



EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 07/14
September 2014

**Identification and Estimation
of Outcome Response with
Heterogeneous Treatment Externalities**

by

Tiziano Arduini

(Università di Roma “La Sapienza”)

Eleonora Patacchini

(Cornell University, EIEF and CEPR)

Edoardo Rainone

(Banca d’Italia and Università di Roma “La Sapienza”)

Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities *

Tiziano Arduini[†] Eleonora Patacchini [‡] Edoardo Rainone[§]

Abstract

This paper studies the identification and estimation of treatment response with heterogeneous spillovers in a network model. We generalize the standard linear-in-means model to allow for multiple groups with between and within-group interactions. We provide a set of identification conditions of peer effects and consider a 2SLS estimation approach. Large sample properties of the proposed estimators are derived. Simulation experiments show that the estimators perform well in finite samples. The model is used to study the effectiveness of policies where peer effects are seen as a mechanism through which the treatments could propagate through the network. When interactions among groups are at work, a shock on a treated group has effects on the non-treated. Our framework allows for quantifying how much of the indirect treatment effect is due to variations in the characteristics of treated peers (treatment contextual effects) and how much is because of variations in peer outcomes (peer effects).

Keywords: Networks, Heterogeneous Peer Effects, Spatial Autoregressive Model, Two-Stage Least Squares, Efficiency, Policy Evaluation, Treatment Response, Indirect Treatment Effect.

JEL Classification Codes: C13, C21, D62

*We are grateful to Badi Baltagi, Giacomo De Giorgi, Herman K. van Dijk, Chihwa Kao, Alfonso Flores-Lagunes, Kajal Lahiri, Yoonseok Lee, Xiaodong Liu, Bin Peng, Fa Wang, Ian Wright, to participants at the Econometrics Study Group at Syracuse University and at the 2014 Camp Econometrics NY for helpful comments and stimulating discussions.

[†]Sapienza University of Rome. E-mail: tiziano.arduini@uniroma1.it.

[‡]Corresponding author. Cornell University, EIEF and CEPR. E-mail: ep454@cornell.edu

[§]Banca d'Italia and Sapienza University of Rome. E-mail: edoardo.rainone@uniroma1.it. The views expressed here do not necessarily reflect those of Banca d'Italia.

1 Introduction

The program evaluation literature focuses on estimating the program effects without externalities. There is a growing awareness, however, that there may be indirect effects that are important to measure (see Manski, 2013). Existing methodological contributions as well as studies collecting empirical evidence are still scarce. In particular, while there are a few papers about the identification and estimation of treatment response with interactions (Hudgens and Halloran, 2008, Miguel and Kremer, 2004 and Sinclair et al., 2012), to the best of our knowledge there are no studies that consider the presence of *heterogeneous interactions*.

Angelucci and De Giorgi (2009) estimates the indirect effects of the flagship Mexican welfare program, PROGRESA, on the consumption of ineligible households. This study finds that cash transfers to eligible households indirectly increase the consumption of ineligible households living in the same village. These findings are clearly very important for designing policies as well as developing experiments to evaluate them.¹ The framework, however, does not determine how much of the spillover is due to effects from eligible to ineligible subjects, effects within ineligible (eligible) subjects and feedback effects. It identifies the presence of indirect effects by comparing outcomes between untreated household in untreated villages and untreated households in treated villages. When network data are available, the analysis can be pushed forward and the heterogeneous impact of policies can be modeled and quantified.

Heterogeneity can be conceived in different ways. First, *treatment heterogeneity*, when the intensity or type of treatment can differ depending on the treated unit. Second, *treatment effect heterogeneity* when the treatment is the same for each agent but its effect is different depending on her characteristics. Third, *interaction-driven heterogeneity*, when the diffusion of the treatment effect through interactions generates an heterogeneous individual response. This may be due to both differences in interaction strengths within and between groups and to network structure, if data on connections are available. Several papers have focused on the first two types of heterogeneity.² In this paper we focus on the third kind of heterogeneity.

Using a network approach, our analysis brings three contributions to this literature. First, we derive analytically the bias that arises if spillovers are ignored. Second, we provide estimands for understanding whether different types of untreated - eligible or ineligible- are differently impacted by the treatment. Finally, our framework allows us to distinguish between different sources of treatment transmission - in particular, how much of the treatment response is generated by variations in the characteristics of treated peers(*treatment contextual effects*) and how much is due to spillovers through outcomes (*peer effects*). More specifically, our paper provides a network-based approach to estimate the average effects of the treatment in the presence of spillovers on subjects both eligible and ineligible for a program, accounting for heterogeneous within and between-group spillover effects. We show that heterogeneity in

¹More specifically, policy interventions should internalize the externalities that they engender, and experiments to evaluate their effectiveness should consider the effects on the entire local economy (e.g. the school, the village, the city), rather than focusing on differences between treatment and control group from the same local entity. When spillovers are at work, both groups' performance may change.

²See Imbens and Woolridge (2009) for a revision of recent studies using matching and non-parametric methodologies to address the second type of heterogeneity. Remarkably, Crump et al. (2008) proposes a non-parametric test for subpopulation heterogeneity in the effect of the treatment. Firpo (2007) proposed a quantile treatment estimation where the heterogeneity is given by the position of unit in the pre-treatment outcome distribution. Other papers employ more complex techniques to allow both the first and second type of heterogeneity. Among the others, generalized cross-validation statistic (Imai, Ratkovic, et al., 2013), boosting (LeBlanc and Kooperberg, 2010), Bayesian Additive Regression Trees (Chipman et al., 2010) have been used.

the effects is both helpful in terms of identification and harmful for traditional estimation methods. We develop an estimation approach able to provide reliable estimates of all the cascade effects that stem from a given network topology.

Interaction among agents can be modeled in several ways. When the exact topology of connections is known, one possibility is to consider the peer effects that stem from the given network structure. There is a large and growing literature on peer effects in economics using network data.³ The popular model employed in empirical work is the Manski-type linear-in-means model (Manski, 1993). Three assumptions underlie this statistical model: *(i)* the network is exogenous, *(ii)* the effects of all peers are equal, *(iii)* peer status is measured without error. Although these assumptions may be restrictive in empirical analyses, only a few recent papers consider alternative models and methods in which some of these assumptions are relaxed. Point *(i)* has been recently studied by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2011) who propose parametric modelling assumptions and Bayesian inferential methods to integrate a network formation model with the study of behavior over the formed networks. Point *(iii)* belongs to another strand of recent literature which looks at the consequences of peer-group misspecification, focusing in particular on sampling issues (see Chandrasekhar and Lewis, 2011, Liu, Patacchini, and Rainone, 2013 and Liu, 2013a). In this paper, we consider the specification and estimation of a peer effects model when assumption *(ii)* is removed. Lee and Liu (2010) considers a peer effects model with one endogenous variable and one adjacency matrix in a multiple network context, with no between-network interactions. Liu (2013) extends this model to the case of two endogenous variables and one adjacency matrix. In this paper, we allow the model to have two endogenous variables, two adjacency matrices, and both within and between-group interactions. We also consider the generalization to the case of multiple endogenous variables.⁴ To the best of our knowledge, we are the first to consider models of peer effects where different peers are allowed to exert a different influence and where both within and between groups interactions can be at work.⁵ We maintain assumptions *(i)* and *(iii)*.

We show that the multiple group structure of the model requires modifying the conventional identification conditions (Bramoullé et al., 2009 and Cohen-Cole et al., 2012) and has interesting connections with the concepts of *chains* and *Tree-indexed Markov chains* (see Benjamini and Peres, 1994).

We propose efficient 2SLS estimators using instruments based on the two reduced forms. We show that the standard IV approximation (Kelejian and Prucha, 1998, Kelejian and Prucha, 1999 and Liu and Lee, 2010) involves a huge number of IVs, even if we use a low degree approximation of the optimal instruments.⁶ For this reason, we consider many-instrument asymptotics (Bekker, 1994) allowing the number of IVs to increase with the sample size.

Differently from Lee and Liu (2010) and Liu (2013b) where the many instruments derive from the multiple network framework, in our model the many instruments derive from the (approximation of the) multiple adjacency-matrix framework. A multiple matrix framework does not only result in an increasing number of instruments but also yields multiple approximations

³See Jackson and Zenou, forthcoming 2013, part III, for a collection of recent studies.

⁴There is a long tradition in spatial econometrics looking at spatial autoregressive models with multiple endogenous variables (see Kelejian and Prucha, 2004). In the spatial econometrics context, however, the adjacency matrix is the same for all endogenous variables, and no groups are considered.

⁵Goldsmith-Pinkham and Imbens (2013) also estimate a model with two peer effects, but without cross effects, using a Bayesian estimation method.

⁶See Prucha (2013) for a review of Generalized Method of Moments estimators in a spatial framework.

of the optimal instruments. As a result, we show that the form of the many-instrument bias differs, though the leading order remains unchanged. We also propose a bias-correction procedure. Simulation experiments show that the bias-corrected estimator performs well in finite samples. When the number of endogenous variables is allowed to grow, our estimator remains consistent and asymptotically normal if the number of endogenous variables grows more slowly than the sample size. Finally, we investigate the bias occurring when the interaction structure is misspecified. We derive analytically the bias that occurs when only within-group peer effects are considered, i.e when interactions between groups are at work but ignored by the econometrician. We then use a simulation experiment to evaluate this bias in finite samples.

In the last part of the paper we show the empirical salience of our model for policy purposes. As highlighted by Manski (2013), the policy maker can rarely manipulate peer outcomes. Peer effects, however, can be seen as a mechanism through which the treatment could propagate through the networks. If peer effects are at work, then the policy intervention has not only a direct effect on outcomes but also an indirect one through the outcomes of connected agents (i.e. the so called "social multiplier"). We show via Monte Carlo simulations that the presence of *heterogeneous peer effects* and *between-group interactions* may create unexpected, or sometimes paradoxical results if the policy maker ignores the heterogeneity of interactions among groups. Our results can be helpful to explain why several policy programs do not accomplish the expected goals.

The paper is organized as follows. The next section introduces the econometric model. Identification conditions are derived in Section 3, and in Section 4 we consider 2SLS estimation for the model. Section 5 investigates the bias occurring when the interaction structure is misspecified. We devote Section 6 to show the importance of our analysis for the identification of treatment response with spillovers. We first derive estimands for direct, indirect and total effects of treatment strategies in network settings with interactions. Then we use a simulation experiment to show the extent to which the heterogeneity of the endogenous effects can affect the outcome response for different groups. Section 7 concludes.

2 The Network Model with Heterogeneous Peer Effects

A general network model has the specification

$$Y = \phi GY + X\beta + G^*X\gamma + \epsilon, \quad (1)$$

where $Y = (y_1, \dots, y_n)'$ is an n -dimensional vector of outcomes, $G = [g_{ij}]$ is an $n \times n$ adjacency matrix, g_{ij} is equal to 1 if i and j are connected, 0 otherwise. G^* is the row-normalized version of G , where $g_{ij}^* = g_{ij}/\sum_j g_{ij}$. X is a $n \times p$ matrix of exogenous variables capturing individual characteristics. $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ is a vector of errors whose elements are i.i.d. with zero mean and variance σ^2 for all i . For model (1), ϕ represents the *endogenous effect*, where an agent's choice/outcome may depend on those of his/her peers on the same activity, and γ represents the *contextual effect*, where an agent's choice/outcome may depend on the exogenous characteristics of his/her peers. Let $X^* = (X, G^*X)$ and $\beta^* = (\beta, \gamma)$.

Let A and B be two countable sets (types) of individuals (e.g. males and females, blacks and whites) such that $A \cap B = \emptyset$ and $n = n_a + n_b$ is the cardinality of $A \cup B$, with n_a and n_b being respectively the cardinalities of A and B . Let us define $Y = (Y'_a, Y'_b)', X = (X'_a, X'_b)',$ and $G = \begin{bmatrix} G_a & G_{ab} \\ G_{ba} & G_b \end{bmatrix}$. For instance, the subscript a denotes that $Y, X \in A$, G is formed only

among nodes of type A and the subscript ab denotes the fact that links are directed from b to a .⁷ Appendix A defines regularity conditions.

Model (1) can be written as

$$Y_a = \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + X_a^* \beta_a^* + G_{ab}^* X_b \gamma_{ab} + \epsilon_a, \quad (2)$$

$$Y_b = \phi_b G_b Y_b + \phi_{ba} G_{ba} Y_a + X_b^* \beta_b^* + G_{ba}^* X_a \gamma_{ba} + \epsilon_b, \quad (3)$$

where $\beta_a^* = (\beta_a, \gamma_a)$, $X_a^* = (X_a, G_a^* X_a)$, $X_b^* = (X_b, G_b^* X_b)$, $\beta_b^* = (\beta_b, \gamma_b)$, and ϵ_a and ϵ_b are i.i.d errors with variance σ_a^2 and σ_b^2 , respectively. Let us suppose for simplicity that $\sigma_a^2 = \sigma_b^2 = \sigma$. Model (2) - (3) is a generalization of the standard framework in the sense that it allows endogenous effects to be different within and between groups. If we stack up equations (2) - (3) and restrict the endogenous effect parameters of the two equations to be the same (i.e. $\phi_a = \phi_b = \phi_{ab} = \phi_{ba}$), then we obtain model (1).

Let us define the following matrices

$$A\delta_a = X_a^* \beta_a^* + G_{ab}^* X_b \gamma_{ab} + \epsilon_a,$$

$$B\delta_b = X_b^* \beta_b^* + G_{ba}^* X_a \gamma_{ba} + \epsilon_b,$$

where $A = (X_a, G_{ab}^* X_b, \epsilon_a)$, $\delta_a = (\beta_a^*, \gamma_{ab}, 1)$, $B = (X_b, G_{ba}^* X_a, \epsilon_b)$ and $\delta_b = (\beta_b^*, \gamma_{ba}, 1)$. By plugging Y_b in equation (2) we have

$$\begin{aligned} Y_a &= \phi_a G_a Y_a + \phi_{ab} G_{ab} (J_b(\phi_{ba} G_{ba} Y_a + B\delta_b)) + A\delta_a \\ &= (\phi_a G_a + \phi_{ab} \phi_{ba} C_a) Y_a + \phi_{ab} G_{ab} J_b B\delta_b + A\delta_a, \end{aligned} \quad (4)$$

where $J_b = (I - \phi_b G_b)^{-1} = \sum_{k=1}^{\infty} (\phi_b G_b)^k$ provided $\|\phi_b G_b\|_{\infty} < 1$, where $\|\cdot\|_{\infty}$ is the row-sum matrix norm. The ij_{th} element of J_b sums all k-distance paths from j to i when $i, j \in B$ scaling them by ϕ_b^k and $C_a = G_{ab} J_b G_{ba}$.⁸ Therefore the reduced form of model (2) is

$$Y_a = M_a (\phi_{ab} G_{ab} J_b B\delta_b + A\delta_a), \quad (5)$$

where $M_a = (I - \phi_a G_a - \phi_{ab} \phi_{ba} C_a)^{-1}$.⁹ A sufficient condition for the non singularity of $(I - \phi_a G_a - \phi_{ab} \phi_{ba} C_a)$ is $\|\phi_a G_a\|_{\infty} + \|\phi_{ab} \phi_{ba} C_a\|_{\infty} \leq 1$. This condition also implies that M_a is uniformly bounded in absolute value.¹⁰

We note that: (i) we present an aggregate model specification (i.e. G which multiplies y in model (1) is not row-normalized), but the approach applies also to an average model (i.e.

⁷More formally, $Y_a = R_a Y$, $X_a = R_a X$, $G_a = R_a G R'_a$ and $G_{ab} = R_a G R'_b$, where $R_a = (I_{n_a}, O_{n_a, n_b})$ and $R_b = (O_{n_b, n_a}, I_{n_b})$ are matrices that select the nodes in group a and b respectively. $O_{k,l}$ is a $k \times l$ matrix of zeros.

⁸ C_a is a matrix which captures all the indirect connections among nodes of type A passing through one or more nodes of type B . Note that the ij_{th} generic element of $G_{ab} G_{ba}$ is equal to the number of length-2 paths directed from $j \in A$ to $i \in A$ passing through a node $l \in B$. This matrix accounts only for distance-2 indirect connections while $C_a = G_{ab} J_b G_{ba}$ captures all the paths starting from $j \in A$ and ending to a generic node in B , eventually passing through other nodes of type B and finally arriving in $i \in A$ scaling them by ϕ_b .

⁹This matrix captures all direct and indirect paths among type A nodes passing through others type A nodes and type B nodes.

¹⁰The assumption is crucial for identification of the model and asymptotic normality of the estimator (see Appendix A).

when G which multiplies y in model (1) is row-normalized);¹¹ (ii) our model specification has two groups, but all the assumptions, propositions and proofs can be naturally extended to a finite number of groups; (iii) we consider a single network, but the approach can be extended to the case of multiple networks(i.e. a network with several components) with the addition of network fixed effects in the model specification; (iv) we can also add a heterogeneous spatial lag in the error term $\epsilon_a = \rho_a W_a \epsilon_a + \rho_{ab} W_{ab} \epsilon_b$.¹²

3 Identification

Let us define $Z_a = (G_a Y_a, G_{ab} Y_b, X_a^*, G_{ab}^* X_b)$. Equation (2) is identified if $E(Z_a)$ has full column rank for large n.¹³ In this section, we find sufficient conditions for $E(Z_a)$ to have full column rank.¹⁴ The detailed proof is given in Appendix C.

Proposition 1. *Let X_a and X_b have full column rank. If the sequences of $\{M_a\}$, $\{M_b\}$, $\{J_a\}$ and $\{J_b\}$ are UB matrices,¹⁵ then $E(Z_a)$ has full column rank in the following cases*

- 1. [(I)]
- 2. (a) i. $\beta_a \phi_a + \gamma_a \neq 0$,
ii. I_a , G_a and G_a^2 are linearly independent.
[and]
- (b) i. $\beta_b \phi_b + \gamma_b \neq 0$,
ii. G_{ab} and $G_{ab} G_b$ are linearly independent.
[or]
- 3. (a) i. $\gamma_{ab} \neq 0$,
ii. G_{ab} and $G_a G_{ab}$ are linearly independent.
[and]
- (b) i. $\gamma_{ba} \neq 0$,
ii. I_a , G_a and $G_{ab} G_{ba}$ are linearly independent.

Note that conditions (2a) are exactly the same identification conditions found by Bramoullé et al. (2009) in the case of homogeneous effects (i.e. only one group). Proposition 1 here is more general as it provides alternative possibilities. When more than one group is considered we do not need linear independence of a particular set of matrices - we have multiple sufficient conditions. Even if I_a , G_a and G_a^2 are linearly dependent we can still identify ϕ_a , and the other

¹¹Aggregate and average models are different in terms of behavioral foundations, contextual effects are supposed to be averages over peers in both cases w.l.o.g. (see Liu, Patachini, and Zenou, 2014forthcoming).

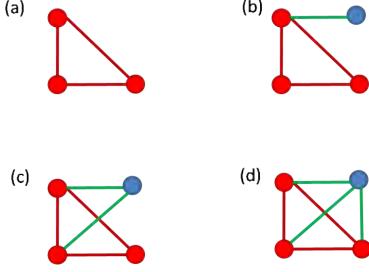
¹²The resulting model is a *SARARMAG*($p; q; g$) with $p = 1$, $q = 1$ and $g = 2$, where p and q are respectively the number of spatial lags for outcome and error, and g is the number of groups (see Kelejian and Prucha, 2007).

¹³This implies that Assumption 4 in Appendix A holds.

¹⁴Symmetric conditions and results hold for equation (3).

¹⁵In practice we need a series expansion to approximate the inversion of the matrices. We are grateful to Chihwa Kao for pointing it out.

Figure 1: Identification with heterogeneous nodes



parameters, relying on linear independence of chains passing through type B nodes.¹⁶ The set of adjacency matrices' combinations can be represented as a *Tree-indexed Markov chain*- the parameters can be identified because of the multiple branches of the tree (see Appendix B). Obviously, if G_a , G_{ba} , G_{ab} and G_b are complete and consequently all products among them are linearly dependent, then the model remains not identified. However, if group A nodes are in a complete network, but the matrices representing between-group interactions are sparse (i.e G_{ab} and G_{ba} are not complete), then identification can be achieved and ϕ_a can be estimated even if G_a is complete. Systems in panels (b) and (c) in Figure 1 can be identified because the adjacency matrix of type B nodes (blue nodes in Figure 1) is sparse, whereas systems in panel (a) and (d) cannot. The additional parameters' restrictions (condition (2b, 3a or 3b)) are due to an additional vector in the full rank condition (i.e. $E(G_{abyb})$).

Proposition 1 has a natural interpretation in terms of instrumental variables. A multiple group framework adds an extra layer of exclusion restrictions. In fact, multiple sets of matrices provide additional instruments. The intuition is that when we distinguish nodes in different types, a higher number of possible network intransitivities are formed. Appendix B provides technical details on the connection between identification in a single group model and a multiple group one.

4 The 2SLS estimator

Equation (2) cannot be consistently estimated by OLS because G_ay_a and G_{abyb} are correlated with ϵ_a .¹⁷ We consider 2SLS estimation for the model in the spirit of Lee-Liu (2010). Following the standard technique used in spatial econometrics literature, we have the following optimal instruments from the two (symmetric) reduced forms

$$E(G_ay_a) = G_a(M_a(\phi_{ab}G_{ab}J_bE(B)\delta_b + E(A)\delta_a)), \quad (6)$$

$$E(G_{abyb}) = G_{ab}(M_b(\phi_{ba}G_{ba}J_aE(A)\delta_b + E(B)\delta_a)). \quad (7)$$

¹⁶For example, we can take advantage of linear independence of I_a , G_a and $G_{ab}G_{ba}$ (instead of I_a , G_a and G_a^2); and G_{ab} and G_aG_{ba} .

¹⁷From equation (5), $G_ay_a = S_a(\phi_{ab}G_{ab}J_bB\delta_b + A\delta_a)$ where $S_a = G_aM_a$. OLS is not consistent because we have $E((G_ay_a)', \epsilon_a) = E((S_a(\phi_{ab}G_{ab}J_bB\delta_b + A\delta_a))', \epsilon_a) = E((S_a\epsilon_a)', \epsilon_a)$, since we assume that the $\text{cov}(\epsilon_a, \epsilon_b) = 0$ and $E(\epsilon_a) = E(\epsilon_b) = 0$. It follows that $E((S_a\epsilon_a)', \epsilon_a) = \sigma_a^2 \text{Tr}(S_a) \neq 0$. A similar argument holds for G_{abyb} .

Recalling that $Z_a = [G_a y_a, G_{ab} y_b, E(A)]$ is a $n \times (k + 2)$ matrix, we have $f_a = E(Z_a) = [E(G_a y_a), E(G_{ab} y_b), E(A)]$, so from equations (6) - (7) we have

$$Z_a = f_a + v_a = f_a + [(\phi_{ab} S_a G_{ab} J_b \epsilon_b + S_a \epsilon_a), (\phi_{ba} S_{ab} G_{ba} J_a \epsilon_a + S_{ab} \epsilon_b)][e_1, e_2]',$$

where e_1 is a first unit vector of dimension $(k + 2)$, $S_a = G_a M_a$ and $S_{ab} = G_{ab} M_b$. These instruments are infeasible given the embedded unknown parameters. f_a can be considered a linear combination of IVs in $H_\infty^* = (S_a(G_{ab} J_b E(B), E(A)), S_{ab}(G_{ba} J_a E(A), E(B)), E(A))$. Furthermore, since $S_a = G_a M_a$ and $S_{ab} = G_{ab} M_b$ provided $\|\phi_a G_a\|_\infty + \|\phi_{ab} \phi_{ba} C_a\|_\infty \leq 1$ and $\|\phi_b G_b\|_\infty < 1$, we have $S_a = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^j = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab} \phi_{ba} G_{ab} \sum_{j=0}^{\infty} (\phi_b^j G_b^j) G_{ba})^j$.

The same approximation holds for S_{ab} . It follows that

$$C_a = G_{ab} J_b G_{ba} = G_{ab} \left(\sum_{j=0}^{\infty} \phi_b^j G_b^j \right) G_{ba} = G_{ab} \left(\sum_{j=0}^p \phi_b^j G_b^j + (\phi_b G_b)^{p+1} J_b \right) G_{ba}.$$

This implies $\|C_a - \sum_{j=0}^p \phi_b^j G_b^j\|_\infty \leq \|(\phi_b G_b)^{p+1}\|_\infty \|C_a\|_\infty = o(1)$ as $p \rightarrow \infty$.

$$S_a = G_a M_a = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^j = G_a \left[\sum_{j=0}^p (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^j + (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^{p+1} S_a \right] \rightarrow \|\sum_{j=0}^p (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^j\|_\infty \leq \|(\phi_a G_a + \phi_{ab} \phi_{ba} o(1))^{p+1}\|_\infty \|S_a\|_\infty = o(1)$$
 as $p \rightarrow \infty$. Hence, the approximation error by series expansion diminishes very quickly in a geometric rate, as long as the degree of approximation (p) increases as n increases. We can also replace S_a and S_{ab} by a linear combination. The instruments become

$$H_\infty^a = (G_a(I, G_a, G_a^2, \dots (G_{ab}(I, G_b, G_b^2, \dots) G_{ba})) \dots (G_{ab}(I, G_b, G_b^2, \dots) E(B), E(A)),$$

$$H_\infty^{ab} = (G_{ab}(I, G_b, G_b^2, \dots (G_{ba}(I, G_a, G_a^2, \dots) G_{ab})) \dots (G_{ba}(I, G_a, G_a^2, \dots) E(A), E(B)),$$

with an approximation error diminishing very quickly when K (or p) goes to infinity, where K denotes the number of instruments. Let us define $H_\infty = [H_\infty^a, H_\infty^{ab}, X_a^*, G_{ab} X_b]$ as the matrix of instruments and select an $n_a \times K$ submatrix H_K based on a p -order approximation of H_∞ .¹⁸ For instance, if we use the second order approximation of the infinite sums, $H_K = (H_2^a, H_2^{ab}, X_a^*, G_{ab} X_b)$ will be the first step best projector. The feasible 2SLS estimator for model (2) is

$$\hat{\mu} = (Z'_a \hat{P}_K Z_a)^{-1} Z'_a \hat{P}_K Y_a, \quad (8)$$

where $\hat{\mu} = (\phi_a, \phi_{ab}, \beta_a^*, \gamma_{ab})$ and $\hat{P}_K = H_K (H'_K H_K)^{-1} H'_K$.

4.1 Asymptotic Properties

This section derives the asymptotic properties of the many-instrument 2SLS estimator for heterogeneous network models. Cohen-Cole et al. (2012) and Liu (2013b) consider a network model with two endogenous variables and one adjacency matrix with multiple networks.¹⁹ Our

¹⁸Note that K is a function of the degree of approximations p .

¹⁹Kelejian and Prucha (2004) considers SAR models with multiple endogenous variables and a unique weights matrix.

network model requires two endogenous variables, and two different adjacency matrices.²⁰ In Lee and Liu (2010) and Liu (2013b), the asymptotic approximation of the 2SLS estimator is based on many-instrument asymptotics, where the many instruments derive from the multiple network framework. In our model the many instruments derive from the (approximation of the) multiple adjacency-matrix framework. A multiple matrix framework results in an increasing number of instruments due to multiple approximations of the optimal instruments.²¹ This complicates the derivations of the asymptotic properties of the many-instrument 2SLS estimator.

The following propositions establish the consistency and asymptotic normality of the many-instrument 2SLS estimator in equation (8). Regularity conditions together with some discussion can be found in Appendix A. Some useful Lemmas are provided in Appendix B. All the proofs are listed in Appendix C. Let

$$F_a = \lim_{n \rightarrow \infty} \frac{1}{n} f'_a f_a, \quad ^{22}$$

$P_K S_a = \Psi_a$ and $P_K T_{ba} = \Xi_{ba}$, where $T_{ab} = S_{ab} G_{ba} J_b$.²³

Proposition 2. *Under assumptions 1-5, if $K/n \rightarrow 0$, then $\sqrt{n}(\hat{\mu} - \mu_0 - b) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$, where $b = (Z'_a P_K Z_a)^{-1} [e_1, e_2] \sigma_a^2 [tr(\Psi_a), \phi_{ba} tr(\Xi_{ba})]' = O_p(K/n)$.*

From Proposition 2, when the number of instruments K grows at a slower rate than the sample size n , the 2SLS estimator is consistent and asymptotically normal. However, the asymptotic distribution of the 2SLS estimator may not be centered around the true parameter value due to the presence of many-instrument bias of order $O_p(K/n)$ (see, e.g., Lee and Liu, 2010). We note that the leading order of the bias is the same as in Lee and Liu (2010) and Liu (2013b). However, the structure of the bias differs. Here, it depends on multiple approximations of the optimal instruments (see the beginning of Section 4). The condition that $K/n \rightarrow 0$ is crucial for the 2SLS estimator to be consistent. This appears evident if we look at the normal equation of our estimator: $\frac{1}{n} Z'_a P_K (Y_a - Z_a \hat{\mu})$. When $\hat{\mu} = \mu_0$ we have that $E(\frac{1}{n} Z'_a P_K (Y_a - Z_a \mu_0)) = [e_1, e_1] \sigma_a^2 [tr(\Psi_a), \phi_{ba} tr(\Xi_{ba})]' = O_p(K/n)$ by Lemma B.2 in the Appendix. This converges to 0 only if the number of instruments grows more slowly than the sample size.²⁴ The following corollary characterizes different scenarios for different rates in which K diverges from n .

Corollary 1. *Under assumptions 1-5, (i) if $K^2/n \rightarrow 0$, $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$; (ii) if $K^2/n \rightarrow c < \infty$, $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(\bar{b}, \sigma_a^2 F_a^{-1})$, where $\bar{b} = \lim_{n \rightarrow \infty} \sqrt{n}b$.*

²⁰We consider the analysis with one network only. The extension to multiple networks extremely complicates the notation burden, but the theoretical results remain basically unchanged.

²¹See Section 4.

²²This is a crucial assumption. See the discussion in Appendix A after Assumption 4.

²³To simplify the notation, we assume that $n \rightarrow \infty$ implies $n_a \rightarrow \infty$ and $n_b \rightarrow \infty$.

²⁴Indeed, if we use a fixed number of instruments given by \bar{H} , the asymptotic distribution will be $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 (\lim_{n \rightarrow \infty} \frac{1}{n} f'_a \bar{P} f_a)^{-1})$. Note that $(F_a - \lim_{n \rightarrow \infty} \frac{1}{n} f'_a \bar{P} f_a) = \lim_{n \rightarrow \infty} f'_a (I - \bar{P}) f_a$, which is positive semi-definite in general. The 2SLS estimator with fixed number of instrument is generally not efficient. In order to have efficiency, we need to index our matrix of instruments with K and let K grow more slowly than the sample size.

The many-instrument bias of the 2SLS estimator can be corrected by the estimated leading-order bias (b) given in Proposition 2. Given consistent estimates of $\hat{\phi}_a$, $\hat{\phi}_b$, $\hat{\phi}_{ab}$, $\hat{\phi}_{ba}$, $\hat{\sigma}_a$ and $\hat{\sigma}_b$, the bias-corrected 2SLS estimator is

$$\hat{\mu}_c = (Z'_a P_K Z_a)^{-1} [Z'_a P_K Y_a - \hat{\sigma}_a^2 [e_1, e_2] [tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]']. \quad (9)$$

The following proposition shows that the bias-corrected estimator is properly centered around the normal distribution.

Proposition 3. *Under assumptions 1-5, if $K/n \rightarrow 0$ and $\hat{\phi}_a$, $\hat{\phi}_b$, $\hat{\phi}_{ab}$, $\hat{\phi}_{ba}$, $\hat{\sigma}_a$ and $\hat{\sigma}_b$ are \sqrt{n} -consistent initial estimators, then $\sqrt{n}(\hat{\mu}_c - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$.*

In the next subsection we discuss the case in which the number of endogenous variables (groups) grows with the sample size.

4.2 Estimation with Many Groups

So far, we have assumed that group numerosity does not depend on the sample size. We believe that, in practice, such an assumption is virtually always satisfied. For instance, if we increase the size of the sample, we will always have two genders: male and female. However, for completeness, it is interesting to explore whether having the number of groups growing together with the sample size affects the estimator properties.

In the many-instrument literature, Anatolyev (2013) and Imbens, Kolesar, et al. (2011) have relaxed the assumption of a fixed number of *exogenous regressors*. To the best of our knowledge, the implications of relaxing the assumption of a fixed number of *endogenous regressors* have not been investigated yet.

Let us define g as the number of endogeneous variables and p as the degree of approximation (see Appendix C for an intuition of p as length of chains).

The following proposition characterizes the rate of divergence of g from n .

Proposition 4. *if $K/n \rightarrow 0$, we have that $g = o(n^{1/p})$.*

This means, that for our estimator to be consistent and asymptotically normal in this framework with many instruments and many endogenous variables we need g to grow more slowly than $n^{1/p}$.

For completeness, let us consider the link between the number of groups (i.e. endogenous variables) and the many-instrument asymptotics.

In our framework we have that $g/K \rightarrow 0$. In order to have a good performance of the estimator we need $K/n \rightarrow 0$. This implies $g/n = 1/s_g \rightarrow 0$, where s_g is the average size of groups. In words, in order to have a good performance of the estimators, we need the size of groups to be large enough. Furthermore, in order to have the estimator properly centered, we need $K^2/n \rightarrow 0$. This implies $g^2/n = g/s_g \rightarrow 0$. Therefore, for asymptotic efficiency, the average size of groups needs to be large enough compared to the number of groups. These results are similar to those in Lee and Liu (2010). However, the framework in Lee and Liu (2010) considers multiple networks embedded in a block-diagonal adjacency matrix (i.e. $G = diag(G_a, G_b)$) with the restriction that the within peer effects are the same for each network, (i.e. $\phi_a = \phi_b$) and there are no interactions between networks. If a network is defined as a group, then our framework can be considered as a generalization. We have different groups, with both within and between-group interactions. Our adjacency matrix is thus not block-diagonal.

4.3 Finite sample performance

In this section, we use simulation experiments to investigate the performance of the proposed estimator in small samples.

We conduct a Monte Carlo simulation study based on the following model

$$y_a = \phi_a G_a y_a + \phi_{ab} G_{ab} y_b + X_a^* \beta_a + G_{ab}^* X_b \gamma_{ab} + \epsilon_a,$$

$$y_b = \phi_b G_b y_b + \phi_{ba} G_{ba} y_a + X_b^* \beta_b + G_{ba}^* X_a \gamma_{ba} + \epsilon_b,$$

where X_a , X_b and $\epsilon \sim N(0, 1)$. Borrowing from Lee and Liu (2010), we generate the G matrix as follows. First, for the i_{th} row of G , we generate an integer $d_i \in [0, 1, \dots, m]$ with a uniform probability function, where $m = 10, 20, 30$. Then we set the $(i+1)_{th}, \dots, (i+d_i)_{th}$ elements of the i_{th} row of G to be ones. If $(i+d_i)_{th} < n_a$, the other elements in that row are zeros; otherwise, the entries of ones will be wrapped around such that the number of $d_i - n_a$ entries of the i_{th} row will be ones. We partition the matrix into four submatrices G_a , G_b , G_{ab} and G_{ba} with a random selection of rows and correspondent columns. The identifier variable used to select the two groups is generated by a Bernoulli distribution with $p=0.5$. The number of replications is 1000 and $n_a = n_b = 500$. We perform two experiments that are summarized in Table 1 and Table 2. Each column reports mean and standard error (in parenthesis) of the empirical distributions of different estimators. The first column shows 2SLS few IVs. It is based on equation (8) with the IV matrix H_K derived by the first order approximation of the best instruments ($K=24$). The second column reports the 2SLS many IVs, it is derived by the second order approximation of the best instruments ($K=84$). Finally, Column 3 shows the 2SLS bias-corrected. It is based on equation (9) with consistent estimates derived from the 2SLS few IVs.

Table 1 reports on the performance of the estimators when changing the density of the network, i.e. the number of connections. Each panel represents a different value of m , which indicates the maximum number of connections. The data are generated with $\beta_a = \beta_b = \gamma_a = \gamma_b = \gamma_{ab} = \gamma_{ba} = 0.5$. The peer effects parameters are set to: $\phi_a = \phi_b = 0.1$ and $\phi_{ab} = \phi_{ba} = 0.2$. The results show that all estimators perform well, with different nuances. In particular, one can observe the trade-off between bias and efficiency for the 2SLS many IVs when network density increases- the higher the density, the higher the gain in terms of efficiency with respect to the 2SLS few IVs. However, the bias (due to the many instruments) increases as well. The bias correction that we propose is thus particularly beneficial when the network is dense.

[INSERT TABLE 1 HERE]

Table 2 reports on the performance of the estimators when changing the heterogeneity within and between-group parameters. The simulation setup remains unchanged, but we now set the maximum number of connections to 20 and let the ϕ parameters vary. In the first panel, we consider $\phi_a = \phi_{ab} = \phi_b = \phi_{ba} = 0.1$ This is the benchmark framework in which peer effects are homogeneous. In the second panel, we introduce some heterogeneity in the within-group interaction effects. We set $\phi_a = \phi_b = 0.1$ and $\phi_{ab} = \phi_{ba} = 0.3$. In the third panel, peer effects are different both within and between groups. We set $\phi_a = 0.1$, $\phi_b = 0.2$ $\phi_{ab} = 0.4$ and $\phi_{ba} = 0.05$. Table 2 shows that the performance of the estimators does not depend on the values of the parameters- the ranking of the estimators in terms of efficiency and bias remains unchanged.

[INSERT TABLE 2 HERE]

To test the robustness of our results, we have also performed two additional exercises.²⁵ First, instead of using randomly generated networks, we have used the Add Health's sociomatix²⁶ as an adjacency matrix, thus replicating features of real-world social networks. Our aim is to understand whether the results of Table 1 are driven by the random generation of links. Second, we use uniform and gamma distributions to generate the errors of the data generating process. In doing so, our aim is to investigate whether and to what extent our i.i.d. assumption for the error terms in the derivation of large sample properties affects the finite sample Monte Carlo results. In both cases, the simulation results are very similar to those reported here.

5 Model Misspecification Bias

In this section, we investigate the bias occurring when the interaction structure is misspecified.

First, we analytically derive the bias that occurs when only within-group peer effects are considered, i.e. when interactions between groups are at work but ignored by the econometrician. We then use a simulation experiment to evaluate this bias in finite samples.

Second, we derive the mapping between the parameters of a model with homogeneous peer effects and those of a model with heterogeneous peer effects. We then use a simulation experiment to give an example of parameter mapping when peer effects are believed to be homogeneous but are actually heterogeneous in the data generating process (DGP).

Let us suppose the econometrician estimates the following model

$$y_a = (I - \phi_a G_a)^{-1} (X_a \beta_a + G_a^* X_a \gamma_a + \epsilon), \quad (10)$$

whereas the real DGP is

$$y_a = G_a (M_a (\phi_{ab} G_{ab} J_b B \delta_b + A \delta_a)), \quad (11)$$

$$y_b = G_{ab} (M_b (\phi_{ba} G_{ba} J_a A \delta_b + B \delta_a)). \quad (12)$$

This model misspecification results in an estimator of the endogenous effect ϕ_a that is inconsistent. First, we are omitting the influence of the outcome of type B agents. Second, we do not consider the indirect connections among type A nodes passing through type B nodes. As a result, G_a^k , with $k \geq 2$, is misspecified. Therefore, the commonly used instrument G_a^2

²⁵Results available upon request.

²⁶A matrix derived from observed connections among students in the Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwistle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

might not be valid as the exclusion restrictions might be violated. Third, we misspecify the contextual effects ($G_a^*X_a$) by ignoring the characteristics of other-type peers.²⁷

Analytically, the bias is

$$E(\hat{\mu}_a) = \mu_a + E(Z'_a P_a Z_a)^{-1} Z_a P_a (\phi_{ab} G_{ab} y_b + G_{ab}^* X_b \gamma_{ab}),$$

where $Z_a = [G_a y_a, X_a, G_a^* X_a]$ and $\mu_a = (\phi_a, \beta_a, \gamma_a)$. The bias is positively correlated with the direct influence of type B on type A, as captured by the peer effects from B to A and the influence of the characteristics of type B on type A.

Table 3 shows the extent of this bias in finite samples through a Monte Carlo simulation. Table 3 represents the performance of the 2SLS few IVs following the same experiment design as in the previous section.²⁸ We report on the case in which the maximum number of connections is 10 for each node (as in panel 2 of table 1).²⁹ The first column reports the real value of the parameters. The second column shows the performance of the 2SLS estimator in the misspecified model. When interactions between groups are at work but ignored by the econometrician it results in the size of the bias derived above. The third column shows the results of the estimator when the econometrician considers the correct DGP (equations (11) and (12)), but does not use the approximation of optimal instruments (in equation (8)). In other words, we consider the case where the traditional network IV approach is applied mechanically, thus $G_a^2 X_a$ and $G_{ab} G_b X_b$ are used as instruments respectively for $G_a Y_a$ and $G_{ab} Y_b$. In short, only within-group instruments are considered. The resulting 2SLS estimator is consistent but not efficient. The fourth column reports the performance of our 2SLS few IVs (in equation (8)), which considers the H_k matrix derived in Section 4(i.e which also includes between group instruments).³⁰ Mean values for each coefficient's empirical distribution and standard errors (in parenthesis) are reported.

Table 3 shows that the bias is large in the second column, especially for the β coefficients. In the second column the bias is not large, but the problem is efficiency. Our approach (third column) reveals no bias and improved efficiency.

[INSERT TABLE 3 HERE]

In our second exercise, we consider the case in which the econometrician estimates a standard network model (model (1)) when the real DGP is characterized by heterogeneous peer effects (model (11) - (12)).

Let us define the following $n \times n$ matrices

$$\begin{aligned} G(a) &= \begin{bmatrix} G_a & O_{ab} \\ O_{ba} & O_b \end{bmatrix}, & G(ab) &= \begin{bmatrix} O_a & G_{ab} \\ O_{ba} & O_b \end{bmatrix}, \\ G(ba) &= \begin{bmatrix} O_a & O_{ab} \\ G_{ba} & O_b \end{bmatrix}, & G(b) &= \begin{bmatrix} O_a & O_{ab} \\ O_{ba} & G_b \end{bmatrix}, \end{aligned}$$

²⁷This issue also arises when full information about node characteristics and network structure is not available. See Chandrasekhar and Lewis, 2011, Liu, Patacchini, and Rainone, 2013 and Liu, 2013a for problems related to the use of sampled network data.

²⁸We use the 2SLS few IVs to ease the comparison of 2SLS estimators with the misspecified set of instruments. Observe that the bias considered here is due to the misspecification of the model rather than to the many-instrument issue.

²⁹The simulation results in the other cases, i.e. when the maximum number of simulations is 20 or 30, are very similar.

³⁰First order approximation of optimal instruments is considered.

where O_l is a $l \times l$ matrix of zeros and O_{lk} is a $l \times k$ matrix of zeros. Let us suppose for simplicity that $\beta = \beta_a = \beta_b$ and $\gamma = \gamma_a = \gamma_{ab} = \gamma_{ba} = \gamma_b$ and focus our attention on the peer effects parameters. In this case model (1) can be written as

$$\begin{aligned} Y &= \phi GY + X\beta + G^*X\gamma + \epsilon \\ &= (\phi_a G(a) + \phi_{ab}G(ab) + \phi_{ba}G(ba) + \phi_b G(b))Y + X\beta + G^*X\gamma + \epsilon. \end{aligned} \quad (13)$$

Hence, the peer effects parameter, ϕ , is the following non-linear function of heterogeneous peer effects

$$\phi = \phi_a G^{-1}G(a) + \phi_{ab}G^{-1}G(ab) + \phi_{ba}G^{-1}G(ba) + \phi_b G^{-1}G(b).$$

If $\phi_a = \phi_b = \phi_{ab} = \phi_{ba} = \phi$, then

$$\phi_a G(a) + \phi_{ab}G(ab) + \phi_{ba}G(ba) + \phi_b G(b) = \phi(G(a) + G(ab) + G(ba) + G(b)) = \phi G.$$

Table 4 contains the results of a simulation experiment in which we estimate model (13), for different values of ϕ_a , ϕ_b , ϕ_{ab} and ϕ_{ba} . The simulation set-up is as before- the data generating process remains as in equations (11) and (12)). The estimator considered is the 2SLS few IVs.

In the first column, we set all the ϕ parameters equal to 0.1. In fact, the 2SLS few IVs consistently estimates $\phi = \phi_a = \phi_b = \phi_{ab} = \phi_{ba}$. In the second column, we add some heterogeneity. We set $\phi_{ab} = 0.3$ and $\phi_{ba} = 0.3$, leaving the other parameters unchanged. The third column corresponds to the case in which all the ϕ parameters are different. As expected, as soon as some heterogeneity is introduced, the estimated value of ϕ is not informative at all.

[INSERT TABLE 4 HERE]

6 Impact Evaluation and Treatment Effect

Let us now highlight the importance of our analysis for the identification of treatment response with spillovers. Let A be the set of *eligible* recipients and B the set of *ineligible* recipients of a treatment (respectively *eligibles* and *ineligibles* hereafter). The treatment is administrated using a *randomized controlled experiment*. Having in mind policy interventions such as conditional cash transfer or microfinance subsidies can be useful. Let T_a be the binary treatment vector whose i^{th} element is $T_{a,i} = \{0, 1\}$, which indicates whether i is treated or not (among the *eligibles*).³¹ Model (2) and (3) can be written as

$$Y_a = \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + X_a^* \beta_a^* + \delta_a T_a + \rho_a G_a T_a + G_{ab}^* X_b \gamma_{ab} + \epsilon_a, \quad (14)$$

$$Y_b = \phi_b G_b Y_b + \phi_{ba} G_{ba} Y_a + X_b^* \beta_b^* + G_{ba}^* X_a \gamma_{ba} + G_{ba} T_a \rho_{ba} + \epsilon_b. \quad (15)$$

In this model, the Stable Unit Treatment Value Assumption (SUTVA)³² doesn't hold because (i) spillovers are at work and (ii) spillovers are heterogeneous. To the best of our knowledge,

³¹Our analysis can be easily adapted to the case of continuous or multinomial treatment. It is also useful to recall an assumption already listed in the previous sections for estimator properties, $G \perp T_a$, which here states that the treatment does not change the network topology. This assumption relates to Manski (2013) which assumes that reference groups are person-specific and treatment-invariant (unable to be manipulated by the policy maker).

³²Following Rubin (1986), SUTVA states that potential outcomes depend on the treatment received, and not on what treatments other units receive and that there are no "hidden treatments".

there are no studies that consider violations of the SUTVA because of (ii). This is surprising given that heterogeneity in spillovers is naturally implied by differences between eligibles and ineligibles.

Our results in Sections 4 provide consistent and efficient estimators for the parameters of model (2) - (3).³³

6.1 Average Treatment Effect with Heterogeneous Spillovers

The Average Treatment Effect in our context can be written as³⁴

$$ATE = E(Y_i | i \in A, T_{a,i} = 1, X, G) - E(Y_i | i \in A, T_{a,i} = 0, X, G). \quad (16)$$

From the reduced form of equation (14)

$$ATE = \delta_a E_{E_T}(m_{a,ii}), \quad (17)$$

where $m_{a,ii}$ is the ii_{th} element of M_a and $E_{E_T}(\cdot) = E(\cdot | i \in A, T_{a,i} = 1, X, G)$ indicates the expected value over the treated eligibles. The Average Treatment Effect is thus equal to the direct impact of the treatment on the individual i (i.e. δ_a) plus the indirect effect of other agents' spillovers on i triggered by i 's treatment (but not triggered by other nodes' treatment)

$$\delta_a M_a = \delta_a I_a + \delta_a \sum_{k=1}^{\infty} (\phi_a G_a + \phi_{ab} \phi_{ba} C_a)^k.$$

Observe that $m_{a,ii}$ is a function of $(G_a, G_b, G_{ab}, G_{ba}, \phi_a, \phi_b, \phi_{ab}, \phi_{ba})$. This implies that when network interactions are at work, the ATE depends on network topology and strength of outcome spillovers among agents. As a result, an individual can have a high increase in outcome even if she has a low treatment direct impact (a low δ_a) but she is *central* in the network.³⁵ Observe that even if $\delta_{a,i} = \delta_a$ (i.e. the treatment effect is homogenous) the ATE can be heterogeneous because of the different position of nodes in the network. Indeed, the ATE can be decomposed into two parts

$$ATE = \underbrace{\delta_a}_{DTE} + \underbrace{\delta_a E_{E_T}[(diag(M_a - I_a))]}_{FLTE}. \quad (18)$$

The first part is the *Direct Treatment Effect* (hereafter *DTE*), whereas the second part is the effect of the treatment due to the interactions among agents, i.e. the effect of i 's treatment that impact i through other nodes. We denote the latter effect as *Feedback Loop Treatment Effect* (hereafter *FLTE*). The sample counterpart of equation (18) is

$$\hat{ATE} = \mu'_t [\hat{\delta}_a diag(\hat{M}_a)] \mu_t \frac{1}{n_a^t} = \hat{\delta}_a \frac{1}{n_a^t} \sum_{i \in N_a^T} m_{a,ii}, \quad (19)$$

³³As mentioned in the Introduction, we do not consider *direct* treatment effect heterogeneity. This assumption can be relaxed, allowing for a double form of heterogeneity: one coming from individual characteristics, the other from the interactions. The identification becomes much more complex. We leave this extension for future research. Following Manski (2013), we also assume here that the treatment does not change the network topology, i.e. that the policy maker cannot manipulate reference groups.

³⁴When the treatment is a randomized control experiment, the average treatment effect is equal to the average treatment effect on treated.

³⁵Of course the centrality itself is not a sufficient condition, a high level of spillovers is required.

where N_a^T is the set of treated individuals which has cardinality $n_a^t < n_a$, μ_t is the $n_a^t \times 1$ selector vector for that units and $\hat{M}_a = M_a(\hat{\phi}_a, \hat{\phi}_b, \hat{\phi}_{ab}, \hat{\phi}_{ba})$ is the estimated counterpart of M_a .

Treatment Effect Misinterpretation and Bias When SUTVA holds, $ATE = DTE$. If interferences are at work, then $ATE \neq DTE$. However, the problem is not only about interpretation. We show below that if spillovers are ignored, then the parameter estimates can be inconsistent. Suppose that a treatment is administered to $n_a^t < n_a$ subjects and we ignore interactions among them. Estimation of δ_a is based on the following regression

$$Y_a = X_a \beta_a + T_a \delta_a + \epsilon_a^*, \quad (20)$$

where $\epsilon_a^* = \rho_a G_a T_a + \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + \epsilon_a$ contains the three relevant spillover effects omitted:³⁶ (i) the direct treatment spillover from other eligibles $\rho_a G_a T_a$, (ii) the endogenous outcome spillover from other treated eligibles $\phi_a G_a Y_a$ and (iii) the endogenous outcome spillover from ineligibles $\phi_{ab} G_{ab} Y_b$. Misinterpretation occurs because the estimate of δ_a is interpreted as a DTE while, if the data generating process is given by equations (14) and (15), it is an ATE . Bias can occur if the treatment is correlated with the three components listed above

$$\begin{aligned} \hat{\delta}_a = \delta_a + bias &= \delta_a + (T_a' T_a)^{-1} T_a \{ \rho_a G_a T_a \\ &+ \phi_a G_a M_a [\phi_{ab} G_{ab} J_b (\rho_{ba} G_{ba} T_a) + T_a \delta_a + \rho_a G_a T_a] \\ &+ \phi_{ab} G_{ab} M_b [\phi_{ba} G_{ba} J_a (\delta_a T_a + \rho_a G_a T_a) + \rho_{ba} G_{ba} T_a] \}. \end{aligned}$$

The bias is due to the spillover effects coming from the three omitted components listed before. By correctly specifying the interaction structure we can consistently estimate the direct treatment effect purged of the influence of the three omitted components.

It should appear clear from our discussion that, if the spillovers' coefficients and the direct treatment effect are positive, neglecting between and within-group interactions result in an overestimation of the direct treatment effect. Manski (2013) defines this scenario as *Reinforcing Interactions*. Of course one can imagine different scenarios where interactions are not reinforcing and, on the contrary, are *Opposing Interactions*.

Our approach has an advantage from this point of view- it allows interactions between and within groups to be heterogeneous (e.g. *Reinforcing Interactions* within groups members and *Opposing Interactions* between groups). We also note that, using again Manski (2013)'s terminology, our framework can be adapted to the estimation of social interaction with *leaders* and *followers*, labeling those agents as groups A and B.

6.2 Indirect Treatment Effect

As mentioned before, the Indirect Treatment Effect (hereafter *ITE*) has been an object of interest in several papers. Most of the existing papers focus attention on the indirect effect on ineligibles (see, e.g. Angelucci and De Giorgi, 2009). However, when the population is split into two sets, it is also natural and interesting from a policy perspective to understand whether different types of untreated (eligible or ineligible), are differently impacted by the

³⁶The other omitted terms, $X_a^* \beta_a^*$ and $G_{ab}^* X_b \gamma_{ab}$, are independent from the treatment.

treatment. Let us define $ITEE$ and $ITEI$ as the Indirect Treatment Effect on Eligibles and the Indirect Treatment Effect on Ineligibles, respectively.

The Indirect Treatment Effect on Eligibles in our model is

$$ITEE = E(Y_i | i \in A, M_i T \neq 0 \cap T_{a,i} = 0, X, G) - E(Y_i | i \in A, M_i T = 0 \cap T_{a,i} = 0, X, G), \quad (21)$$

whereas the Indirect Treatment Effect on Ineligibles can be defined as

$$ITEI = E(Y_i | i \in B, M_i T \neq 0, X, G) - E(Y_i | i \in B, M_i T = 0, X, G), \quad (22)$$

where M_i is the i_{th} row of $M = \left(\begin{bmatrix} I_a \\ I_b \end{bmatrix} - \begin{bmatrix} G_a \phi_a & G_{ab} \phi_{ab} \\ G_{ba} \phi_{ba} & G_b \phi_b \end{bmatrix} \right)^{-1}$, $T = [T_a, 0_b]$, and 0_b is a $n_b \times 1$ vector of zeros. $M_i T = 0$ indicates that i is not affected by any of the treated nodes (i.e. that there are no direct and indirect paths in the networks between i and a treated node).

Let us now suppose that, given our data generating process (equations (14) and (15)) we are asked by a policy maker to evaluate the Indirect Treatment Effects after a treatment administered to the eligibles (i.e. to a subset of A). From model (14) - (15) we can derive the following formulas

$$ITEE = E_{E_u}[M_{a_i}(\phi_{ab}G_{ab}J_b(\rho_{ba}G_{ba}T_a) + \delta_a T_a + \rho_a G_a T_a)],$$

$$ITEI = E_I[M_{b_i}(\phi_{ba}G_{ba}J_a(\delta_a T_a + \rho_a G_a T_a) + \rho_{ba}G_{ba}T_a)],$$

where M_{a_i} is the i_{th} row of M_a , M_{b_i} is the i_{th} row of M_b , $E_I(\cdot) = E_I(\cdot | i \in B, M_i T \neq 0, X, G)$ indicates the expected value over the (indirectly treated) ineligibles, and $E_{E_u}(\cdot) = E(\cdot | i \in A, M_i T \neq 0 \cap T_{a,i} = 0, X, G)$ indicates the expected value over (indirectly treated) untreated eligibles. Observe that these estimands depend on direct and indirect connections because of network-based spillovers. More formally, they can be decomposed into different parts. For instance, $ITEE$ may be decomposed into three effects. The first term, $\delta_a M_a$, captures propagation of the treatment via outcome spillovers.³⁷.

The nice feature of this derivation of $ITEI$ and $ITEE$ is that instead of simply addressing the question whether an ITE is different from zero, we can also decompose it into different sources of treatment's transmission. For instance, one can find that the treated population has a strong reaction to the treatment (δ_a and ρ_a are high) and transmits it to ineligibles through low magnitude peer effects (ϕ_{ab} is low). The same level of $ITEI$, however, can also arise from a scenario where there is a low reaction to the treatment within group (δ_a and ρ_a are low) and a large transmission between groups (ϕ_{ab} is high).

Understanding these different channels is paramount for policy purposes. Most importantly, our framework enables the researcher to distinguish the role of *contextual effects* from

³⁷Given that $M_a = (I - \phi_a G - \phi_{ab} \phi_{ba} C_a)^{-1}$, we have $M_a \delta_a = I_a \delta_a + [(I_a - \phi_a G - \phi_{ab} \phi_{ba} C_a)^{-1} - I] \delta_a$. The first term is the diagonal matrix of treatment direct effects which has (by definition) no impact on the untreated, while the second term represents the propagation of those effects through the network via endogenous spillovers (i.e. changes in outcomes due to treatment). The second term, $\rho_a M_a G_a$, measures the spillover arising from the treatment given to other units ($\rho_a G_a$), as well as its amplification through interactions (as captured by M_a). Finally, $\phi_{ab} \rho_{ba} M_a G_{ab} J_b G_{ba}$, denotes the spillover accruing to ineligibles distinguished between outcome amplification ($M_a G_{ab} J_b$) and (indirect) treatment effect ($\rho_{ba} G_{ba}$). A similar decomposition can be applied to $ITEI$

peer effects in transmitting the treatment. In other words, one can quantify how much of the effect is generated by the direct effect of the treatment through exogenous variables (as captured by δ_a , ρ_a and ρ_{ba}) and how much is due to spillovers through outcomes (as captured by ϕ_{ba} , ϕ_a , ϕ_b and ϕ_{ab}). Note also that having these estimates at hand, one can understand which effects (within eligibles, within ineligibles and between them) are the dominant ones in spreading out the policy's beneficial effect.

We can thus simply use the sample counterpart to estimate the $ITEE$ and $ITEI$

$$\begin{aligned} \hat{ITEE} &= \mu'_u [\hat{M}_a (\hat{\phi}_{ab} G_{ab} \hat{J}_b (G_{ba} \hat{\gamma}_{ba}) + \hat{\beta}_a + G_a \hat{\gamma}_a)] \mu_u \frac{1}{n_a^u}, \\ \hat{ITEI} &= \iota'_b [\hat{M}_b (\hat{\phi}_{ba} G_{ba} \hat{J}_a (\hat{\beta}_a + G_a \hat{\gamma}_a) + G_{ba} \hat{\gamma}_{ba})] \iota_a \frac{1}{n_b}, \end{aligned}$$

where $n_a^u < n_a$ is the number of eligibles who are untreated, μ_u is the $n_a \times 1$ selector vector for that units and ι_l is an $n_l \times 1$ vector of ones.

6.3 Total Treatment Effect

One can also be interested in evaluating the treatment effect on the entire population (or network). As the SUTVA has been removed and spillovers are in place, it is useful to derive the Total Treatment effect (hereafter TTE). Following our previous notation we have the following definition for TTE

$$TTE = E(Y_i | i \in A \cup B, M_i T \neq 0, X, G) - E(Y_i | i \in A \cup B, M_i T = 0, X, G).$$

This represents the treatment effect on a generic individual in the network (eligible or ineligible). Its sample counterpart is

$$\hat{TTE} = \iota' \left(\begin{bmatrix} I_a \\ I_b \end{bmatrix} - \begin{bmatrix} G_a \hat{\phi}_a & G_{ab} \hat{\phi}_{ab} \\ G_{ba} \hat{\phi}_{ba} & G_b \hat{\phi}_b \end{bmatrix} \right)^{-1} \left(\hat{\delta}_a \begin{bmatrix} T_a \\ O_b \end{bmatrix} + \begin{bmatrix} G_a \hat{\rho}_a & G_{ab} \\ G_{ba} \hat{\rho}_{ba} & G_b \end{bmatrix} \begin{bmatrix} T_a \\ O_b \end{bmatrix} \right) \iota \frac{1}{n},$$

where ι is an $n \times 1$ vector of ones. Note that the \hat{TTE} is basically the weighted average of \hat{ATE} , \hat{ITEE} , and \hat{ITEI} .

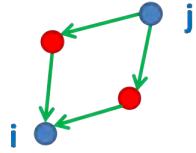
6.4 Control Group

It is well-known that the ATE , $ITEE$, $ITEI$ and TTE are identified if we have a control group, i.e. if we can distinguish sample of units who are not treated (directly or indirectly). This can be quite challenging when estimating the indirect treatment effects. In a network context, we have two possibilities: (i) a *multiple network-based approach* and (ii) *topology-driven approach*.

In the first case, we have multiple networks, some of which are randomly treated and others which are not- offering a valid control group. A similar scheme is often followed for policy design and evaluation in a non-network context.³⁸

³⁸For example, in PROGRESA, a set of treated and untreated villages are surveyed (see Angelucci and De Giorgi (2009) for more details on the program design).

Figure 2: *Topology-driven* policy evaluation design



The second possibility is unique to a network approach and exploits the architecture of networks. When information on actual connections is available and the *direction* is known, it may be possible to estimate *ATE*, *ITEE*, *ITEI* and *TTE* using only one network. The network topology determines the possibility of having the control population if there are some nodes in the network that are not influenced by a treatment to other nodes. For example, let us consider the network in Figure 2, where the red nodes are treated and the blue ones are not. According to the directions of the edges (arrows in the picture), the blue node i is influenced by red nodes whereas the blue node j is not. Therefore, the direction of the links between nodes stemming from this network topology allows us to distinguish between indirectly treated nodes (node i) and control group nodes (node j).³⁹

6.5 Policy Experiments

Manski (2013) studies treatment response in settings with endogenous effects. In this framework, endogenous effects are seen as a mechanism in which the treatments could propagate.⁴⁰ The main objects of interest are $P[Y]$ and $P[Y(T_a)]$, the outcome distributions respectively without and with a treatment T_a administered to the population. Policy makers are usually interested in comparing these two distributions since interventions are often finalized to reduce inequality between a disadvantaged cluster and the rest of the population. The marginal effect of T on Y accounts for the adjustment of the outcome after a policy intervention.

In this paper, we consider a network framework with heterogeneous peer effects similar to Manski (2013).⁴¹ In this section, we numerically study the empirical density functions $P[Y_a] - P[Y_a(T_a)]$ and $P[Y_b] - P[Y_b(T_a)]$, where the subscripts indicate the reference to eligible and ineligible populations.

We perform a numerical simulation to assess the extent to which the underlining heterogeneity of the endogenous effects can affect the outcome response for different groups. Our goal is to provide evidence about the individual and aggregate implications of this heterogeneity. In the simulation experiment below, we show that for some values of ϕ_a , ϕ_b , ϕ_{ba} and ϕ_{ab} it may be (paradoxically) more convenient to treat a group other than the target one. This has implications for the study of socio-economic inequality. Importantly, by allowing estimation of all the different parameters of interest, our model specification can be used to understand what nodes (or which type of nodes) should be targeted by a social planner whose final goal is to maximize an aggregate outcome or to converge to a desired distribution of individual

³⁹Note also that we need these two kinds of nodes to be comparable in terms of characteristics.

⁴⁰If a dynamic model is at work, then a social multiplier may also arise in terms of expectations.

⁴¹The framework in Manski (2013) considers only one group, thus homogeneous peer effects (and no between group interactions).

outcomes.

We present an experiment where we treat a random sample of nodes and simulate the treatment's propagation through a network characterized by *heterogeneous peer effects*. More specifically, we look at the increase of type A and type B nodes' outcomes once a certain set of nodes receives a treatment⁴².

Two exercises are implemented. In the first, we evaluate aggregate effects, i.e. the change in the sum of outcomes (for both type A and type B individuals) which follows a treatment for different values of peer effects (i.e. ϕ_a , ϕ_b , ϕ_{ba} , ϕ_{ab}). In the second exercise, we look at distributional effects, i.e. at changes in the empirical distribution of individual node outcomes for different sets of peer effects parameters following the policy intervention .

Figures 5 and 6 report on the first exercise. Figure 5 depicts the results when fixing $\phi_b = \phi_{ba} = 0.1$ and varying ϕ_a and ϕ_{ab} . We generate a grid of values for parameters resulting from two sequences: $\phi_a = 0.02, 0.04, \dots, 0.50$ and $\phi_{ab} = 0.02, 0.04, \dots, 0.50$. For each couple (ϕ_a, ϕ_{ab}) we generate one hundred independent replications using the same DGP as described in Section 4.3 and compute $Y_a^s = \sum_{i \in A} y_i$ and $Y_b^s = \sum_{i \in B} y_i$. We then select a random sample of one hundred type A nodes to be treated. This treatment is represented by an $n_a \times 1$ vector T_a of zeros for non treated nodes and ones for treated nodes. Finally, we compute $Y_a^{s*} = \sum_{i \in A} y_i^*$ and $Y_b^{s*} = \sum_{i \in B} y_i^*$, where $y_i^* = y_i + \frac{\partial y_i}{\partial T_a} T_a$. This exercise represents the case where group A nodes are treated and there are low interactions between nodes A and nodes B ($\phi_{ba} = 0.1$).

From equations (14) and (15) we have

$$\Delta y_i = y_i^* - y_i = \frac{\partial y_i}{\partial T_a} T_a = \begin{cases} M_{a_i}(\phi_{ab}G_{ab}J_b(\rho_{ba}G_{ba}T_a) + (\delta_a + G_a\rho_a)T_a) & \text{if } i \in A \\ M_{b_i}(\phi_{ba}G_{ba}J_a(\delta_a + \rho_aG_a)T_a + \rho_{ba}G_{ba}T_a) & \text{if } i \in B \end{cases}.$$

Figure 5 represents the differences $\Delta Y_a^s = Y_a^{s*} - Y_a^s = \sum \frac{\partial Y_a}{\partial T_a} T_a$ and $\Delta Y_b^s = Y_b^{s*} - Y_b^s = \sum \frac{\partial Y_b}{\partial T_a} T_a$ for all the possible combinations (ϕ_a, ϕ_{ab}) .⁴³

Figure 5 shows that ΔY_a^s increases steadily with ϕ_a (and slightly with ϕ_{ab}), whereas ΔY_b^s remains roughly unchanged. These results are not surprising. If there are no interactions (or low interactions) between the two groups, then there is no reason why the outcome of group B should change. The variation in the outcome of the group A depends on the extent of the endogenous effects (ϕ_a). If instead there are interactions between the two groups, then the treatment response depends on both ϕ_a and ϕ_{ab} . For example, assuming a positive effect of the independent variable, if a policy intervention targets a group when the two groups have the same outcome profile, we expect an increase in *inequality* in terms of outcomes between

⁴²We compute the marginal effect matrix of T_a on Y_a multiplied by the treatment vector

$$\frac{\partial E(Y_a|G, X)}{\partial T_a} T_a = M_a(\phi_{ab}G_{ab}J_b(\rho_{ba}G_{ba}T_a) + \delta_a T_a + \rho_a G_a T_a).$$

Note that when there are no interactions between the two groups (or only type A nodes are considered in the analysis), we have $\frac{\partial E(Y_a|G, X)}{\partial T_a} T_a = S_a(\delta_a T_a + \rho_a G_a T_a)$, where $S_a = (I - \phi_a)^{-1}$. This is the marginal effect matrix in a standard peer effects model.

The marginal effect matrix of T_a on Y_b is

$$\frac{\partial E(Y_b|G, X)}{\partial T_a} T_a = M_b(\phi_{ba}G_{ba}J_a(\delta_a T_a + \rho_a G_a T_a) + \rho_{ba}G_{ba}T_a). \quad (23)$$

Observe that the marginal effect of T_a on Y_a is different from the marginal effect of T_a on Y_b -an increase in T_a differently affects nodes depending on their type.

⁴³Some combinations are missing in the grid because it is unlikely to draw G_a and G_{ab} such that $\|\phi_a G_a\|_\infty + \|\phi_{ab} \phi_{ba} G_a\|_\infty \leq 1$. These combinations are at the edge of the parameter space.

the two types when the within-peer effects (ϕ_a) are high and the between-peer effects (ϕ_{ab}) are low.

[INSERT FIGURE 5 HERE]

Figure 6 depicts the results when fixing $\phi_b = \phi_{ab} = 0.1$, and varying ϕ_a and ϕ_{ba} . The experiment design remains unchanged. This exercise represents the case where group A nodes are treated and there are increasing influences within nodes A and from nodes A and nodes B (ϕ_{ba} increasing up to 0.5). Figure 6 shows that an increase of ϕ_{ba} is beneficial for ΔY_b^s , as type B nodes receive an impulse from type A nodes. Interestingly, type B nodes may actually benefit even more than A nodes (the treated group). Our results shows that when $\phi_{ba} > 0.20$, we observe $\Delta Y_b^s > \Delta Y_a^s$. In terms of policy effects, this means that if a policy targets one group but peer effects between groups are high, then we can observe increasing *inequality* between the two groups, rather than the expected decrease (assuming that the targeted group has a lower starting outcome). In terms of the estimands derived in Section 6 , note that the blue surfaces in Figures 5 and 6 are simply $ITEI \times n_b$ while the red ones are $ITEE \times n_a^u + ATE \times n_a^t$, plotted for different combinations of parameters.

[INSERT FIGURE 6 HERE]

In the second exercise, we consider four points from the grid formed by ϕ_a and ϕ_{ba} and look at the empirical distributions of $\Delta y_{i \in A}$ and $\Delta y_{i \in B}$. We estimate these distributions using a normal kernel density. We consider the case where $\phi_a = 0.1$ and $\phi_{ba} = 0.1$ as a benchmark and then increase the strength of peer effects among agents in different ways.

In Figure 7 we increase the effect within group A only ($\phi_a = 0.3$). While this change is irrelevant for type B nodes, it has interesting implications for the distribution of outcomes among type A nodes (Panel a). While in the benchmark model (the single line), the distribution is quasi-bimodal (due to the treated and non-treated A nodes), an increase of ϕ_a smooths the distribution (the bold line). In other words, the higher the endogenous effects, the more evenly the benefits of the policy intervention are shared among nodes (individuals).

[INSERT FIGURE 7 HERE]

In Figure 8 we increase the between-group effect only ($\phi_{ba} = 0.3$). Type A density remains basically unchanged (Panel a) . The impact is instead apparent on the outcome distribution of type B nodes (Panel b). One can observe an important shift to the right. This means that non treated type B nodes benefit more than non-treated type A nodes (from the treatment to type A nodes).

The red and blue curves in Figures 7 and 8 are the empirical density functions $P[Y_a(T_a) - Y_a]$ and $P[Y_b(T_a) - Y_b]$, respectively. They have $\frac{ITEE \times n_a^u + ATE \times n_a^t}{n_a}$ and $ITEI$ as expected values.

[INSERT FIGURE 8 HERE]

7 Concluding Remarks

We generalize the linear-in-means model to the presence of two groups and between-group interactions. We derive the sufficient conditions to identify the model and propose efficient

2SLS estimators. We characterize the bias which arises when interactions are ignored and evaluate it in finite sample using simulation experiments. We illustrate the relevance of these issues for policy purposes. If peer effects are seen as a mechanism in which the treatments could propagate through the networks, then accounting for *heterogeneous peer effects* and *between-group interactions* can be helpful in designing and evaluating policy interventions that alter the outcome distribution. We show that when between-group interactions are strong, an impulse to a given group can engender benefits to another group which are even higher than those accruing to the target group. Examples of types of interventions where the local non-target population may also be indirectly affected by the treatment through social and economic interaction with the target population are widely varied. For example, the recipients of conditional cash transfers may share resources with ineligible households who live in the same community or with extended family members, which could affect the incentives to accumulate human capital (Angelucci et al., 2010). School vouchers or other incentives (such as equipment provision) to increase schooling of indigent children may increase the learning ability of untreated children if, for example, textbooks or computers are shared among classmates. A number of organizations promote the deworming of children in the developing world as a public health and development strategy. Supplying deworming drugs to a group of children may benefit untreated children by reducing disease transmission, thus lowering infection rates for both groups.

In sum, our paper contributes to the literature by providing a framework able to decompose the treatment response into different components, including the crucial difference between endogenous effects and effects stemming from exogenous variations in the characteristics of the treated.

References

- Anatolyev, Stanislav (2013). “Instrumental variables estimation and inference in the presence of many exogenous regressors”. In: *The Econometrics Journal* 16.1, pp. 27–72.
- Angelucci, M. et al. (2010). “Family networks and school enrolment: Evidence from a randomized social experiment”. In: *Journal of Public Economics* 94.3–4, pp. 197–221.
- Baltagi, B. H., C. Kao, and L. Liu (2012). “On the Estimation and Testing of Fixed Effects Panel Data Models with Weak Instruments”. In: *Advances in Econometrics* 30, pp. 199–235.
- Bekker, P. A. (1994). “Alternative approximations to the distributions of instrumental variable estimators”. In: *Econometrica*, pp. 657–681.
- Benjamini, I. and Y. Peres (1994). “Markov chains indexed by trees”. In: *The annals of probability* 22, pp. 219–243.
- Bramoullé, Y., H. Djebbari, and B. Fortin (2009). “Identification of peer effects through social networks”. In: *Journal of Econometrics* 150, pp. 41–55.
- Chandrasekhar, A. and R. Lewis (2011). “Econometrics of sampled networks”. In: *Unpublished manuscript, MIT.[422]*.
- Chipman, Hugh A, Edward I George, Robert E McCulloch, et al. (2010). “BART: Bayesian additive regression trees”. In: *The Annals of Applied Statistics* 4.1, pp. 266–298.
- Cohen-Cole, E., X. Liu, and Y. Zenou (2012). “Multivariate Choice and Identification of Social Interactions”. In: *CEPR Discussion Paper No. DP9159*.
- Crump, Richard K et al. (2008). “Nonparametric tests for treatment effect heterogeneity”. In: *The Review of Economics and Statistics* 90.3, pp. 389–405.
- Donald, S. and W. Newey (2001). “Choosing the Number of Instruments”. In: *Econometrica* 69.5, pp. 1161–1191.
- Firpo, Sergio (2007). “Efficient Semiparametric Estimation of Quantile Treatment Effects”. In: *Econometrica* 75.1, pages.
- Goldsmith-Pinkham, P. and G. Imbens (2013). “Social networks and the identification of peer effects”. In: *Journal of Business and Economic Statistics* 31, pp. 253–264.
- Hsieh, C. S. and L. F. Lee (2011). *A Social Interactions Model with Endogenous Friendship Formation and Selectivity*. Tech. rep. Working paper.
- Hudgens, M. and E. Halloran (2008). “Toward Causal Inference With Interference”. In: *Journal of American Statistical Association* 103, pp. 832–842.
- Imai, Kosuke, Marc Ratkovic, et al. (2013). “Estimating treatment effect heterogeneity in randomized program evaluation”. In: *The Annals of Applied Statistics* 7.1, pp. 443–470.
- Imbens, G., M. Kolesar, et al. (2011). “Identification and Inference with Many Invalid Instruments”. In:
- Imbens, G. and G. Woolridge (2009). “Recent Developments in the Econometrics of Program Evaluation.” In: *Journal of Economic Literature* 47(1), pp. 5–86.
- Jackson, M.O. and Y. Zenou (forthcoming 2013). *Economic Analyses of Social Networks*.
- Kelejian, H. and I. R. Prucha (1998). “A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”. In: *The Journal of Real Estate Finance and Economics* 17.1, pp. 99–121.
- (1999). “A generalized moments estimator for the autoregressive parameter in a spatial model”. In: *International economic review* 40.2, pp. 509–533.

- Kelejian, H. and I. R. Prucha (2001). "On the asymptotic distribution of the Moran I test statistic with applications". In: *Journal of Econometrics* 104.2, pp. 219–257.
- (2004). "Estimation of simultaneous systems of spatially interrelated cross sectional equations". In: *Journal of Econometrics* 118.1, pp. 27–50.
- (2007). "HAC estimation in a spatial framework". In: *Journal of Econometrics* 140.1, pp. 131–154.
- LeBlanc, Michael and Charles Kooperberg (2010). "Boosting predictions of treatment success". In: *Proceedings of the National Academy of Sciences* 107.31, pp. 13559–13560.
- Liu, X. (2013a). "Estimation of a local-aggregate network model with sampled networks". In: *Economics Letters* 118.1, pp. 243–246.
- Liu, X. and L. F. Lee (2010). "GMM estimation of social interaction models with centrality". In: *Journal of Econometrics* 159, pp. 99–115.
- Liu, X., E. Patacchini, and E. Rainone (2013). "The Allocation of Time in Sleep: A Social Network Model with Sampled Data". In: *CEPR Discussion Paper No. DP9752*.
- Liu, X., E. Patacchini, and Y. Zenou (2014forthcoming). "Endogenous peer effects: Local Aggregate or Local Average?" In: *Journal of Economic Behavior and Organization*.
- Manski, C. F. (1993). "Identification of endogenous social effects: The reflection problem". In: *The review of economic studies* 60.3, pp. 531–542.
- (2013). "Identification of treatment response with social interactions". In: *The Econometrics Journal* 16.1, S1–S23. ISSN: 1368-423X.
- Miguel, Edward and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities". In: *Econometrica* 72, pp. 159–217.
- Rubin, D. (1986). "Which ifs have causal answers? Discussion of Hollands Statistics and causal inference". In: *Journal of the American Statistical Association* 81, pp. 961–962.
- Sinclair, Betsy, Margaret McConnell, and Donald P. Green (2012). "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments". In: *American Journal of Political Science* 56, pp. 1055–1069.
- Staiger, D. and J. H. Stock (1997). "Instrumental Variables Regression with Weak Instruments". In: *Econometrica* 65.3, pp. 557–586.

APPENDIX

Appendix A: Assumptions and Discussions

Let us introduce some notation and assume the following regularity conditions: a sequence of square matrices $\{A\}$, where $A = [A_{ij}]$, is defined "uniformly bounded in absolute value" (UB) if there exists a constant $c_b < \infty$ (that does not depend on n) such that $\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^n |A_{ij}| < c_b$ and $\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^n |A_{ij}| < c_b$. We indicate that $\{A\}$ is bounded only in row (column) sum absolute value as UBR (UBC). For the sake of simplicity we will assume that $n \rightarrow \infty$ implies $n_a \rightarrow \infty$ and $n_b \rightarrow \infty$.

Assumption 1. *The elements of ϵ_a and ϵ_b are iid with zero mean, variance σ_a^2 and σ_b^2 respectively, and zero covariance. Moments higher than the fourth exist.*

Assumption 2. *The elements of X_a and X_b are uniformly bounded constants, X_a and X_b have full rank k , and $\lim_{n_a \rightarrow \infty} \frac{1}{n_a} X_a' X_a$ and $\lim_{n_b \rightarrow \infty} \frac{1}{n_b} X_b' X_b$ are finite and non singular.*

Assumption 3. *The sequences of matrices $\{G_a\}$, $\{G_{ab}\}$, $\{G_b\}$, $\{G_{ba}\}$, $\{M_a\}$, $\{M_b\}$, $\{J_b\}$, and $\{J_a\}$ are UB.*

The first assumption is needed in order to apply the Kelejian and Prucha (2001) Central Limit Theorem (CLT) of a linear and quadratic form. Assumption 2 is standard in the literature. Assumption 3 is exploited in Kelejian and Prucha (1999) to limit the spatial dependence among the units. It rules out any spatial unit root case. As Lee (2004) pointed out, it plays an important role in the derivation of asymptotic properties of the estimators for spatial econometric models. It guarantees that the variance of Y_a and Y_b is bounded as n goes to infinity. Observe that this assumption is also crucial for the identification of the heterogeneous network model, as shown in Proposition 1.

Assumption 4 is a sufficient condition for identification of the social network model. For assumption 4 to hold, $E(Z_a)$ must be full column rank for large enough n_a .

Assumption 4. *$F_a = \lim_{n \rightarrow \infty} \frac{1}{n} f_a' f_a$ is finite and a full rank matrix, $F_b = \lim_{n \rightarrow \infty} \frac{1}{n} f_b' f_b$ is finite and a full rank matrix.*

Since the variance of the structural error is $\text{var}(v_a)$ and the concentration parameter (which measures the instrument's strength) is $f_a' f_a / \text{var}(v_a)$, this assumption implies that the concentration parameter grows at the same rate as the sample size. Such a rate is assumed in Bekker (1994). Hence, we assume that the instruments are stronger than assumed in the weak-instrument literature.⁴⁴ For the sake of brevity we focus on equation (2), and we imply the same argument holds for equation (3).

Assumption 5. *There exists a $K \times (k + 2)$ matrix Θ_K such that $\frac{1}{n} \|E(Z_a) - H_K \Theta_K\|^2 \rightarrow 0$ as $n, K \rightarrow \infty$.*

⁴⁴See Staiger and Stock, 1997 or Baltagi et al., 2012 for a panel data version of weak-instrument asymptotics. Another interesting extension could be to derive the estimator's asymptotic properties under many weak instruments. In doing so, we are allowing the rate of concentration parameter to be different than the rate of the sample size. Consequently, we can compare it with the rate in which K increases. See for example, Chao and Swanson (2005)

Following Lee and Liu(2010), assumption 5 requires that the (infeasible) best IV matrix can be well approximated by a certain linear combination of the feasible IV matrix H_K as the number of instruments increases with the sample size. Once we assume this, we can deal with the approximation of S_a and S_{ab} . We have to approximate this matrix since we cannot use it as matrix of instruments because it is formed by unknown parameters. If H_K has the following structure then assumption 5 holds and we can obtain efficiency under certain conditions.

Proposition 5. *If $\|\phi_a G_a\|_\infty + \|\phi_{ab} \phi_{ba} C_a\|_\infty < 1$, let us define $H_K^{(p)} = (H_K^{a(p)}, H_K^{b(p)}, X_a^*, G_{ab} X_b)$ where*

$$H_K^{a(p)} = (G_a(G_a, (G_{ab}(G_b, \dots, G_b^{p+1})G_{ba}), \dots, (G_a(G_a, (G_{ab}(G_b, \dots, G_b^{p+1})G_{ba})^{p+1}(E(A), G_{ab}J_bB)),$$

$$H_K^{b(p)} = (G_{ab}(G_b, (G_{ba}(G_a, \dots, G_a^{p+1})G_{ab}), \dots, (G_b(G_b, (G_{ba}(G_a, \dots, G_a^{p+1})G_{ab})^{p+1}(E(B), G_{ba}J_aA),$$

where p is an increasing integer valued function of K , there exists a $K \times (k+2)$ matrix $\Theta_K^{(p)}$ such that $\|f_a - H_K^{(p)} \Theta_K^{(p)}\|_\infty \rightarrow 0$ as $n, K \rightarrow \infty$.

Therefore, the 2SLS estimator can be asymptotically efficient when we use an increasing number of instruments.

Appendix B: Some Useful Lemmas

Lemma 1. Recall that $Z_a = f_a + v_a$. Let $e_f = \frac{1}{n}f'_a(I - P_K)f_a$. As $K/n = O(1)$: (i) $\text{Tr}(e_f) = o(1)$ (ii) $v'_a P_K v_a = O_p(K)$. (iii) $f'_a P_K v_a = O_p(\sqrt{nK})$. (iv) $e_f = O(\text{tr}(e_f))$. (v) $\frac{1}{n}Z'_a P_K Z_a = \frac{1}{n}f'_a f_a - e_f + \frac{1}{n}f'_a P_K v_a + \frac{1}{n}v'_a P_K f_a + \frac{1}{n}v'_a P_K v_a = O_p(1)$.

Proof. (i) See lemma B.3 (i) in Lee and Liu (2010).

(ii) Let us write $v'_a P_K v_a = \epsilon'_a S'_a P_K S_a \epsilon_a + \epsilon'_b J'_b G'_{ba} S'_{ab} P_K S_{ab} G_{ba} J_b \epsilon_b$. Let us focus on the first term of the sum, since $E|\epsilon'_a S'_a P_K S_a \epsilon_a| = E[\text{tr}(|\epsilon'_a S'_a P_K S_a \epsilon_a|)] = \sigma_a^2 \text{tr}(|P_K S_a S'_a P_K|) = O(K)$ by lemma B.2 (ii) Lee and Liu (2010), then by Markov's Inequality $\text{Pr}(|\epsilon_a S'_a P_K S_a \epsilon_a| \geq \alpha) \leq \frac{E(|\epsilon_a S'_a P_K S_a \epsilon_a|)}{\alpha} = O_p(K)$.

For the second part of the sum, given also that $S_{ab} G_{ba} J_b = T_{ab}$ where T_{ab} is UB, we can apply the same proof and obtain the same order of probability. We then have $O(f(x)) + O(f(x)) = O(f(x))$.

(iii) For each j we have by Cauchy-Schwarz inequality $|e'_j f'_a P_K v_a| \leq \sqrt{e'_j f'_a f_a e_j} \sqrt{\epsilon_a S'_a P_K S_a \epsilon_a} = O(\sqrt{n}) O_p(\sqrt{K}) = O_p(\sqrt{nK})$.

(iv) By lemma A.3 (ii) in Donald and Newey (2001).

(v) $\frac{1}{n}Z'_a P_K Z_a = \frac{1}{n}f'_a f_a - O(\text{tr}(e_f)) + O_p(K/n) + O_p(\sqrt{K/n}) = O_p(1)$. \square

Lemma 2. Recall that $Z_a = f_a + v_a$, let $P_K S_a = \Psi_a$ and $P_K T_{ba} = \Xi_{ba}$. As $K/n = O(1)$: (i) $E(v'_a P_K \epsilon_a) = \sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]'$. (ii) $E(v'_a P_K \epsilon_a \epsilon'_a P_K v_a) = \sigma_a^4 \text{tr}^2([\Psi_a, \Xi_{ba}]) + O(K)$. (iii) $[Z'_a P_K \epsilon_a - [\sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba} \text{tr}(\Xi_{ba})]'] / \sqrt{n}] = f'_a \epsilon_a / \sqrt{n} + O_p(\sqrt{K/n}) + O_p(\sqrt{\text{tr}(e_f)}) = f'_a \epsilon_a / \sqrt{n} + O_p(1)$.

Proof. (i) $E(v'_a P_K \epsilon_a) = [e_1, e_2] E([(f_a S_a G_{ab} J_b \epsilon_b + S_a \epsilon_a), (\phi_{ba} S_{ab} G_{ba} J_a \epsilon_a + S_{ab} \epsilon_b)]') P_K \epsilon_a = [e_1, e_2] [E(\epsilon'_a S'_a P_K \epsilon_a), E(\phi_{ba} \epsilon'_a J'_a G'_{ba} S'_{ab} P_K \epsilon_a)]' = \sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba} \text{tr}(\Xi_{ba})]'$.

(ii) By lemma A.2 in Lee (2001), $E(v'_a P_K \epsilon_a \epsilon'_a P_K v_a) = E([\epsilon'_a (\Psi_a) \epsilon_a \epsilon'_a (\Psi_a) \epsilon_a, \epsilon'_a \Xi_{ba} \epsilon_a \epsilon'_a \Xi_{ba} \epsilon_a]) =$

$$\begin{aligned} & (\mu_4^a - 3\sigma_a^4) \sum_i [(\Psi_a), \phi_{ba}(\Xi_{ba})]_{ii}^2 + \sigma_0^4 [[\text{tr}^2(\Psi_a), \phi_{ba} \text{tr}(\Xi_{ba})] + \\ & \quad \text{tr}([\Psi_a, \phi_{ba}(\Xi_{ba})]' [\Psi_a, \phi_{ba}(\Xi_{ba})] + \text{tr}([\Psi_a]^2, [\Xi_{ba}]^2))] \\ & = \sigma_a^4 \text{tr}^2([\Psi_a, \Xi_{ba}]) + O(K), \end{aligned}$$

where the last equality holds by Lemma B.2 (ii) in Lee and Liu (2010).

(iii) Since $Z'_a P_K \epsilon = f_a \epsilon_a - f'_a (I - P_K) \epsilon + v_a P_K \epsilon$, then $(Z'_a P_K \epsilon_a - \sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba} \text{tr}(\Xi_{ba})]') / \sqrt{n} =$

$$f'_a \epsilon_a / \sqrt{n} - f'_a (I - P_a) \epsilon / \sqrt{n} + [v'_a P_K \epsilon_a - \sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba} \text{tr}(\Xi_{ba})]] / \sqrt{n}.$$

By Lemma 1 above and by Lemma B.2 (ii) in Lee and Liu (2010) $\sqrt{n} f'_a (I - P_a) \epsilon_a = O_p(\sqrt{\text{tr}(e_f)})$. By Lemma 2 (i), (ii) and Markov's inequality for variance we have $\frac{1}{\sqrt{n}} [v'_a P_K \epsilon_a - \sigma_a^2 [e_1, e_2] [\text{tr}(\Psi_a), \hat{\phi}_{ba} \text{tr}(\Xi_{ba})]]' = O_p(\sqrt{K/n})$. \square

Appendix C: Proofs

Proof of proposition 1. We need to prove that $E(Z_a) = (E(G_a y_a), E(G_{ab} y_b), X_a, G_a X_a, G_{ab} X_b)$ is full column rank. This means that if $E(G_a y_a) d_1 + E(G_{ab} y_b) d_2 + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0$ then $d_1 = d_2 = d_3 = d_4 = d_5 = 0$, where d_1, d_2, d_3, d_4, d_5 are parameters.

By inserting the definitions of $E(G_a y_a)$ and $E(G_{ab} y_b)$ we have:

$$G_a(M_a(\phi_{ab}G_{ab}J_bE(B)\delta_b + E(A)\delta_a)d_1 + G_{ab}(M_b(\phi_{ba}G_{ba}J_aE(A)\delta_b + E(B)\delta_a)d_2 + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0.$$

More explicitly,

$$\begin{aligned} & G_a(M_a(\phi_{ab}G_{ab}J_b(X_b\beta_b + GX_b\gamma_b + G_{ba}X_a\gamma_{ba}) + X_a\beta_a + GX_a\gamma_a + G_{ab}X_b\gamma_{ab})d_1 \\ & + G_{ab}(M_b(\phi_{ba}G_{ba}J_a(X_a\beta_a + G_aX_a\gamma_a + G_{ab}X_b\gamma_{ab}) + X_b\beta_b + G_bX_b\gamma_b + G_{ba}X_a\gamma_{ba})d_2 \\ & \quad + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0 \end{aligned} \quad (24)$$

Let us assume that J_a , J_b , M_a and M_b are invertible and thus

$$J_a = (\phi_a G_a)^{-1} = \sum_{k=0}^{\infty} (\phi_a G_a)^k,$$

$$J_b = (\phi_b G_b)^{-1} = \sum_{k=0}^{\infty} (\phi_b G_b)^k,$$

$$M_a = (\phi_a G_a + \phi_{ab}\phi_{ba}G_{ab}J_bG_{ba})^{-1} = \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba}G_{ab} \sum_{k=0}^{\infty} (\phi_b G_b)^k G_{ba})^j,$$

$$M_b = (\phi_b G_b + \phi_{ba}\phi_{ab}G_{ba}J_aG_{ab})^{-1} = \sum_{j=0}^{\infty} (\phi_b G_b + \phi_{ba}\phi_{ab}G_{ba} \sum_{k=0}^{\infty} (\phi_a G_a)^k G_{ab})^j.$$

Going back to equation (24), we obtain

$$\begin{aligned} & G_a(\sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba}G_{ab} \sum_{j=0}^{\infty} (\phi_b^j G_b^j) G_{ba})^j (\phi_{ab}G_{ab} \sum_{k=0}^{\infty} (\phi_b G_b)^k (X_b\beta_b + G_bX_b\gamma_b + G_{ba}X_a\gamma_{ba})) \\ & \quad + X_a\beta_a + G_a X_a \gamma_a + G_{ab} X_b \gamma_{ab}) d_1 \\ & + G_{ab}(\sum_{j=0}^{\infty} (\phi_b G_b + \phi_{ba}\phi_{ab}G_{ba} \sum_{k=0}^{\infty} (\phi_a G_a)^k G_{ab})^j (\phi_{ba}G_{ba} \sum_{k=0}^{\infty} (\phi_a G_a)^k (X_a\beta_a + G_a X_a \gamma_a + G_{ab} X_b \gamma_{ab})) + \\ & \quad X_b\beta_b + G_b X_b \gamma_b + G_{ba} X_a \gamma_{ba}) d_2 \\ & \quad + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0 \end{aligned} \quad (25)$$

The left side of the previous equation is the sum of products of the matrices G_a , G_{ba} , G_{ab} and G_b times X_a or X_b weighted by different parameters.⁴⁵

Let us define $J = (k, p, m)$ and $C(c(1) \in A, \dots, c(l) \in B, J)$. C is a *set of paths*, hereafter called a *chain*,⁴⁶ of length l which starts from A and ends at B , having k links from a type

⁴⁵The matrices sequence is multiplied by X_a or X_b depending on the last interaction matrix. For instance $G_a^2 G_{ab}$ is multiplied by X_b while $G_b G_{ba}$ is multiplied by X_a .

⁴⁶In this notation a *chain* includes all possible *paths* that have common features. For instance, all of *paths* starting from A and arriving to B are in the same *chain*.

B node to another type B node, p links from a type A node to another type A node, and m links between nodes of different types. The concept of *chain* is particularly useful in our context. Indeed, the product of adjacency matrices contains the same information of a chain. For instance $G_a \equiv C(c(1) \in A, c(2) \in A, k = 0, p = 1, m = 0)$ and $G_a G_{ab} \equiv C(c(1) \in A, c(2) \in A, c(3) \in B, k = 0, p = 1, m = 1)$. A similar characterization can be written for all combinations (products) of adjacency matrices considered in equation (25).

Taking advantage of this notation, the system in equation (25) can be characterized by the following two matrices

$$C = \left[\frac{C_a}{C_{ab}} \right] = \begin{array}{c} I_a \\ G_a \\ G_a^2 \\ \vdots \\ G_a^k \\ \vdots \\ G_{ab} G_{ba} \\ \vdots \\ G_{ab} G_b G_{ba} \\ \vdots \\ G_{ab} G_b^k G_{ba} \\ \vdots \\ G_a G_{ab} G_b G_{ba} \\ \vdots \\ G_a^l G_{ab} G_b^k G_{ba} \\ G_a G_{ab} G_b G_{ba} \\ \vdots \\ G_a^k G_{ab} G_b^l G_{ba} \\ \vdots \\ C(c(1) \equiv A, \dots, c(l) \equiv A, J) \end{array}, \Theta = \left[\frac{\Theta_a}{\Theta_{ab}} \right] = \begin{array}{c} d_1 & d_2 & d_5 & d_4 & d_3 \\ \hline 0 & 0 & 0 & 0 & 1 \\ \beta_a & 0 & 0 & 1 & 0 \\ \beta_a \phi_a + \gamma_a & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (\beta_a \phi_a + \gamma_a) \phi_a^{k-1} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \gamma_{ba} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \gamma_{ba} \phi_b & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \gamma_{ba} \phi_b^k & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_a \phi_{ba} \phi_{ab} \phi_b \phi_a & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_a \phi_{ba} \phi_{ab} \phi_b^k \phi_a^l & 0 & 0 & 0 & 0 \\ \beta_b \phi_{ba} \phi_{ab} \phi_b^k & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_b \phi_{ba} \phi_{ab} \phi_b^k & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(\theta, J) g(\Phi, J) & f(\theta, J) g(\Phi, J) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \beta_b & 1 & 0 & 0 \\ \gamma_{ab} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{ab} \phi_a^{k-1} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & (\gamma_b + \beta_b \phi_b) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & (\gamma_b + \beta_b \phi_b) \phi_b^{k-1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{ba} \phi_{ba} \phi_{ab} \phi_b & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{ba} \phi_a^{l-1} \phi_{ba} \phi_{ab} \phi_b^k & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \beta_b \phi_{ba} \phi_{ab} \phi_a \phi_b & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \beta_b \phi_{ba} \phi_{ab} \phi_a^l \phi_b^k & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(\theta, J) g(\Phi, J) & f(\theta, J) g(\Phi, J) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array},$$

where C represents different products of G_a , G_{ba} , G_{ab} and G_b (*chains*) appearing in the left side of equation (25) and Θ collects the relative coefficients. Θ has five columns which distinguish the elements that are multiplied by d_1, d_2, d_3, d_4 or d_5 .

The lower panel represents *chains* starting from A and arriving to B (labeled as C_{ab}), while the upper panel collects *chains* starting from A and coming back to A (labeled as C_a). The generic element of Θ is defined by the following objects, $\theta = (\theta_a, \theta_b)$, where $\theta_a = (\beta_a, \gamma_a, \gamma_{ab})$ and $\theta_b = (\beta_b, \gamma_b, \gamma_{ba})$; $\Phi = (\phi_{ba}, \phi_{ab}, \phi_b, \phi_a)$ and $f(\theta, J) = \beta_a I_{J, \beta_a}(\beta_a) + \gamma_a I_{J, \gamma_a}(\gamma_a) + \gamma_{ab} I_{J, \gamma_{ab}}(\gamma_{ab}) + \beta_b I_{J, \beta_b}(\beta_b) + \gamma_b I_{J, \gamma_b}(\gamma_b) + \gamma_{ba} I_{J, \gamma_{ba}}(\gamma_{ba})$ is a set of indicator functions that take value one if the argument appears in the corresponding chain and zero otherwise. The function $g(\Phi, J) = \prod_k \phi_b^k \prod_l \phi_a^l \prod_m (\phi_{ba} \phi_{ab})^m$ keeps track of the number of times the relative chain

passes from one type of node to another scaled by the respective interaction parameters. Observe that $(d_1, d_2, d_3, d_4, d_5)' \Theta' H_\infty = 0$ is equal to the condition $E(G_a y_a) d_1 + E(G_{ab} y_b) d_2 + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0$. The elements of H_∞ are equal to the elements of C multiplied by X_a or X_b depending on the last interaction matrix.⁴⁷

From C and Θ one can argue that the model is identified in the cases listed in Proposition 1.

Let us focus on case (2). For $E(Z_a)$ to have full rank, it suffices that Θ has full rank. This means that we need the linear independence of at least five *chains* (rows of C), translating to the linear independence of I_a , G_a , G_a^2 , G_{ab} and $G_{ab}G_b$.⁴⁸ The corresponding five rows of Θ are thus linear independent. Additionally we need to have five linear independent columns of Θ , so having $\beta_a \phi_a + \gamma_a \neq 0$ and $\beta_b \phi_b + \gamma_b \neq 0$ suffices to reach the full rank condition for Θ and consequently $E(Z_a)$. The same argument applies for case (3). \square

Relationship with *chains* and *trees*. In the proof of Proposition 1 we have established the equivalence between sequences of products of adjacency matrices and the concept of *chains*. In order to provide a better intuition behind the multiple sufficient conditions argument note that, according to the proof of Proposition 1 notation, a set of chains with a certain length p can be divided in g^{p+1} number of chains, where g is the number of node types. For instance, chains of length 1 can be classified in four categories when nodes are split into two types. following proof notation we can define $C(1) \equiv C(c(1)) \equiv a, c(2) \in A, 0, 1, 0 \cup C(c(1)) \in B, c(2) \in B, 1, 0, 0 \cup C(c(1)) \in A, c(2) \in B, 0, 0, 1 \cup C(c(1)) \in B, c(2) \equiv a, 0, 0, 1$ (e.g. $G_a \cup G_b \cup G_{ab} \cup G_{ba} = G$).

We can see this system of chains as a tree, more specifically as a *Tree-indexed Markov chain*. A tree is a graph with a distinguished vertex $x_0 \in g$ (here a type A node, the starting point) and the degree of each vertex is at least two (in our case the number of types, g). Its structure is basically determined by a countable set of states (in our case the number of types, g) characterized by a transition probability ($\{p(x, y) | x, y \in g\}$ in our case).⁴⁹

Let $T_a := \cup_{l,J} C(c(1) \in A, \cdot, c(l), J)$ (Figure 3), it is simply the collection of all possible chains of all possible lengths starting in a type A node. For identification purposes, we simply need that G_a , G_b , G_{ab} and G_{ba} are not empty (and not full).⁵⁰ In words, it means that there are no reasons why two randomly drawn nodes cannot be connected for each combination type (or that each node is connected with all of other nodes).⁵¹

⁴⁷Note that H_∞ is the IV matrix considered in Section 4, which is approximated by H_K in the feasible 2SLS estimation.

⁴⁸Note that here we need at least three chains from C_a and two from C_{ab} because we are considering the outcome equation for type A nodes, i.e. the staring point of chains is always a type A node.

⁴⁹Given that here we are not interested in determining the transition probability law of a chain, even if it is simple to estimate and is basically the link formation probability considered for all of the possible combinations of nodes' type. Benjamini and Peres (1994) give a detailed discussion on *Tree-indexed Markov chain*.

⁵⁰It is equivalent to say that the probability $0 < P(g_{ij} = 1) < 1$, $i \in A, B$ and $j \in A, B$. Note that transition probability can be derived from G_a , G_b , G_{ab} and G_{ba} . Here we are simply excluding the classical linear in mean framework (when the matrices are complete) and the case in which there are no connections (when the matrices are empty).

⁵¹From a Markov Chain perspective again, a more restrictive condition consists in assuming that the underlying Markov Chain is irreducible and aperiodic. This means that type A are connected with type B or type A with the same probability (and the same holds for type B). Thus, in this case tree branches with the same length have the same probability of being observed. The aperiodicity and irreducibility are not necessary for

Figure 3: Tree of Chains when type A and B nodes are considered (or have connections)

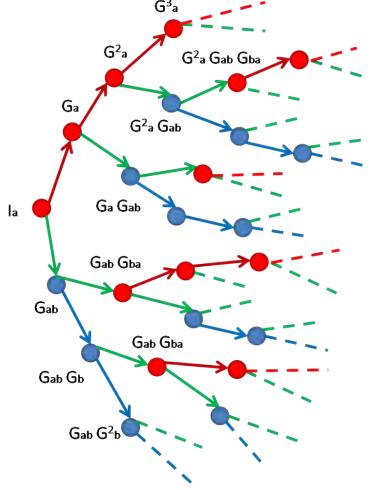
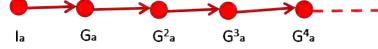


Figure 4: Non-tree of chains when only type A nodes are considered (or have connections)



An interesting feature of this framework is that, as case (3) tells us, even if I_a , G_a , and G_a^2 are linearly dependent, we can still identify ϕ_a and the other parameters relying on linear independence of chains passing through type B nodes.⁵² In other words, we can identify the parameters because of the multiple branches of the tree (see Figure 3).⁵³

Comparison with the identification conditions for *homogeneous* models. Let us conclude this discussion by further highlighting the connection between identification in a single group model and in a multiple group one. Let us reproduce a single group model by considering only type A nodes. The model is

$$y_a = \phi_a G_a y_a + \beta_a X_a + \gamma_a G_a X_a + \epsilon_a. \quad (26)$$

In order to obtain identification we want $(E(G_a y_a), X_a, G_a X_a)$ to have full rank. Given that $E(G_a y_a) = G_a(I - \phi_a G_a)^{-1}(\beta_a X_a + \gamma_a G_a X_a) = G_a \sum_{j=0}^{\infty} (\phi_a G_a)^j (\beta_a X_a + \gamma_a G_a X_a)$, the matrices used in the proof of Proposition 1 can be written in the following way

the identification condition to hold, but of course are sufficient.

⁵²Holding condition (3) instead of (2). We basically take advantage of linear independence of I_a , G_a and $G_{ab}G_{ba}$ instead of G_a^2 .

⁵³The additional parameter restrictions (conditions (2b, 3a or 3b) in Proposition 1) are basically due to an additional vector in the full rank condition (i.e. $E(G_{ab} y_b)$).

$$C = [C_a] = \begin{bmatrix} I_a \\ G_a \\ G_a^2 \\ \vdots \\ G_a^k \\ \vdots \end{bmatrix}, \Theta = [\Theta_a] = \begin{bmatrix} 0 & 0 & 1 \\ \beta_a & 1 & 0 \\ \beta_a \phi_a + \gamma_a & 0 & 0 \\ \vdots & \ddots & \ddots \\ (\beta_a \phi_a + \gamma_a) \phi_a^{k-1} & 0 & 0 \\ \vdots & \ddots & \ddots \end{bmatrix}.$$

As before, these two matrices respectively represent the chains and their coefficients. According to the proof of Proposition 1, the full rank condition for $(E(G_a y_a), X_a, G_a X_a)$ depends on C and Θ . From C and Θ one can argue that the model is identified if (see proof of Proposition 1 for details)

1. $\beta_a \phi_a + \gamma_a \neq 0$,
2. I_a, G_a and G_a^2 are linear independent.

These are exactly the conditions of Proposition 1 in Bramoullé et al. (2009). Note that if I_a, G_a and G_a^2 are linearly dependent, then G_a^k is also linearly dependent $\forall k$. Given that here we cannot differentiate nodes, we have $I_a \equiv C(0)$, $G_a \equiv C(1)$, and $G_a^2 \equiv C(2)$, where $C(k)$ represent the set of chains with length k . In terms of chains it means that $C(k) \equiv \{C(k-1), C(1)\} \equiv \{C(k-2), C(2)\} \equiv \dots \equiv \{C(2), C(k-2)\} \equiv \{C(1), C(k-1)\}$. In words it means that each chain's set can be represented by at least two sets of chains. So each G_a^k can be represented by the product of two matrices, $G_a G_a^{k-1}, G_a^2 G_a^{k-2}$, and so on. This is the connection to the linear independence of I_a, G_a , and G_a^2 as a condition for identification. In this case, a length l set of chains cannot be separated by node type, and thus T_a is composed only of one chain (Figure 4) instead of multiple chains (Figure 3).⁵⁴ Therefore, we need I_a, G_a , and G_a^2 to be linearly independent in order to have at least three independent chains in C and consequently identify the model's parameters satisfying the restriction $\beta_a \phi_a + \gamma_a \neq 0$.

Proof of proposition 2. By the classical expansion the estimator is

$\sqrt{n}(\hat{\mu} - \mu_0) = \frac{1}{n}(Z'_a P_K Z_a)^{-1} \frac{1}{\sqrt{n}} Z'_a P_K \epsilon_a$. As $Z_a = f_a + v_a$, by Lemma 1 (v), we have $\frac{1}{n}(Z'_a P_K Z_a) = F_a + o_p(1)$. By Lemma 2 (iii) $[Z'_a P_K \epsilon_a - \sigma_a^2 [e_1, e_2] [tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]]' / \sqrt{n} = f'_a \epsilon_a / \sqrt{n} + o_p(1) \xrightarrow{d} N(0, \sigma_a^2 F_a)$ by CLT.

Hence, the proposition is derived by Slutsky theorem

$$\frac{1}{n}(Z'_a P_K Z_a)^{-1} \frac{1}{\sqrt{n}} Z'_a P_K \epsilon \xrightarrow{d} F_a^{-1} \cdot N(0, \sigma_a^2 F_a) = N(0, \sigma_a^2 F_a^{-1}).$$

□

Proof of proposition 3. Given the proof of Proposition 2, it is sufficient to show that

$$\hat{\sigma}_a^2 [e_1, e_2] [tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]' / \sqrt{n} = o_p(1).$$

⁵⁴Borrowing from Markov chains vocabulary again, this is because the state that characterizes the chain is only one (A).

If we fix C_a , then by Lemma C.12 in Lee and Liu (2008) $\hat{M}_a - M_a = \hat{M}_a(\hat{\phi}_a - \phi_a)G_a + \hat{M}_a((\hat{\phi}_{ab}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})C_a)$. So we can write, $tr(\hat{\Psi}_a) - tr(\Psi_a) = tr(P_K(S_a(\hat{\phi}) - S_a)) = tr(P_K(G_a(\hat{M}_a - M_a)) = tr(P_K(G_a(\hat{M}_a(\hat{\phi}_a - \phi_a)G_a + \hat{M}_a((\hat{\phi}_{ab}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})C_a)) = (\hat{\phi}_a - \phi_a)tr(P_K(G_a(\hat{M}_aG_a)) + (\hat{\phi}_{ba}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})tr(P_K(G_a(\hat{M}_aC_a)))$.

Since the product of UB matrices is still UB (Kelejian and Prucha 1998), using the lemma B.2 (ii) in Lee and Liu (2010) and the initial \sqrt{n} consistency assumption, we obtain

$$\sqrt{n}(\hat{\phi}_a - \phi_a)tr(P_K(G_a(\hat{M}_aG_a)) + \sqrt{n}(\hat{\phi}_{ba}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})tr(P_K(G_a(\hat{M}_aC_a))) / n = o_p(1)O(K/n) + o_p(1)O(K/n) = o_P(K/n).$$

Finally, we have $\sqrt{n}(\hat{\sigma}_a^2 - \sigma_a^2)(tr(\hat{\Psi}_a) - tr(\Psi_a)) / n = o_p(1)o_p(K/n) = o_p(K/n) = o_P(1)$ as $K/n \rightarrow 0$. The same procedure can be applied for $\sqrt{n}(\hat{\sigma}_b^2 - \sigma_b^2)(tr(\hat{\Xi}_a) - tr(\Xi_a)) / n$ and for the second element of the stacked vector v . \square

Proof of proposition 4. Let p be a finite integer. Let us define the number of instruments equal to

$$K = \sum_{p=1}^P g^p + o(1),$$

so that we have $\sum_{n=1}^P g^p = O(g^p)$. Since we assume $K/n \rightarrow 0$, we have $\sum_{p=1}^P g^p = o(n)$ by assumption. This implies that $g^p = o(n)$. It follows that $g = o(n^{1/p})$. \square

Proof of proposition 5. We prove this proposition for $H_K^{a(p)}$. The same applies for $H_K^{b(p)}$. Let $\Theta_K^{(p)}$ be the matrix of true parameters derived from the p -order expansion of Θ (see Section 3). If $\sup\|\phi_a G_a\|_\infty + \sup\|\phi_{ab}\phi_{ba}C_a\|_\infty < 1$, then

$$H_K^{(p)}\Theta_K^{(p)} = G_a \sum_{j=0}^p (\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^j (\phi_{ab}G_{ab}J_b E(B)\delta_b + E(A)\delta_a).$$

It follows that $\|f_a - H_K^{(p)}\Theta_K^{(p)}\|_\infty = \|(\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^{p+1}S_a(\phi_{ab}G_{ab}J_b E(B)\delta_b + E(A)\delta_a)\|_\infty \leq \|(\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^{p+1}\|_\infty \|S_a\|_\infty \|\phi_{ab}G_{ab}J_b E(B)\delta_b + E(A)\delta_a\|_\infty = o(1)$ as $p \rightarrow \infty$. \square

Appendix D: Tables and Figures

Table 1: Monte Carlo Simulation: 1000 obs., 1000 replications

	(1)	(2)	(3)
10 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.100(0.032)	0.100(0.026)	0.099 (0.027)
$\phi_{ab} = 0.2$	0.201(0.031)	0.201(0.026)	0.197(0.069)
$\beta_a = 0.5$	0.501(0.047)	0.501(0.047)	0.501(0.048)
$\gamma_a = 0.5$	0.503(0.081)	0.502(0.078)	0.503(0.076)
$\gamma_{ab} = 0.5$	0.496(0.079)	0.496(0.075)	0.500(0.097)
20 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.098(0.025)	0.098(0.020)	0.100(0.020)
$\phi_{ab} = 0.2$	0.197(0.023)	0.195(0.019)	0.200(0.019)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.501(0.048)
$\gamma_a = 0.5$	0.506(0.096)	0.506(0.093)	0.503(0.093)
$\gamma_{ab} = 0.5$	0.500(0.097)	0.497(0.093)	0.496(0.093)
30 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.020)	0.098(0.016)	0.099(0.016)
$\phi_{ab} = 0.2$	0.198(0.019)	0.195(0.015)	0.199(0.015)
$\beta_a = 0.5$	0.500(0.048)	0.501(0.047)	0.501(0.047)
$\gamma_a = 0.5$	0.506(0.110)	0.507(0.107)	0.505(0.107)
$\gamma_{ab} = 0.5$	0.500(0.112)	0.498(0.109)	0.497(0.109)

Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.2$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$

Table 2: Monte Carlo Simulation: 1000 obs., 1000 replications

	(1)	(2)	(3)
20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.040)	0.095(0.032)	0.097(0.032)
$\phi_{ab} = 0.1$	0.101(0.038)	0.101(0.031)	0.099(0.031)
$\beta_a = 0.5$	0.501(0.047)	0.505(0.047)	0.501(0.047)
$\gamma_a = 0.5$	0.504(0.084)	0.506(0.081)	0.506(0.080)
$\gamma_{ab} = 0.5$	0.497(0.083)	0.497(0.078)	0.498(0.078)
Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.1$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$			
20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.097(0.029)	0.097(0.022)	0.103(0.022)
$\phi_{ab} = 0.3$	0.298(0.027)	0.298(0.021)	0.302(0.021)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.501(0.048)
$\gamma_a = 0.5$	0.506(0.077)	0.506(0.075)	0.501(0.075)
$\gamma_{ab} = 0.5$	0.500(0.076)	0.500(0.072)	0.496(0.072)
Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.3$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$			
20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.016)	0.097(0.010)	0.091(0.452)
$\phi_{ab} = 0.4$	0.390(0.012)	0.370(0.008)	0.401(0.064)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.502(0.079)
$\gamma_a = 0.5$	0.505(0.074)	0.504(0.073)	0.502(0.073)
$\gamma_{ab} = 0.5$	0.498(0.070)	0.478(0.069)	0.498(0.582)
Note: y_b is generated with $\phi_b = 0.2$, $\phi_{ba} = 0.05$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$			

Table 3: Monte Carlo Simulation: 1000 obs., 1000 replications, 10 max connections, $\phi_b = \phi_{ab} = \phi_{ba} = 0.3$, $\beta_b = 0.5$, $\gamma_b = \gamma_{ab} = \gamma_{ba} = 0.5$

	2SLS/Misspecified model	2SLS/misspecified IVs	2SLS/correct model-correct IVs
$\phi_a = 0.3$	0.3684 (0.0335)	0.2971 (0.0465)	0.2999 (0.0164)
$\beta_a = 0.5$	0.3856 (0.1804)	0.4865 (0.2233)	0.5097 (0.1489)
$\gamma_a = 0.5$	-0.0016 (0.2315)	0.4929 (0.2102)	0.4963 (0.1765)

Table 4: Monte Carlo Simulation: 1000 obs., 1000 replications, 10 max connections

	(1) $\phi_a = \phi_b = \phi_{ba} = \phi_{ab} = 0.1$	(2) $\phi_a = \phi_b = 0.1, \phi_{ba} = \phi_{ab} = 0.3$	(3) $\phi_a = 0.1, \phi_b = 0.2, \phi_{ba} = 0.05, \phi_{ab} = 0.4$
ϕ	0.100(0.020)	0.178(0.021)	0.205(0.022)
β	0.500(0.031)	0.499(0.034)	0.499(0.033)
γ	0.446(0.022)	0.445(0.024)	0.442(0.025)

Figure 5: Policy experiment: varying ϕ_a and ϕ_{ab}

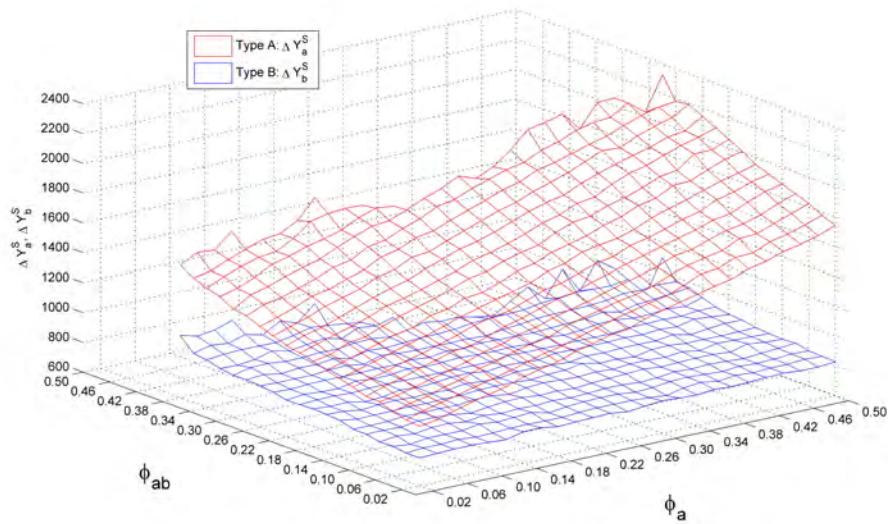


Figure 6: Policy experiment: varying ϕ_a and ϕ_{ba}

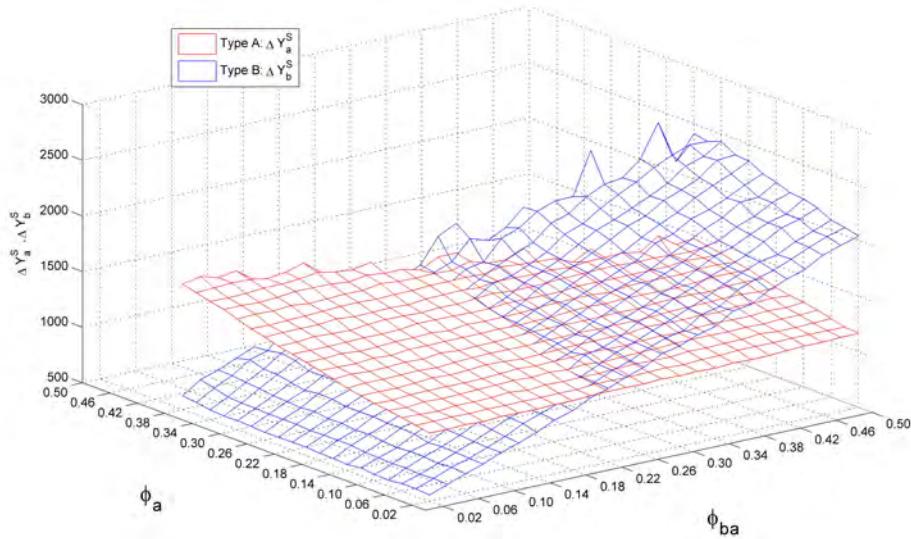


Figure 7: Kernel density estimation of empirical distributions of $\Delta y_{i \in A}$ and $\Delta y_{i \in B}$, increasing ϕ_a .

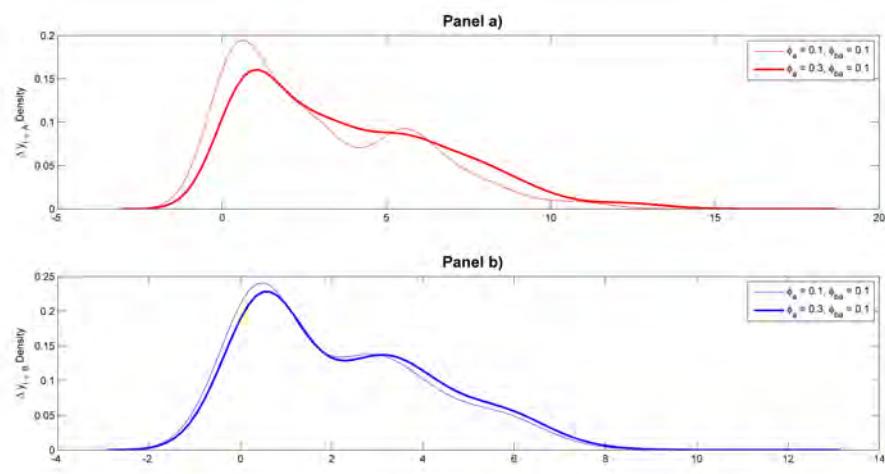


Figure 8: Kernel density estimation of empirical distributions of $\Delta y_{i \in A}$ and $\Delta y_{i \in B}$, increasing ϕ_{ba} .

