0.0958	0.1492	0.0873

6 Conclusion

The effects of networks on policy implications are twofold in such a treatment model. Networks present a social multiplier such that the effect of policy on one individual is spread to their peers and subsequently throughout their network. This policy may therefore induce a greater number of individuals into treatment than first targeted. Given that the number of individuals induced into treatment increases, it is natural to consider the benefits to these individuals of receiving the treatment. Networks play a key role in determining who receives treatment.

At first glance the initial addition of networks, with no assumptions on the network, has little effect on policy. However, once this initial result is broken down and the network model augmented, we find that networks do in fact contribute intrinsically to the benefits of treatment. Firstly, if the characteristics of peers are significant in defining not just treatment, but also outcomes, the effect of treatment and policy increases with the introduction of networks.

Furthering this analysis, we have found that homophily is a notable player in the interaction of networks with policy. Homophily establishes friendships defined by similar characteristics. An increase in this homophilic effect is shown to diminish the effect of policy and the role of networks in policy. If individuals relate only to those with similar characteristics, they have little influence on those with dissimilar characteristics. Those more likely to be in the treatment group will be linked to others in the treatment group, benefiting those who are already receiving treatment. However those not receiving treatment are likely to have friends with less desirable characteristics who also do not receive treatment. The outcome for these individuals from receiving treatment is significantly reduced due to the less favourable characteristics of their peers.

We have found that policy in our model becomes most effective when the probability of attending college is increased, without influencing the wage of the individual. To maximise the effect of policy, positive environmental effects should be encouraged, impacting the returns of all individuals and ignoring the negative homophily effects.

Additionally, increasing the exogenous peer effect leads to increased positive outcomes, such that encouraging positive characteristics in peers has a significant social multiplier effect.

Optimal networks, in terms of positive policy and educational returns, are networks characterised by high density and high levels of randomness. Increasing the number of friends and the randomness of the network formation both work to reduce the negative effects of homophily.

In conclusion, networks are found to generally hold a positive role in education, increasing the effect of policy and promoting the returns to treatment. Homophilic relationships diminish these initial returns, to the extent that the benefit to these networks is in fact minimal.

As demonstrated, involving the exogenous, endogenous and correlated effects in the educational treatment model must be managed carefully. The non-linear estimation of the treatment effects model is sensitive to the specification of these peer effects within the treatment and outcome equations due to the complex interactions of the components of peer effect with the estimation strategy. Nevertheless, by considering the nature of each form of peer influence, we establish the fundamental role peer effects play in the returns to education, and the flexibility of the network model when brought to the existing econometric literature.

Incorporating peer effects into a treatment effects model has numerous applications. Most obviously we can consider the effect of peers on the returns to attending college, the context around which this paper is discussed. Similar settings may be found whenever we consider the role of peers in making a binary decision. For example we could consider the effects of peers on the decision to smoke, take high school mathematics or get married, investigating the outcomes and the role of networks in each case. The treatment effects model is robust to the addition of networks and provides a vehicle through which the complex influences of peers can be explored. This paper has demonstrated this interaction, drawing out the significant impact of exogenous, endogenous and correlated peer effects on the treatment model, treatment effects and policy.

Network models such as those proposed by Bramoullé et al. (2009) and Goldsmith-Pinkham & Imbens (2013) hold important potential when considering policy and optimal outcomes for individuals. The results of this paper demonstrate that these models, when integrated with existing economic research and econometric methods, allow us to consider carefully the interrelated and complex networks which are fundamentally embedded in the way we interact. The integral role of networks and peers must therefore be fully contemplated when analysing critical decisions of individuals and evaluating key policy proposals.

Bibliography

- [1] Lucia Barbone and Peter Dolton. Key Players? Identifying High School Networking Effects on Earnings. CESinfo Area Conference on the Economics of Education, 2015.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55, 2009.
- [7] 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.

- [8] Pedro Carneiro, James J Heckman, and Edward J Vytlačil. Estimating marginal returns to education. *The American Economic Review*, 101(6):2754–2781, 2011.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.
- [13] Jason Fletcher. Social interactions and college enrollment: evidence from the national education longitudinal study. *Contemporary Economic Policy*, 31(4):762–778, 2013.
- [14] 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, 2012.
- [15] Jason M Fletcher. Similarity in peer college preferences: New evidence from Texas. *Social science research*, 41(2):321–330, 2012.
- [16] Andrea Galeotti and Gerrit Mueller. Friendship relations in the school class and adult economic attainment. IZA discussion paper, 2005.
- [17] 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.

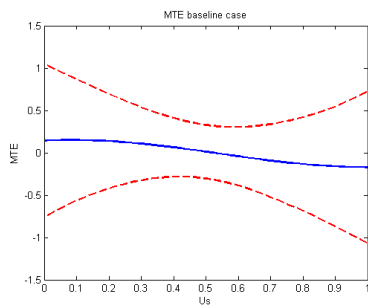
- [18] Paul Goldsmith-Pinkham and Guido W Imbens. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.
- [19] Kathleen Mullan Harris. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007–2009 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. DOI: 10.3886/ICPSR27021.v9, 2009.
- [20] James J Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005.
- [21] 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.
- [22] James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part I: Causal models, structural and econometric policy evaluation. *Handbook of econometrics*, 6:4779–4874, 2007.
- [23] 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, 2007.
- [24] Chih-Sheng Hsieh and Lung Fei Lee. A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, 31(2):301–319, 2016.
- [25] 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.

- [26] Matthew O Jackson. Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives*, 28(4):3–22, 2014.
- [27] Dean Karlan, Markus M Mobius, Tanya S Rosenblat, and Adam Szeidl. Measuring trust in Peruvian shantytowns. *Unpublished manuscript*, 2005.
- [28] 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.
- [29] Michael Kremer and Dan Levy. Peer effects and alcohol use among college students. *The Journal of Economic Perspectives*, 22(3):189–189, 2008.
- [30] 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.
- [31] Lung-fei Lee, Xiaodong Liu, and Xu Lin. Specification and estimation of social interaction models with network structures. *Econometrics Journal*, 157(1):34–52, 2010.
- [32] 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.
- [33] Charles F Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.
- [34] 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.
- [35] Robert A Moffitt et al. Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 4(45-82):6–17, 2001.

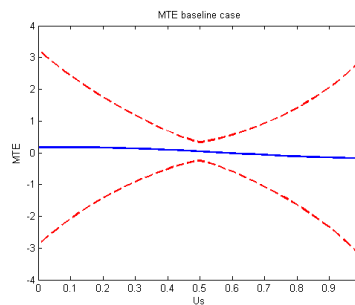
- [36] OECD. Education at a Glance 2016. /content/book/eag-2016-en, DOI: <http://dx.doi.org/10.1787/eag-2016-en>, OECD Publishing.
- [37] Eleonora Patacchini, Edoardo Rainone, and Yves Zenou. Dynamic aspects of teenage friendships and educational attainment. CEPR Discussion Paper No. DP8223, 2011.
- [38] Eleonora Patacchini and Yves Zenou. Juvenile delinquency and conformism. *Journal of Law, Economics, and Organization*, 28(1):1–31, 2012.
- [39] 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.
- [40] Richard E. Quandt. A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972.
- [41] Peter M Robinson. Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.
- [42] A. D. Roy. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers New Series*, 3(2):135–146, 1951.
- [43] Bruce Sacerdote. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Review of Economics and statistics*, 116(2):681–704, 2001.
- [44] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [45] Zebing Wu. *Adolescent friendship network and college enrollment: a longitudinal network analysis of selection and influence processes*. PhD thesis, University of Iowa, 2015.
- [46] David J Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, 85(1):9–23, 2003.

Appendix A

Additional Figures and Tables

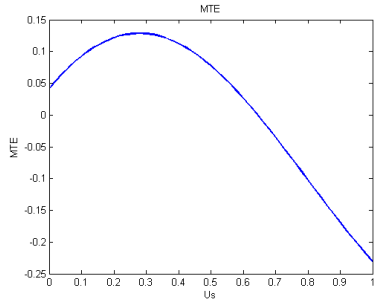


(a) MTE - no networks

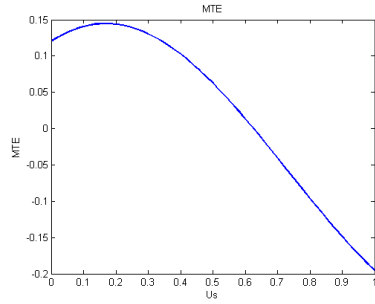


(b) MTE with networks

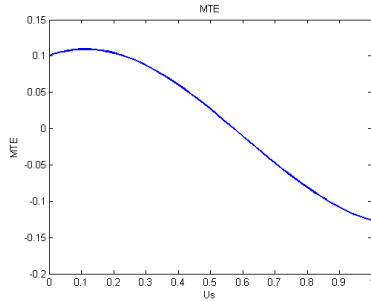
Figure A.1: MTE - Bootstrapped errors



(a) Perfect links with \mathbf{X}

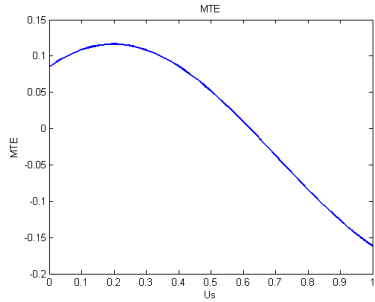


(b) Imperfect links Prob(link)=0.8

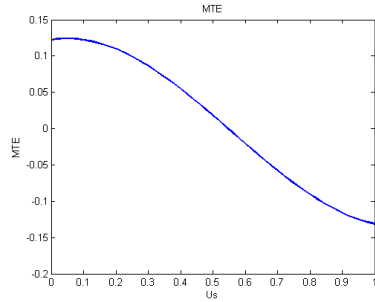


(c) Watts-Strogatz correlated with \mathbf{X}

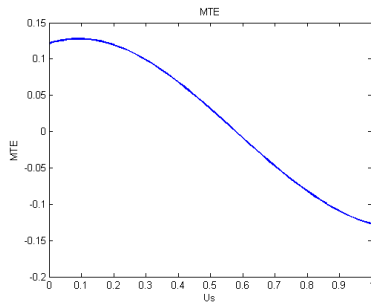
Figure A.2: MTE - Networks dependent on \mathbf{X}



(a) Correlated network with $k=1$



(b) Correlated network with $p=0.2$



(c) Watts-Strogatz $p=0.5, k=1$ (baseline)

Figure A.3: MTE - Correlated network

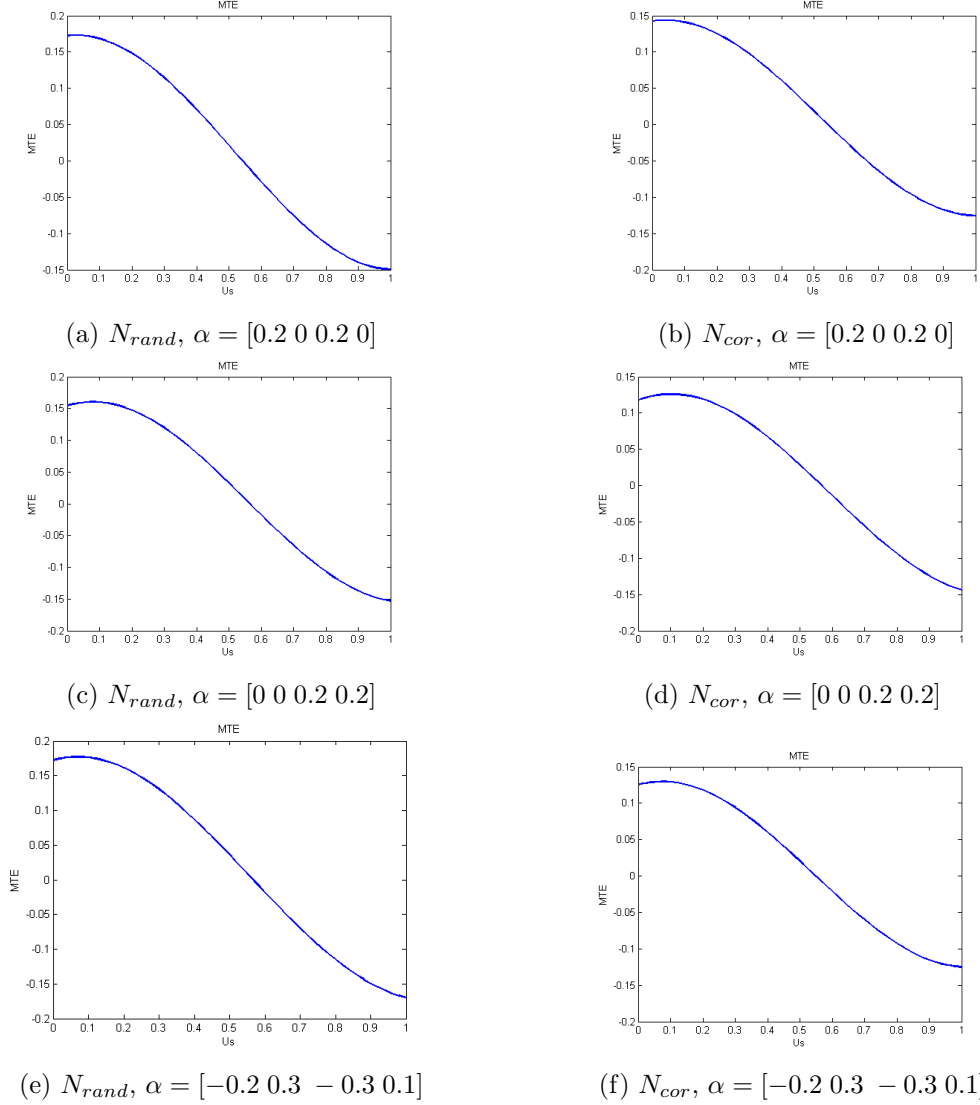
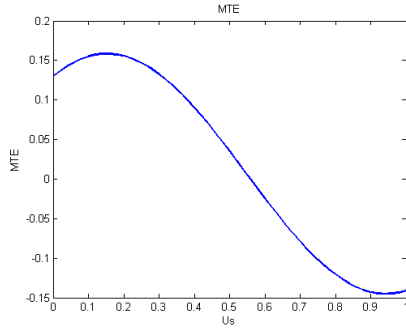


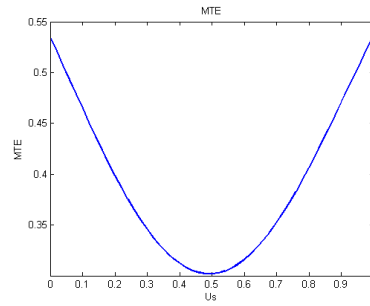
Figure A.4: MTE - Network size=[300 150 40 10]

Table A.1: Network size=[300 150 40 10]

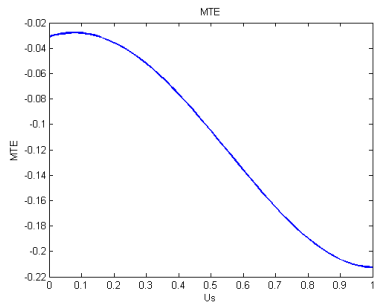
Policy Change (MPRTE)	$N_{rand}, \alpha = [0.2 \ 0 \ 0.2 \ 0]$	$N_{cor}, \alpha = [0.2 \ 0 \ 0.2 \ 0]$	$N_{rand}, \alpha = [0 \ 0 \ 0.2 \ 0.2]$
$Z_2 = Z_2 + 0.01$	0.0951	0.1077	0.1266
$GS = GS + 0.01$	0.0782	0.0981	0.1171
$Z = Z + \epsilon$	0.1090	0.0953	0.1105
$P = P + \epsilon$	0.1000	0.0888	0.1021
$P = (1 + \epsilon)P$	0.0743	0.0677	0.0808
$N_{cor}, \alpha = [0 \ 0 \ 0.2 \ 0.2]$	$N_{rand}, \alpha = [-0.2 \ 0.3 \ -0.3 \ 0.1]$	$N_{cor}, \alpha = [-0.2 \ 0.3 \ -0.3 \ 0.1]$	Baseline N_{cor}
0.076132	0.1430	0.0852	0.0917
0.084032	0.1292	0.1050	0.0958
0.093023	0.1236	0.0904	0.0949
0.084944	0.1159	0.0839	0.0882
0.070242	0.0919	0.0654	0.0725



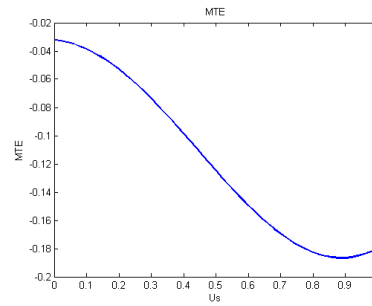
(a) N_{rand} , $G\mathbf{X}$ removed



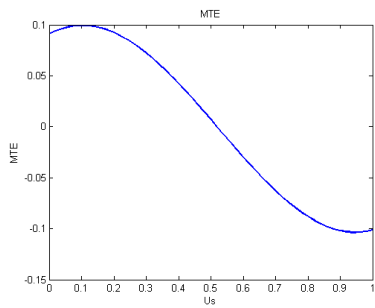
(b) N_{cor} , $G\mathbf{X}$ removed



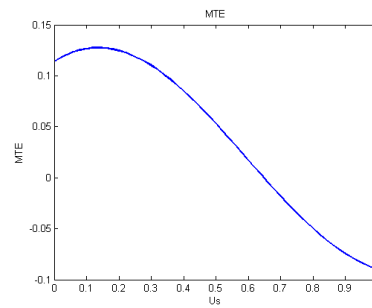
(c) N_{rand} , GS included



(d) N_{cor} , GS included



(e) N_{rand} , η_l included



(f) N_{cor} , η_l included

Figure A.5: MTE - Alternate model specifications