Estimating Network Effects without Network Data

Pedro CL Souza*

November 2014

Job Market Paper

Abstract

Empirical research on social and economic networks has been constrained by the limited availability of data regarding such networks. This paper develops a method that does not rely on network data to estimate network effects. The proposed method also estimates the probability that pairs of individuals form connections, which may depend on exogenous factors such as common gender. The method may incorporate imperfect network data, such as with self-reported data, with the dual purpose of refining the estimates and testing whether the reported connections positively affect the probability that a link is formed. To achieve those goals, I derive a maximum likelihood estimator for network effects that is not conditional on network observation. Networks are treated as a source of unobserved heterogeneity and eliminated based on data collected from observing many groups. This is accomplished with recourse to a spatial econometric model with unobserved and stochastic networks. I then apply the model to estimate network effects in the context of a program evaluation. I demonstrate theoretically and empirically that including network effects has important implications for policy assessments.

Keywords: social networks, spillovers, spatial econometrics.
JEL Codes: C21, C49, O12, D85.

*Department of Economics and STICERD, London School of Economics. I am very grateful to Javier Hidalgo, Steve Pischke, Robin Burgess and Oriana Bandiera for extensive support and feedback on this project. I also thank Clare Balboni, Francisco Costa, Jan-Emmanuel De Neve, Samuel Marden, Priscilla Negreiros, Taisuke Otsu, João Pessoa, Markus Riegler, Munir Squires, Dimitri Szerman, and participants at Econometric Society European Winter Meeting 2014, 61st North American Meeting of the Regional Science Association International 2014, LSE Econometrics, Development, Labour and Summer seminars for extremely helpful comments. Contact: p.souza@lse.ac.uk.

1 Introduction

Personal interconnectedness is an important and pervasive feature of human life. Social and economic networks enhance learning in classrooms (Angrist and Lang, 2004; Ammermuller and Pischke, 2009), influence decisions regarding technology adoption (Foster and Rosenzweig, 1995; Conley and Udry, 2010) and serve as mechanisms for informal contractual enforcement (Ambrus et al., 2014). In recent years, the many ways in which social networks affect choices and behavior have been the subject of extensive research (Jackson, 2010). However, incorporating these mechanisms in applied research remains challenging because of the limited availability of network data. Even when networks are able to be observed, these observations are often imperfect, such as when data are self-reported or subject to measurement errors.

This paper develops a method for estimating network effects when network data are either unobserved or imperfectly observed. The method does not rely on network data and derives network effects using only individuals' dependent and explanatory variables data. I specifically propose an estimator that accomplishes three objectives. First, I estimate network spillovers – the difference between expected outcomes when networks are and are not relevant – without network data.[1] Spillovers also capture the extent to which social networks amplify the effect of explanatory variables on outcomes (Miguel and Kremer, 2004). Second, I illuminate structural mechanisms that give rise to network spillovers. I separately identify and estimate Manski's (1993) endogenous effects (the dependence of one's own choices on the choices of others) from exogenous effects (the dependence of one's own choices on the exogenous variables of others), controlling for correlated effects (the similarity of peers in terms of unobservable characteristics).[2] The method also estimates and predicts the probability that pairs of individuals form a connection, which is allowed to depend on exogenous factors such as common gender. Third, I incorporate imperfect network data, such as self-reported network data, with the dual purpose of refining the estimates and providing a test for whether reported connections positively affect the probability that a connection is formed. Rejection of the null demonstrates self-reported network data validity.

To achieve these goals, I propose a spatial econometric model with unobserved and stochastic networks that is coupled with a model for random network formation. I derive a likelihood for the model which is not conditional on network. This likelihood is equivalent to integrating the likelihood conditonal on observing the true network with respect to the probability density function of the stochastic network. Observation of data on individuals' outcomes and explanatory variables in many self-contained groups, such as classrooms in a school, then provides the identifying condition to estimate the model that serves as a substitute for network observation. In essence, networks are treated as a source of unobserved heterogeneity. I allow for time and fixed effects at the individual or group level when panel data are available and when networks are invariant over time.

The estimator for network spillovers is consistent and asymptotically normally distributed under weak identification assumptions because in this case it is not necessary to separately identify endogenous and exogenous effects. In other words, the parameters of the model are identified up to a set and, as I will show, the network

[1] This is also important because OLS estimates are often inconsistent for individual reaction parameters when networks are irrelevant if network spillovers are not included in the regression, and the size of inconsistency depends on the unobserved network.

[2] Endogenous effects are the autoregressive component of a spatial model. Exogenous effects is exogenous component of a spatial model. Correlated effects are captured by fixed effects at the individual level. These are precisely defined with recourse to the model in Section 2. The reflection problem is solved if there are asymmetries in the expected network (Kelejian and Prucha (1998), Bramoullé et al. (2009) and De Giorgi et al. (2010) explore similar assumptions when networks are observed) or observation of groups with distinct sizes is available (see also Lee, 2007).

spillovers are constant if evaluated at a parameter that belongs to the identified set. Consistency and confidence regions for the structural parameters are provided making use of the set identification framework.[3] To provide point identification for structural parameters of the model, I explore the difference between observed second moments of the data and those implied by the model. I utilize the fact that the presence of social interactions creates dispersion in average outcomes across groups that cannot be explained by independent variables or peer group heterogeneity alone. Such "excess" variance is explored to build an additional moment restriction and to solve a Generalized Method of Moments (GMM) problem which also includes the score conditions implied by the maximizing the likelihood. This completes the requirements for point identification and consistent estimatation of the structural parameters of the model.[4]

To illustrate how this method can be applied in practice, I employ the estimator developed herein to investigate treatment effects both on treated and their peers in a setting potentially conducive to spillovers. The randomized intervention of Bandiera et al. (2013)[5] studies the effect on the treated of the provision of livestock and training to low-income households in Bangladesh and finds that the lack of capital and skills is a strong determinant of the occupational choices of the poor. Targeted households begin new livestock-rearing businesses, increase self-employment hours and reduce wage hours. Due to village-level randomization, a large portion of the individuals in the selected villages are treated, which raises the possibility that network effects are important in determining these outcomes, particularly for peers of those who are treated.

Without using network data, I first demonstrate that network spillovers are economically and statistically significant in determining certain outcomes, especially food expenditure and food security. In these cases, spillovers amount to half of the original treatment for both treated households and their peers. Spillovers of occupational choice and livestock are either insignificant or of a small magnitude. To analyze the structural mechanisms that lead to these results, I then decompose spillovers into exogenous and exogenous effects. I demonstrate that, regarding occupational choice and assets, a marginal connection to a treated household has an effect in opposite direction to the effect on the treated: an additional connection decreases self-employment hours, increases wage hours and decreases livestock value.[6] On the other hand, a marginal connection to the treated increases food per capita expenditure and food security to a significant extent. These results are consistent with the phenomenon in which peers of treated households partially fill the vacancies left by those who begin new livestock-rearing businesses and suggests a specialization at the village level, where treated households gain comparative advantage in livestock rearing. Estimating the network structure also demonstrates that network densities are fairly low in the majority of cases, suggesting local interactions via personal contacts as opposed to changes in prices in village-level markets. Finally, inclusion of self-reported network data indicates that family links convey meaningful interactions between households, whereas economic (i.e., non-family) links are much less capable of explaining these social dynamics. This result thus reinforces the idea that families are natural *loci* for sharing information and conducting business.

The methods developed in this paper contribute to the spatial econometrics literature, which has to date

[3] Chernozhukov et al. (2007), Bugni (2010) and Romano and Shaikh (2010).

[4] Graham (2008) uses a similar idea in the context where networks are observed, within the linear-in-means model.

[5] I thank the authors for sharing data.

[6] The magnitudes of the estimates imply that peers of treated households compensate around 25-30% of the reduction in treated households' wage hours due to exogenous effects. Endogenous effects move in opposite direction reducing the size of the overall spillover effects. Additional details can be found in Section 5.

considered estimation only when networks are observed, non-stochastic and measured without error. The role of randomness in network formation has also received scant attention in spatial models, despite its importance in social networks (Diestel, 2010). The dependence of existing methods on acquiring knowledge of true networks has been stressed as a limitation of the previous literature (Anselin, 2010; Plümper and Neumayer, 2010).[7] Representative papers in the spatial econometrics literature include those by Anselin (1988) and Kelejian and Prucha (1998, 1999, 2001, 2010). Lee (2004, 2007) and Lee et al. (2010) also consider a maximum likelihood estimator. The case in which networks are not observed is explored by Lam and Souza (2013a, b, 2014) and Manresa (2013), who consider the estimation of networks when one group is observed for many periods of time and, as a consequence, clearly suit different applications. It is useful to highlight that the latter papers estimate networks as a collection of pairwise links. In contrast, the current paper is concerned with the probability that a link is formed and the role of exogenous factors therein. The identification results reported by Manski (1993), Graham (2008), Bramoullé et al. (2009) and De Giorgi et al. (2010) are also derived under the assumption that networks are observed. In another strand of the literature, stochastic network formation models, such as those described by Holland and Leinhardt (1981), Frank and Strauss (1986) and Strauss and Ikeda (1990), also consider the estimation of network structure only when network observations are available.

Beyond its contribution to the spatial econometric literature, this paper provides a method for systematically investigating network effects, with potential applications in many fields, such as peer effects in education (Sacerdote, 2001; Angrist and Lang, 2004; Ammermuller and Pischke, 2009; Bramoullé et al., 2009; De Giorgi et al., 2010), information diffusion and technology adoption (Foster and Rosenzweig, 1995; Bandiera and Rasul, 2006; Conley and Udry, 2010), social networks and labor outcomes (Rees, 1966; Granovetter, 1973; Montgomery (1991); Conley and Topa, 2002; Munshi, 2003; Pellizzari, 2004; Calvó-Armengol and Jackson, 2004) and crime and delinquent behavior (Glaeser et al., 1996; Dell, 2012). In the macroeconomic and trade literature, these methods can be used to study networks as sources of aggregate fluctuations (Acemoglu et al., 2012) and to estimate parameters of gravity equations (Anderson and van Wincoop, 2003). These approaches are particularly relevant when obtaining data on networks is difficult, time-consuming or expensive, which frequently occurs with social network data because reported links are frequently subjective and prone to behavioral biases.

The remainder of the paper is structured as follows. In Section 2, I introduce the model, define network spillovers and illustrate the inconsistencies that arise when networks are not accounted for. In Section 3, I present the estimator for network effects in the absence of network data and explore its asymptotic properties. Section 4 provides a simulation to validate the performance of the estimator in small samples. Section 5 compares the methods in this paper with existing alternatives for estimating spillovers. It also provides an application to treatment spillovers based on the study of Bandiera et al. (2013). Section 6 concludes.

[7] Plümper and Neumayer (2010) show that misspecification of the networks causes serious bias in parameters of the model, which should be a particular concern for the study of social interactions, where these issues frequently appear. Another facet of the same problem emerges in estimation techniques that proposes using peers of peers' exogenous variables as instruments for one's own endogenous variable, such as Kelejian and Prucha (1998, 1999), Bramoullé et al. (2009) and De Giorgi et al. (2010). To the extent that network data suffers from measurement errors, one risks violating relevance or validity assumptions without awareness.

2 Model

The model consists of two parts: a model for stochastic network formation and, given a network, a spatial econometric model that connects explanatory variables to outcomes. The former is sufficiently flexible to allow the probability link formation to depend on exogenous characteristics, such as sharing race or gender or the distance between households.[8] This model may also incorporate individual-level characteristics that attract links or, conversely, that make an individual more inclined to form links with others. In this Section, I assume a simple Bernoulli model for network formation; a full account is provided in Appendix B.[9] Given a network, the spatial econometric model has been extensively considered in the literature, such as by Anselin (1988), Lee (2004), Bramoullé et al. (2009), Lee et al. (2010) and De Giorgi et al. (2010); however, in contrast to previous papers, I consider the estimation of network effects in the absence of network data.

I assume that data are available for groups $j = 1, \ldots, v$ and individuals $i = 1, \ldots, n_j$. Individuals interact within groups with observed boundaries, but data with respect to networks within groups are not available. For example, information is available on classes that students belong to but information regarding intra-classroom networks is not available; households are known to be located in villages, but the researcher does not have information regarding the pattern of interaction between households.

For each group j, a network is described with a *directed graph* G_j, an unordered collection of ordered pairs of individuals among n_j individuals. This set lists links along with their associated directions: $\{i, k\} \in G_j$ implies individual i affects individual k in group j. For example, if individual 1 affects 2, 2 affects 3 and 3 affects 2, then $G_j = \{\{1, 2\}, \{2, 3\}, \{3, 2\}\}$. As noted by Wasserman and Faust (1994, Ch. 4), Diestel (2010, Ch. 1), Jackson (2010, Ch. 2), Ballobás (2013, Ch. 1) and others, this representation is quite general. For example, Figure 1 portrays estimated links between United States senators, as described by Lam and Souza (2014), based on their 2013 voting records. It is also convenient to express the graph with a so-called *neighboring* or *spatial matrix* W_j, of $n_j \times n_j$ dimensions, a representation of G_j with $\{W_j\}_{ik} = 1$ if $\{i, k\} \in G_j$ and $\{W_j\}_{ik} = 0$ otherwise. It is assumed that no individual affects him or herself; thus $\{W_j\}_{ii} = 0$, for all $i \in \{1, \ldots, n_j\}$.[10]

Network formation is random with a probability law, indexed by parameters of interest θ_g. I use a simple model for clarity of explanation only. Suppose a link between individuals is formed with probability δ_1 when the pair shares a characteristic and δ_0 otherwise. To write the probability distribution function, allow $n_j \times n_j$ matrix Q_j to register the commonality of this individual characteristic. If i and k have the same gender, for example, let the elements of the matrix $\{Q_j\}_{ik} = \{Q_j\}_{ki} = 1$ and zero otherwise. Matrix Q_j could also capture if i self-reported a connection with k. In these cases, $P\{\{W_j\}_{ik} = 1 | \{Q_j\}_{ik}\} = \delta_0(1 - \{Q_j\}_{ik}) + \delta_1\{Q_j\}_{ik}$. The vector of parameters of interest, carried to estimation, is $\theta_g = (\delta_1, \delta_0)'$. Under the assumptions that link formation is homogenous and independent across pairs of individuals, the probability distribution function is

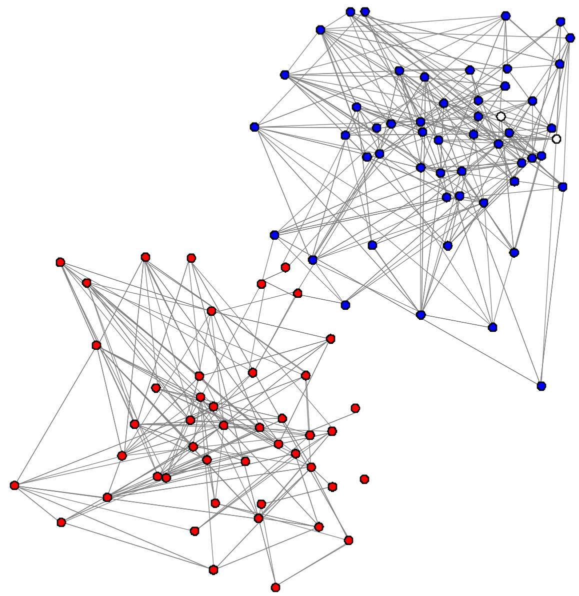
$$P\{W_j = w_j | Q_j\} = \prod_{i,k<n_j} (\delta_1^{\{Q_j\}_{ik}} \delta_0^{1-\{Q_j\}_{ik}})^{\{W_j\}_{ik}} ((1-\delta_1)^{\{Q_j\}_{ik}} (1-\delta_0)^{1-\{Q_j\}_{ik}})^{1-\{W_j\}_{ik}}. \quad (1)$$

[8] The model also falls into the Exponential Random Markovian Graphs category. See Holland and Leinhardt (1981), Frank and Strauss (1986) and Strauss and Ikeda (1990).

[9] See also Wasserman and Faust (1994) and Jackson (2010).

[10] G_j and W_j are arrays which depend on the group sizes n_j. In order to keep notation concise, I adopt $G_j \equiv G_{n_j,j}$ and $W_j \equiv W_{n_j,j}$.

Figure 1: Graph example from Lam and Souza (2014).



Note: Red nodes are Senators that belong to the Republican party, blue are Democrats and white are independents.

Model (1) is a simple but arguably truthful representation of situations where differential patterns of associations dominates coalition or strategic behavior, cases in which independence of link formation is violated. A classroom divided along gender or racial lines is possibly an example that satisfies assumption above.

Given a network, it remains to describe a model linking explanatory variables to outcomes. Denote W_j^0 and M_j^0 as two *random* and *unobserved* realizations of a network-generating process, such as the one introduced above. This network is embedded is a spatial econometric model, which incorporates dependence of one's own outcome variable on others' outcome variables and others' exogenous variables. For a particular group $j = 1, \ldots, v$ composed of n_j individuals, the model is given by

$$y_j \quad = \quad \lambda_0 W_j^0 y_j + x_j \beta_{10} + W_j^0 x_j \beta_{20} + v_j \tag{2}$$

where y_j is a column vector of dimension $n_j \times 1$, x_j is $n_j \times k$, and v_j is the $n_j \times 1$ disturbance vector. Disturbance term v_j is assumed to follow a structure that allows for spatial dependence, $v_j = \rho_0 M_j^0 v_j + \epsilon_j$, where ϵ_j is $n_j \times 1$, independent and normally distributed with variance σ_0^2. As a particular example, this includes group-level clustering and heteroskedasticity that arises from heterogeneous exposure to disturbances of others.

In Manski's (1993) taxonomy, the term $W_j^0 y_j$ corresponds to the *endogenous effects*, or the dependence of one's own behavior on the behavior of others through link strength scalar parameter λ_0. Parameter β_1, of dimension $k \times 1$, captures the direct effect of one's own exogenous variables on one's own dependent variables. Parameter β_2, of the same dimension, describes the effects of others' exogenous variables on one's own dependent variable. Thus, $W_j^0 x_j$ is denoted as *contextual* or *exogenous effects. Correlated effects* are represented by the error $v_j = \rho_0 M_j^0 v_j + \epsilon_j$ and fixed effects, which I describe in Section 3.4. This model is similar to the model in Bramoullé et al. (2009) and Lee et al. (2010), among other studies, and is known as the "mixed regressivespatial autoregressive model" in the spatial econometrics literature (Anselin, 1988). I am then interested in the estimation of usual spatial parameters $\theta_s = (\lambda_0, \beta_{10}', \beta_{20}', \rho_0, \sigma_0^2)'$ and θ_g. Hence, the complete set of structural parameters of interest is $\theta = (\theta_s', \theta_g')'$.

Dependence of one's own outcomes on other's outcomes and exogenous variables often means that the overall response to exogenous variation exceeds β_{10}. As a consequence, to the extent that individual network spillovers depend on one's own exogenous variation, estimators for β_{10} that do not account for network spillovers are frequently inconsistent, as I demonstrate immediately below.

Using the series decomposition[11] $(I_{n_j} - \lambda_0 W_j^0)^{-1} = \sum_{s=0}^{\infty} \lambda_0^s (W_j^0)^s$ to obtain the reduced-form model, the expected outcomes are separated into two components: the individual reaction or elasticity with respect to x_j and its effect through the network,

$$\mathbb{E}y_j \quad = \quad x_j\beta_{10} + W_j^0 x_j \beta_{20} + \sum_{s=1}^{\infty} \left(\lambda_0 W_j^0\right)^s \left(x_j\beta_{10} + W_j^0 x_j \beta_{20}\right). \tag{3}$$

The term $x_j\beta_{10}$ is understood as the individual-level elasticity with respect to x_j if networks were irrelevant, whereas the second and third terms jointly denote network spillovers, the additional effect on the mean exclusively due to individual interconnectedness:

$$\varphi\left(x_j, \theta_0\right) \quad \equiv \quad W_j^0 x_j \beta_{20} + \sum_{s=1}^{\infty} \left(\lambda_0 W_j^0\right)^s \left(x_j\beta_{10} + W_j^0 x_j \beta_{20}\right) \;=\; \sum_{s=1}^{\infty} \lambda_0^{s-1} \left(W_j^0\right)^s x_j \left(\lambda_0 \beta_{10} + \beta_{20}\right). \tag{4}$$

Clearly, if $\lambda_0 = 0$ and $\beta_{20} = 0_{k\times 1}$, or $\delta_1 = \delta_0 = 0$, then $\varphi\left(x_j, \theta_0\right) = 0$. Spillover $\varphi(x_j, \theta_0)$ is a $n_j \times 1$ vector because each individual accrues his or her own spillover.

Separate identification of the individual reaction and network spillovers is relevant in at least two scenarios. Provided that the ultimate goal is to consistently estimate β_{10}, $\varphi\left(x_j, \theta_0\right)$ is a confounding factor. As shown in Subsection 2.1, when networks are unaccounted for, consistent estimating β_{10} requires an underlying network structure such that one's own network spillovers are independent of one's own exogenous variation, a condition that breaks down in simple counterexamples.

Moreover, network spillovers are of interest in their own right, as shown by the plethora of examples in the literature. Glaeser et al. (1996) argue that social interactions explain petty criminal behavior very well, but are also of moderate importance in explaining more serious offenses. Hence, crime prevention policies have indirects effects by reducing of others' proclivity toward criminal activity, and the effect's magnitude then shapes and informs the public policy debate. In another example, Foster and Rosenzweig (1995) reason that farmers' decision to adopt high-yielding seed varieties depends on other farmers' decisions regarding adoption and their accrued profit; consequently, a single farmer's adoption decision multiplies itself by inducing others to adopt also. Finally, note that parameter $\varphi\left(x_j, \theta\right)$ can be explored to optimize treatment effects under a given budget of resources. To the extent that network spillovers are prevalent and positive, often average treatment effects can frequently be maximized by concentrating treatment in fewer groups.

Remark 1. Panel or spatiotemporal models can be naturally introduced from equation (2). Index explanatory

[11] Conditions for existence of this decomposition are derived in Section 3.

variables and outcomes at time $t = 1, \ldots, T$ and the complete model reads

$$y_{jt} = \lambda_0 W_j^0 y_{jt} + x_{jt}\beta_{10} + W_j^0 x_{jt}\beta_{20} + \alpha_j + \gamma_t + v_{jt} \quad (5)$$

where α_j is a vector of $n_j \times 1$ time-invariant coefficients (but allowed to vary at the group or individual levels), which are also denoted, folllowing Manski (1993), as *correlated effects.* The vector γ_t represents time effects. Under the invariance of networks with respect to time, I propose a data transformation that eliminates these nuisance parameters in Subsection 3.4. When x_{jt} is a treatment indicator, model (5) can be described as a differences-in-differences estimator supplemented with a network component. In the absence of network effects ($\lambda_0 = 0$ and $\beta_{20} = 0_{k\times 1}$), the model is reduced to a standard differences-in-differences. In this context, the terms $\lambda_0 W_j^0$ and $W_j^0 x_{jt}\beta_{20}$ measure the treatment spillovers through the network. □

2.1 Inconsistency when Networks are Unaccounted for

Equations (3) and (4) immediately imply that the aggregate group response to a shock is the sum of one's own variation in the absence of networks (β_{10}) and network spillovers (φ),

$$y_j = x_j\beta_{10} + \varphi(x_j, \theta_0) + \epsilon_j. \quad (6)$$

On the one hand, disentangling the two components provides insights into the mechanisms that determine the responses to the shock. In particular, the role of networks is separated from the response in its absence; this construct is useful for example to provide external validity to randomized controlled trials prior to reimplementation in settings in which networks might differ. On the other hand, the omission of $\varphi(x_j, \theta_0)$ biases OLS estimates when one's own spillover is not orthogonal to one's own shock.

Consistency for β_{10} requires that $\mathbb{E}(\varphi(x_j, \theta_0)|x_j) = 0$ for all $i = 1, \ldots, n_j$, the case in which the researcher would be oblivious to network spillovers. At the other extreme, only under perfect correlation between x_j and $\varphi(x_j, \theta_0)$ the OLS estimates are consistent for the sum of β_{10} and full spillovers. In general, however, independence is not generally attained, failing in particular under reciprocated networks or correlation between x_{ij} and x_{kj} for $i \neq k$[12]. In this case, the biasing term $(x_j'x_j)^{-1}x_j'\mathbb{E}(\varphi(x_j, \theta_0)|x_j)$ depends on the network structure, which is unknown; thus, the size and presence of bias are also unknown. I now provide some examples.

Example 1. *(Classrooms and the linear-in-means model).* Manski (1993) proposes the linear-in-means network model in which individuals interact with all others in a given classroom and

$$W_j^0 = \begin{bmatrix} 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \end{bmatrix} = \frac{1}{n-1}\iota_n\iota_n' - \frac{1}{n-1}I_n$$

[12]This type of violation would occur in the case in which individuals who are eligible for a treatment are also more likely to have other eligible individuals in their social networks. Snowballing a treatment is another clear example of violation of the no self-spillover condition $\mathbb{E}(\varphi(x_j, \theta_0)|x_j) = 0$.

where I_n is the $n \times n$ identity matrix and ι_n is the $n \times 1$ vector of ones. Suppose x_j is a treatment dummy and α is the proportion of the individuals in the group that were treated. The expectation of response conditional on treatment is obtained via the reduced-form model

$$y_j = \left(S_j^0\right)^{-1} x_j\beta_{10} + \left(S_j^0\right)^{-1} W_j^0 x_j \beta_{20} + \left(S_j^0\right)^{-1} \left(R_j^0\right)^{-1} \epsilon_j$$

where $S_j = I_n - \lambda_0 W_j^0$, $(S_j^0)^{-1} = \frac{n-1}{n-1+\lambda_0} I_n + \frac{\lambda_0}{(n-1+\lambda_0)(1-\lambda_0)} \iota_n \iota_n'$ and $(S_j^0)^{-1} W_j^0 = -\frac{1}{n-1+\lambda_0} I_n + \frac{1+\lambda_0}{(n-1+\lambda_0)(1-\lambda_0)} \iota_n \iota_n'$. The expectation of the outcome of individual i in group j, conditional on not receiving a treatment, is

$$\mathbb{E}\left[y_{ij} \middle| x_{ij} = 0\right] = \alpha n \frac{\lambda_0 \beta_{10} + (1+\lambda_0)\beta_{20}}{(n-1+\lambda_0)(1-\lambda_0)}$$

and describes the network spillovers to untreated individuals. Conditioned on receiving a treatment,

$$\mathbb{E}\left[y_{ij} \middle| x_{ij} = 1\right] = \frac{(n-1)\beta_{10} - \beta_{20}}{n-1+\lambda_0} + \alpha n \frac{\lambda_0 \beta_{10} + (1+\lambda_0)\beta_{20}}{(n-1+\lambda_0)(1-\lambda_0)} \tag{7}$$

thus, in general, the population difference $\mathbb{E}[y_{ij}|x_{ij} = 1] - \mathbb{E}[y_{ij}|x_{ij} = 0]$ is approximately β_{10} for a typical classroom size, such as $n = 25$. This result implies that OLS estimates are consistent for β_{10} and oblivious to network spillovers. □

Example 2. *(Households and local interaction).* Households typically interact with few others, and relations are generally reciprocated. For the sake of example, suppose a network is consists of isolated subgroups of five households, in which interaction across subgroups is negligible in comparison with interactions within. In this setting, W_j^0 is a block-diagonal matrix with $\frac{n}{5}$ blocks[13], or $W_j^0 = I_{\frac{n}{5}} \otimes \left(\frac{1}{4}\iota_5\iota_5' - \frac{1}{4}I_5\right)$. Suppose a proportion α receive a treatment. In contrast to the previous example, the difference $\mathbb{E}[y_{ij}|x_{ij} = 1] - \mathbb{E}[y_{ij}|x_{ij} = 0]$ is no longer approximately β_{10}, which can be shown by replacing $n = 5$ in equation (7). As a consequence, OLS estimates are biased for β_{10} and capture the portion of one's own spillovers that correlate with one's own treatment status. □

Generally, OLS is only consistent for β_{10} in particular network structures. When networks remain unobserved, the implementation of such a strategy depends on hypotheses that rule out feedback mechanisms. In Section 3, I provide a method for consistently estimating $\varphi(x_j, \theta)$ under few identifying assumptions that address both motivating elements. The method is based on a maximum likelihood integrated with respect to unobserved networks, resulting in a likelihood that is independent of network observation. In essence, I deal with the networks as unobserved heterogeneity. As will be shown, although the point identification of θ is not obtained without additional assumptions, spillover $\varphi(x, \theta)$ is constant within the identified set and thus point-identified. Section (3.3) uses additional identifying information to sort through the identified set and reestablish point identification for the structural parameters.

[13] For simplicity, assume n is a multiple of 5.

3 Estimation of Network Effects

Spatial econometric models dealt with the case of known W_0 and M_0. Under certain conditions, including network observation, Lee (2004) and Lee et al. (2010) show consistency and asymptotic normality of a quasi-maximum likelihood estimator for θ_s. In this scenario, accounting for network effects would not pose a challenge. However, these results are of no use if W_0 and M_0 are unobserved or imperfectly observed, such as when there are measurement errors[14] or data are self-reported.

In contrast, I deal with networks as a form of unobserved heterogeneity. Networks are randomly formed with certain probability law, homogenous across groups, and observation of many groups is available. More formally, I propose an integrated likelihood approach. The likelihood unconditional on network observation is the integral of the likelihood given a network (from a spatial model soon introduced) with respect to the probability density function for a stochastic network model:

$$\ln \mathcal{L}\left(\theta | y_n, x_n, Q_n\right) = \int \ln \mathcal{L}\left(\theta | y_n, x_n, W_n, M_n\right) dP\left(W_n, M_n | Q_n, x_n, \theta\right) \tag{8}$$

where $y_n = (y_1', \ldots, y_v')'$, $x_n = (x_1', \ldots, x_v')'$, W_n and M_n are a random block matrix with $W_1, \ldots, W_v$ and $M_1, \ldots, M_v$ along the main diagonal. Therefore W_n and M_n have dimension $n \times n$, $n = \sum_{j=1}^{v} n_j$. Likelihood $\ln \mathcal{L}\left(\theta | y_n, x_n, W_n, M_n\right)$ is derived from a spatial model and for simplicity it is assumed independent of Q_n. The probability density function of networks, $P(W_n, M_n | Q_n, x_n, \theta)$, depends on exogenous variables Q_n and x_n and parameters θ. In this way, the probability that peers form a link is affected by individual characteristics Q_n which do not directly affect the mean and exogenous variables x_n. For example, connections may depend on a treatment status dummy[15].

Since there is a finite number of possible graphs, labelled $s = 1, \ldots, g_{nv}$, with $g_{nv} = 2^{\sum_{j=1}^{v} n_j(n_j-1)}$, the full likelihood can be exactly approximated by

$$\ln \mathcal{L}\left(\theta | y_n, x_n, Q_n\right) = \sum_{s=1}^{g_{nv}} \ln \mathcal{L}\left(\theta | y_n, x_n, W_n^s\right) P\left(W_n^s | Q_n, x_n, \theta\right). \tag{9}$$

Even for relatively small numbers of n_j and v, g_{nv} is an enormous number. Taking $v = 5$ and $n_j = 10$ for $j = 1, \ldots, v$, the total of number of graphs g_{nv} exceeds 10^{135}. Therefore, evaluation of this integral is computationally costly and burdensome.

I propose a modification that implements a computationally efficient estimator. I substitute W_0 and M_0 for their expected values[16] $W_n^e\left(Q_n, \theta\right) = \int W_n dP\left(W_n | Q_n, x_n, \theta\right)$ and $M_n^e\left(Q_n, \theta\right) = \int M_n dP\left(M_n | Q_n, x_n, \theta\right)$. Estimation of network spillovers and structural parameters is based on the likelihood of the model

$$y_j = \lambda W_j^e\left(Q_j, \theta\right) y_j + x_j \beta_1 + W_j^e\left(Q_j, \theta\right) x_j \beta_2 + v_j^e \tag{10}$$

[14] Observation of networks with measurement errors constitute a challenge for methods that are, directly or indirectly, based on network-generated instruments, as validity assumptions are often violated. This is the case of Kelejian and Prucha (1998, 1999), Bramoullé et al. (2009) and others. Also see Plümper and Neumayer (2010).

[15] I rule out endogeneity with respect to outcomes y_n. This is the topic of a future extension to the current paper.

[16] For simplicity of explanation, momentarily assuming $\tilde{W}$ and $\tilde{M}$ are independent, which does not hold for the rest of the paper.

with $v_j^e(Q_j,\theta) = \rho M_j^e(Q_j,\theta)v_j + \epsilon_j$. The term "pseudo-likelihood" is used to distinguish the likelihood of this model from the likelihood of the model with known networks.

Model (10) is equivalent to the model if networks were observed in addition to mispecification terms that are close[17] to zero when $\theta = \theta_0$,

$$y_j = \lambda W_j^0 y_j + x_j\beta_{10} + W_j^0 x_j\beta_{20} + \lambda\left\{W_j^e\left(Q_j,\theta\right) - W_j^0\right\}y_j + \left\{W_j^e\left(Q_j,\theta\right) - W_j^0\right\}x_j\beta_{20} + v_j^e. \quad (11)$$

Intuitively, the misspecification terms containing $\{W_j^e(Q_j,\theta) - W_j^0\}$ are of small relevance when a large number of groups is observed. This point is best exemplified if group sizes are constant, condition that is not carried for the remainder of the paper. Under certain conditions, a Law of Large Numbers ensures that $v^{-1}\sum_{j=1}^{v} W_j^0 \xrightarrow{p} W_j^e(Q_j,\theta)$. Averaging the model across groups then implies that misspecification terms are small when $v \longrightarrow \infty$.

The substitution of true networks for expected networks has two consequences. First, the fact that model is inherently misspecified implies that the equality between information matrix and expected hessian does not hold, which will have implications for the expression of the asymptotic variance. Second, the introduction of expected networks implies that pointwise identification of parameters θ is generally not achieved. There are multiple combinations of λ, θ and β_2 such that the model (10) is observationally equivalent.

Subsections 3.1 to 3.3 discuss identification in three scenarios. In Subsection 3.1, I show that knowledge of one parameter (I arbitrarily focus the discussion on λ_0) restores identification under the mild additional assumption that there are at least three distinct group sizes. I will show that variation in group sizes allows me to separately identify endogenous and exogenous effects.[18] Knowledge of λ_0 separately identifies the case of a weak connections with high probability (low λ, high δ_0 and δ_1) from the case of strong connections with low probability (high λ, low δ_0 and δ_1). This is then sufficiently to fully identify the model.

In Subsection 3.2 considers the estimation of θ when λ_0 is unknown and no additional information is provided. In this case, the true parameter θ_0 is identified up to a set Θ_0. Importantly, I demonstrate that parameters in the identified set yield network spillovers equal to the spillovers evaluated at the true parameter. That is, for all $\theta \in \Theta_0$, $\varphi(x_j,\theta) = \varphi(x_j,\theta_0)$. Hence, network spillovers are point-identified. I provide the set estimator and confidence regions for the parameters. In the interest of generality, the test for network data validity is also proposed in this context. I adapt the ideas of Chernozhukov et al. (2007), Romano and Shaikh (2010) and Bugni (2010) to provide confidence regions for the structural parameter θ.

The problem with unknown λ_0 can be analogously interpreted as an under-identified Generalized Method of Moments (GMM) problem in which moment conditions are given by the score of the likelihood. The previous non-identification result manifests itself as the absence of one moment condition relative to the number of parameters. In Subsection 3.3, I then make full use of the model to obtain one additional moment condition which restores point identification of θ.

Earlier work on identification of social interactions observed that the presence of social interactions generates

[17] Comparison between likelihood computed with expected network and true networks can be found in Tables 7 and 8 in the Appendix.

[18] As also shown by Lee (2007) for the case in which networks are known. Asymmetries in the network, such as those considered by Kelejian and Prucha (1998,1999), Bramoullé et al. (2009) and De Giorgi et al. (2010) could also be used to provide identification. These would in turn require asymmetries in Q_n.

dispersion of average group outcomes beyond what can be explained by variance of explanatory variables of peer group heterogeneity alone (Glaeser et al., 1996; Graham, 2008). I implement this idea in the case where networks are unknown. This introduces an additional moment condition: the difference between observed and model-implied across-group outcome variance. As I will show, this restores identification. Consistency and asymptotic normality of the GMM estimator follows. Before proceeding, I formally derive the likelihood.

Define $S_j^e(Q,\theta) \equiv I - \lambda W_j^e(Q_j,\theta)$, $S_j^0(\lambda) \equiv I - \lambda W_j^0$, $S_j^0 \equiv S_j^0(\lambda_0)$, $R_j^e(\theta) \equiv I - \rho M_j^e(Q_j,\theta)$, $R_j^0(\rho) \equiv I - \rho M_j^0$, $R_j^0 \equiv R_j^0(\rho_0)$, $Z_j^e(Q_j,\theta_c) = (x_j, W_j^e(Q_j,\theta_c)x_j)$ and the block matrices $W_n^0(Q_n,\theta_c) = \mathrm{diag}(W_1^0(Q_1,\theta_c), \ldots, W_v^0(Q_v,\theta_c))$, $W_n^e(Q_n,\theta_c) = \mathrm{diag}(W_1^e(Q_1,\theta_c), \ldots, W_v^e(Q_v,\theta_c))$, $M_n^e(Q_n,\theta_c) = \mathrm{diag}(M_1^e(Q_1,\theta_c), \ldots, M_v^e(Q_v,\theta_c))$, $S_n^e(Q_n,\theta_c) = \mathrm{diag}(S_1^e(Q_1,\theta_c), \ldots, S_v^e(Q_1,\theta_c))$, and $Z_n^e(Q_n,\theta_c) = (Z_1^{e'}(Q_1,\theta_c), \ldots, Z_v^{e'}(Q_v,\theta_c))'$. Model (2) can be denoted $y_n = \lambda_0 W_n^0 y_n + x_n\beta_{10} + W_n^0 x_n \beta_{20} + v_n$, where $v_n = (v_1', \ldots, v_v')'$. The pseudo-likelihood is

$$\ln \mathcal{L}_n^e(\theta | y, x, Q) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) + \ln|S_n^e(Q_n,\theta)| + \ln|R_n^e(Q_n,\theta)| - \frac{1}{2\sigma^2}\epsilon_n^{e'}(Q_n,\theta)\,\epsilon_n^e(Q_n,\theta) \tag{12}$$

with $\epsilon_n^e(Q_n,\theta) = R_n^e(Q_n,\theta)(S_n^e(Q_n,\theta)y_n - Z_n^e(Q_n,\theta)\beta)$ for $\beta = (\beta_1', \beta_2')'$. Parameters β and σ^2 are concentrated out of the likelihood, simplifying derivations and implementation. Denote $\theta_c = \theta \setminus \{\beta, \sigma^2\}$ the non-concentrated parameters. At each θ_c, the closed-form solutions for the concentrated parameters are

$$\hat\beta(Q,\theta_c) = (Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^e(Q_n,\theta_c))^{-1}Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n$$

and

$$\begin{aligned}\hat\sigma^2(Q,\theta_c) &= \frac{1}{n}(S_n^e(Q_n,\theta_c)y_n - Z_n^e(Q_n,\theta_c)\hat\beta(\theta_c))'R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)(S_n^e(Q_n,\theta_c)y_n - Z_n^e(Q_n,\theta_c)\hat\beta(\theta_c))\\ &= \frac{1}{n}y_n'S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n\end{aligned}$$

where P_n^e is the projection matrix $P_n^e(Q_n,\theta_c) = I_n - R_n^e(Q_n,\theta_c)Z_n^e(Q_n,\theta_c)(Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)$ $Z_n^e(Q_n,\theta_c))^{-1}Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)$ and $P_n^e \equiv P^e(Q_n,\theta_c^0)$. The final form for the concentrated pseudo-likelihood brought to maximization is

$$\ln \mathcal{L}_n^c(\theta_c | y_n, x_n, Q_n) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat\sigma^2(Q_n,\theta_c) + |S_n^e(Q_n,\theta_c)| + |R^e(Q_n,\theta_c)|. \tag{13}$$

The final estimator is $\hat\theta = (\hat\theta_c', \hat\beta(\hat\theta_c)', \hat\sigma^2(\hat\theta_c))'$, where $\hat\theta_c \equiv \arg\max_{\theta\in\Theta_c}\ln\mathcal{L}_n^c(\theta_c | y_n, x_n, Q_n)$. I now lay formal hypothesis to guarantee asymptotic properties of the estimator.

3.1 Pointwise identification of θ when λ_0 is known

In this subsection, I present the basic assumptions for consistent estimation and pointwise identification of the parameters in the model. Identification Assumption 6, required for pointwise identification of θ, holds only if λ_0 is known to the researcher[19]. Assumptions 1-5 are maintained throughout the remaning subsections.

[19] In fact, Assumption 6 holds in the case where one parameter among λ_0, β_{20} and θ_g^0 is known. For simplicity, I arbitrarily focus the argument on λ_0.

The first assumption defines the true model, properties of the networks and homogeneity of the probability law that generates (unobserved) networks across groups. The zero main diagonal is essentially an identification condition and implies that no individual affects him or herself. The independence of P with respect to β and σ^2 allows me to concentrate these parameters, as described previously, and is taken for simplicity only as results do not depend crucially on it.

Assumption 1. *For each group $j = 1, \ldots, v$, data are generated according to the model*

$$y_j \;\; = \;\; \lambda_0 W_j^0 y_j + x_j \beta_{10} + W_j^0 x_j \beta_{20} + v_j$$

with $v_j = \rho_0 M_j^0 v_j + \epsilon_j$ and $\epsilon_j \sim \mathbb{N}\left(0, \sigma^2 I\right)$. The elements of x_n and Q_n are uniformly bounded constants. Let $mat_{n_j}\left(\{0,1\}\right)$ denote the space of n_j-by-n_j-by-2 matrices with entries in $\{0,1\}$ and zero main diagonal, a $(\Omega, \mathcal{F}, P)$ be a probability space with $\mathcal{F}$ as σ-algebra of subsets of Ω and P as probability measure. $\{W_j^0, M_j^0\}$ is particular realization from a random matrix[20]*, a measurable map from $(\Omega, \mathcal{F})$ to $mat_{n_j}\left(\{0,1\}\right)$, with probability distribution function $P\left(W_j, M_j \middle| \theta, x_j\right)$ with common functional form across groups. P does not depend on β or σ^2.*

In some applications, it is customary to conduct a row-sum normalization of W_j, the operation consisting of replacing W_j by a W_j^* with $\{W_j^*\}_{ik} = \{W_j\}_{ik} / \sum_{s=1}^{n_j} \{W_j\}_{is}$ (Anselin, 1988, Kelejian and Prucha, 1998, 1999, 2001, 2010, Lee, 2004, 2007, Lee et al., 2010). This implies that all individuals in the group are affected by and affect others to the same extent: row sums of W_j^* add to one. This assumption is avoided here on the basis of anectodal observation that individuals are generally not homogenous in terms of their connection to others in the group. In classrooms, for example, some students may be more affected by peers than others. I leave networks to be, more simply, a collection of binary numbers.

It well-known that under row-sum normalization condition, $|\lambda_0| < 1$ suffices for uniform boundedness of W_j^0 and $(S_j^0)^{-1}$, with $S_j^0 \equiv I_{n_j} - \lambda_0 W_j^0$ (Anselin, 1988). In the current setting, I propose the following notion of boundedness: let $\max_i \left|\lambda_0 \sum_{k=1}^{n} \{W_j^0\}_{ik}\right| \leq 1$, and so no row multiplied by λ_0 in absolute value exceeds one. This includes row-sum normalization as a special case; for constant row sums W_j^0 across rows, $\lambda_0 \sum_{k=1}^{n} \{W_j^0\}_{ik} = \lambda_0^* \sum_{k=1}^{n} \{W_j^{*0}\}_{ik}$ with $\lambda_0^* = \lambda_0 \sum_{s=1}^{n_j} \{W_j\}_{1s}$. In this case, it is clear that letting W_j^0 as a collection of binary numbers and $|\lambda_0|$ closer to zero is only a normalization option. Formally,

Assumption 2. *The sequence of n-by-n realized matrices $\lambda_0 W_n^0$ and $(S_n^0)^{-1}$ and of expected matrices $\lambda W_n^e\left(Q_n, \theta\right)$ and $\left(S_n^e(Q_n, \theta)\right)^{-1}$ are uniformly bounded. $W_n^e(Q_n, \theta)$ exists for all $\theta \in \Theta$.*

The next assumption guarantees y_j has an equilibrium and its mean and variance are well defined.

Assumption 3. *For all $j = 1, \ldots, v$, the eigenvalues of the realized matrix S_j^0 are smaller than one in absolute value.*

Asymptotics on v and n_j, without any specific order of divergence, is necessary to guarantee that the misspecification term goes to zero asymptotically and variance terms are consistently estimated in the limit.

[20] In fact, $\{W_j^0, M_j^0\}$ are arrays and full notation should include respective dimensions, $\{W_{n_j,j}^0, M_{n_j,j}^0\}$. This is supressed for simplicity.

Assumption 4. $n \to \infty$ *where* $n = \sum_{j=1}^{v} n_j$.

As a minor technical point, it is only necessary that non-concentrated parameters belong to a compact parameter set Θ_c.

Assumption 5. *The parameter set* Θ_c *is compact and the true parameter* $\theta_c^0 \in \Theta_c^0$.

Next, I lay out the identification condition required for point identification of parameters, followed by immediately by easy-to-interpret sufficient conditions as demonstrated in Proposition 1.

Assumption 6. *(Identification). Define*

$$\gamma\left(Q_n, \theta_c\right) \;\; = \;\; \frac{1}{n}\mathbb{E}\{\beta_0' Z_n^{0'} (S_n^{0'})^{-1} \tilde{P}_n^e\left(Q_n, \theta_c\right) (S_n^0)^{-1} Z_n^0 \beta_0\}$$

with $\tilde{P}_n^e\left(Q_n, \theta_c\right) = S_n^{e'}\left(Q_n, \theta_c\right) R_n^{e'}\left(Q_n, \theta_c\right) P_n^e\left(Q_n, \theta_c\right) R_n^e\left(Q_n, \theta_c\right) S_n^e\left(Q_n, \theta_c\right)$. *For every point* $\theta_c \in \Theta_c$, *the condition* $\gamma\left(Q_n, \theta_c\right) > 0$ *holds.*

The reduced-form of the model evaluated at the true vector of parameter θ_0 is

$$y \;\; = \;\; \left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1} Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 + \left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1}\left(R_n^e\left(Q_n, \theta_c^0\right)\right)^{-1}\epsilon_n^e. \tag{14}$$

As $\left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1} = I_n + \lambda_0 G_n^e\left(Q_n, \theta_c^0\right)$, where $G_n^e\left(Q_n, \theta_c^0\right) \equiv W_n^e\left(Q_n, \theta_c^0\right)\left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1}$, the expression above can also be written as

$$y_n \;\; = \;\; Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 + \lambda_0 G_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 + \left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1}\left(R_n^e\left(Q_n, \theta_c^0\right)\right)^{-1}\epsilon_n^e. \tag{15}$$

For separate identification of λ_0 and $\beta_0 = (\beta_{10}', \beta_{20}')'$, it is necessary to guarantee that matrices $Z_n^e\left(Q_n, \theta_c^0\right)$ and $G_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 = W_n^e\left(Q_n, \theta_c^0\right)(S_n^e\left(Q_n, \theta_c^0\right))^{-1} Z_n^e\left(Q_n, \theta_c^0\right)\beta_0$ are not dependent asymptotically. In turn, asymptotic independence of the concerned matrices is a necessary and sufficient condition for Assumption 6, as I now show. Following Lemma 3, $\gamma\left(Q_n, \theta_c\right)$ is well approximated by $\gamma^e\left(Q_n, \theta_c\right)$, where

$$\gamma^e\left(Q_n, \theta_c\right) \;\; = \;\; \frac{1}{n}\beta_0' Z_n^{e'}\left(Q_n, \theta_c^0\right)(S_n^{e'}\left(Q_n, \theta_c^0\right))^{-1}\tilde{P}_n^e\left(Q_n, \theta_c\right)(S_n^e\left(Q_n, \theta_c^0\right))^{-1} Z_n^e\left(Q_n, \theta_c^0\right)\beta_0.$$

Given that $\tilde{P}_n^e\left(Q_n, \theta_c\right) = S_n^{e'}\left(Q_n, \theta_c\right) R_n^{e'}\left(Q_n, \theta_c\right) P_n^e\left(Q_n, \theta_c\right) R_n^e\left(Q_n, \theta_c\right) S_n^e\left(Q_n, \theta_c\right)$ is positive definite, then $\gamma\left(Q_n, \theta_c\right) = 0$ if, and only if, $\left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1} Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 = 0$, which is equivalent to $Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 + \lambda_0 G_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0 = 0$ using $\left(S_n^e\left(Q_n, \theta_c^0\right)\right)^{-1} = I_n + \lambda_0 G_n^e\left(Q_n, \theta_c^0\right)$ or, essentially, that $Z_n^e\left(Q_n, \theta_c^0\right)$ and $G_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0$ are asymptotically independent. The next Proposition, which resemble similar results of Bramoullé et al. (2009) and Lee et al. (2010), states the desired conditions.

Proposition 1. *If* λ_0 *is known,* $\beta_{20} \neq \lambda_0\beta_{10}$, x, $W_n^e\left(Q_n, \theta_c^0\right) x_n$ *and* $\left(W_n^e\left(Q_n, \theta_c^0\right)\right)^2 x_n$ *are linearly independent, then* $Z_n^e\left(Q_n, \theta_c^0\right)$ *and* $G_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0$ *are asymptotically independent, and therefore Assumption 6 holds.*

It is useful to note that variation in group sizes is often sufficient to assure independence between x_n, $W_n^e\left(Q_n, \theta_c^0\right) x_n$ and $\left(W_n^e\left(Q_n, \theta_c^0\right)\right)^2 x_n$. This is also seen in the subgroup model of Lee (2007) where individuals are sorted in many groups. In particular, let the probabilistic model for network formation be the pure Bernoulli, where links are formed with probability δ_0, independent of exogenous characteristic. Then $W_j^e(Q_j,\theta_c^0) = \delta_0(\iota_{n_j}\iota_{n_j}' - I_{n_j})$ and $(W_j^e(Q_j,\theta_c^0))^2 = \delta_0^2(n_j-2)(\iota_{n_j}\iota_{n_j}' + I_{n_j})$. With at least three distinct values of n_j, independence condition in the previous Proposition is guaranteed[21].

Under the conditions introduced above, I present the basic Theorem. Proofs are found in the Appendix D.

Theorem 1. *Under assumptions 1-6, $\hat{\theta}$ is a consistent estimator for θ_0, i.e., $\hat{\theta}\xrightarrow{p}\theta_0$.*

Asymptotic distribution can be obtained from a Taylor expansion around the point $\frac{\partial \ln \mathcal{L}^e(\hat{\theta}|y_n,x_n,Q_n)}{\partial\theta} = 0$. For a point $\tilde{\theta}$ between $\hat{\theta}$ and θ_0,

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \left[\frac{1}{n}\frac{\partial \ln \mathcal{L}^e(\tilde{\theta}\big| y_n,x_n,Q_n)}{\partial\theta\partial\theta'}\right]^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}^e\left(\theta_0|\, y_n,x_n,Q_n\right)}{\partial\theta}. \tag{16}$$

The variance matrix of the score vector is $\Sigma_n(\lambda_0) \equiv \mathbb{E}[\frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial\theta}\cdot\frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial\theta'}]$. In the limit, $\hat{\theta}\xrightarrow{p}\theta_0$, which implies $\tilde{\theta}\xrightarrow{p}\theta_0$ and so the Hessian matrix converges to $\Omega_n(\lambda_0) = \mathbb{E}[\frac{1}{n}\frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial\theta\partial\theta'}]$. As the model is inherently misspecified, the Hessian is not equal to the expected outer product of the gradient. The asymptotic variance-covariance matrix converges instead to the usual sandwich estimator. That is,

Theorem 2. *Under assumptions 1-5, $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{p}N(0,\Sigma^{-1}(\lambda_0)\Omega(\lambda_0)\Sigma^{-1}(\lambda_0))$, where $\Sigma(\lambda_0) = \lim_{n\to\infty}\Sigma_n(\lambda_0)$ and $\Omega(\lambda_0) = \lim_{n\to\infty}\Omega_n(\lambda_0)$.*

3.2 Set identification of θ when λ_0 is unknown

There is one simple way asymptotic independence of the matrices is violated. Any path $\{\lambda_+,\beta_{2+},\theta_c^+\}$ such that $W_n^e\left(Q_n,\theta_c^+\right)x_n\beta_{2+} = W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{20}$ and $\lambda_+W_n^e\left(Q_n,\theta_c^+\right) = \lambda_0 W_n^e\left(Q_n,\theta_c^0\right)$ results in a similar reduced-form, constituting a breakdown of Assumption 6. Parameters are not individually identified, which is compatible with the difficulty in separately identifying a large number of weak connections from a small number of strong connections. I now turn to the problem of estimation and inference on the identified set.

Using assumptions 1-5 only, I employ methods of estimation and inference on set-identified models of Chernozhukov et al. (2007), Romano and Shaikh (2010) and Bugni (2010) to establish desired results. The point of departure from classic asymptotic analysis is the observation that the identified set $\Theta_0 = \{\tilde{\theta}\in\Theta : F_n(\tilde{\theta}) = F_n(\theta_0)\}$, for $F_n(\theta) = \mathbb{E}\ln\mathcal{L}_n^e(\theta)$, and the estimated set $\hat{\Theta} = \{\tilde{\theta}\in\Theta : \ln\mathcal{L}_n^e(\tilde{\theta}) = \inf_{\theta\in\Theta}\ln\mathcal{L}_n^e(\theta)\}$ are not singletons.

In the current case, the identified set is of considerable importance because for any $\theta\in\Theta_0$, network spillovers are constant and equal to network spillovers evaluated at the true parameter vector, $\varphi(x_n,\theta_0)$. In order to establish this result, define the subset $\Phi\left(\theta|\, y_n,x_n\right)\subseteq\Theta$ as the parameters such that spillovers are equal to $\varphi(x_n,\theta)$, that is,

$$\Phi\left(\theta|\, y_n,x_n\right) = \{\theta^+\in\Theta : \lambda_+W_n^e(Q_n,\theta_c^+) = \lambda W_n^e\left(Q_n,\theta_c\right), W_n^e(Q_n,\theta_c^+)x_n\beta_2^+ = W_n^e\left(Q_n,\theta_c\right)x_n\beta_2\}. \tag{17}$$

The next Proposition states that θ_0 belongs to the identified set Θ_0 and that it is fully characterized by the subset of Θ such that spillovers are equal to $\varphi(x_n,\theta_0)$.

[21] That is, if there are three distinct values of n_j, the only conformable vectors c_1, c_2 and c_3 such that $xc_1 + \delta_0(\mathrm{diag}(\iota_{n_1}\iota_{n_1}',\ldots,\iota_{n_j}\iota_{n_j}') - I_n)xc_2 + (\mathrm{diag}((n_1-2)\iota_{n_1}\iota_{n_1}',\ldots,(n_j-2)\iota_{n_j}\iota_{n_j}') + I_n)^2xc_3 = 0$ are $c_1 = c_2 = c_3 = 0$.

Proposition 2. *For any* $\theta \in \Phi(\theta^0 | y_n, x_n)$*, the network spillovers evaluated at* θ *are equal to network spillovers evaluated at* θ_0*,* $\varphi(x_n, \theta) = \varphi(x_n, \theta_0)$*. Also, this is the identified set,* $\Phi(\theta^0 | y_n, x_n) = \Theta_0$*.*

The objective then is to produce a sequence of sets such that: (i) in the limit, they are consistent estimates of Θ_0, in a sense that the Hausdorff set distance metric[22] d_h converges to zero in probability, and (ii) select a set $\hat{\Theta}_\alpha$ such that the coverage probability is asymptotically controlled, that is, $\lim_{n\to\infty} P\{\Theta_0 \subseteq \hat{\Theta}_\alpha\}) = 1 - \alpha$ for $\alpha \in [0, 1]$.

These objectives can be fulfilled with the definition of contour sets of the rescaled likelihood $L_n(\theta | y_n, x_n, Q_n) = -n^{-1}[\ln \mathcal{L}_n^e(\theta | y_n, x_n, Q_n) - \inf_{\theta\in\Theta} \ln \mathcal{L}_n^e(\theta | y_n, x_n, Q_n)]$ and $\hat{\Theta}(c_n) = \{\theta \in \Theta : L_n(\theta | y_n, x_n, Q_n) \leq c_n\}$. The next Theorem proves that the estimator $\hat{\Theta} = \hat{\Theta}(0)$ is consistent for Θ_0, i.e, $d_h(\hat{\Theta}, \Theta_0) \xrightarrow{p} 0$. In fact, this result can be obtained if any sequence c_n such that $n^{-1}c_n \xrightarrow{p} 0$ is used to produce an alternative estimator $\hat{\Theta}(c_n)$. For the construction of a set that covers Θ_0 with probability α, it is necessary to select $c_n = \hat{c}_n(\alpha)$ such that $\hat{\Theta}(\hat{c}_n(\alpha))$ possesses the desired property.

Notice the event $\{\Theta_0 \subseteq \hat{\Theta}(c_n)\}$ is equivalent to the event $\{\sup_{\theta\in\Theta_0} L_n(\theta | y_n, x_n, Q_n) \leq c_n\}$, and hence, in order to build coverage regions for the identified set Θ_0 with predetermined probability α, it suffices to input a $c_n = \hat{c}_n(\alpha)$ such that $\hat{c}_\alpha$ consistently estimates the α-quantile of the test statistic $\sup_{\theta\in\Theta_0} L_n(\theta | y_n, x_n, Q_n)$. That is, for any set $K \subseteq \Theta$, use

$$\hat{c}_n(\alpha) = \inf\left\{\tilde{c} : P\left\{\sup_{\theta\in K} L_n(\theta | y_n, x_n, Q_n) \leq \tilde{c}\right\} \geq 1 - \alpha\right\}.$$

Given the probability is not known, I will use a bootstrap algorithm to produce usable estimates of $\hat{c}_n(\alpha)$. For the moment, assume $\hat{c}_n(\alpha)$ is known. The next Theorem shows asymptotic properties of the estimated contour sets $\hat{\Theta}(c_n)$ for the various choices of c_n.

Theorem 3. *Let* c_n *be such that* $n^{-1}c_n \xrightarrow{p} 0$*. (1) Under Assumptions 1-5, if* $\Theta_0 \neq \Theta$ *and* Θ *compact,* $\Theta_0 \subseteq \hat{\Theta}(c_n)$ *with probability approaching one,* $d_h(\hat{\Theta}(c_n), \Theta_0) = o_p(1)$ *and* $d_h(\hat{\Theta}(c_n), \Theta_0) = O_p(n^{-\frac{1}{2}})$*. (2) For* $c = \hat{c}_n(\alpha)$ *consistent estimator of the* α*-quantile of* $\sup_{\theta\in\Theta_0} L_n(\theta | y_n, x_n, Q_n)$*,* $\lim_{n\to\infty} P\{\Theta_0 \subseteq \hat{\Theta}(\hat{c}_n(\alpha))\}) = 1 - \alpha$*. (3) Given Proposition 2, the network spillover is point-identified. (4) Point-identification for* β_{10} *and* σ_0^2 *is obtained and* $(\hat{\beta}_1, \hat{\sigma}^2) \xrightarrow{p} (\beta_{10}, \sigma_0^2)$*.*

Obtaining confidence regions for known functions of the identified set is important at least in two circumstances. First, it provides confidence regions for the network spillovers, i.e., confidence regions for Φ_0, the image of Θ_0 under the known function $\varphi(x, \theta)$ for given $\theta \in \Theta_0$. Second, I will show it provides a framework for validation of network data, when it is available. I now develop these points.

Following Romano and Shaikh (2010), in general terms, let f be a known function with $f : \Theta \to \Upsilon$, with Υ_0^f being the image of Θ_0 under f, and also let $f^{-1}(v) = \{v \in \Upsilon : f(\theta) = v\}$. This suggests a modification of the inferential test statistic in the following way: note $v \in \Upsilon_0^f$ if, and only if, there exists some $\theta \in f^{-1}(v)$

[22] The Hausdorff set distance metric is defined

$$d_h(A, B) = \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\}$$

with $d(b, A) = \inf_{a\in A} \|b - a\|$ and $d_h(A, B) = \infty$ if A or B are empty.

subject to $Q_n(\theta) = 0$, which in turn implies that $\inf_{\theta \in f^{-1}(v)} Q_n(\theta) = 0$. As before, the objective is to construct a set $\hat{\Upsilon}_\alpha$ such that coverage probability is $1-\alpha$, i.e., $\lim_{n\to\infty} P\{\Upsilon_0 \subseteq \hat{\Upsilon}_\alpha\} = 1-\alpha$ and, in analogy to the previous case, this set can be defined by selecting $c_n^f(\alpha)$ such that the event $\{\Upsilon_0^f \subseteq \hat{\Upsilon}_\alpha\}$ is equivalent to $\{\sup_{v\in\Upsilon_0^f} \inf_{\theta\in f^{-1}(v)} L_n(\theta) \leq c_n^f(\alpha)\}$.

Again, if the α-quantiles of the test statistic $\sup_{v\in\Upsilon_0^f} \inf_{\theta\in f^{-1}(v)} L_n(\theta)$ were available, coverage region with asymptotically controlled error probability α would be obtained directly. Appendix E details a bootstrap algorithm for obtaining consistent estimates $\hat{c}_n^f(\alpha)$ of $c_n^f(\alpha)$. For the moment, I now describe the two important applications of this procedure for the context of inference on the network spillovers and network effects.

Remark 2. *(Confidence region for network spillovers).* The procedure above can be applied directly replacing function f with known function $\varphi(x;\theta)$. In this case, because $\varphi(x_n;\theta)$ is a function from Θ to $\mathbb{R}^1$, and given Proposition 2 states the network spillovers are constant in the identified set, the image Υ_0^φ is a scalar in $\mathbb{R}$ and the confidence region is actually a confidence *interval*, a subset of $\mathbb{R}^1$. □

Remark 3. *(Testing for reported network connections).* Introduce reporting of network data with recourse to matrix Q_j, making $\{Q_j\}_{ik} = 1$ if individual i in group j reports a link with individual k in the same group, through which it is believed that i affects k. In this case, a reasonable network model is given by a collection of Bernoulli trials with probability link formation depending on link observed reports, that is, model (1) with Q_n as described above. In this setting, structural parameter δ_1 is the the estimated probability *given* observation of link reports, and δ_0 otherwise. The null hypothesis of interest is $\mathcal{H}_0 : \delta_1 - \delta_0 = 0$, with alternative $\mathcal{H}_A : \delta_1 - \delta_0 \neq 0$. In the setting above, suffices to take $\tilde{f} : \Theta \to \mathbb{R}^1$ as $\tilde{f}(\theta) = \delta_1 - \delta_0$ and build appropriate confidence intervals. □

3.3 Pointwise identification when λ_0 is unknown using outcome dispersion

In the previous subsection, I showed that parameters of interest are identified up to a set and network spillovers are constant within the identified set. A theoretically feasible restriction to fully identify the model is to assume λ_0 is known: under certain conditions, Theorem 1 proves consistency. Nevertheless, this assumption is unlikely to be satisfied in practice, as λ_0 is rarely observed. In this Section, I increment the problem with one additional restriction which restores point identification, selecting a parameter in the identified set.

This restriction is derived from matching the observed to the model-implied variance of the group-average outcome. The intuition is straightforward. When social interactions are not present, sufficiently large group sizes implies that group averages should be relatively close to population averages conditional on observables. Introduction of social interactions affects dispersion in the following way. Since individuals mirror the choices of the others, outcomes within a group positively correlate. In other words, a positive shock to the group affects individuals not only through individual decision, but also through peer composition. As a consequence, average of group outcome increases to greater extent than in the counterfactual in which social interactions are irrelavant. A similar reasoning applies to a bad shock. It follows that average outcome across groups are more disperse relative to the case in which social interactions are irrelevant.

It has been observed elsewhere[23] that group outcomes are substantially dispersed across groups even when similar along observable characteristics. This anecdotal observation has been denoted as "excess variance" and

[23] Hanushek (1971), Rivkin et al. (2005), Glaeser et al. (1996).

used to provide identification when networks are known (Graham, 2008). Other papers have contributed to identification using covariance restrictions in the context of social interactions, such as in the survey paper by Blume et al. (2011, p. 872) and references therein.

Since network formation depends on a model described in Section 2, the dispersion across groups provides a restriction that includes link strength, probability of link formation and dependence on exogenous characteristics of the others. The relation is usually non-linear and I will show it is sufficient to provide identification. The main idea is that, accounting for variance originating from explanatory variables and the individual or group heterogeneity, the remaining variance can only be explained by social interactions and pattern of association therein. Define, from the outset, the within and between group variance,

$$V_{W,j}(y_n) \;=\; n_j^{-1}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2 \qquad ; \qquad V_{B,j}(y_n)=(\bar{y}_j-\bar{y})^2$$

where $\bar{y}_j = n_j^{-1}\sum_{i=1}^{n_j} y_{ij}$ and $\bar{y} = v^{-1}\sum_{j=1}^{v}\bar{y}_j$. It is useful to derive the expectation of these quantities in terms of the variance of outcomes as predicted by the model. Then, $\mathbb{E}V_{W,j}(y) = n_j^{-1}\sum_{i=1}^{n_j}[\mathbb{V}(y_j)]_{ii}$ and $\mathbb{E}V_{B,j}(y) = n_j^{-2}\iota_{n_j}'\mathbb{V}(y_j)\iota_{n_j}$. From the reduced-form of model (2), the covariance matrix of outcomes for group j is given by[24]

$$\mathbb{V}(y_j) \;=\; \mathbb{E}(s_j x_j\beta_{10}\beta_{10}' x_j' s_j' + 2s_j^* x_j\beta_{10}\beta_{20}' x_j' s_j' + s_j^* x_j\beta_{20}\beta_{20}' x_j' s_j^{*'}) + \mathbb{E}((S_j^0)^{-1}\epsilon_j\epsilon_j'(S_j^0)^{-1'}) \tag{18}$$

for $s_j = (S_j^0)^{-1} - \mathbb{E}((S_j^0)^{-1})$ and $s_j^* = (S_j^0)^{-1}W_j^0 - \mathbb{E}((S_j^0)^{-1}W_j^0)$. In absence of networks, $s_j = I_{n_j}$ and $s_j^* = 0_{n_j\times n_j}$ and, therefore, outcome variance is increased when social interactions are considered. As pointed out above, in applications it is usually the case that the latter is larger than the former in the positive semi-definite sense although the reverse relation is theoretically possible for certain parameters. The distance between variances $V_{B,j}$ and $V_{W,j}$ and their theoretical expected counterparts as implied by the model, $\mathbb{E}V_{B,j}(y_n)$ and $\mathbb{E}V_{W,j}(y_n)$, is used to distinguish between competing parameters that belong to the identified set. Given $V_{B,j}$ and $V_{W,j}$ are observed from data, we only need to generate predictions from the model (18). Naturally, this strategy depends on the thorectical calculation of $\mathbb{V}(y_j)$, which are often difficult to evaluate analytically but straightforward to compute. I now introduce one particular example where identification is throughoutly proven only with between-variance of outcomes.

Example 3. *(Bernoulli network model).* In a simple setting where link formation is independent and equal to δ_1, I conduct a Series Expansion and take a first-order approximation. That is, $(S_j^0)^{-1} - \mathbb{E}(S_j^0)^{-1} = \lambda_0(W_j^0 - \mathbb{E}W_j^0) + \cdots$ which is approximately $\lambda_0(W_j^0 - \mathbb{E}W_j^0)$ as remaining terms decay in exponential rates. Using independence of the Bernoulli trials that generate links, equation (18) simplifies to

$$\mathbb{V}\{y_j\} \;=\; \text{diag}\left(\mathbb{V}\{W_j\}\left(\lambda^2\text{diag}\left(x_j^{11}\right) + 2\lambda\text{diag}\left(x_j^{12}\right) + \lambda^2\sigma^2\iota_{n_j}\right)\right) + \sigma^2 I_{n_j} \tag{19}$$

where $\mathbb{V}\{W_j^0\}$ is the variance of W_j^0, $x_j^{11} = \text{diag}(x_j\beta_{10}\beta_{10}'x_j')$, $x_j^{12} = \text{diag}(x_j\beta_{10}\beta_{20}'x_j')$ and $x_j^{22} = \text{diag}(x_j\beta_{20}\beta_{20}'x_j')$ extracts the main diagonal of a matrix into a column vector or vice-versa, as appropriate. Off-diagonal terms

[24] For the panel data with fixed effects, proceed as described in Subsection 3.4. In this Section, for simplicity I assume $\rho_0 = 0$. This is not substantial as all results are maintained in the more general case.

are zero. In the Bernoulli model without dependence on exogenous characteristics, $\mathbb{V}\{W_j\} = \delta_1(1-\delta_1)\iota_{n_j}\iota_{n_j}'$ and, in this case,

$$\begin{aligned}\mathbb{V}\{y_j\} &= \text{diag}\left(\delta_1(1-\delta_1)\iota_{n_j}\iota_{n_j}'\left(\lambda^2\text{diag}\left(x_j^{11}\right)+2\lambda\text{diag}\left(x_j^{12}\right)+\text{diag}\left(x_j^{22}\right)+\lambda^2\sigma^2\iota_{n_j}\right)\right)+\sigma^2 I_{n_j}\\ &= \delta_1(1-\delta_1)\left(\lambda^2\iota_{n_j}'\text{diag}\left(x_j^{11}\right)+2\lambda\iota_{n_j}'\text{diag}\left(x_j^{12}\right)+\text{diag}\left(x_j^{22}\right)+n_j\lambda^2\sigma^2\right)I_{n_j}+\sigma^2 I_{n_j}\end{aligned}$$

and the between-group variance is

$$V_{B,j} = n_j^{-1}\delta_1\left(1-\delta_1\right)\left(\lambda^2\iota_{n_j}'\text{diag}\left(x_j^{11}\right)+2\lambda\iota_{n_j}'\text{diag}\left(x_j^{12}\right)+\iota_{n_j}'\text{diag}\left(x_j^{22}\right)+n_j\lambda^2\sigma^2\right)+n_j^{-1}\sigma^2.$$

This provides the additional restriction required for the identification of θ. Formally, the Jacobian of the matrix formed by stacking restrictions, including those originating from reduced-form estimation, has full rank, and then Theorem 6 of Rothenberg (1971, p. 585) is applied. Proofs can be found in Appendix D.8. □

The approach suggests a Genaralized Method of Moments estimator with moment conditions given by $q_{1,j}(y_j,x_j,\theta) = \mathbb{E}V_{B,j}(y_j,x_j,\theta) - V_{B,j}(y_j,x_j,\theta)$ and $q_{2,j}(y_j,x_j,\theta) = \mathbb{E}V_{W,j}(y_j,x_j,\theta) - V_{W,j}(y_j,x_j,\theta)$ minimized on the estimated set $\hat{\Theta}$,

$$\hat{\theta} \quad = \quad \arg\min_{\theta\in\hat{\Theta}}\left(\sum_{j=1}^{v} q_j(y_j,x_j,\theta)\right)'\Omega\left(\sum_{j=1}^{v} q_j(y_j,x_j,\theta)\right)$$

where $q_j(y_j,x_j,\theta) = [q_{1,j}(y_j,x_j,\theta), q_{2,j}(y_j,x_j,\theta)]'$ and 2×2 weight matrix Ω. It is equally possible to estimate the same GMM problem on the unrestricted parameter set Θ and introduce score conditions given by the solution of the pseudo-likelihood and assigning arbitrarily large weights to them. Unfortunately, the expected variances are generally difficult to compute. Even in simple examples, one has to rely on very crude approximations of to obtain the expectation of $(S_j^0)^{-1}$. Next, I outline a general procedure for simulating the moment conditions (Gouriéroux and Monfort, 1997) and prove the desired asymptotic properties, including consistency for $\hat{\theta}$. The final estimator is the solution to

$$\hat{\theta} \quad = \quad \arg\min_{\theta\in\hat{\Theta}}\left(\sum_{j=1}^{v} S^{-1}\sum_{s=1}^{S} q_{s,j}(y_j,x_j,\theta)\right)'\Omega\left(\sum_{j=1}^{v} S^{-1}\sum_{s=1}^{S} q_{s,j}(y_j,x_j,\theta)\right) \tag{20}$$

where $q_{s,j}(y_j,x_j,\theta) = [V_{B,j}(y_j,x_j,\theta) - V_{B,j}(\hat{y}_{j,s},x_j,\theta); V_{W,j}(y_j,x_j,\theta) - V_{B,j}(\hat{y}_{j,s},x_j,\theta)]$ with $\hat{y}_{j,s} = (S_j^s)^{-1}(x_j\beta_1 + W_j^s x_j\beta_2 + e_j^s)$, $S^s = (I_{n_j} - \lambda W_j^s)^{-1}$, W_j^s sampled from the distribution of the network-generating model with parameters θ and ϵ_j^s is sampled from a normal distribution with variance σ^2. If the simulator is unbiased, one can expect that $S^{-1}\sum_{s=1}^{S} q_{s,j}(y_j)\xrightarrow{p} q_j(y_j)$ as $S\longrightarrow\infty$ and asymptotic properties follow. In addition, given $\hat{\Theta}$ is $\sqrt{n}$-consistent for Θ_0 on the Hausdorff metric, one might expect minimizing on the set $\hat{\Theta}$ is asymptotically equivalent to minimizing on the identified set Θ_0.

Theorem 4. *If parameters are identified, (i) estimator (20), minimized on the estimated set $\hat{\Theta}$, as defined in Section 3.2, is consistent for θ_0, $\hat{\theta}\xrightarrow{p}\theta_0$, and (ii) if $S\to\infty$ sufficiently fast, $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d} N(0,\Sigma^*)$, where $\Sigma^* = (G'(\Omega^*)^{-1}G)^{-1}$, $G = \mathbb{E}\nabla_\theta q_n(y_n,x_n,\theta_0)$ and $\Omega^* = (\mathbb{E}q_n(y_n,x_n,\theta_0)q_n(y_n,x_n,\theta_0)')^{-1}$ with optimal choice of* *weight matrix* Ω^* *and* $q_n(y_n, x_n, \theta_0) = \sum_{j=1}^{v} q_j(y_j, x_j, \theta_0)$.

3.4 Fixed and Time Effects

In this subsection, I propose a data transformation to eliminate fixed effects, along with corresponding treatment of the variance-covariance matrix induced by this transformation. This is of considerable importance given that explanatory variables x_j may correlate with unobserved components that vary at the group or individual-level, for example an unobserved "good teacher" shock in a classroom or unobserved peer characteristic that may affect learning.

Bramoullé et al. (2009) and Lee (2007) propose eliminating fixed effects subtracting average of connected peers (local differencing) or average of all individuals in a group in a given time period, regardless of connection status (global differencing). Neither approach is available in the current setting: by definition of the problem in the current paper, networks are unobserved, and hence local differencing is not defined. Yet, global differencing cannot be applied in the absence of row-sum normalization. Group fixed effects with the row-sum normalization condition implies that all individuals are affected to the same degree by network spillovers originating for them. When the row-sum normalization condition is removed, heterogeneity of individual responses to fixed effects through networks implies that no data manipulation possibly removes them in the absence of network observation.

For this purpose, I introduce time dimension and time-difference data in order to remove fixed effects. This approach also has the advantage of allowing for individual fixed effects. Let the spatio-temporal model be, for $t = 1, \ldots, T$,

$$y_{jt} \quad = \quad \lambda W_j y_{jt} + x_{jt}\beta_1 + W_j x_{jt}\beta_2 + \alpha_j + \gamma_t + v_{jt} \tag{21}$$

where $v_{jt} = \rho M_j v_{jt} + \epsilon_{jt}$. Here, α_j represents a $n_j \times 1$ vector of individual or group fixed effects, or both. In the classical fixed effects case, α_j is allowed to vary over individuals; the group effect case is when $\alpha_j = \dot{\alpha}_j \iota_{n_j}$, with constant scalar $\dot{\alpha}_j$ throughout individuals in group j and does not vary over time. Notation is left sufficiently general to incorporate both cases. Group effects, in Manski's (1993) terminology, are denominated *correlated effects*.

Define $\dot{y}_{jt} = y_{jt} - \bar{y}_{j\cdot}$, $\bar{y}_{j\cdot} = T^{-1}\sum_{t=1}^{T} y_{jt}$, $\dot{x}_{jt} = x_{jt} - \bar{x}_{j\cdot}$, $\bar{x}_{j\cdot} = T^{-1}\sum_{t=1}^{T} x_{jt}$, $\bar{\gamma}_t = \gamma_t - \dot{\gamma}_\cdot$ and $\bar{\gamma}_\cdot = T^{-1}\sum_{t=1}^{T}\gamma_t$. The transformed model is

$$\dot{y}_{jt} \quad = \quad \lambda W_j \dot{y}_{jt} + \dot{x}_{jt}\beta_1 + W_j \dot{x}_{jt}\beta_2 + \dot{\gamma}_t + \dot{v}_{jt}. \tag{22}$$

which is a consequence of (21) because the time-differenced $W_j y_{jt}$ is equal to $W_j \dot{y}_{jt}$, and similarly for the $W_j \dot{x}_{jt}\beta$, under the hypothesis of invariance of the network over time. Explicitly, the k-th line of the time-differenced

$W_j y_{jt}$ is

$$\begin{aligned}\sum_{i=1}^{n_j}\{W_j\}_{ki}\{y_{jt}\}_i - T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n_j}\{W_j\}_{ki}\{y_{jt}\}_i &= \sum_{i=1}^{n_j}\{W_j\}_{ki}\{y_{jt}\}_i - \sum_{i=1}^{n_j}\{W_j\}_{ki}T^{-1}\sum_{t=1}^{T}\{y_{jt}\}_i \\ &= \sum_{i=1}^{n_j}\{W_j\}_{ki}\left(\{y_{jt}\}_i - \{\bar{y}_{j\cdot}\}_i\right) \qquad (23)\end{aligned}$$

Letting $\dot{y}_{nT} = (\dot{y}'_{11}, \ldots \dot{y}'_{1T}, \ldots, \dot{y}'_{v1}, \ldots, \dot{y}'_{vT})'$ and $\dot{x}_{nT} = (\dot{x}'_{11}, \ldots, \dot{x}'_{1T}, \ldots, \dot{x}'_{v1}, \ldots, \dot{x}'_{vT})'$, and similarly for $\dot{v}$ and $\dot{\gamma}$, the full model can be rewritten $\dot{y}_{nT} = \lambda W_{nT}\dot{y}_{nT} + \dot{x}_{nT}\beta_1 + W_{nT}\dot{x}_{nT}\beta_2 + \dot{\gamma}_T + \dot{v}_{nT}$, where $W_{nT} = \operatorname{diag}\{I_T \otimes W_1, \ldots, I_T \otimes W_v\}$. Remaining matrices are defined in a similar way and carry the subscript nT for clarity. The variance-covariance matrix of $\dot{v}_{nT}$ is $\mathbb{E}(\dot{v}_{nT}\dot{v}'_{nT}) = \sigma_0^2 (R^0_{nT})^{-1}\dot{\Sigma}_{nT}(R^{0\prime}_{nT})^{-1}$, where $\dot{\Sigma}_{nT} = \sigma_0^2 I_{nT} - \sigma_0^2 T^{-1}\cdot \operatorname{diag}(\iota_T\iota'_T \otimes I_{n_1}, \ldots, \iota_T\iota'_T \otimes I_{n_v})$. This more complicated form recognizes the dependence in $\dot{v}_{nT}$ introduced by time-average subtraction. Finally, likelihood (12) is adjusted to

$$\begin{aligned}\ln \mathcal{L}^e_{nT}(\theta|\, y_{nT}, x_{nT}, Q_{nT}) &= -\frac{nT}{2}\ln\left(2\pi\sigma^2\right) + \ln|S^e_{nT}(Q_{nT},\theta)| + \ln|R^e_{nT}(Q_{nT},\theta)| \\ &\quad -\frac{1}{2\sigma^2}\epsilon^e_{nT}(Q_{nT},\theta)'\,\dot{\Sigma}_{nT}\epsilon^e_{nT}(Q_{nT},\theta) \qquad (24)\end{aligned}$$

where $\epsilon^e_{nT}(Q_{nT},\theta) = R^e_{nT}(Q_{nT},\theta)(\dot{y}_{nT} - \lambda W^e_{nT}(Q_{nT},\theta)\dot{y}_{nT} - \dot{x}_{nT}\beta_1 - W^e_{nT}(Q_{nT},\theta)\dot{x}_{nT}\beta_2 - \dot{\gamma}) = R^e_{nT}(Q_{nT},\theta)$ $(S^e_{nT}(Q_{nT},\theta)\dot{y}_{nT} - \dot{Z}^e_{nT}(Q_{nT},\theta)\tilde{\beta})$ and $\dot{Z}^e_{nT}(Q_{nT},\theta)$ now also incorporate time effects: $\dot{Z}^e_{jt}(Q_j,\theta) = (x_{jt},$ $W^e_j(Q_j,\theta)x_{jt}, \mathbf{1}\{t=1\}\iota_{n_j}, \ldots, \mathbf{1}\{t=T\}\iota_{n_j})$ and $\tilde{\beta} = (\beta', \gamma_1, \ldots, \gamma_T)'$. In fact, any variable not subject to exogenous effects can be incorporated by adding columns to $\dot{Z}^e_{jt}(Q_{nT},\theta)$. The concentrators are now

$$\begin{aligned}\hat{\tilde{\beta}}(Q_{nT},\theta) &= (Z^{e\prime}_{nT}(Q_{nT},\theta)\ddot{\Sigma}_{nT}Z^e_{nT}(Q_{nT},\theta))^{-1}Z^{e\prime}_{nT}(Q_{nT},\theta)\ddot{\Sigma}_{nT}S^e_{nT}(Q_{nT},\theta)y_{nT} \\ \hat{\tilde{\sigma}}^2(Q_{nT},\theta) &= \frac{1}{n}(S^e_{nT}(Q_{nT},\theta)y - Z^e_{nT}(Q_{nT},\theta)\hat{\tilde{\beta}})\ddot{\Sigma}_{nT}(S^e_{nT}(Q_{nT},\theta)y_{nT} - Z^e_{nT}(Q_{nT},\theta)\hat{\tilde{\beta}})\end{aligned}$$

where $\ddot{\Sigma}_{nT} = R^{e\prime}_{nT}(Q_{nT},\theta)\dot{\Sigma}_{nT}R^e_{nT}(Q_{nT},\theta)$. Concentrated likelihood (12) remains unchanged with $\hat{\sigma}^2(Q_{nT},\theta)$ substituted for $\hat{\tilde{\sigma}}^2(Q_{nT},\theta)$. Preceding theorems are applied with obvious modifications.

4 Simulations and Implementation

In this Section, I conduct a simulation exercise to demonstrate the small-sample empirical properties of the estimator. MATLAB codes are available upon request[25]. The algorithms are presented in Appendix E.

Four simulations sets are performed: purely cross-sectional model (2), under $T = 1$ and absence of fixed effects; the panel (5) with $T = 5$ and fixed effects but no time effects; with time effects but no fixed effects; and, finally, with both time and fixed effects. Sample sizes are $(n = 25, v = 250)$, $(n = 100, v = 250)$, $(n = 25, v = 1000)$ and $(n = 100, v = 1000)$. Simulations with smaller n and v can be found in Appendix F.1. In every case, I allow for heterogeneity in group sizes, by sampling n_j from a standard normal distribution with mean n and standard error 5, rounded to the nearest integer.

[25]STATA codes will soon be available.

True parameters are $\theta_s = (0.0125, 1, 1, 0.04, 0.04, 1)'$ and $\theta_g = (0.75, 0.30)'$. In a row-normalized model and with this combination of parameters, $\lambda = 0.0125$ would roughly correspond to an autoregressive parameter of 0.16 for $n = 25$, 0.32 for $n = 50$ and 0.65 for $n = 75$. The probability of common exogenous characteristic is 50%. Finally, x and ϵ are drawn from a normal distribution with mean 0 and variance 1. The simulation is composed of 500 repetitions. The average of the estimated standard errors, following the procedure outlined in 3.2, is shown in parentheses, while standard deviations of the point estimates computed across replications is shown in square brackets. Simulations are conducted in the absence of information on λ_0.

Simulated results are largely satisfactory in all cases. Convergence to spatial parameters and those that underpin the randomness in networks, is observed, even with small $n = 25$ and $v = 25$. Moreover, the network spillover is correctly estimated. In Table 9 of Appendix F.1, I show that OLS estimates would be inconsistent at averages $\hat{\beta}_{OLS} = 1.0670$ for $n = 25$ and $\hat{\beta}_{OLS} = 1.1127$ for $n = 50$. This bias is eliminated with the proposed method. Introduction of time dimension and fixed effects do not change the results, despite the fact that estimates of σ^2 now take into account that cross-section and time variation has been eliminated as the consequence of data transformation (Subsection 3.4). For the case without time and fixed effects, estimates of disturbance variance is, in most cases, larger than the true value, but this is expected as it captures the misspecification component due to the fact that the observed model is considered under expected networks – naturally different from the true networks. It is also noteworthy that estimated standard errors are very close in most cases to standard errors of point estimates across iterations, demonstrating good performance of the hypothesis testing procedure.

I also show results on three additional cases in Appendix F.1. Tables 10 and 11 shows the performance of the estimator with very low sample sizes. It shows that even with small samples up to $n = 25$ and $v = 50$, estimates are acceptably close to true parameters and confidence intervals are correctly estimated. Then, I introduce across-group connections by randomly assigning value 1 to off-block elements of matrix W_j^0 with probability δ_A. Although not explicitly incorporated in theory, it is shown that a small amount of violation from the isolated-group assumption does not deteriorate empirical performance of the estimator. Performace was good up to $\delta_A = 0.05$ or $\delta_A = 0.075$. Finally I conduct estimation and hypothesis testing when λ_0 is known but misspecified, shown in Table 13 of Appendix F.1. I assume incorrectly $\lambda = 0.0250$, twice the true value. As expected, I observe halved $\hat{\delta}_1$ and $\hat{\delta}_0$ and $\hat{\beta}_2$ estimated twice the true parameter. Associated standard errors followed the same expected pattern.

I also implement the multivariate network model described in example 4 of Subsection B, where probability of link formation is described by

$$P\left\{\{W_j\}_{ik} = 1 | Q_j\right\} = Q^1_{jik}\delta_1 + Q^0_{jik}\delta_0$$

where Q^1_{jik} is the distance between individuals i and k who belong to group j, and respectively for Q^0_{jik}. Distances are sampled from a uniform distribution between -2.5 and 2.5, and probabilities are cut such they do not exceed 1 or fall below 0. True values $\delta_1 = 0.25$ and $\delta_0 = 0.50$, and remaining parameters remain unchanged from previous setting. Results are shown in Table 14 of Appendix F.1 and are also satisfactory with convergence to true parameters and standard errors also being observed at small values of n and v. Estimation of λ using second moments is also satisfactory.

Table 1: Simulations.

	$T=1$.				$T=5$, fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	25	100	25	100	25	100	25	100
v	250	250	1000	1000	250	250	1000	1000
	0.0117	0.0122	0.0126	0.0119	0.0119	0.0120	0.0118	0.0119
$\hat{\lambda}$	(0.002)	(0.006)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.000)
	[0.002]	[0.005]	[0.002]	[0.001]	[0.003]	[0.001]	[0.000]	[0.000]
	0.9999	1.0001	0.9998	0.9999	1.0006	1.0000	1.0002	0.9999
$\hat{\beta}_1$	(0.009)	(0.005)	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)	(0.001)
	[0.009]	[0.005]	[0.005]	[0.003]	[0.004]	[0.002]	[0.002]	[0.001]
	0.0461	0.0402	0.0400	0.0401	0.0428	0.0398	0.0408	0.0400
$\hat{\beta}_2$	(0.018)	(0.003)	(0.008)	(0.001)	(0.008)	(0.001)	(0.003)	(0.001)
	[0.018]	[0.003]	[0.006]	[0.002]	[0.007]	[0.001]	[0.003]	[0.001]
	0.7166	0.7497	0.7605	0.7492	0.7247	0.7510	0.7389	0.7499
$\hat{\delta}_1$	(0.164)	(0.024)	(0.097)	(0.012)	(0.085)	(0.012)	(0.031)	(0.004)
	[0.162]	[0.025]	[0.081]	[0.013]	[0.073]	[0.011]	[0.036]	[0.006]
	0.2892	0.2995	0.3015	0.2992	0.2918	0.3010	0.2966	0.3002
$\hat{\delta}_0$	(0.062)	(0.007)	(0.031)	(0.004)	(0.032)	(0.003)	(0.015)	(0.002)
	[0.063]	[0.007]	[0.030]	[0.004]	[0.027]	[0.003]	[0.014]	[0.002]
	1.0571	1.2199	1.0547	1.2228	0.8421	0.9778	0.8451	0.9774
$\hat{\sigma}^2$	(0.018)	(0.003)	(0.008)	(0.002)	(0.008)	(0.001)	(0.003)	(0.001)
	[0.019]	[0.011]	[0.009]	[0.006]	[0.007]	[0.004]	[0.003]	[0.002]
	-0.0007	0.0137	0.0007	-0.0099	0.0008	-0.0096	0.0005	0.0020
$\varphi(x,\hat{\theta})$	(0.023)	(0.092)	(0.008)	(0.048)	(0.009)	(0.050)	(0.006)	(0.020)
	[0.006]	[0.007]	[0.001]	[0.002]	[0.001]	[0.001]	[0.000]	[0.000]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

Table 2: Simulations.

	$T = 5$, time effects.				$T = 5$, time and fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	25	100	25	100	25	100	25	100
v	250	250	1000	1000	250	250	1000	1000
	0.0121	0.0119	0.0123	0.0118	0.0132	0.0121	0.0120	0.0117
$\hat{\lambda}$	(0.006)	(0.005)	(0.002)	(0.000)	(0.002)	(0.005)	(0.006)	(0.001)
	[0.005]	[0.005]	[0.001]	[0.000]	[0.002]	[0.005]	[0.005]	[0.001]
	1.0001	0.9996	0.9995	0.9999	1.0005	0.9998	1.0000	1.0002
$\hat{\beta}_1$	(0.004)	(0.002)	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)	(0.001)
	[0.004]	[0.002]	[0.002]	[0.001]	[0.004]	[0.002]	[0.002]	[0.001]
	0.0413	0.0400	0.0396	0.0402	0.0411	0.0399	0.0403	0.0402
$\hat{\beta}_2$	(0.007)	(0.001)	(0.003)	(0.001)	(0.007)	(0.002)	(0.004)	(0.001)
	[0.006]	[0.001]	[0.003]	[0.001]	[0.006]	[0.001]	[0.003]	[0.001]
	0.7387	0.7508	0.7588	0.7486	0.7445	0.7524	0.7479	0.7483
$\hat{\delta}_1$	(0.070)	(0.009)	(0.034)	(0.007)	(0.074)	(0.013)	(0.046)	(0.006)
	[0.071]	[0.011]	[0.036]	[0.006]	[0.071]	[0.011]	[0.047]	[0.006]
	0.2958	0.2992	0.3018	0.3003	0.2980	0.2982	0.2969	0.3008
$\hat{\delta}_0$	(0.028)	(0.004)	(0.014)	(0.002)	(0.029)	(0.005)	(0.017)	(0.002)
	[0.027]	[0.003]	[0.013]	[0.002]	[0.027]	[0.003]	[0.015]	[0.002]
	0.0114	0.0195	0.0110	0.0603	0.0423	0.0029	0.0002	0.0019
$\hat{\sigma}^2$	(0.007)	(0.001)	(0.003)	(0.001)	(0.007)	(0.002)	(0.005)	(0.001)
	[0.014]	[0.008]	[0.007]	[0.001]	[0.012]	[0.003]	[0.007]	[0.001]
	0.0005	-0.0050	0.0011	0.0036	0.0012	0.0082	-0.0002	0.0003
$\varphi(x, \hat{\theta})$	(0.009)	(0.039)	(0.005)	(0.016)	(0.008)	(0.035)	(0.005)	(0.021)
	[0.001]	[0.001]	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]	[0.000]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

5 Application

Empirical research has led to substantial interest in evaluating the effects of randomized policies on targeted individuals. Much less progress has been made on evaluating the spillovers related to those policies, possibly because of problems associated with observing and defining interactions among people. The method developed in the present paper provides a comprehensive evaluation of programs when networks are unknown or unreliable, and information on a large number of groups is available and network effects are suspected.

The importance of assessing spillovers is further highlighted when a large proportion of individuals are subject to a shock. This effect raises the possibility that spillovers or externalities play a key role in overall program results (Angelucci et al., 2010). As an example of this setting, I analyze the effect of a randomized intervention in which a large proportion of individuals was simultaneously targeted. This example also illustrates that randomization in treatment variables can be used to estimate network effects, as opposed to randomization in the group formation (Sacerdote, 2001).

I employ data for a large-scale randomized intervention, which provided compelling evidence that occupational choice of the world's poor is determined by a lack of capital and skills (Bandiera et al., 2013). The intervention consisted of the assignment of livestock and skills training, both relevant in terms of the outlay (at approximately USD $140) and duration (training lasted for two years). The authors found significant changes in the occupational choices of the poor, who moved from wage jobs toward self-employment associated with livestock rearing. The program was instituted in 1409 communities, which consisted of clusters of 84 households on average. In each community, households belonging to the bottom quintile of the wealth distribution were identified, and all were eligible for treatment, with certain exceptions. In total, 7953 beneficiaries were surveyed, and all eligible households in the randomly selected communities were treated.

The baseline results comparing the treatment group in selected villages against the treatment group in non-selected villages indicate a dramatic change in the occupational status of targeted households. Four years after treatment, poor women dedicated 92% additional hours to self-employment running their livestock-rearing businesses and moved away from wage hours that were frequently insecure and temporary. This lasting change in occupational status was also associated with higher earnings, higher per capita expenditure, better general wellbeing and higher measures of life satisfaction. After treatment, poor households were classified between near-poor and middle class according to a host of economic indicators.

With recourse to the estimation method developed in this paper, and without network data, I supplement these results with several network-dependent findings. I show that specific program effects are not contained to targeted individuals. Network spillovers affect food expenditure and food security at magnitude around half of the original treatment, but are either insignificant or small determinants of occupational choice and livestock assets. I also shed light on the underlying network structural mechanisms that give rise to these externalities. By separately identifying endogenous and exogenous effects, I am able to estimate the marginal effects of a connection to treated households. I find that the occupational choice of peers of the treated households move in an opposite direction to the treated households: a marginal connection to treated households reduces self-working hours, increases wage hours and decreases livestock value. The magnitudes of the effects are such that exogenous effects counteract 25-30% of the reduction in treated households' wage hours.[26] However, connections

[26]This is the ratio between the increase of wage hours due to exogenous effects and the direct effect of reduction of wage hours.

to the treated households strongly increase food expenditure and food security. These results are consistent with the interpretation that the treated households gained comparative advantage in livestock rearing, which partially changed the occupational choices of their peers. Overall, network effects are shown to form an integral component of the program evaluation.

There is wide consensus that capital, opportunities, income, information and choices affect the outcomes of peers (Jackson, 2010). In fact, the opportunities of others have been regarded as a form of social capital (Glaeser et al., 2002). In this way, a shock to one's peers can be interpreted in the same fashion as a shock to one's self, and the example described here provides evidence of this mechanism. Now, I turn to a description of the program, followed by the identification strategy and the results.

5.1 Program Description

Selection of targeted individuals proceeded in stages. In collaboration with BRAC, a local non-profit organization, the most vulnerable districts were selected based on food-security measures, as described by the World Food Program. Second, BRAC employees selected the poorest communities within each district. Finally, within each community, a combination of a participatory rural appraisal exercise and survey data were used to allocate households to one of five wealth bins. Households belonging to the poorest wealth bins were selected as a potential beneficiary if other eligibility criteria were met, such as not participating as microfinance borrowers and owning no productive assets. Randomization was conducted at the local BRAC branch level, among its 40 offices in Bangladesh, and stratified at the subdistrict level to ensure balance between treated and control groups. Within each subdistrict, one branch was randomly allocated to treatment and another to the control group, and asset transfer was conducted for all selected individuals within the communities covered by the treated BRAC branches. Consequently, a substantial fraction of the community population was treated, raising the possibility that aggregate community-level effects are substantially larger than the sum of isolated individual treatment effects, including, for example, as a consequence of learning, insurance and informal skills reinforcement from neighbors, who in turn may or may not be in the treatment group themselves. If eligible and selected through the randomization process, households received a transfer of live animals (valued at approximately USD \$140) and subsequent skills training for two years that were specifically designed for the chosen asset. Program beneficiaries could select among cows, goats or chickens that added up to the same face value; the large majority chose cows. Participants were required to keep possession of the asset for a minimum of two years, but in practice there were no sanctions in case of noncompliance. All potential beneficiaries of the program and a sample of households across the village wealth distribution were surveyed just before the intervention in 2007 and in two additional waves in 2009 and 2011. The comprehensive survey consisted of household members' sociodemographic characteristics, business assets and activities, land holdings and transfers, financial assets and liabilities, non-business assets, homestead ownership status and improvements, women's empowerment and vulnerability (such as earnings seasonality and food security), and a health module. Network self-reported links were registered when applicable, and data included family outside the household, their business activities, land transfers (through inheritance, mortgage, rent, share, received as dowry or gift, bought or sold), business as-

These are numbers are averages across all individuals in treated villages, considering the number of treated households in each village and the network parameters which affect the number of expected connections. In this case, endogenous effects counteract exogenous effects which combined produce spillovers of smaller magnitudes. See also Subsection 5.2.1.

set transfers (same possibilities as above), finance links (loans, outstanding lending or transfers) and letting of house ownerships. The questionnaire was applied to all selected and a sample of nonselected households in both treatment and control groups.

5.2 Evaluation and Identification Strategies

Treatment effects on the treated could be evaluated comparing the change before and after treatment in the outcomes of selected households who live in a treated village against similar changes in the outcomes of selected households who live in nontreated villages. However, this approach would be unsuitable for estimating the network effects due to two reasons.

First, exclusion of nontreated households in treated and control villages prevents wider evaluation of policy for those groups. Second, as I showed in Subsection 2.1, the outcome of the differences-in-differences estimator is unclear when network effects are present because it may or may not capture network spillovers (φ). The extent to which the spillovers are estimated depends on the degree of reciprocation in the network, which is unobserved. When reciprocation is not present or interaction groups are large enough, Example 1 shows that the estimator is consistent for the individual elasticity in the counterfactual in which households are unconnected (β_{10}). On the other hand, separately estimating network-independent β_{10} from network-dependent φ is also important when the researcher desires to evaluate the policy impact in a setting where networks might considerably differ.

To tackle these issues, I consider a triple differences-in-differences with all households in treated and non-treated villages regardless of selection status. Momentarily ignoring network effects, one could specify a double differences-in-differences which would compare changes in outcomes of the selected households before and after treatment against similar changes in outcomes of the nonselected households. However, this strategy would not be sufficient because randomization was conducted at the village level: selection of potential beneficiaries within the villages was determined according to wealth at the baseline. I take two remedial actions. I introduce household fixed effects and I use the control villages to account for different trends in absence of treatment. The third difference eliminates the change before and after treatment in the outcomes of selected households who live in a nontreated village against similar changes in outcomes of nonselected households who also live in nontreated villages.

The final model is then a triple differences-in-differences with household fixed effects. The identification assumption is that trends as observed in the nontreated villages are a good counterfactuals for trends in treated villages. I denote $S_{ij} = 1$ if individual i of village j was selected as a potential beneficiary of the program and $T_{ij} = 1$ if village j was randomly selected for treatment. The model without networks is

$$y_{ijt} = \sum_{s=2}^{3} \beta_{1s} S_{ij} T_{ij} \mathbf{1}\{s=t\} + \sum_{s=2}^{3} \eta_{1s} S_{ij} \mathbf{1}\{s=t\} + \sum_{s=2}^{3} \eta_{2s} T_{ij} \mathbf{1}\{s=t\} + \gamma_t + \alpha_{ij} + \epsilon_{ijt} \qquad (25)$$

where y_{ijt} represents the outcome for individual i in village j at time t, $s = 2$ and 3 are the second and third survey wave (two and four years after treatment, respectively), α_{ij} is a fixed effect at the individual level, γ_t is a full set of time effects, $\mathbf{1}\{\cdot\}$ is an indicator function, and ϵ_{ijt} is the disturbance term, clustered at the village level. The program impact on the treated in the counterfactual in which households are unconnected are β_{12} and β_{13}.

I next introduce network spillovers, which take the form of two additional network-dependent terms attached to equation (25). Identification in the network setting follows after identification of the treatment effects on the treated, as introduced above, with added assumptions on variability of group sizes and moment condition based on outcome dispersion, as explained in Section 3. The full model in vector notation is

$$\begin{aligned} y_{jt} &= \lambda W_j^0 y_{jt} + \sum_{s=2}^{3} \beta_{1s} ST_j \mathbf{1}\{s=t\} + \sum_{s=2}^{3} W_j^0 ST_j \mathbf{1}\{s=t\}\beta_{2s} + \\ &\quad + \sum_{s=2}^{3} \eta_{1s} S_j \mathbf{1}\{s=t\} + \sum_{s=2}^{3} \eta_{2s} T_j \mathbf{1}\{s=t\} + \gamma_t + \alpha_j + \epsilon_{jt} \end{aligned} \tag{26}$$

where W_j^0 is the unobserved household-level network and ST_j is a column vector with the ith line indicating whether individual i was selected and lives in treated village j. Vector $\alpha_j = [\alpha_{1j}, \dots, \alpha_{n_j j}]$ are household-level fixed effects. The term $\lambda W_j^0 y_{jt}$ represents the endogenous effects – the fact that one's own choice depends on others' choices – and $W_j^0 ST_j \mathbf{1}\{s=t\}\beta_{2s}$ represents exogenous effects, i.e., the dependence of one's own choices on others' treatment status. As explained in Subsection 3.4, the correlated effects are captured by the fixed effects and eliminated via the subtraction of time averages. Coefficients β_{22} and β_{23} are interpreted as the marginal effect of treating a peer. Finally, I average network spillovers $\varphi(x_{jt}, \hat{\theta})$ for treated individuals after two and four years (denoted $\hat{\varphi}_{T,2}$ and $\hat{\varphi}_{T,4}$, respectively) and similarly for nontreated individuals (denoted as $\hat{\varphi}_{NT,2}$ and $\hat{\varphi}_{NT,4}$, respectively). It is notable that the overall treatment effect for the treated individuals is the sum of the program effect and spillovers. The construction of the confidence intervals and standard errors is described in Subsection 3.2.

5.2.1 Alternative Methods for Estimating Network Effects

There are a variety of methods in the literature to estimate network effects. For example, a possibility in the current setting is to compare nonselected households in treated villages against nonselected households in control villages. Other alternatives explored in the literature introduce variation in the fraction of the population assigned to treatment across groups (Crépon et al., 2012). There are two reasons why the current method improves on these approaches.

The first reason is related to precision of the estimates. Consider two polar cases: general equilibrium effects in which social interactions are intermediated solely by the markets (decrease in the supply of wage hours increases wage in the market) and local interactions (wage jobs left by treated households are occupied through network acquaintances). General equilibrium effects means that all individuals are affected to a small extent by the decisions of others. Networks are dense with weak links. In contrast, local interactions imply strong network spillovers only for those connected to treated households and null for unconnected individuals. The latter case generates large variation in individual outcome which then affects the precision of the estimates.

Second, comparison of nonselected households estimates network spillovers only, which can originate from a combination of endogenous and exogenous effects. In the current setting, for instance, the marginal effect of a connection requires separately identifying endogenous and exogenous effects, which is not possible by comparing nonselected households in treated villages against nonselected households in nontreated villages.

5.3 Empirical Results

I consider four sets of outcomes: occupational choice indicators (self-working hours, wage employment hours and specialization in self-employment in Table 3), earnings and seasonality (household earnings, in thousands of Bangladeshi Takas, share of income originating from seasonal and regular activities in Table 4), livestock assets (number of cows, poultry and livestock value in thousands of Takas in Table 5) and per capita expenditures (nonfood and food items and food security in Table 6). As an indicator of differential patterns of association, I allow the probability of link formation to depend on the proximity of household identifiers, registered as $Q_{ij} = 1$ and zero otherwise. It has been anecdotally observed that identifiers were allocated while field surveyors followed local streets and roads, and therefore serve as a proxy for geographical distance. This pattern is only a generalization from the purely naive network in which the probability of link formation is constant and independent of any variable.[27]

For each outcome, I show the triple differences-in-differences estimates of the program effects for the treated households ignoring networks, as in equation (25). These are shown in odd numbered columns in Tables 3-6. For example, column 1 of Table 3 indicates that treated increased self-working hours in 468.9 and 465.1 hours per year, two and four years after treatment respectively, and these results are significant at the 1% confidence level. Even columns display the results of the triple differences-in-differences augmented with the network module, as in equation (26). For example, column 2 of Table 3 also indicates treated increased self-working hours in 469.8 and 460.0 hours per year, two and four years after treatment respectively. These numbers are not significantly different from the cases in which networks were ignored in column 1. Therefore, in this particular case, inconsistency due to omission of networks was not a relevant problem.

The following four rows display the results for the network spillovers. Results in this case are not significant at 10% level two years after treatment for treated and nontreated, and point estimates are -6.3 and -3.2 hours per year. However, spillovers are positive and significant four years after treatment at 28.8 and 14.7 self-working hours per year for treated and nontreated respectively, indicating a slight increase in the supply of self-working hours due to spillovers for both types of households. The estimates for the program effect on treated and spillovers, as discussed above, does not depend on separately identifying endogenous and exogenous effects and, hence, do not rely on the presence of group size asymmetries and the moment condition based on outcome dispersion.

Breaking down spillovers in endogenous and exogenous effects then allows me to estimate the marginal effect of the connection to a treated household. These rows are labelled "Link to T". A marginal connection reduces working hours in 24.6 and 17.9 hours per year two and four years after treatment respectively, and are significant at the 1% confidence level. The probabilities of link formation are very high, at 98.3% if individuals live in close vicinity, and 39.6% otherwise indicating that, in this case, network effects operate via general equilibrium. The hypothesis that these numbers are equal is rejected at the 1% level.

I present the remaining results in three stages. First, I describe the results for network spillovers for all outcomes. These are followed by the estimates of the network structure and the marginal effect of a connection

[27] Estimation with naive model for probability of link formation is conducted as a robustness in Table 19 in Appendix F.2. In addition, estimation without fixed effects, time effects and both are also shown to highlight that in their absence network estimates are highly biased.

to a treated household. Finally, I incorporate network data directly into the procedure and demonstrate that the main conclusions remain unchanged. I also show that family self-reported links convey meaningful interaction and mixed results for economic (non-family) links.

5.3.1 Network Spillovers.

As shown in Subsection 3.2, it is not necessary to identify the parameters that underpin network formation or those that link explanatory variables to outcomes in a given network, and it is also not necessary to separate endogenous and exogenous effects. It is sufficient that Proposition 2 ensures that spillovers are constant within the identified set.

The current application shows that spillovers on treated and non-treated individuals determined outcomes to a relevant degree. The effect of spillovers was particularly salient in explaining food per capita expenditures. For example, spillovers amounted to 207.0 Takas per year for non-treated individuals after two years, compared with an estimated program effect of 423.9 Takas for treated individuals over the same period. This difference corresponds to a 6.9% increase on top of baseline levels of consumption, or 48.8% of the treatment effect on the treated individuals. The spillover effect is even larger for the treated subpopulation. After two years, spillovers from the treated households to themselves were responsible for an expenditure increase of 380.0 Takas, or 89.6% of the treatment effects. Notably, column 3 of Table 3 shows that estimates of treatment effects when networks are not included in the analysis are approximately 40% higher. This difference is attributed to the fact that OLS estimates, as presented in Subsection 2.1, may be inconsistent when networks effects are not accounted for.

This result is further confirmed by estimates of food security that are measured by respondents that reported having at least two meals on most days, indicating a positive effect for both the treated and the non-treated groups, across two and four years, ranging from 2.7 percentage points for the non-treated group two years after treatment to 7.1 percentage points for the treated group at the same time. The direct program effect is estimated at 16.9 and 7.6 percentage points (after two years and four years, respectively). Nonfood expenditures are either constant or exhibit a slight increase for the treated group, whereas the non-treated group reduced nonfood consumption after four years. As discussed below, this result can be explained by the reduction in productive assets following the specialization of the peers of the treated group in terms of wage labor.

Spillovers were significant to a small extent in determining self-employment and wage hours, specialization in self-employment, the share of seasonal and regular activities and asset holdings. As discussed above, network spillovers are reduced-form estimates that consist of endogenous effects, or the fact that one's own choice depends on others' choices, and exogenous effects, the fact that one's own choices depend on the treatment of others. Disentangling these structural mechanisms is useful in shedding light on the causes of these results, and this is undertaken in the next Subsection.

5.3.2 Endogenous and Exogenous Effects (or Marginal Value of Connections to the Treated)

I now provide point estimates of structural parameters. Given a network, its full set consists of link strength (λ), one's own response to one's own treatment after two and four years (β_{11} and β_{12}) and exogenous effects (or, in the current setting, the effect of one additional connection to a treated individual, β_{21} and β_{22}). The parameters that capture the network link are the probability of link formation if households are located in close proximity (δ_1), such that $Q_{ij} = 1$ if the difference in household identifiers is less than two[28], and if households are not in close proximity (δ_0). These parameters discriminate between the polar cases in which interactions occur on a localized scale, through personal interconnections and without intermediation of the markets (equivalent to low-density networks, or low δ_0 and δ_1) or through general equilibrium effects in which one's own choices affect all others to a small degree and result in dense networks (high δ_0 and δ_1). As demonstrated in Theorem 4, identification is achieved using the comparison between observed and theoretical across-group dispersion of outcomes as implied by the model. In a social setting, the across-group variation of outcomes cannot be explained by outcome dispersion, peer group heterogeneity or disturbance variance alone. This indicates a moment condition and suggests the use of a GMM criterion that is capable of sorting among structural parameters within the identified set.

In the current application, the estimates show that, whereas treated individuals reduced wage hours (113.5 and 141.9 hours per year, two and four years after treatment, respectively) and increased self-employment hours (469.8 and 460.0 hours per year) associated with livestock rearing, a marginal connection to a treated household had the opposite effect, increasing wage hours (24.6 and 17.9 hours per year for each treated peer) and decreasing self-working hours (13.9 and 13.0 hours per year for each treated peer). Treated individuals specialize in self-employment, and connected peers modestly decrease specialization. Individuals who received treatment left vacancies on wage jobs that were partially filled by individuals located in close geographic proximity[29]. The density of estimated networks is high only for self-employment and wage hours; above 90% for households that live in close proximity and approximately 40% otherwise. The interaction patterns of all other outcomes are much more localized, with densities of approximately 20% or lower in most cases.

The results demonstrate that treated individuals increased their livestock assets by more than the original treatment. Meanwhile, non-treated individuals reduced their stock of assets. This outcome was not observed for poultry, which is consistent with the low takeover rate of this type of asset. Livestock value followed the same pattern for both groups. Since the treatment also consisted of skills training – specifically targeted for the type of assets provided – and was of long duration (2 years), treated individuals were endowed with a stronger comparative advantage in livestock rearing, whereas connected peers tended to specialize in wage jobs instead.

The final component of the analysis involves the food staples. A marginal connection to a treated peer significantly increases food consumption per capita and food security. In fact, one connection may be responsible for an effect on food expenditures that is equivalent to the direct effect of treatment on the treated individual (443.6 versus 423.9 Takas) and a 9.6 percentage point increase in food security. This finding shows that comovements of occupational choices of the treated and their peers were largely beneficial to all.

5.3.3 Including Network Data

Finally, I make use of network data collected in the survey to reassess the conclusions obtained in their absence. Inclusion of network data serves two primary purposes. First, I show that the main conclusions summarized above remain unchanged (Tables 15 to 18 of Appendix F.2). Second, allowing link formation to depend on

[28] Robusteness checks are conducted in Table 19 of Appendix F.2.

[29] The null hypothesis of no differential association is rejected at the 5% level for all specifications, as shown in Tables (3)-(6). Given the estimated parameters and the number of treated households in each households, a simple simulation exercise shows that exogenous effects counterbalanced 25-30% of the reduction in wage hours of treated households.

link reporting enables me to test whether the associated coefficient is significant, which constitutes as a test of

Table 3: Occupational Choice.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Self hours.		Wage hours.		Self emp. only.	
	Method	OLS.	Network.	OLS.	Network.	OLS.	Network.
Not function of $\hat{\lambda}$.	Program effect	468.928***	469.774***	−110.799***	−113.531***	0.107***	0.114***
	after 2 years ($\hat{\beta}_{11}$).	(28.62)	(23.20)	(31.07)	(10.61)	(0.02)	(0.01)
	Program effect	465.075***	460.039***	−137.255***	−141.918***	0.112***	0.120***
	after 4 years ($\hat{\beta}_{12}$).	(31.32)	(23.21)	(34.10)	(8.63)	(0.02)	(0.01)
	Spillover on T	—	−6.347	—	26.855***	—	−0.032***
	after 2 years ($\hat{\varphi}_{T,2}$).		(10.55)		(8.45)		(0.01)
	Spillover on T	—	28.847***	—	19.369**	—	−0.025***
	after 4 years ($\hat{\varphi}_{T,4}$).		(9.68)		(8.54)		(0.00)
	Spillover on NT	—	−3.229	—	14.491***	—	−0.018***
	after 2 years ($\hat{\varphi}_{NT,2}$).		(5.37)		(4.55)		(0.00)
	Spillover on NT	—	14.676***	—	10.452***	—	−0.013***
	after 4 years ($\hat{\varphi}_{NT,4}$).		(1.09)		(0.75)		(0.00)
Function of $\hat{\lambda}$.	Link to T	—	−24.604***	—	13.904***	—	−0.050***
	after 2 years ($\hat{\beta}_{21}$).		(2.76)		(2.52)		(0.01)
	Link to T	—	−17.932***	—	13.030***	—	−0.043***
	after 4 years ($\hat{\beta}_{22}$).		(2.76)		(1.59)		(0.01)
	Link probability	—	0.983***	—	0.639***	—	0.192***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).		(0.03)		(0.03)		(0.01)
	Link probability	—	0.396***	—	0.331***	—	0.106***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).		(0.01)		(0.01)		(0.00)
	Link strength	—	0.05***	—	0.05***	—	0.15***
	($\hat{\lambda}$).		(0.01)		(0.00)		(0.01)
	p-value $\mathcal{H}_{NV}$.	—	< 0.001	—	< 0.001	—	< 0.001
	Avg treated outcome.	421.8	421.8	646.7	646.7	0.303	0.303
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes: *, ** and *** indicates significance at 10%, 5% and 1% levels. All regressions have household fixed effects. Standard errors clustered at the village level. "Spillover on T" refers to the average $\varphi(x_t, \hat{\theta})$ on the treated only. "Spillovers on NT" refers to equivalent calculation on the non-treated only. "Link to T" refers to the marginal effect of a connection to a treated individual. "Avg treated outcome" refers to the mean outcome of treated at the baseline. "p-value $\mathcal{H}_{NV}$" is the p-value of testing the null hypothesis that household proximity does not affect the probability of link formation. Estimates dependent on the identification strategy for $\hat{\lambda}$ are denoted under the tab "Function of $\hat{\lambda}$". "Self hours" refers to self-working hours per year. "Wage hours" refers to wage working hours per year. "Self emp. only" is a dummy variable if individual is specialized in self-employment.

Table 4: Earnings and Seasonality.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Earnings.		Share Seas.		Share Reg.	
	Method	OLS.	Network.	OLS.	Network.	OLS.	Network.
Not function of $\hat{\lambda}$.	Program effect	0.475	0.506***	0.012	−0.028***	0.201***	0.181***
	after 2 years ($\hat{\beta}_{11}$).	(0.46)	(0.12)	(0.02)	(0.01)	(0.02)	(0.01)
	Program effect	2.598***	2.729***	−0.089***	−0.074***	0.191***	0.165***
	after 4 years ($\hat{\beta}_{12}$).	(0.54)	(0.31)	(0.02)	(0.01)	(0.02)	(0.01)
	Spillover on T	—	−0.045	—	−0.051***	—	0.023**
	after 2 years ($\hat{\varphi}_{T,2}$).		(0.10)		(0.02)		(0.01)
	Spillover on T	—	0.008	—	−0.005	—	0.029**
	after 4 years ($\hat{\varphi}_{T,4}$).		(0.11)		(0.02)		(0.01)
	Spillover on NT	—	−0.025	—	−0.023***	—	0.012**
	after 2 years ($\hat{\varphi}_{NT,2}$).		(0.06)		(0.01)		(0.00)
	Spillover on NT	—	0.004	—	−0.002	—	0.015***
	after 4 years ($\hat{\varphi}_{NT,4}$).		(0.09)		(0.01)		(0.00)
Function of $\hat{\lambda}$.	Link to T	—	−0.447	—	−0.010***	—	−0.022***
	after 2 years ($\hat{\beta}_{21}$).		(0.46)		(0.01)		(0.01)
	Link to T	—	−0.326	—	−0.016***	—	−0.015**
	after 4 years ($\hat{\beta}_{22}$).		(0.29)		(0.01)		(0.00)
	Link probability	—	0.075***	—	0.272***	—	0.238***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).		(0.00)		(0.01)		(0.00)
	Link probability	—	0.023***	—	0.136***	—	0.106***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).		(0.00)		(0.00)		(0.00)
	Link strength	—	0.50***	—	0.20***	—	0.20***
	($\hat{\lambda}$).		(0.17)		(0.08)		(0.05)
	p-value $\mathcal{H}_{NV}$.	—	< 0.001	—	< 0.001	—	< 0.001
	Avg treated outcome.	4.607	4.607	0.674	0.674	0.478	0.478
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes: Earnings in thousand of Takas per year. "Share Seas." refers to the share of seasonal earnings relative to total earnings. "Share Reg." refers to share of regular earnings, as reported by the respondent, relative to total earnings. See also Table 3.

network data validity. I combine network reports into two categories: family and economic (non-family) links. Non-family links include an ensemble of many categories of self-reported links, such as business and labor relationships, financial assets and liabilities and household ownership. The null hypothesis of no network validity was rejected at the 1% level for all specifications regarding occupational choice, earnings and seasonality. The results for livestock holding and expenditures are more nuanced. Whereas for most specifications, the null of no validity was rejected for family links, economic links are much less capable of conveying interactions that influence the outcomes of others. This result suggests that families are natural *loci* that favor asset transactions, particularly when those transactions involve cows, and through which food consumption and expenditures flow.

Table 5: Livestock.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Cows.		Poultry.		Livestock Value.	
	Method	OLS.	Network.	OLS.	Network.	OLS.	Network.
Not function of $\hat{\lambda}$.	Program effect	1.119***	1.131***	2.147***	2.120***	10.326***	10.417***
	after 2 years ($\hat{\beta}_{11}$).	(0.04)	(0.03)	(0.42)	(0.50)	(0.56)	(0.39)
	Program effect	1.078***	1.102***	1.294**	1.326***	10.984***	11.175***
	after 4 years ($\hat{\beta}_{12}$).	(0.03)	(0.03)	(0.62)	(0.50)	(0.64)	(0.40)
	Spillover on T	—	−0.033***	—	0.099	—	−0.221***
	after 2 years ($\hat{\varphi}_{T,2}$).		(0.01)		(0.17)		(0.07)
	Spillover on T	—	−0.057***	—	−0.087	—	−0.459***
	after 4 years ($\hat{\varphi}_{T,4}$).		(0.00)		(0.20)		(0.07)
	Spillover on NT	—	−0.020***	—	0.059	—	−0.132***
	after 2 years ($\hat{\varphi}_{NT,2}$).		(0.01)		(0.10)		(0.04)
	Spillover on NT	—	−0.033***	—	−0.052	—	−0.274***
	after 4 years ($\hat{\varphi}_{NT,4}$).		(0.01)		(0.08)		(0.04)
Function of $\hat{\lambda}$.	Link to T	—	−0.996***	—	1.277	—	−10.456***
	after 2 years ($\hat{\beta}_{21}$).		(0.16)		(4.12)		(1.90)
	Link to T	—	−1.285***	—	−2.725	—	−16.464***
	after 4 years ($\hat{\beta}_{22}$).		(0.17)		(4.11)		(2.33)
	Link probability	—	0.024***	—	0.007**	—	0.013***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).		(0.00)		(0.00)		(0.00)
	Link probability	—	0.012***	—	0.009***	—	0.007***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).		(0.00)		(0.00)		(0.00)
	Link strength	—	0.50***	—	0.50	—	0.50***
	($\hat{\lambda}$).		(0.03)		(0.38)		(0.16)
	p-value $\mathcal{H}_{NV}$.	—	< 0.001	—	< 0.001	—	< 0.001
	Avg treated outcome.	0.083	0.083	1.79	1.79	0.940	0.940
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes: "Cows" refers to the number of cows held by the household, and similarly for poultry. Livestock value evaluates in thousands of Takas at market value. See also Table 3.

Table 6: Expenditures.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Nonfood PCE.		Food PCE.		Food Security.	
	Method	OLS.	Network.	OLS.	Network.	OLS.	Network.
Not function of $\hat{\lambda}$.	Program effect	-242.239	-220.509	585.304^{**}	423.929^{***}	0.189^{***}	0.169^{***}
	after 2 years ($\hat{\beta}_{11}$).	(293.34)	(164.53)	(247.19)	(134.22)	(0.03)	(0.01)
	Program effect	175.022	278.277	585.415^{***}	445.063^{***}	0.010^{***}	0.076^{***}
	after 4 years ($\hat{\beta}_{12}$).	(375.16)	(174.72)	(227.38)	(134.27)	(0.03)	(0.01)
	Spillover on T	—	-8.526	—	380.002^{***}	—	0.017^{***}
	after 2 years ($\hat{\varphi}_{T,2}$).		(68.25)		(55.82)		(0.00)
	Spillover on T	—	-171.985^{**}	—	243.172^{***}	—	0.071^{***}
	after 4 years ($\hat{\varphi}_{T,4}$).		(68.15)		(56.88)		(0.02)
	Spillover on NT	—	-5.039	—	206.992^{***}	—	0.027^{***}
	after 2 years ($\hat{\varphi}_{NT,2}$).		(40.34)		(30.14)		(0.00)
	Spillover on NT	—	-101.655^{*}	—	132.459^{***}	—	0.032^{***}
	after 4 years ($\hat{\varphi}_{NT,4}$).		(52.65)		(40.73)		(0.01)
Function of $\hat{\lambda}$.	Link to T	—	-14.185	—	443.619^{***}	—	0.096^{***}
	after 2 years ($\hat{\beta}_{21}$).		(988.46)		(85.36)		(0.01)
	Link to T	—	-2649.43^{***}	—	249.126^{***}	—	0.087^{***}
	after 4 years ($\hat{\beta}_{22}$).		(980.96)		(84.79)		(0.01)
	Link probability	—	0.032^{***}	—	0.132^{***}	—	0.128^{***}
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).		(0.00)		(0.01)		(0.00)
	Link probability	—	0.009^{***}	—	0.080^{***}	—	0.052^{***}
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).		(0.00)		(0.00)		(0.00)
	Link strength	—	0.50^{***}	—	0.20^{**}	—	0.50^{**}
	($\hat{\lambda}$).		(0.14)		(0.11)		(0.21)
	p-value $\mathcal{H}_{NV}$.	—	< 0.001	—	< 0.001	—	< 0.001
	Avg treated outcome.	1054.5	1054.5	2953.7	2953.7	0.457	0.457
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes: "Nonfood PCE" refers to non-food per capita expenditure in thousands of Takas per year, and similarly for food per capita expenditures. Food security is a dummy equal to one if households have at least two meals in most days. Estimates of the program impact on nonfood per capita expenditure on the treated using the triple differences model (column 1) was the only case which does not match well the estimates obtained from the double differences which compares the selected individuals in treated villages against selected in nontreated villages. See Bandiera et al. (2013) and Table 3.

6 Conclusion

Social and economic networks are useful for understanding many aspects of individual choice, decisions and behavior. Although there has recently been substantial progress on the theoretical underpinnings of network formation, empirical research frequently remains constrained by the availability of network data. The contribution of this paper is then to provide a method for estimating network effects in the absence of network data. The method also estimates the probability that pairs of individuals form a connection based on individual characteristics such as common gender. I also incorporate imperfect network data with the dual purpose of refining the estimates and providing a test for its validity.

The key contribution of the paper was to derive a maximum likelihood estimator that is not conditional on network data. It is obtained by integrating a likelihood conditional on networks which originates from a spatial econometric model with respect to the probability density function of the stochastic network. In this setting, I showed how the observation of outcomes and explanatory variables for many groups such as classrooms serves as a substitute for the network observation. This approach then offers a procedure for estimating network effects using datasets that were previously not suited for this purpose.

Empirical research has led to substantial interest in evaluating the effects of randomized policies on targeted individuals. Much less progress has been made on evaluating the spillovers related to those policies. To illustrate how the method can be applied in practice, I employed the estimator to investigate the impact of a large-scale randomized intervention on the peers of those who were treated. This is the intervention of Bandiera et al. (2013), which consisted of the provision of livestock and skill training to low-income households in Bangladesh.

The proposed estimator met three objectives and yielded useful insights on the wider effects of the policy. The first objective was to provide – in the absence of network data – a consistent and asymptotically normal estimator of network spillovers. In the application, I found that network spillovers were economically and statistically significant in determining some outcomes, especially food per capita expenditure and food security. Network spillovers were responsible for an increase of 206.9 Takas in yearly food per capita expenditure compared with a treatment effect of 423.9 Takas on the treated.[30]

The second objective of the paper was to elucidate the structural mechanisms that gave rise to these spillovers. I derived a method to separately identify endogenous and exogenous effects, controlling for correlated effects, in the absence of network data by using the variability in group sizes. I further solved the problem of separately identifying a few strong links from a large number of weak links by using the "excess" outcome variance that cannot be explained by independent variables or peer group heterogeneity alone.[31] For this purpose, I reinterpreted the estimator as the solution of a Generalized Method of Moments problem in which moment conditions were given by the score of the likelihood. In this case, the earlier identification difficulty originated from the absence of one moment condition relative to the number of parameters. I then explored the difference between observed second moments of the outcomes and those implied by the model to provide an additional restriction which completes the identification requirements. I am then able to show that the solution of this problem is a consistent and asymptotically normal estimator to the structural parameters of the model.

[30] Respectively an 14% and 7% increase relative to food consumption levels at the baseline.

[31] These are similiar in essence to the identification ideas in Lee (2007) and Graham (2008), which explore the case in which networks are observed.

In the application studied herein, I found that a marginal connection to the treated led to effects in opposite direction to the treatment effect on the treated. Regarding occupational choice and livestock value, one additional connection to a treated household decreased self-employment by 24.6 hours per year, added 13.9 wage hours per year and decreased livestock value by 10.4 thousand Takas. Treated households increased their self-employment hours, decreased their wage hours and increased the value of their livestock. In constrast, regarding food per capita expenditure and food security, a marginal connection to the treated was in the same direction to the treatment effect on the treated, and often of strong magnitudes. A marginal connection to the treated increased food per capita expenditure by 443.6 Takas per year and increased food security by 9.6 percentage points, compared with direct treatment effects of, respectively, 424.0 Takas per year and 16.9 percentage points. With the exception of self-employment and wage hours, I also found that network densities were fairly low, which suggested local interactions through personal contacts rather than through prices and markets. These results are consistent with the interpretation that treated individuals gained comparative advantage in livestock rearing. The randomized policy then generated a village-level occupational specialization in which treated households were employed in rearing the livestock, partially changing the occupational choice and well-being of their peers as measured by food consumption.

The third objective of this paper was to incorporate imperfect network data, such as when data are self-reported, with the dual purpose of refining the estimates and proposing a test for whether reported connections positively affect the estimated connection probability. In the application, I found that reported family links have a greater effect than the reported economic (non-family) links in determining the outcomes of others. The test rejected the null hypothesis that family links do not influence the number of cows but failed to reject the similar influence of economic links. The same holds true for livestock value, indicating that family ties facilitated asset transactions.

The method developed in the present paper contributes to the spatial econometrics literature that has to date considered only models for which networks are accurately known (Anselin (2010) and references therein). Similarly, the literature on the identification of network models addressed a number of techniques only when networks could be observed (Manski (1993), Bramoullé et al. (2009), De Giorgi et al. (2010) and others). This novel method can be applied in many fields, from peer effects (Ammermuller and Pischke, 2009), crime and delinquent behavior (Glaeser et al., 1996) to the estimation of parameters of gravity equations (Anderson and van Wincoop, 2003).

The interest in networks to this date has not been matched with availability of network data, possibly because of problems associated with observing and defining interactions among people. The method developed in the present paper provided a systematic procedure for estimating network effects when networks are unknown or unreliable and information on a large number of groups is available. This ability has shown to be particularly relevant in estimating effects of exogenous variation policy through randomized controlled trials both on treated and their peers. In this way, the paper demonstrated both theoretically and empirically that including network effects may have important implications for policy assessments. Estimating network spillovers and distinguishing among endogenous, exogenous and correlated effects in the absence of network data is certainly a useful empirical tool for future applied research.

References

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, September 2012.

A. Ambrus, A. G. Chandrasekar, and M. Elliott. Social investments, informal risk sharing and inequality. Unpublished Memo., September 2014.

A. Ammermuller and J.-S. Pischke. Peer effects in european primary schools: Evidence from pirls. *Journal of Labor Economics*, 27 (3):315–348, 2009.

J. Anderson and E. van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1): 170–192, March 2003.

J. D. Angrist and K. Lang. Does school integration generate peer effects? evidence from boston's metco program. *American Economic Review*, 94(5):1613–1634, December 2004.

L. Anselin. *Spatial Econometrics: Methods and Models*. Springer, August 1988.

L. Anselin. Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25, 2010.

B. Ballobás. *Modern Graph Theory*. Springer, corrected 2nd edition edition, 2013.

O. Bandiera and I. Rasul. Social networks and technology adoption in northern mozambique. *The Economic Journal*, 116:869–902, October 2006.

O. Bandiera, R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman. Can basic entrepreneurship transform the economic lives of the poor. *STICERD Working Paper Series*, (EOPP 43), April 2013.

L. Blume, W. A. Brock, S. N. Durlauf, and Y. Ioannides. Identification of social interactions. In J. Behabib, A. Bisin, and M. O. Jackson, editors, *Handbook of Social Economics*, volume 1B. Noth-Holland, 2011.

Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1): 41–55, 2009.

F. Bugni. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, 78(2):735–753, March 2010.

A. Calvó-Armengol and M. O. Jackson. The effects of social networks on employment and inequality. *American Economic Review*, 94(3):426–454, June 2004.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrican*, 75(5):1243–1284, September 2007.

T. G. Conley and G. Topa. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17: 303–327, 2002.

T. G. Conley and C. R. Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, 2010.

B. Crépon, E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. Technical Report 18597, NBER Working Paper, 2012.

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

M. Dell. Trafficking networks and the mexican drug war. Unpublished Memo., December 2012.

R. Diestel. *Graph Theory*. Graduate Texts in Mathematics. Springer, 2010.

A. D. Foster and M. R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209, December 1995.

O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

E. Glaeser, B. Sacerdote, and J. A. Scheinkman. Crime and social interactions. *The Quarterly Journal of Economics*, 111(2): 507–548, May 1996.

E. Glaeser, D. Laibson, and B. Sacerdote. An economic approach to social capital. *The Economic Journal*, 112(483):F437–F458, November 2002.

C. Gouriéroux and A. Monfort. *Simulation-Based Econometric Methods*. Core Lectures. Oxford University Press, 1997.

B. S. Graham. Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):645–660, May 2008.

M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

E. Hanushek. Teacher characteristics and gains in student achievement: estimation using micro data. *American Economic Review*, 61(2), May 1971.

P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, March 1981.

T. Hsing and W. B. Wu. On weighted u-statistics for stationary processes. *The Annals of Probability*, 32(2):1600–1631, 2004.

M. O. Jackson. *Social and Economic Networks*. Princeton University Press, November 2010.

H. H. Kelejian and I. R. Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

H. H. Kelejian and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533, 1999.

H. H. Kelejian and I. R. Prucha. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104:219–57, 2001.

H. H. Kelejian and I. R. Prucha. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157:53–67, 2010.

C. Lam and P. C. Souza. Detection and estimation of block structure in spatial weight matrix. *Econometric Reviews*, August 2014.

L.-F. Lee. Asumptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72 (6):25, November 2004.

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

L.-F. Lee and J. Yu. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154: 165–185, 2010.

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

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

C. F. Manski. Economic analysis of social interactions. *The Journal of Economic Perspectives*, 14(3):115–136, Summer 2000.

E. Miguel and M. Kremer. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, January 2004.

J. D. Montgomery. Social networks and labor-market outcomes: toward an economic analysis. *American Economic Review*, 81(5): 1408–1418, December 1991.

K. Munshi. Networks in the modern economy: Mexican migrants in the u.s. labor market. *Quarterly Journal of Economics*, 118(2): 549–599, May 2003.

W. Newey and D. McFadden. Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, 1994.

M. Pellizzari. Do friends and relatives really help in getting a good job? Technical Report 623, Centre for Economic Performance Working Paper, March 2004.

T. Plümper and E. Neumayer. Model specification in the analysis of spatial dependence. *European Journal of Political Research*, 49(3):418–442, 2010.

A. Rees. Information networks in labor markets. *American Economic Review*, 56(1/2):559–566, March 1966.

S. G. Rivkin, E. Hanushek, and J. F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, March 2005.

J. R. Romano and A. M. Shaikh. Inference for the identified set in partially identified econometric models. *Econometrica*, 78(1): 169–211, January 2010.

T. J. Rothenberg. Identification in parametric models. *Econometrica*, 39(3):577–591, 1971.

B. Sacerdote. Peer effects with random assingment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116 (2):681–704, May 2001.

T. A. B. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 37:131–53, 2011.

D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409): 204–212, 1990.

S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, 1994.

S. Wasserman and P. Pattison. Logit models and logistic regression for social networks: An introduction to markov graphs and p*. *Psychometrika*, 61(3):401–425, September 1996.

A Summary of Notation.

$\beta = (\beta_1', \beta_2')$, $\theta_c = \theta \setminus \{\beta, \sigma^2\}$, $n = \sum_{j=1}^{v} n_j$.

I_n an identity matrix of dimensions $n \times n$, ι_n is a $n \times 1$ vector of ones.

$y_n = \left(y_1', \ldots, y_j', \ldots, y_v'\right)'$, $y_j = y_{n_j,j} \left(y_{1j}, \ldots, y_{ij}, \ldots, y_{n_j j}\right)'$, $j = 1, \ldots v$, $i = 1, \ldots, n_j$.

$x_n = \left(x_1', \ldots, x_j', \ldots, x_v'\right)'$, $x_j = x_{n_j,j} = \left(x_{1j}', \ldots, x_{ij}', \ldots, x_{n_j j}\right)'$, $j = 1, \ldots v$, $i = 1, \ldots, n_j$.

$\epsilon_n = \left(\epsilon_1', \ldots, \epsilon_j', \ldots, \epsilon_v'\right)'$, $\epsilon_{n_j,j} = \left(\epsilon_{1j}, \ldots, \epsilon_{ij}, \ldots, \epsilon_{n_j j}\right)'$, $j = 1, \ldots v$, $i = 1, \ldots, n_j$.

$y_j = \lambda_0 W_j^0 + x_j \beta_{10} + W_j^0 x_j \beta_{20} + v_j$, $v_j = \rho_0 M_j^0 v_j + \epsilon_j$.

$Z_j^0 = Z_{n_j,j}^0 = \left(x_j, W_j^0 x_j\right)$, $Z_n^0 = \left(Z_1^{0'}, \ldots, Z_v^{0'}\right)'$.

$y_j = \lambda W_j^e \left(Q, \theta_c\right) + x_j \beta_1 + W_j^e \left(Q, \theta_c\right) x_j \beta_2 + \hat{v}_j$.

$Z_j^e \left(Q_j, \theta_c\right) = Z_{n_j,j}^e \left(Q, \theta_c\right) = \left(x_j, W_j^e \left(Q, \theta_c\right) x_j\right)$, $Z_n^e \left(Q, \theta_c\right) = \left(Z_1^{e'} \left(Q_1, \theta_c\right), \ldots, Z_v^{e'} \left(Q_v, \theta_c\right)\right)'$.

$W_n^0 = \text{diag}\left(W_1^0, \ldots, W_v^0\right)$, $W_n^e \left(Q_n, \theta_c\right) = \text{diag}\left(W_1^e \left(Q_1, \theta_c\right), \ldots, W_v^e \left(Q_v, \theta_c\right)\right)$.

$S_j^0 (\lambda) = S_{n_j,j}^0 (\lambda) = I_{n_j} - \lambda W_j^0$, $S_j^0 = S_j^0 (\lambda_0)$, $S_n^0 = \text{diag}\left(S_1^0, \ldots, S_v^0\right)$.

$\left(S_n^0\right)^{-1} = \lambda_0 G_n^0 + I_n$, $G_n^0 = W_n^0 \left(S_n^0\right)^{-1}$.

$S_j^e \left(Q_j, \theta\right) = I_{n_j} - \lambda W_j^e \left(Q_j, \theta\right)$, $S_n^e \left(Q_j, \theta_c\right) = \text{diag}\left(S_1^e \left(Q_1, \theta_c\right), \ldots, S_v^e \left(Q_1, \theta_c\right)\right)$.

$\left(S_n^e \left(Q, \theta_c\right)\right)^{-1} = I_n + \lambda G_n^e \left(Q_n, \theta_c\right)$, $G_n^e \left(Q_n, \theta_c\right) \equiv W_n^e \left(Q_n, \theta_c\right) \left(S_n^e \left(Q_n, \theta_c\right)\right)^{-1}$.

$R_j^0 (\rho) = I_{n_j} - \rho M_j^0$, $R_j^0 = R_j^0 (\rho_0)$, $R_n^0 = \text{diag}\left(R_1^0, \ldots, R_v^0\right)$.

$R_j^e (\theta) = I_{n_j} - \rho M_j^e \left(Q_j, \theta\right)$, $R_n^e \left(Q_n, \theta_c\right) = \text{diag}\left(R_1^e \left(Q_1, \theta_c\right), \ldots, R_v^e \left(Q_v, \theta_c\right)\right)$.

$M_n^0 = \text{diag}\left(M_1^0, \ldots, M_v^0\right)$, $M_n^e \left(Q_n, \theta_c\right) = \text{diag}\left(M_1^e \left(Q_1, \theta_c\right), \ldots, M_v^e \left(Q_v, \theta_c\right)\right)$.

$P_n^e \left(Q_n, \theta_c\right) = I_n - R_n^e \left(Q_n, \theta_c\right) Z_n^e \left(Q_n, \theta_c\right) \left[Z_n^{e'} \left(Q_n, \theta_c\right) R_n^{e'} \left(Q_n, \theta_c\right) R_n^e \left(Q_n, \theta_c\right) Z_n^e \left(Q_n, \theta_c\right)\right]^{-1} Z_n^{e'} \left(Q_n, \theta_c\right) R_n^{e'} \left(Q_n, \theta_c\right)$.

$B_n \left(Q_n, \theta_c\right) = \lambda \left(W_n^0 - W_n^e \left(Q, \theta_c\right)\right) \left(I_n + \lambda_0 G_n^0\right)$.

$\mu [A]$ and $\Sigma [A]$ denote the expectation and variance-covariance matrix of vector A.

B Alternative network models.

I previously described the probability of link formation as dependent on a dummy for sharing exogenous characteristic with independence link formation. I now expand the classes of models in two different directions: I first allow the probability of link formation to depend on a continuous measure, such as distance between households location. Because many modes of social interactions can occur in parallel, it is also important to allow for a multivariate network formation model. In second place, I drop link independence assumption with recourse to the Exponential Random Markovian Graphs (ERMG) family of models, as introduced by Frank and Strauss (1986) and expanded by Wasserman and Pattison (1996). These are presented in form of examples.

Example 4. *(Multivariate network model).* Several forms of relations coexist; arguably, a truthful representation of the probability of link formation will then depend on a number of factors. Allow then Q_{ji}^I as $1 \times k^I$ to be a matrix of individual's i characteristics that underpin probability of link formation and depend exclusively on individual, non-relational, characteristics. For example, this may encompass testing whether males may tend to form more connections than the rest of the population, or personal income may have a relation to social interactions. Let Q^R_{jk} be characteristics of the potential recipient of the link that may generate attraction, of dimension $1 \times k^R$ and, finally, Q^B_{jik} common, shared characteristics, such as belonging to the same gender, or continuous geographic distance between households, with dimension $1 \times k^B$. Coefficients are captured with recourse to θ^I_g, θ^R_g and θ^B_g of compatible dimensions.

$$P\left\{\{W_j\}_{ik} = 1 | Q_j\right\} \quad = \quad Q^I_{ji}\theta^I_g + Q^R_{jk}\theta^R_g + Q^B_{jik}\theta^B_g. \tag{27}$$

Because probabilities should stay in the range $[0, 1]$, it is plausible to use, instead, $P\left\{\{W_j\}_{ik} = 1|Q_j\right\} = \text{logit}(Q^I_{ji}\theta^I_g + Q^R_{jk}\theta^R_g + Q_{jik}\theta^B_g)$ or the equivalent probit version. It is important to note that, even without using the second moments to provide identification, it is still possible to conduct hypothesis testing in the partial identification framework, as long as there is no collinearity among Q^I_{ji}, Q^R_{ji} and Q^B_{jik} for all i, k and j. More specifically, suppose one is interested in whether race commonality affects the probability of link formation. The researcher can then test $\mathcal{H}_0 : \theta^B_g = 0$, with the procedure outlined in Subsection 3.2, although it will not be possible to identify the magnitude of the effect unless as a solution to equation (20) is provided. □

Example 5. *(ERMG family).* Models of statistic network formation have a long tradition in the literature of estimation of network structure given observations from random graphs generators (Holland and Leinhardt (1981), Frank and Strauss (1986), Strauss and Ikeda (1990) and Snijders (2011)) and are of considerable generality, including the case where link formation are not independent. In particular, Frank and Strauss (1986) proved that, if the graph is such that edges without common nodes are independent conditional on all remaining edges (that is, the graph is Markovian[32]) and homogeneous[33], and all isomorphic graphs have same probability,

$$P\{W_j = w_j\} \quad = \quad \frac{1}{\kappa(\theta_g)} \cdot \exp\left\{\theta^0_g T(w_j) + \sum_{s=1}^{n-1} \theta^s_g S_s(w_j)\right\} \tag{28}$$

where $T(w_j) = \sum_{i,k,l} \{w_j\}_{ik} \{w_j\}_{kl} \{w_j\}_{li}$ is the number of triangles, and $S_s(w_j)$ is the number of s-stars in w_j. $\kappa(\theta_g)$ is a normalization constant that depends on parameters $\theta_g = (\theta^0_g, \theta^1_g, \ldots, \theta^{n-1}_g)'$. The Markovian assumption is a relatively mild hypothesis and states that, although dependence between the existence of edges may happen, this cannot be so for edges which do not possess a common node. This formulation is particularly appealing as it provides a probability law for network formation under minimal hypothesis, along with its sufficient statistics. Wasserman and Pattison (1996) expand the class of models to incorporate any set of sufficient statistics $Z(w_j)$, such that

$$P\{W_j = w_j\} \quad = \quad \frac{1}{\kappa(\theta_g)} \cdot \exp\left\{\theta'_g Z(w_j)\right\}. \tag{29}$$

Note that, as a consequence of homogeneity, edges have equal probability of being formed with expected network $W^e_j(\theta_g) = p\iota_{n_j}\iota'_{n_j} - pI_{n_j}$. This is the same expectation as the one obtained in the simple Bernoulli model. The class of models considered in when using this expectation in equations (10) and likelihood (13) is much larger than might initially appear. □

[32]Let D be a graph whose nodes are all possible edges of G, that is, all pairs of nodes of G, containing therefore $n!\,(n-1)!$ nodes. If the existence of an edge between $\{a, b\}$ in G depends on the existence of an edge between $\{c, d\}$, conditional on all rest of the graph, then $\{a, b\}$ and $\{c, d\}$ are neighbours in D. The Markovian assumption means, therefore, that all $\{s, t\}$ and $\{u, v\}$ are nonneighbours for different s, t, u and v.

[33]That is, nodes are a priori indistinguishible.

C Score Vector and Hessian Matrix.

The likelihood is $\ln \mathcal{L}^e\left(\theta | y, x, Q_n\right) = -\frac{n}{2} \ln\left(2\pi\sigma^2\right) + \ln\left|S_n^e\left(Q_n,\theta\right)\right| + \ln\left|R_n^e\left(Q_n,\theta\right)\right| - \frac{1}{2\sigma^2} \epsilon_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right)$ where $\epsilon_n^e\left(Q_n,\theta\right) = R_n^e\left(Q_n,\theta\right)$ $\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right)$. First-order derivatives are

$$
\begin{aligned}
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \lambda} &= -\mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} W_n^e\left(Q_n,\theta\right)\right] + \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \beta_1} &= -\frac{1}{\sigma^2} x_n' R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \beta_2} &= -\frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \epsilon_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \rho} &= -\mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1} M_n^e\left(Q_n,\theta\right)\right] + \frac{1}{\sigma^2}\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta - W^e\left(Q_n,\theta\right) x_n\beta_2\right)' M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial \ln \mathcal{L}^e(\theta)}{\partial \theta_{gi}} &= -\lambda \mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\right] - \rho \mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1} \nabla_{\theta_{gi}} M_n^e\left(Q_n,\theta\right)\right] \\
&\quad + \frac{1}{2\sigma^2} \rho \nabla_{\theta_{gi}} M_n^e\left(Q_n,\theta\right)' \left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right)' \epsilon_n^e\left(Q_n,\theta\right) \\
&\quad + \frac{1}{2\sigma^2} R_n^e\left(Q_n,\theta\right)' \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)' \epsilon^e\left(Q_n,\theta\right)
\end{aligned}
$$

and second-order derivatives

$$
\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \lambda} &= -\mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} W_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} W_n^e\left(Q_n,\theta\right)\right] - \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) W_n^e\left(Q_n,\theta\right) y_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \beta_1'} &= -\frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) x_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \beta_2'} &= -\frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) W_n^e\left(Q_n,\theta\right) x_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \sigma^2} &= -\frac{1}{\sigma^4} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \rho} &= -\frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q,\theta\right) - \\
&\quad \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) M_n^{e'}\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \lambda \partial \theta_{gi}} &= -\lambda \mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} W_n^e\left(Q_n,\theta\right)\right] - \mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1} \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\right] \\
&\quad + \frac{1}{\sigma^2} y_n' \nabla_{\theta_{gi}} W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_b,\theta\right) - \rho \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) \nabla_{\theta_{gi}} M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
&\quad - \rho \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
&\quad - \frac{1}{\sigma^2} y_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_1 \partial \beta_1'} &= \frac{1}{\sigma^2} x_n' R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) x_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_1 \partial \beta_2'} &= \frac{1}{\sigma^2} x_n' R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) W_n^e\left(Q_n,\theta\right) x_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_1 \partial \sigma^2} &= \frac{1}{\sigma^4} x_n' R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_1 \partial \rho} &= \frac{1}{\sigma^2} x_n' M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) - \frac{1}{\sigma^2} x_n' R_n^{e'}\left(Q_n,\theta\right) M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_1 \partial \theta_{gi}} &= \rho \frac{1}{\sigma^2} x_n' \nabla_{\theta_{gi}} M_n^{e'}\left(Q,\theta\right) \epsilon_n^e\left(Q,\theta\right) + \rho \frac{1}{\sigma^2} x_n R_n^{e'}\left(Q_n,\theta\right) \nabla_{\theta_{gi}} M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
&\quad + \frac{1}{\sigma^2} x_n' R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right). \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_2 \partial \beta_2'} &= \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) W_n^e\left(Q_n,\theta\right) x_n \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_2 \partial \sigma^2} &= \frac{1}{\sigma^4} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_2 \partial \rho} &= \rho \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) + \\
&\quad \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \beta_2 \partial \theta_{gi}} &= -\frac{1}{\sigma^2} x_n' \nabla_{\theta_{gi}} W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) + \rho \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) \nabla_{\theta_{gi}} M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
&\quad \rho \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
&\quad \frac{1}{\sigma^2} x_n' W_n^{e'}\left(Q_n,\theta\right) R_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \sigma^2 \partial \sigma^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \epsilon_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \sigma^2 \partial \rho} &= -\frac{1}{\sigma^4}\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) M_n^{e'}\left(Q_n,\theta\right) \epsilon_n^e\left(Q_n,\theta\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \sigma^2 \partial \theta_{gi}} &= \frac{1}{\sigma^4} \epsilon_n^{e'}\left(Q_n,\theta\right) R_n^e\left(Q_n,\theta\right) \nabla_{\theta_{gi}} W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right) \\
&\quad - \rho \frac{1}{\sigma^4} \epsilon_n^{e'}\left(Q_n,\theta\right) \nabla_{\theta_{gi}} M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right) y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right) x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial \rho \partial \rho} &= \mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1} M_n^e\left(Q_n,\theta\right)\left(R_n^e\left(Q_n,\theta\right)\right)^{-1} M_n^e\left(Q_n,\theta\right)\right]
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial\rho\partial\theta_{gi}} &= \rho \mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}M_n^e\left(Q_n,\theta\right)\right] - \mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)\right] \\
&\quad -\frac{1}{\sigma^2}\left(\lambda y_n + x_n\beta_2\right)'\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)'M_n^{e'}\left(Q_n,\theta\right)\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad +\frac{1}{\sigma^2}\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'\nabla_{\theta_{gi}}M_n^{e'}\left(Q_n,\theta\right)\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad -\rho\frac{1}{\sigma^2}\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'M_n^{e'}\left(Q_n,\theta\right)\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right) \\
&\quad -\lambda\frac{1}{\sigma^2}\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'M_n^{e'}\left(Q_n,\theta\right)R_n^e\left(Q_n,\theta\right)\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right) \\
\frac{\partial^2 \ln \mathcal{L}^e(\theta)}{\partial\theta_{gi}\partial\theta_{gk}} &= \lambda^2\mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gk}}W_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)\right] - \lambda\mathrm{tr}\left[\left(S_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}\theta_{gk}}W_n^e\left(Q_n,\theta\right)\right] \\
&\quad +\rho^2\mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gk}}M_n^e\left(Q_n,\theta\right)\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)\right] - \rho\mathrm{tr}\left[\left(R_n^e\left(Q_n,\theta\right)\right)^{-1}\nabla_{\theta_{gi}\theta_{gk}}M_n^e\left(Q_n,\theta\right)\right] \\
&\quad +\rho\frac{1}{2\sigma^2}\nabla_{\theta_{gi}\theta_{gk}}M_n^e\left(Q_n,\theta\right)'\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad -\rho\frac{1}{2\sigma^2}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)'\nabla_{\theta_{gk}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)'\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad -\rho^2\frac{1}{2\sigma^2}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)'\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'\nabla_{\theta_{gk}}M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right) \\
&\quad -\rho\frac{1}{2\sigma^2}\nabla_{\theta_{gi}}M_n^e\left(Q_n,\theta\right)'\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right)'R_n^e\left(Q_n,\theta\right)\nabla_{\theta_{gk}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right) \\
&\quad -\rho\frac{1}{2\sigma^2}\nabla_{\theta_{gk}}M_n^e\left(Q_n,\theta\right)'\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)'\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad +\frac{1}{2\sigma^2}R_n^e\left(Q_n,\theta\right)'\nabla_{\theta_{gi}\theta_{gk}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)'\epsilon_n^e\left(Q_n,\theta\right) \\
&\quad -\rho\frac{1}{2\sigma^2}R_n^e\left(Q_n,\theta\right)'\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)'\nabla_{\theta_{gk}}M_n^e\left(Q_n,\theta\right)\left(S_n^e\left(Q_n,\theta\right)y_n - x_n\beta_1 - W_n^e\left(Q_n,\theta\right)x_n\beta_2\right) \\
&\quad -\frac{1}{2\sigma^2}R_n^e\left(Q_n,\theta\right)'\nabla_{\theta_{gi}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)'R_n^e\left(Q_n,\theta\right)\nabla_{\theta_{gk}}W_n^e\left(Q_n,\theta\right)\left(\lambda y_n + x_n\beta_2\right)
\end{aligned}$$

Derivatives $\nabla_{\theta_{gi}}W_j^e\left(Q_j,\theta\right) = \frac{\partial W_j^e(Q_j,\theta)}{\partial\theta_{gi}}$, $\nabla_{\theta_{gi}\theta_{gk}}W_j^e\left(Q_j,\theta\right) = \frac{\partial^2 W_j^e(Q_j,\theta)}{\partial\theta_{gi}\partial\theta_{gk}}$ and similarly for derivatives of $M_j^e\left(Q_j,\theta\right)$ and model-dependent and so are omitted here.

D Proofs.

D.1 Useful Lemmas.

Lemmas without proofs can be found in Kelejian and Prucha (2001), Lee (2004) or Lee et al. (2010).

Lemma 1. *For any $n\times n$ matrix Λ_n with uniformly bounded column sums in absolute value, uniformly bounded $n\times k$ matrix Z_n, and if $u_n \sim N\left(0,\sigma^2 I\right)$ of dimension $n\times 1$, then $\frac{1}{\sqrt{n}}Z_n'\Lambda_n u_n = O_p(1)$.*

Lemma 2. $\mathbb{E}\left(u_n'\Lambda_n u_n\right) = \sigma^2 tr\left(\Lambda_n\right)$ *and* $Var\left(u_n'\Lambda_n u_n\right) = \left(\mu_4 - 3\sigma^4\right)vec_D'\left(\Lambda_n\right)vec_D\left(\Lambda_n\right) + \sigma^4\left[tr\left(\Lambda_n\Lambda_n'\right) + tr\left(\Lambda_n^2\right)\right]$.

Lemma 3. *Define $\Lambda_n^{-1} \equiv (S_n^{0'})^{-1}\tilde{\Lambda}_n^{-1}\left(S_n^0\right)^{-1}$, $(\Lambda_n^e)^{-1} \equiv (S_n^{e'}\left(Q_n,\theta_c^0\right))^{-1}\tilde{\Lambda}_n^{-1}\left(S_n^e\left(Q_n,\theta_c^0\right)\right)^{-1}$ and $\tilde{\Lambda}_n \equiv (S_n^{e'}\left(Q_n,\theta_c\right)R_n^{e'}\left(Q_n,\theta_c\right)P_n^e\left(Q_n,\theta_c\right)$ $R_n^e\left(Q_n,\theta_c\right)S_n^e\left(Q_n,\theta_c\right))^{-1}$. Then, for any randomly distributed vector ϵ_n of dimension $n\times 1$ such that $\mathbb{E}\epsilon_i\epsilon_j = 0$ for $i\neq j$ with $\mathbb{E}\epsilon_i^2 < \infty$ and if link formation is independent, $\frac{1}{n}\mathbb{E}(\epsilon_n'\Lambda_n^{-1}\epsilon_n) = \frac{1}{n}\mathbb{E}(\epsilon_n'\Lambda_e^{-1}\epsilon_n) + o_p(1)$.*

Proof. For simplicity, consider $R_n^e\left(Q_n,\theta_c^0\right) = R_n^0 = I_n$. Proof generalizes immediately otherwise. Then $\frac{1}{n}\mathbb{E}\left\{\epsilon_n'\left[\Lambda_n^{-1} - \left(\mathbb{E}\Lambda_n\right)^{-1}\right]\epsilon_n\right\} = \frac{1}{n}\mathbb{E}\left\{\epsilon_n'\Lambda_n^{-1}\left[\mathbb{E}\Lambda_n - \Lambda_n\right]\left(\mathbb{E}\Lambda_n\right)^{-1}\epsilon_n\right\} = \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^n\sum_{j=1}^n\epsilon_i\epsilon_j\left[\Lambda_n^{-1}\left(\mathbb{E}\Lambda_n - \Lambda_n\right)\mathbb{E}\Lambda_n^{-1}\right]_{ij}\right\} = \frac{1}{n}\sum_{i=1}^n\mathbb{E}\left\{\epsilon_i^2\right\}\mathbb{E}\left[\Lambda_n^{-1}\left(\mathbb{E}\Lambda_n - \Lambda_n\right)\mathbb{E}\Lambda_n^{-1}\right]_{ii}$ as ϵ_i is independent of ϵ_j for $i\neq j$. Because $\mathbb{E}\left[\Lambda_n^{-1}\left(\mathbb{E}\Lambda_n - \Lambda_n\right)\left(\mathbb{E}\Lambda_n\right)^{-1}\right]_{ij} \xrightarrow{p} 0$ and $\mathbb{E}\left\{\epsilon_i^2\right\} < \infty$, then $\frac{1}{n}\mathbb{E}\left\{\epsilon_n'\left[\Lambda_n^{-1} - \left(\mathbb{E}\Lambda_n\right)^{-1}\right]\epsilon_n\right\} = o_p(1)$. Remains to show $\mathbb{E}\Lambda_n = \Lambda_n^e$. By definition, $\Lambda_n = \left(I_n - \lambda_0 W_n^0\right)\tilde{\Lambda}_n(I_n - \lambda_0 W_n^{0'}) = \tilde{\Lambda}_n - \lambda_0 W_n^0\tilde{\Lambda}_n - \lambda_0\tilde{\Lambda}_n W_n^{0'} + \lambda_0^2 W_n^0\tilde{\Lambda}_n W_n^{0'}$. It follows that $\mathbb{E}\Lambda_n = \tilde{\Lambda}_n - \lambda_0 W_n^e\left(Q_n,\theta_c^0\right)\tilde{\Lambda}_n - \lambda_0\tilde{\Lambda}_n W_n^{e'}\left(Q_n,\theta_c^0\right) + \lambda_0^2\mathbb{E}W_n^0\tilde{\Lambda}_n W_n^{0'} = \tilde{\Lambda}_n - \lambda_0 W_n^e\left(Q_n,\theta_c^0\right)\tilde{\Lambda}_n - \lambda_0\tilde{\Lambda}_n W_n^{e'}\left(Q_n,\theta_c^0\right) + \lambda_0^2 W_n^e\left(Q_n,\theta_c^0\right)\tilde{\Lambda}_n W_n^{e'}\left(Q_n,\theta_c^0\right) = \Lambda_e$ where the second equality holds only if link formation is independent, i.e., if $\mathbb{E}\{W_j^0\}_{ik}\{W_j^0\}_{i'k'} = \mathbb{E}\{W_j^0\}_{ik}\mathbb{E}\{W_j^0\}_{i'k'}$ if either $i\neq i'$ or $k\neq k'$. □

Lemma 4. *Let ϵ_n be a $n \times 1$ stationary, ergodic process with $\mathbb{E}\epsilon_n = 0$. Then* $\frac{1}{n}\mathbb{E}(\epsilon_n' \Lambda_n^{-1} \epsilon_n) = \frac{1}{n}\mathbb{E}(\epsilon_n' \left(\Lambda_n^e\right)^{-1} \epsilon_n) + o_p(1)$.

Proof. Lemma 3 applies with the following modification. Given $\sum_{i=1}^n \sum_{j=1}^n \epsilon_i \epsilon_j \left[\Lambda_n^{-1}\left(\mathbb{E}\Lambda_n - \Lambda_n\right)\mathbb{E}\Lambda_n^{-1}\right]_{ij}$ is a weighted U-statistic, with summable weights, Theorem 3 of Hsing and Wu (2004) is applied to obtain convergence in probability to zero.

□

Lemma 5. $\frac{1}{n}\mathbb{E}\{\beta_0' Z_n^{0'} (S_n^{0'})^{-1} S_n^{e'}\left(Q_n, \theta_c^0\right) R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right) S_n^e\left(Q_n, \theta_c^0\right)\left(S_n^0\right)^{-1} Z_n^0 \beta_0\} = o_p(1)$.

Proof. Apply Lemma 4 with minor modifications twice. First, note that $\frac{1}{n}\mathbb{E}\{\beta_0' Z_n^{0'} (S_n^{0'})^{-1} S_n^{e'}\left(Q_n, \theta_c^0\right) R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right)$ $S_n^e\left(Q_n, \theta_c^0\right)\left(S_n^0\right)^{-1} Z_n^0 \beta_0\} = \frac{1}{n}\mathbb{E}\{\beta_0' Z_n^{0'} R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right) Z_n^0 \beta_0\} + o_p(1)$. Secondly, $\frac{1}{n}\mathbb{E}\{\beta_0' Z_n^{0'} R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right)$ $R_n^e\left(Q_n, \theta_c^0\right) Z_n^0 \beta_0\} = \frac{1}{n}\mathbb{E}\left\{\beta_0' Z_n^{e'}\left(Q_n, \theta_c^0\right) R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right)\beta_0\right\} + o_p(1)$. Properties of projection matrix ensures $P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right) Z_n^e\left(Q_n, \theta_c^0\right) = 0$. □

Lemma 6. $\frac{1}{n}\mathbb{E}\{\epsilon_n' (R_n^{0'})^{-1} (S_n^{0'})^{-1} S_n^{e'}\left(Q_n, \theta_c^0\right) R_n^{e'}\left(Q_n, \theta_c^0\right) P_n^e\left(Q_n, \theta_c^0\right) R_n^e\left(Q_n, \theta_c^0\right) S_n^e\left(Q_n, \theta_c^0\right)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1} \epsilon_n\} = \sigma_0^2 + o_p(1)$.

Proof. Direct consequence of Lemma 4 taken with $\theta_c = \theta_c^0$. □

D.2 Derivation of pdf of networks.

For the *p1-reciprocity model*, the probability that random matrix W^* takes a particular value w^* is

$$
\begin{aligned}
P(W = w) &= \prod_{i<j} \delta_F^{w_{ij} w_{ji}} \prod_{i<j} \delta_A^{w_{ij}(1-w_{ji}) + (1-w_{ij}) w_{ji}} \prod_{i<j} \delta_N^{(1-w_{ij})(1-w_{ji})} \\
&= \exp\left\{\ln \delta_F \sum_{i<j} w_{ij} w_{ji} + \ln \delta_A \sum_{i<j} w_{ij}\left(1 - w_{ji}\right) + \left(1 - w_{ij}\right) w_{ji} + \ln \delta_N \sum_{i<j}\left(1 - w_{ij}\right)\left(1 - w_{ji}\right)\right\} \\
&= \frac{1}{\kappa} \exp\left\{\theta_g^1 \sum_{i \neq j} w_{ij} + \theta_g^2 \sum_{i<j} w_{ij} w_{ji}\right\}
\end{aligned}
$$

where $\theta_g^1 = \ln \frac{\delta_A}{\delta_N}$, $\theta_g^2 = \frac{\delta_F \delta_N}{\delta_A^2}$ and $\kappa = \left(\prod_{i<j} \delta_N\right)^{-1}$. Introducing dependence on sharing exogenous characteristics, the pdf is

$$
\begin{aligned}
P\left(W = w \mid Q = q\right) &= \prod_{i<j}\left(\delta_{1F}^{q_{ij}} \delta_{0F}^{1-q_{ij}}\right)^{w_{ij} w_{ji}} \prod_{i<j}\left(\delta_{1A}^{q_{ij}} \delta_{0A}^{1-q_{ij}}\right)^{(1-w_{ij}) w_{ji} + w_{ij}(1-w_{ji})} \prod_{i<j}\left(\delta_{1N}^{q_{ij}} \delta_{0N}^{1-q_{ij}}\right)^{(1-w_{ij})(1-w_{ji})} \\
&= \exp\left\{\ln\left\{\prod_{i<j}\left(\delta_{1F}^{q_{ij}} \delta_{0F}^{1-q_{ij}}\right)^{w_{ij} w_{ji}} \prod_{i<j}\left(\delta_{1A}^{q_{ij}} \delta_{0A}^{1-q_{ij}}\right)^{(1-w_{ij}) w_{ji} + w_{ij}(1-w_{ji})} \prod_{i<j}\left(\delta_{1N}^{q_{ij}} \delta_{0N}^{1-q_{ij}}\right)^{(1-w_{ij})(1-w_{ji})}\right\}\right\} \\
&= \exp\left\{\sum_{i<j} w_{ij} w_{ji}\left(q_{ij} \ln \delta_{1F} + (1 - q_{ij}) \ln \delta_{0F}\right) + \sum_{i<j}\left(1 - w_{ij}\right) w_{ji}\left(q_{ij} \ln \delta_{1A} + (1 - q_{ij}) \ln \delta_{0A}\right)\right. \\
&\quad \left. + \sum_{i<j} w_{ij}\left(1 - w_{ji}\right)\left(q_{ij} \ln \delta_{1A} + (1 - q_{ij}) \ln \delta_{0A}\right) + \sum_{i<j}\left(1 - w_{ij}\right)\left(1 - w_{ji}\right)\left(q_{ij} \ln \delta_{1N} + (1 - q_{ij}) \ln \delta_{0N}\right)\right\} \\
&= \frac{1}{\kappa} \exp\left\{\theta_g^1 \sum_{i \neq j} w_{ij} + \theta_g^2 \sum_{i \neq j} w_{ij} q_{ij} + \theta_g^3 \sum_{i<j} w_{ij} w_{ji} + \theta_g^4 \sum_{i<j} w_{ij} w_{ji} q_{ij}\right\}
\end{aligned}
$$

where $\theta_g^1 = \ln\frac{\delta_{0A}}{\delta_{0N}}$, $\theta_g^2 = \ln\frac{\delta_{0N}\delta_{1A}}{\delta_{0A}\delta_{1N}}$, $\theta_g^3 = \ln\frac{\delta_{0F}\delta_{0N}}{\delta_{0A}^2}$, $\theta_g^4 = \ln\frac{\delta_{1F}\delta_{0A}^2\delta_{1N}}{\delta_{0F}\delta_{1A}^2\delta_{0N}}$ and $\kappa^{-1} = \exp\left\{\ln\left(\delta_{1N}\delta_{0N}\sum_{i<j} q_{ij}\right)\right\}\prod_{i<j}\delta_{0N}$.

D.3 Proposition 1.

Proof. Under the assumption, full column rank means that the only solutions for the constants c_1, c_2 and c_3 in the equation $x_n c_1 + W_n^e\left(Q_n,\theta_c^0\right)x_n c_2 + G_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{10}c_3 + G_n^e\left(Q_n,\theta_c^0\right)W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{20}c_3 = 0$ are $c_1 = c_2 = c_3 = 0$. Under the assumption that $G_n^e\left(Q_n,\theta_c^0\right) \equiv W_n^e\left(Q_n,\theta_c^0\right)\left(S_n^e\left(Q_n,\theta_c^0\right)\right)^{-1} = \left(S_n^e\left(Q_n,\theta_c^0\right)\right)^{-1}W_n^e\left(Q,\theta_c^0\right)$, i.e., assuming symmetry of $W_n^e\left(Q_n,\theta_c^0\right)$, expression is equal to $x_n c_1 + W_n^e\left(Q_n,\theta_c^0\right)x_n c_2 + \left(S_n^e\left(Q_n,\theta_c^0\right)\right)^{-1}W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{10}c_3 + \left(S_n^e\left(Q_n,\theta_c^0\right)\right)^{-1}\left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n\beta_{20}c_3$, then equivalent to assessing

$$
\begin{aligned}
& S_n^e\left(Q_n,\theta_c^0\right)x_n c_1 + S_n^e\left(Q_n,\theta_c^0\right)W_n^e\left(Q_n,\theta_c^0\right)x_n c_2 + W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{10}c_3 + \left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n\beta_{20}c_3 \\
= \quad & \left(I_n + \lambda W_n^e\left(Q_n,\theta_c^0\right)\right)x_n c_1 + \left(I_n + \lambda W_n^e\left(Q_n,\theta_c^0\right)\right)W_n^e\left(Q_n,\theta_c^0\right)x_n c_2 + W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{10}c_3 + \left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n\beta_{20}c_3 \\
= \quad & x_n c_1 + \lambda W_n^e\left(Q_n,\theta_c^0\right)x_n c_1 + W_n^e\left(Q_n,\theta_c^0\right)x_n c_2 + \lambda\left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n c_2 + W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{10}c_3 + \left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n\beta_{20}c_3 \\
= \quad & x_n c_1 + W_n^e\left(Q_n,\theta_c^0\right)x_n\left(\lambda c_1 + c_2 + \beta_{10}c_3\right) + \left(W_n^e\left(Q_n,\theta_c^0\right)\right)^2 x_n\left(\lambda c_2 + \beta_{20}c_3\right)
\end{aligned}
$$

As x_n, $W_n^e\left(Q_n,\theta_c^0\right)x_n$ and $\left(W_n^e\left(Q_n,\theta_c\right)\right)^2 x_n$ are linearly independent, $c_1 = 0$, then implying $c_2 + \beta_{10}c_3 = 0$ and $\lambda c_2 + \beta_{20}c_3 = 0$. Together, $\left(-\lambda\beta_{10} + \beta_{20}\right)c_3 = 0$. Given $\beta_{20} \neq \lambda\beta_{10}$, $c_3 = c_2 = 0$. If $W_n^e(Q_n,\theta_c^0)$ is not symmetric, premultiply the initial expression by $W_n^e(Q_n,\theta_c^0)S_n^e(Q_n,\theta_c^0)(W_n^e(Q_n,\theta_c^0))^{-1} = I_n + \lambda_0 W_n^e(Q_n,\theta_c^0)$ and same result follows. □

D.4 Theorem 1.

Proof. (Uniform Convergence). The goal is to show that the concentrated log-likelihood $2\left(n\right)^{-1}\left[\ln\mathcal{L}_n^c\left(\theta_c\right) - Q_n\left(\theta_c\right)\right]$ converges uniformly to zero on Θ_c, where $F_n\left(\theta_c\right) = \max_{\beta,\sigma^2}\mathbb{E}\ln\mathcal{L}_n^c\left(\theta_c\right)$, that is,

$$
\sup_{\theta_c\in\Theta_c}\left|\frac{1}{n}\ln\mathcal{L}_n\left(\theta_c\right) - \frac{1}{n}F_n\left(\theta_c\right)\right| = \sup_{\theta_c\in\Theta_c}\left|\ln\tilde{\sigma}^2\left(\theta_c\right) - \ln\hat{\sigma}^2\left(\theta_c\right)\right| = o_p\left(1\right).
$$

In first place, misspecification component in $\hat{\sigma}^2\left(Q,\theta_c\right)$ is made explicit. Given $S_n^e\left(Q_n,\theta_c\right) = I_n - \lambda W_n^e\left(Q_n,\theta_c\right)$ and $\left(S_n^0\right)^{-1} = \lambda_0 G_n^0 + I_n$ where $G_n^0 = W_n^0\left(S_n^0\right)^{-1}$, then $S_n^e\left(Q_n,\theta_c\right)\left(S_n^0\right)^{-1} = \lambda_0 G_n^0 + I_n - \lambda\lambda_0 W_n^e\left(Q_n,\theta_c\right)G_n^0 - \lambda W_n^e\left(Q_n,\theta_c\right)$. Now $\lambda_0 W_n^e\left(Q_n,\theta_c\right) = \lambda_0 W_n^0 + \lambda_0\left(W_n^e\left(Q_n,\theta_c\right) - W_n^0\right) = I_n - S_n^0 + \lambda_0\left(W_n^e\left(Q_n,\theta_c\right) - W_n^0\right)$ and $S_n^e\left(Q_n,\theta_c\right)\left(S_n^0\right)^{-1} = \left(\lambda_0 - \lambda\right)G_n^0 + I_n + B_n\left(Q_n,\theta_c\right)$ where the misspecification term is defined $B_n\left(Q_n,\theta_c\right) \equiv \lambda\left(W_n^0 - W_n^e\left(Q_n,\theta_c\right)\right) + \lambda\lambda_0\left(W_n^0 - W_n^e\left(Q_n,\theta_c\right)\right)G_n^0 = \lambda\left(W_n^0 - W_n^e\left(Q_n,\theta_c\right)\right)\left(I + \lambda_0 G_n^0\right)$. Therefore, using the reduced-form equation $S_n^e\left(Q_n,\theta_c\right)y_n = S_n^e\left(Q_n,\theta_c\right)\left(S_n^0\right)^{-1}Z_n^0\beta_0 + S_n^e\left(Q_n,\theta_c\right)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n$,

$$
\begin{aligned}
P_n^e\left(Q_n,\theta_c\right)R_n^e\left(Q_n,\theta_c\right)S_n^e\left(Q_n,\theta_c\right)y_n \quad = \quad & P_n^e\left(Q_n,\theta_c\right)R_n^e\left(Q_n,\theta_c\right)Z_n^0\beta_0 + \left(\lambda_0 - \lambda\right)P_n^e\left(Q_n,\theta_c\right)R_n^e\left(Q_n,\theta_c\right)G_n^0 Z_n^0\beta_0 \\
& + P_n^e\left(Q_n,\theta_c\right)R_n^e\left(Q_n,\theta_c\right)B_n\left(Q_n,\theta_c\right)Z_n^0\beta_0 + P_n^e\left(Q_n,\theta_c\right)R_n^e\left(Q_n,\theta_c\right)S_n^e\left(Q_n,\theta_c\right)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n.
\end{aligned}
$$

Given that $\hat{\sigma}^2(Q_n,\theta_c)=\frac{1}{n}y_n' S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n$, $\hat{\sigma}^2(Q,\theta_c)=\sum_{i=1}^{10}K_i(Q,\theta_g)$, where

$$
\begin{aligned}
K_1(Q_n,\theta_g) &= \frac{1}{n}\left[R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right]\\
K_2(Q_n,\theta_g) &= \frac{2}{n}(\lambda_0-\lambda)\left[R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)G_n^0 Z_n^0\beta_0\right]\\
K_3(Q_n,\theta_g) &= \frac{2}{n}\left[R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right]\\
K_4(Q_n,\theta_g) &= \frac{2}{n}\left[R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]\\
K_5(Q_n,\theta_g) &= \frac{1}{n}(\lambda_0-\lambda)^2\left[R_n^e(Q_n,\theta_c)G_n^0 Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)G_n^0 Z_n^0\beta_0\right]\\
K_6(Q_n,\theta_g) &= \frac{2}{n}(\lambda_0-\lambda)\left[R_n^e(Q_n,\theta_c)G_n^0 Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right]\\
K_7(Q_n,\theta_g) &= \frac{2}{n}(\lambda_0-\lambda)\left[R_n^e(Q_n,\theta_c)G_n^0 Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]\\
K_8(Q_n,\theta_g) &= \frac{1}{n}\left[R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right]\\
K_9(Q_n,\theta_g) &= \frac{2}{n}\left[R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]\\
K_{10}(Q_n,\theta_g) &= \frac{1}{n}\left[R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]' P_n^e(Q_n,\theta_c)\left[R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]
\end{aligned}
$$

Given Lemma 1, $K_4(Q,\theta_g)$, $K_7(Q,\theta_g)$ and $K_9(Q,\theta_g)$ are $o_p(1)$. Remains to show the problem in expectation. The concentrators are

$$
\begin{aligned}
\tilde{\beta}(Q_n,\theta_c) &= \left[Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^e(Q_n,\theta_c)\right]^{-1} Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\mathbb{E}y_n\\
\tilde{\sigma}^2(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{\left[S_n^e(Q_n,\theta_c)y_n - Z_n^e(Q_n,\theta_c)\tilde{\beta}(\theta_c)\right]' R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)\left[S_n^e(Q_n,\theta_c)y_n - Z_n^e(Q_n,\theta_c)\tilde{\beta}(\theta_c)\right]\right\}.
\end{aligned}
$$

Noticing $P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^e(Q_n,\theta_c)=0$, the expectation

$$
\begin{aligned}
\tilde{\sigma}^2(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{y_n' S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n\right\}\\
&= \frac{1}{n}\mathbb{E}\left\{\left[\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]' S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right\}\\
&\quad +\frac{1}{n}\mathbb{E}\left\{\left[\left(S_n^0\right)^{-1}Z_n^0\beta_0\right]' S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}Z_n^0\beta_0\right\}\\
&= \frac{1}{n}\mathbb{E}\left\{\left[\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right]' S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon\right\}\\
&\quad +\frac{1}{n}\mathbb{E}\left\{\beta_0' Z_n^{0'}\left[(\lambda_0-\lambda)G_n^0+I_n+B(Q_n,\theta_c)\right]' R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)\left[(\lambda_0-\lambda)G_n^0+I_n+B_n(Q_n,\theta_c)\right]Z_n^0\beta_0\right\}
\end{aligned}
$$

and so $\tilde{\sigma}^2(Q,\theta_c)=\sum_{i=1}^{7}\tilde{K}_i(Q,\theta_c)$ with

$$
\begin{aligned}
\tilde{K}_1(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{\epsilon_n'\left(R_n^{0'}\right)^{-1}\left(S_n^{0'}\right)^{-1}S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right\}\\
\tilde{K}_2(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{(\lambda_0-\lambda)^2\beta_0'Z_n^{0'}G_n^{0'}R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)G_n^0Z_n^0\beta_0\right\}\\
\tilde{K}_3(Q_n,\theta_c) &= \frac{2}{n}\mathbb{E}\left\{(\lambda_0-\lambda)\beta_0'Z_n^{0'}G_n^{0'}R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right\}\\
\tilde{K}_4(Q_n,\theta_c) &= \frac{2}{n}\mathbb{E}\left\{(\lambda_0-\lambda)\beta_0'Z_n^{0'}G_n^{0'}R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R^e(Q_n,\theta_c)B(Q_n,\theta_c)Z_n^0\beta_0\right\}\\
\tilde{K}_5(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{\beta_0'Z_n^{0'}R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^0\beta_0\right\}\\
\tilde{K}_6(Q_n,\theta_c) &= \frac{2}{n}\mathbb{E}\left\{\beta_0'Z_n^{0'}R_n^{e'}(Q_n,\theta_c)P^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right\}\\
\tilde{K}_7(Q_n,\theta_c) &= \frac{1}{n}\mathbb{E}\left\{\beta_0'Z_n^{0'}B_n(Q_n,\theta_c)'R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)B_n(Q_n,\theta_c)Z_n^0\beta_0\right\}.
\end{aligned}
$$

By Lemma 2, $\tilde{K}_1(Q_n,\theta_c)=K_{10}(Q_n,\theta_c)+o_p(1)$. Also, $\tilde{K}_2(Q_n,\theta_c)=K_5(Q_n,\theta_c)+o_p(1)$, $\tilde{K}_3(Q_n,\theta_c)=K_2(Q_n,\theta_c)+o_p(1)$, $\tilde{K}_4(Q_n,\theta_c)=K_6(Q_n,\theta_c)+o_p(1)$, $\tilde{K}_5(Q_n,\theta_c)=K_1(Q_n,\theta_c)+o_p(1)$, $\tilde{K}_6(Q_n,\theta_c)=K_3(Q_n,\theta_c)+o_p(1)$ and $\tilde{K}_7(Q_n,\theta_c)=K_8(Q_n,\theta_c)+o_p(1)$. As a consequence, $\hat{\sigma}^2(Q_n,\theta_c)-\tilde{\sigma}^2(Q_n,\theta_c)=o_p(1)$ uniformly on θ_c. Convergence is uniform on the parameter space as λ, ρ and θ_c appear as polynomial factors.

(Identification for $\lambda=\lambda_0$). Consider the non-stochastic auxiliary model $y_j=\lambda_0W_j^e\left(Q_j,\theta_c^0\right)y_j+x_j\beta_1+W_j^e\left(Q_j,\theta_c^0\right)x_j\beta_2+v_j$ where true neighboring matrices are given by expected network at true parameter values, $W_j^0=W_j^e\left(Q,\theta_c^0\right)$ and $M_j^0=M_j^e\left(Q,\theta_c^0\right)$. Its likelihood is

$$\ln\mathcal{L}_n^{**}(\theta)=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)+\ln\left|S_n^e(Q_n,\theta)\right|+\ln\left|R_n^e(Q_n,\theta)\right|-\frac{1}{2\sigma^2}\sum_{j=1}^{v}\epsilon_j^{e'}(Q_j,\theta)\epsilon_j^e(Q_j,\theta)$$

where $\epsilon_j^e(Q_j,\theta)=R_j^e(Q_j,\theta)\left(S_j^e(Q_j,\theta)y_j-x_j\beta_1-W_j^e(Q_j,\theta)x_j\beta_2\right)$. As usual, parameters β and σ^2 can be concentrated out of the likelihood. The concentrators are given by

$$
\begin{aligned}
\hat{\beta}^{**}(Q_n,\theta_c) &= \left[Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)Z_n^e(Q_n,\theta_c)\right]^{-1}Z_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n\\
\hat{\sigma}^{**2}(Q_n,\theta_c) &= \frac{1}{n}\left[S^e(Q_n,\theta_c)y_n-Z^e(Q_n,\theta_c)\hat{\beta}(\theta_c)\right]'R_n^{e'}(Q_n,\theta_c)R_n^e(Q_n,\theta_c)\left[S_n^e(Q_n,\theta_c)y_n-Z^e(Q_n,\theta_c)\hat{\beta}(\theta_c)\right]\\
&= \frac{1}{n}y_n'S_n^{e'}(Q_n,\theta_c)R_n^{e'}(Q_n,\theta_c)P_n^e(Q_n,\theta_c)R_n^e(Q_n,\theta_c)S_n^e(Q_n,\theta_c)y_n
\end{aligned}
$$

The final form for the concentrated likelihood is $\ln\mathcal{L}_n^{c**}(\theta_c)=-\frac{n}{2}(\ln(2\pi)+1)-\frac{n}{2}\ln\hat{\sigma}^2(\theta_c)+\ln\left|S_n^e(Q_n,\theta)\right|+\ln\left|R_n^e(Q_n,\theta)\right|$. The problem in expectation $F_n^{**}(\theta)=\max_{\beta,\sigma^2}\mathbb{E}\ln\mathcal{L}_n^{**}(\theta)$ is $F_n^{**}(\theta)=-\frac{n}{2}(\ln(2\pi)+1)+\ln\left|S_n^e(Q_n,\theta)\right|+\ln\left|R_n^e(Q_n,\theta)\right|-\frac{n}{2}\tilde{\sigma}^{**2}(\theta)$, where $\tilde{\sigma}^{**2}(Q_n,\theta_c)$

is given by

$$
\begin{aligned}
& \frac{1}{n}\mathbb{E}\left\{y_n' S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c) y_n\right\} \\
= & \frac{1}{n}\mathbb{E}\left\{\epsilon_n'\left(R_n^{e'}(Q_n,\theta_c^0)\right)^{-1}\left(S_n^{e'}(Q_n,\theta_c^0)\right)^{-1} S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c)\left(S_n^e(Q_n,\theta_c^0)\right)^{-1}\left(R_n^e(Q_n,\theta_c^0)\right)^{-1}\epsilon_n\right\} \\
& +\frac{1}{n}\mathbb{E}\left\{\beta_0' Z_n^{e'}(Q_n,\theta_c^0)\left(S_n^{e'}(Q_n,\theta_c^0)\right)^{-1} S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c)\left(S_n^e(Q_n,\theta_c^0)\right)^{-1} Z_n^e(Q_n,\theta_c^0)\beta_0\right\} \\
= & \frac{\sigma^2}{n}\operatorname{tr}\left\{\left(R_n^{e'}(Q_n,\theta_c^0)\right)^{-1}\left(S_n^{e'}(Q_n,\theta_c^0)\right)^{-1} S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c)\left(S_n^e(Q_n,\theta_c^0)\right)^{-1}\left(R_n^e(Q_n,\theta_c^0)\right)^{-1}\right\} \\
& +\frac{(\lambda_0-\lambda)^2}{n}\beta_0' Z_n^{e'}(Q_n,\theta_c^0) G_n^{e'}(Q_n,\theta_c^0) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) G_n^e(Q_n,\theta_c^0) Z_n^e(Q_n,\theta_c^0)\beta_0.
\end{aligned}
$$

By Jensen's Inequality, $F_n^{**}(\theta) \le F_n^{**}(\theta_0)$. Identification in the original model follows from

$$
\frac{1}{n}F_n(\theta_c) - \frac{1}{n}F_n\left(\theta_c^0\right) = \frac{1}{n}\left[F_n^{**}(\theta_c) - F_n^{**}\left(\theta_c^0\right)\right] + \frac{1}{2}\left[\ln\sigma^{**2}(\theta_c) - \ln\tilde{\sigma}^2(\theta_c) + \ln\tilde{\sigma}^2\left(\theta_c^0\right) - \ln\sigma^{**2}\left(\theta_c^0\right)\right].
$$

It is immediate that $\sigma^{**2}\left(\theta_c^0\right) = \sigma_0^2$. Lemmas 5 and 6 imply that $\tilde{\sigma}^2\left(\theta_c^0\right) = \sigma_0^2$. Notice also

$$
\begin{aligned}
\tilde{\sigma}^2(\theta_c) = & \frac{1}{n}\cdot\mathbb{E}\left\{\epsilon_n'\left(R_n^{0'}\right)^{-1}\left(S_n^{0'}\right)^{-1} S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right\} \\
& +\frac{1}{n}\mathbb{E}\left\{\beta_0' Z_n^{0'}\left(S_n^{0'}\right)^{-1} S_n^{e'}(Q_n,\theta_c) R_n^{e'}(Q_n,\theta_c) P_n^e(Q_n,\theta_c) R_n^e(Q_n,\theta_c) S_n^e(Q_n,\theta_c)\left(S_n^0\right)^{-1} Z_n^0\beta_0\right\}.
\end{aligned}
$$

Finally, Lemma 3 and Assumption 6 imply $\ln\sigma^{**2}(\theta_c) - \ln\tilde{\sigma}^2(\theta_c) < 0$. This completes the proof. □

D.5 Theorem 2.

Proof. Jacobian and Hessian matrices are given in Appendix C. The asymptotic distribution can be obtained from a Taylor expansion around the point $\frac{\partial\ln\mathcal{L}^e\left(\hat{\theta}\middle|y_n,x_n,Q_n\right)}{\partial\theta} = 0$. For a point $\tilde{\theta}$ between $\hat{\theta}$ and θ_0,

$$
\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \left[-\frac{1}{n}\frac{\partial\ln\mathcal{L}^e\left(\tilde{\theta}\middle|y_n,x_n,Q_n\right)}{\partial\theta\partial\theta'}\right]^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ln\mathcal{L}^e\left(\theta_0|y_n,x_n,Q_n\right)}{\partial\theta}.
$$

(Showing $\frac{1}{n}\frac{\partial\ln\mathcal{L}^e\left(\tilde{\theta}|y_n,x_n,Q_n\right)}{\partial\theta\partial\theta'} \xrightarrow{p} \frac{1}{n}\frac{\partial\ln\mathcal{L}^e\left(\theta_0|y_n,x_n,Q_n\right)}{\partial\theta\partial\theta'}$*).* Convergence is shown explicitly for three terms: $\frac{\partial\ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\beta_1'}$, $\frac{\partial\ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\sigma^2}$ and $\frac{\partial\ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda^2}$; other terms can be shown with little or no modifications. For

$$
\begin{aligned}
\frac{1}{n}\left\{\frac{\partial\ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\beta_1'} - \frac{\partial\ln\mathcal{L}^e(\theta_0)}{\partial\lambda\partial\beta_1'}\right\} = & \frac{1}{n\sigma_0^2}y_n' W_n^{e'}(Q_n,\theta_0) R_n^{e'}(Q_n,\theta_0) x_n - \frac{1}{n\tilde{\sigma}^2}y_n' W_n^{e'}\left(Q_n,\tilde{\theta}\right) R_n^{e'}\left(Q_n,\tilde{\theta}\right) x_n \\
= & \frac{1}{n}\left[\frac{1}{\sigma_0^2}-\frac{1}{\tilde{\sigma}^2}\right] y_n' W_n^{e'}(Q_n,\theta_0) R_n^{e'}(Q_n,\theta_0) x_n + \frac{1}{n\tilde{\sigma}^2}y_n'\left[W_n^{e'}(Q_n,\theta_0) R_n^{e'}(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta}) R_n^{e'}(Q_n,\tilde{\theta})\right] x_n.
\end{aligned}
$$

The argument follows by noticing $W_n^e(Q_n,\theta_0)$ and $R_n^e(Q_n,\theta_0)$ are row and column-sum bounded, so $\frac{1}{n}y_n' W_n^{e'}(Q_n,\theta_0) R_n^{e'}(Q_n,\theta_0)x = O_p(1)$, while by continuity of the inverse, $\left[\frac{1}{\sigma_0^2}-\frac{1}{\tilde{\sigma}^2}\right] = o_p(1)$. The second term converges in probability as $\frac{1}{n\tilde{\sigma}^2}y_n'[W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta})R^{e'}(Q_n,\tilde{\theta})]x_n = \frac{1}{n\tilde{\sigma}^2}\beta_0' Z_n^{0'}\left(S_n^{0'}\right)^{-1}\left[W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta})R_n^{e'}(Q_n,\tilde{\theta})\right]x + o_p(1)$. Given that $Z_n^0 = [x_n; W_n^0 x_n]$, x_n is non stochastic, W_n^0 is row and column-sum bounded, and $\left[W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta})R_n^{e'}(Q_n,\tilde{\theta})\right] = o_p(1)$, it has been shown that $\frac{1}{n}\left\{\frac{\partial \ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\beta_1'} - \frac{\partial \ln\mathcal{L}^e(\theta_0)}{\partial\lambda\partial\beta_1'}\right\} = o_p(1)$. The next term is

$$\begin{aligned}
\frac{1}{n}\left\{\frac{\partial \ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\sigma^2} - \frac{\partial \ln\mathcal{L}^e(\theta_0)}{\partial\lambda\partial\sigma^2}\right\} &= \frac{1}{n\sigma_0^4}y_n' W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0)\epsilon_n^e(Q_n,\theta_0) - \frac{1}{n\tilde{\sigma}^4}y_n' W_n^{e'}(Q_n,\tilde{\theta})R_n^{e'}(Q_n,\tilde{\theta})\epsilon_n^e(Q_n,\tilde{\theta}) \\
&= \frac{1}{n}\left[\frac{1}{\sigma_0^4}-\frac{1}{\tilde{\sigma}^4}\right]y_n' W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0)\epsilon_n^e(Q_n,\theta_0) \\
&\quad + \frac{1}{n\tilde{\sigma}^4}y_n'\left[W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0)\epsilon_n^e(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta})R_n^{e'}(Q_n,\tilde{\theta})\epsilon_n^e(Q_n,\tilde{\theta})\right] \\
&= \frac{1}{n}\left[\frac{1}{\sigma_0^4}-\frac{1}{\tilde{\sigma}^4}\right]y_n' W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0)\epsilon_n^e(Q_n,\theta_0) \\
&\quad + \frac{1}{n\tilde{\sigma}^4}y_n'\left[W_n^{e'}(Q_n,\theta_0)R_n^{e'}(Q_n,\theta_0) - W_n^{e'}(Q_n,\tilde{\theta})R_n^{e'}(Q_n,\tilde{\theta})\right]\epsilon_n^e(Q_n,\theta_0) + o_p(1)
\end{aligned}$$

as $\epsilon_n^e(Q_n,\tilde{\theta}) = R_n^e(Q_n,\tilde{\theta})(S_n^e(Q_n,\tilde{\theta})y_n - x_n\tilde{\beta}_1 - W_n^e(Q_n,\tilde{\theta})x_n\tilde{\beta}_2) - R_n^e(Q_n,\theta_0)(S_n^e(Q_n,\theta_0)y_n - x_n\beta_{10} - W_n^e(Q_n,\theta_0)x_n\beta_{20}) + \epsilon_n^e(Q_n,\theta_0)$ $= R_n^e(Q_n,\tilde{\theta})([S_n^e(Q_n,\tilde{\theta}) - S_n^e(Q_n,\theta_0)]y_n - x_n[\tilde{\beta}_1-\beta_{10}] - W_n^e(Q_n,\tilde{\theta})x_n\tilde{\beta}_2 + W_n^e(Q_n,\theta_0)x_n\beta_{20}) + R_n^e(Q_n,\tilde{\theta})[S_n^e(Q_n,\tilde{\theta}) - S_n^e(Q_n,\theta_0)]y_n - R_n^e(Q_n,\tilde{\theta})x_n[\tilde{\beta}_1-\beta_{10}] - R_n^e(Q_n,\tilde{\theta})W_n^e(Q_n,\tilde{\theta})x_n[\beta_{20}-\tilde{\beta}_2] + R_n^e(Q_n,\tilde{\theta})[W_n^e(Q_n,\theta_0) - W_n^e(Q_n,\tilde{\theta})]x_n\beta_{20} + [R_n^e(Q_n,\tilde{\theta}) - R_n^e(Q_n,\theta_0)]S_n^e(Q_n,\theta_0)y_n$ $-[R_n^e(Q_n,\tilde{\theta}) - R_n^e(Q_n,\theta_0)]x_n\beta_{10} - [R_n^e(Q_n,\tilde{\theta}) - R_n^e(Q_n,\theta_0)]\ W_n^e(Q_n,\theta_0)x_n\beta_{20} + \epsilon_n^e(Q_n,\theta_0)$, $[S_n^e(Q_n,\tilde{\theta}) - S_n^e(Q_n,\theta_0)]$, $[W_n^e(Q_n,\theta_0) - W_n^e(Q_n,\tilde{\theta})]$ and $[R_n^e(Q_n,\tilde{\theta}) - R_n^e(Q_n,\theta_0)] = o_p(1)$, and $R_n^e(Q_n,\tilde{\theta})$ is row and column-sum bounded, then $\frac{1}{n}\left\{\frac{\partial \ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda\partial\sigma^2} - \frac{\partial \ln\mathcal{L}^e(\theta_0)}{\partial\lambda\partial\sigma^2}\right\} = o_p(1)$.

By Mean Value Theorem, defining $G_j(\lambda,\theta_g) = (S_j^e(Q_j,\theta))^{-1}W_j^e(Q_j,\theta)$, $\text{tr}\left\{G_n^2(\bar{\lambda},\bar{\theta}_g)\right\} = \text{tr}\left\{G_n^2(\lambda_0,\theta_g^0)\right\} + 2\text{tr}\left\{G_n^3(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\lambda}-\lambda_0)$ $+2\text{tr}\left\{\nabla_{\theta_g}W_n^e(\bar{\lambda},\bar{\theta}_g)S_n^e(\bar{\lambda},\bar{\theta}_g)^{-1}G_n(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\theta}_g-\theta_0)$ $+2\lambda\text{tr}\left\{W_n^e(\bar{\lambda},\bar{\theta}_g)\nabla_{\theta_g}W_n^e(\bar{\lambda},\bar{\theta}_g)G_n(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\theta}_g-\theta_0)$ then

$$\begin{aligned}
\frac{1}{n}\left\{\frac{\partial \ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda^2} - \frac{\partial \ln\mathcal{L}^e(\theta_0)}{\partial\lambda^2}\right\} &= 2\text{tr}\left\{G_n^3(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\lambda}-\lambda_0) + 2\text{tr}\left\{\nabla_{\theta_g}W_n^e(\bar{\lambda},\bar{\theta}_g)S_n^e(\bar{\lambda},\bar{\theta}_g)^{-1}G_n(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\theta}_g-\theta_0) \\
&\quad + 2\lambda\text{tr}\left\{W_n^e(\bar{\lambda},\bar{\theta}_g)\nabla_{\theta_g}W_n^e(\bar{\lambda},\bar{\theta}_g)G_n(\bar{\lambda},\bar{\theta}_g)\right\}(\bar{\theta}_g-\theta_0) \\
&\quad + \left[\frac{1}{\sigma_0^2}-\frac{1}{\tilde{\sigma}^2}\right]\sum_{j=1}^{v} y_j' W_j^{e'}(Q_j,\theta_0)R_j^{e'}(Q_j,\theta_0)R_j^e(Q_j,\theta_0)W_j^e(Q_j,\theta_0)y_j \\
&\quad - \frac{1}{\tilde{\sigma}^2}\sum_{j=1}^{v} y_j'\left[W_j^{e'}(Q_j,\tilde{\theta})R_j^{e'}(Q_j,\tilde{\theta})R_j^e(Q_j,\tilde{\theta}) - W_j^{e'}(Q_j,\theta_0)R_j^{e'}(Q_j,\theta_0)R_j^e(Q_j,\theta_0)\right]W_j^e(Q_j,\tilde{\theta})y_j.
\end{aligned}$$

By similar arguments, as above, $\frac{1}{n}\left\{\frac{\partial \ln\mathcal{L}^e(\tilde{\theta})}{\partial\lambda^2} - \frac{\partial \ln\mathcal{L}^e(\theta_0)}{\partial\lambda^2}\right\} = o_p(1)$.

(Showing $\frac{1}{n}\frac{\partial \ln\mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial\theta\partial\theta'} \xrightarrow{p} \mathbb{E}\left(\frac{1}{n}\frac{\partial \ln\mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial\theta\partial\theta'}\right)$*)*. Terms that generically fit into the format $\omega_x(\theta) = \frac{1}{n}\varphi'\Delta(\theta)\varphi$, where φ is non-stochastic vector of dimension n and Δ is a stochastic matrix of conformable dimension can be shown to $\mathbb{V}\{\omega_x(\theta)\} \xrightarrow{p} 0$. For example, $-\frac{\sigma^2}{n}\frac{\partial^2 \ln\mathcal{L}^e(\theta_0)}{\partial\lambda\partial\beta_1'} = \frac{1}{n}x_n' R_n^e(Q_n,\theta_0)W_n^e(Q_n,\theta_0)y = \frac{1}{n}x_n' R_n^e(Q_n,\theta_0)W_n^e(Q_n,\theta_0)\left[\left(S_n^0\right)^{-1}Z_n^0\beta_0 + \left(S_n^0\right)^{-1}\left(R_n^0\right)^{-1}\epsilon_n\right] = \frac{1}{n}x_n' R_n^e(Q_n,\theta_0)$

$W_n^e(Q_n,\theta_0)\left(S_n^0\right)^{-1}x_n\beta_{10}+\frac{1}{n}x_n'R^e(Q_n,\theta_0)W_n^e(Q_n,\theta_0)\left(S_n^0\right)^{-1}W_n^0x_n\beta_{20}+o_p(1)$. Defining $x_n^{(l)}$ as the l-th column of x_n,

$$\omega_{xl}(\theta)\equiv\frac{1}{\sigma^2n}x_n^{(l)\prime}R_n^e(Q_n,\theta_0)W_n^e(Q_n,\theta_0)\left(S_n^0\right)^{-1}x_n^{(l)}=\frac{1}{\sigma^2n}\sum_{i=1}^n\sum_{j=1}^n x_{n,i}^{(l)}x_{n,j}^{(l)}\left(R_n^e\left(Q_n,\theta^0\right)W_n^e(Q_n,\theta_0)\left(S_n^0\right)^{-1}\right)_{ij}.$$

If elements of $\Delta(\theta)$ are approximately independent (taking, for example, $\left(S_n^0\right)^{-1}=I_n+\lambda W_n^0$ as the first-order Series Expansion), then

$$\mathbb{V}(\gamma_l)=\left[\frac{1}{\sigma^2n}\right]^2\sum_{i=1}^n\sum_{j=1}^n\left(x_{n,i}^{(l)}x_{n,j}^{(l)}\right)^2\mathbb{V}\left\{\left(R_n^e\left(Q_n,\theta^0\right)W_n^e(Q_n,\theta_0)\left(S_n^0\right)^{-1}\right)_{ij}\right\}$$

Noticing $\mathbb{V}\left(W_n^0\right)$ is a matrix of constants, $R_n^e\left(Q_n,\theta^0\right)W_n^e(Q_n,\theta_0)$ is column and row-sum bounded, then $\mathbb{V}\{\cdot\}$ goes to zero and so does $\mathbb{V}(\gamma_l)$. An equivalent argument goest through if terms in the middle contains matrix of derivatives. Terms that generically fit into $\omega_\epsilon(\theta)=\frac{1}{n}\epsilon_n'\Delta(\theta)\epsilon_n$, for example, $-\frac{\sigma^2}{n}\frac{\partial^2\ln\mathcal{L}^e(\theta)}{\partial\lambda\partial\sigma^2}=\frac{1}{\sigma^2n}y_n'W_n^{e\prime}(Q_n,\theta)R_n^{e\prime}(Q_n,\theta)\epsilon_n^e(Q_n,\theta)=\frac{1}{\sigma^2n}\epsilon_n'(S_n^{0\prime})^{-1}W_n^{e\prime}(Q_n,\theta)R_n^{e\prime}(Q_n,\theta)R_n^e(Q_n,\theta)((S_n^e(Q_n,\theta))^{-1}y_n$ $-Z_n^e(Q_n,\theta)\beta)=\frac{1}{\sigma^2n}\epsilon_n'(S_n^{0\prime})^{-1}W_n^{e\prime}(Q_n,\theta)R_n^{e\prime}(Q_n,\theta)R_n^e(Q_n,\theta)((S_n^e(Q_n,\theta))^{-1}y_n)+o_p(1)=\frac{1}{\sigma^2n}\epsilon_n'(S_n^{0\prime})^{-1}W_n^{e\prime}(Q_n,\theta)R_n^{e\prime}(Q_n,\theta)$ $R_n^e(Q_n,\theta)(S_n^e(Q_n,\theta))^{-1}\left(S_n^0\right)^{-1}\epsilon_n+o_p(1)$ by Lemma 1, and straightforward adaptation of Lemma 3, converges to

$$\mathbb{E}\left\{-\frac{\sigma^2}{n}\frac{\partial^2\ln\mathcal{L}^e(\theta)}{\partial\lambda\partial\sigma^2}\right\}=\frac{1}{n}\mathrm{tr}\left\{\mathbb{E}\left((S_n^{0\prime})^{-1}W_n^{e\prime}(Q_n,\theta)R_n^{e\prime}(Q_n,\theta)R_n^e(Q_n,\theta)(S_n^e(Q_n,\theta))^{-1}\left(S_n^0\right)^{-1}\right)\right\}.$$

(Asymptotic distribution). Given existence of higher order moments of ϵ_n, the Central Limit Theorem in Kelejian and Prucha (2001) can be applied to show that $\frac{1}{\sqrt{n}}\frac{\partial\ln\mathcal{L}^e(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,\Omega_\theta)$. Given non-singularity of the Hessian matrix as guaranteed by global identification condition in Theorem 1, it follows that

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,\Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1}\right).$$

□

D.6 Proposition 2.

Proof. (i). Starting from the definition of the social multiplier,

$$\begin{aligned}\varphi\left(x_n;W_n^e\left(Q_n,\theta_c^+\right),\beta_{10},\beta_+\right)&=\sum_{j=1}^\infty\lambda_+^{j-1}\left(W_n^e\left(Q_n,\theta_c^+\right)\right)^jx_n(\lambda_+\beta_{10}+\beta_{2+})=\\ \sum_{j=1}^\infty\lambda_0\lambda_+^{-1}\lambda_0^{j-1}\left(W_n^e\left(Q_n,\theta_c^0\right)\right)^jx_n(\lambda_+\beta_{10}+\beta_{2+})&=\sum_{j=1}^\infty\lambda_0^{j-1}\left(W_n^e\left(Q_n,\theta_c^0\right)\right)^jx_n\left(\lambda_0\beta_{10}+\lambda_0\lambda_+^{-1}\beta_{2+}\right)=\\ \sum_{j=1}^\infty\lambda_0^{j-1}\left(W_n^e\left(Q_n,\theta_c^0\right)\right)^jx_n(\lambda_0\beta_{10}+\beta_{20})&=\varphi\left(x_n;W^e\left(Q_n,\theta_c^0\right),\lambda_0,\beta_{10},\beta_{20}\right)\end{aligned}\tag{30}$$

where the last equality follows by $W_n^e(Q_n,\theta_c^+)x_n\beta_{2+} - W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{20} = \lambda_0\lambda_+^{-1}W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{2+} - W_n^e\left(Q_n,\theta_c^0\right)x_n\beta_{20} = W_n^e\left(Q_n,\theta_c^0\right)x_n$ $(\lambda_0\lambda_+^{-1}\beta_{2+} - \beta_{20}) = 0$. *(ii)*. Define $\Phi^*\left(\theta|\, y_n, x_n, Q_n\right) = \{\tilde{\theta} \in \Theta : Q_n(\tilde{\theta}) = Q_n\left(\theta\right)\}$. Sets $\Phi\left(\theta^0\middle|\, y_n, x_n\right) = \Phi^*\left(\theta^0\middle|\, y_n, x_n\right)$, as I now show. Inclusion $\Phi\left(\theta^0\middle|\, y_n, x_n\right) \subseteq \Phi^*\left(\theta^0\middle|\, y_n, x_n\right)$ is immediate from the first part. The reverse $\Phi^*\left(\theta^0\middle|\, y_n, x_n\right) \subseteq \Phi\left(\theta^0\middle|\, y_n, x_n\right)$ follows from a contradiction: suppose there is a θ^* such that $\theta^* \in \Phi\left(\theta^*|\, y_n, x_n\right)$ and $\theta^* \notin \Phi^*\left(\theta^*|\, y_n, x_n\right)$. By construction and Jensen's inequality, $Q_n\left(\theta^*\right) < Q_n\left(\theta^0\right)$. Observation of the reduced-form implies $\epsilon_n^e\left(Q_n,\theta_c^*\right) = \epsilon^e\left(Q_n,\theta^0\right)$, $\ln\left|S_n^e\left(Q_n,\theta^*\right)\right| = \ln\left|S_n^e\left(Q_n,\theta^0\right)\right|$ and $\ln\left|R_n^e\left(Q_n,\theta^*\right)\right| = \ln\left|R_n^e\left(Q,\theta^0\right)\right|$, and so $Q_n\left(\theta^*\right) = Q_n\left(\theta^0\right)$, a contradiction. Therefore, given that $\Phi\left(\theta^0\middle|\, y_n, x_n\right) = \Phi^*\left(\theta^0\middle|\, y_n, x_n\right)$, for any $\theta_c \in \Phi^*\left(\theta^0\middle|\, y_n, x_n\right)$, and, by definition, $\Phi^*\left(\theta^0\middle|\, y_n, x_n\right) = \Theta_0$, the result is proven. □

D.7 Theorem 3.

Proof. For parts (1) and (2), see Theorem 3.2 and Lemma 3.1 of Chernozhukov et al. (2007). By construction, and uniform convergence of Theorem 1 conditions C.1 with $a_n = n$, degeneracy property C.3 and condition C.4 therein are satisfied. Condition C.2 is guaranteed by uniform convergence and boundness of the objective function on a compact set Θ. Parts (3) and (4) are immediate corollaries. □

D.8 Example 3.

The full model is $y_j = \lambda_0 W_j^0 y_j + x_j\beta_{10} + W_j^0 x_j\beta_{20} + \epsilon_j$ with reduced form $y_j = (S_j^0)^{-1}x_j\beta_{10} + (S_j^0)^{-1}W_j^0 x_j\beta_{20} + (S_j^0)^{-1}\epsilon_j$. Then

$$y_j - \mathbb{E}y_j \quad = \quad ((S_j^0)^{-1} - \mathbb{E}(S_j^0)^{-1})x_j\beta_{10} + ((S_j^0)^{-1}W_j^0 - \mathbb{E}\{(S_j^0)^{-1}W_j^0\})x_j\beta_{20} + (S_j^0)^{-1}\epsilon_j$$

and $\mathbb{V}y_j = \mathbb{E}((y_j - \mathbb{E}y_j)(y_j - \mathbb{E}y_j)')$ is

$$\begin{aligned}\mathbb{V}y_j \quad = \quad & \mathbb{E}\left\{((S_j^0)^{-1} - \mathbb{E}(S_j^0)^{-1}x_j\beta_{10}\beta_{10}'x_j'((S_j^0)^{-1} - \mathbb{E}(S_j^{0'})^{-1})\right\} + 2\mathbb{E}\left\{((S_j^0)^{-1} - \mathbb{E}\{(S_j^0)^{-1}\})x_j\beta_{10}\beta_{20}'x_j'((S_j^0)^{-1}W_j^0 - \mathbb{E}\{(S_j^0)^{-1}W_j^0\})'\right\} \\ & + \mathbb{E}\left\{((S_j^0)^{-1}W_j^0 - \mathbb{E}\{(S_j^0)^{-1}W_j^0\})x_j\beta_{20}\beta_{20}'x_j'((S_j^0)^{-1}W_j^0 - \mathbb{E}\{(S_j^0)^{-1}W_j^0\})'\right\} + \mathbb{E}\left\{(S_j^0)^{-1}\epsilon_j\epsilon_j'(S_j^{0'})^{-1}\right\}.\end{aligned}$$

Denote these terms sequentially as A_j, B_j, C_j and D_j. $A_j = s_j x_j^{11} s_j'$, where $s_j = ((S_j^0)^{-1} - \mathbb{E}\{(S_j^0)^{-1}\})$ and $x_j^{11} = x_j\beta_{10}\beta_{10}'x_j'$. Then

$$A_j \quad = \quad \begin{bmatrix} \sum_{i,k}\mathbb{E}\left\{s_{1i}s_{1k}\right\}x_{ik}^{11} & \cdots & \sum_{i,k}\mathbb{E}\left\{s_{1i}s_{nk}\right\}x_{ik}^{11} \\ \vdots & \ddots & \vdots \\ \sum_{i,k}\mathbb{E}\left\{s_{ni}s_{1k}\right\}x_{ik}^{11} & \cdots & \sum_{i,k}\mathbb{E}\left\{s_{ni}s_{nk}\right\}x_{ik}^{11} \end{bmatrix}$$

where s_{ik} denotes the (i,k)th element of s_j, and similarly for x_j^{11}. Matrix s_j can be approximated $s = I + \lambda_0 W_j^0 + \lambda_0^2(W_j^0)^2 + \cdots - (I + \lambda_0\mathbb{E}W_j^0 + \lambda_0^2\mathbb{E}(W_j^0)^2 + \cdots) \approx \lambda_0(W_j^0 - W_j^e(\theta_j^0))$. Hence s_{ik} is dependent of $s_{i'k'}$ if, and only if, $i = i'$ and $k = k'$. Take w_{ik} as the (i,k)th element of W_j^0. This simplifies term A_j to

$$A_j = \lambda^2 \begin{bmatrix} \sum_i \mathbb{V}\{w_{1i}\} x_{ii}^{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{V}\{w_{ni}\} x_{ii}^{11} \end{bmatrix}$$

which then implies $A_j = \text{diag}\left(\lambda^2 \mathbb{V}\{W_j\} \text{diag}(x_j^{11})\right)$. Proceeding in a similar fashion, $B_j = s_j x_j^{12} s_j^{*\prime}$ with $x_j^{12} = x_j \beta_{10} \beta_{20}' x_j'$ and $s_j^* = W_j^0 + \lambda_0 (W_j^0)^2 + \lambda_0^2 (W_j^0)^3 + \cdots - (W_j^e(\theta_0) + \lambda_0 \mathbb{E}(W_j^0)^2 + \lambda_0^2 \mathbb{E}(W_j^0)^3 + \cdots) \approx W_j^0 - W_j^e(\theta_0)$

$$B_j = 2 \begin{bmatrix} \sum_{i,k} \mathbb{E}\{s_{1i} s_{1k}^*\} x_{ik}^{12} & \cdots & \sum_{i,k} \mathbb{E}\{s_{1i} s_{nk}^*\} x_{ik}^{12} \\ \vdots & \ddots & \vdots \\ \sum_{i,k} \mathbb{E}\{s_{ni} s_{1k}^*\} x_{ik}^{12} & \cdots & \sum_{i,k} \mathbb{E}\{s_{ni} s_{nk}^*\} x_{ik}^{12} \end{bmatrix} = 2\lambda \begin{bmatrix} \sum_i \mathbb{V}\{w_{1i}\} x_{ii}^{12} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{V}\{w_{ni}\} x_{ii}^{12} \end{bmatrix}$$

and then $B_j = \text{diag}\left(2\lambda \mathbb{V}\{W_j\} \text{diag}(x_j^{12})\right)$. The second equality uses independence between Bernoulli trials. For C_j,

$$\begin{aligned} C_j &= \begin{bmatrix} \sum_i \mathbb{E}\{s_{1i}^{*2}\} x_{ii}^{22} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{E}\{s_{ni}^{*2}\} x_{ii}^{22} \end{bmatrix} = \begin{bmatrix} \sum_i \mathbb{V}\{w_{1i}\} x_{ii}^{22} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{V}\{w_{ni}\} x_{ii}^{22} \end{bmatrix} \\ &= \text{diag}\left(\mathbb{V}\{W_j\} \text{diag}\left(x_j^{22}\right)\right) \end{aligned}$$

Lastly,

$$\begin{aligned} D_j &= \begin{bmatrix} \sum_{i,j} \mathbb{E}\{s_{1i} s_{1j}\} \mathbb{E}\{e_{ij}\} & \cdots & \sum_{i,j} \mathbb{E}\{s_{1i} s_{nj}\} \mathbb{E}\{e_{ij}\} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} \mathbb{E}\{s_{ni} s_{1j}\} \mathbb{E}\{e_{ij}\} & \cdots & \sum_{i,j} \mathbb{E}\{s_{ni} s_{nj}\} \mathbb{E}\{e_{ij}\} \end{bmatrix} = \begin{bmatrix} \sum_i \mathbb{E}\{s_{1i}^2\} \sigma^2 & \cdots & \sum_i \mathbb{E}\{s_{1i} s_{ni}\} \sigma^2 \\ \vdots & \ddots & \vdots \\ \sum_i \mathbb{E}\{s_{ni} s_{1i}\} \sigma^2 & \cdots & \sum_i \mathbb{E}\{s_{ni}^2\} \sigma^2 \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} \sum_i \mathbb{E}\{s_{1i}^2\} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{E}\{s_{ni}^2\} \end{bmatrix} = \lambda^2 \sigma^2 \begin{bmatrix} \sum_i \mathbb{V}\{w_{1i}\} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i \mathbb{V}\{w_{ni}\} \end{bmatrix} + \sigma^2 I_{n_j} \\ &= \lambda^2 \sigma^2 \text{diag}\left(\mathbb{V}\{W_j\} \iota_{n_j}\right) + \sigma^2 I_{n_j}. \end{aligned}$$

The entire expression reads $\mathbb{V} y_j = \text{diag}\left(\mathbb{V}\{W_j\} (\lambda^2 \text{diag}(x_j^{11}) + 2\lambda \text{diag}(x_j^{12}) + \lambda^2 \sigma^2 \iota_{n_j})\right) + \sigma^2 I_{n_j}$. Using Theorem 6 of Rothenberg (1971, p. 585), suffices that the jacobian of matrix of restrictions has rank equal to the unknown parameters. The identified set can be translated, in this case, as $\delta\lambda = \delta_0 \lambda_0$ and $\beta_2 \lambda^{-1} = \beta_{20} \lambda_0^{-1}$, where the combination of the parameters in the right hand side is identified from data; parameters

β_{10} and σ_0^2 are point-identified. The jacobian then reads

$$\mathcal{J}(\theta)=\begin{bmatrix} 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1\\ \delta & 0 & 0 & \lambda & 0\\ -\beta_2\lambda^{-2} & 0 & \lambda^{-1} & 0 & 0\\ J_{K1}(\theta) & J_{K2}(\theta) & J_{K3}(\theta) & J_{K4}(\theta) & J_{K5}(\theta)\end{bmatrix}$$

where

$$\begin{aligned} J_{K1}(\theta) &= 2n_j^{-1}\delta_1(1-\delta_1)\lambda\left(\iota_{n_j}'\operatorname{diag}(x_j^{11})+n_j\sigma^2\right)\\ J_{K2i}(\theta) &= n_j^{-1}\delta_1(1-\delta_1)\left(\lambda^2\iota_{n_j}'\frac{\partial\operatorname{diag}(x_j^{11})}{\partial\beta_{1i}}+2\lambda\iota_{n_j}'\frac{\partial\operatorname{diag}(x_j^{12})}{\partial\beta_{1i}}\right)\\ J_{K3i}(\theta) &= n_j^{-1}\delta_1(1-\delta_1)\left(2\lambda\iota_{n_j}'\frac{\partial\operatorname{diag}(x_j^{12})}{\partial\beta_{2i}}+\iota_{n_j}'\frac{\partial\operatorname{diag}(x_j^{22})}{\partial\beta_{2i}}\right)\\ J_{K4}(\theta) &= n_j^{-1}(1-2\delta_1)\left(\lambda^2\iota_{n_j}'\operatorname{diag}(x_j^{11})+2\lambda\iota_{n_j}'\operatorname{diag}(x_j^{12})+\iota_{n_j}'\operatorname{diag}(x_j^{22})+n_j\lambda^2\sigma^2\right)\\ J_{K5}(\theta) &= \delta_1(1-\delta_1)-n_j\lambda^2+1. \end{aligned}$$

Identification is guarateed with $\operatorname{rank}(\mathcal{J}(\theta))=K$, where K is the number of parameters in the structural model. Given σ_0^2 is identified, the last equation gives a solution for δ_1 and λ. Linear independence is guaranteed if the only column vector c that satisfies $\mathcal{J}(\theta)c=0$ is $c=0$. For the case of one exogenous covariate, this immediately implies $c_2=c_5=0$. We then have $c_1\delta+c_4\lambda=0$, $-c_1\beta\lambda^{-2}+c_3\lambda^{-1}=0$ and $c_1J_{K1}(\theta)+c_3J_{K3}(\theta)+c_4J_{K4}(\theta)=0$. Substituting out c_1 and c_3 in the third equation, one obtains the condition that $c_4\left[-\lambda\delta^{-1}J_{K1}(\theta)-\lambda\delta^{-1}\beta J_{K3}(\theta)+J_{K4}(\theta)\right]=0$. If $\lambda\neq0$, it is equivalent to $-\lambda\delta^{-1}J_{K1}(\theta)-\lambda\delta^{-1}\beta J_{K3}(\theta)+J_{K4}(\theta)\neq0$ at θ_0. This condition is empirically testable for all $\theta\in\Theta_0$, which is sufficient as $\theta_0\in\Theta_0$.

D.9 Theorem 4.

Proof. (Consistency). Because $\hat\Theta$ converges to Θ_0 in the Hausdorff metric, $\hat\Theta\subseteq\Theta_0^\epsilon$ for $\Theta_0^\epsilon=\{\theta\in\Theta: d(\theta,\Theta_0)\leq\epsilon\}$ with $\epsilon=o(1)$ and $\epsilon\geq0$. It follows that

$$\hat\theta = \arg\min_{\theta\in\Theta_0}\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}(y,\theta)\right)'\Omega\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}(y,\theta)\right)+o_p(1)$$

When S and v are going to infinity,

$$v^{-2}\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}(y,\theta)\right)'\Omega\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}(y,\theta)\right)\xrightarrow{a.s.}\left(\mathbb{E}_y^0\mathbb{E}_{W,e}q_{s,j}(y,\theta)\right)'\Omega\left(\mathbb{E}_y^0\mathbb{E}_{W,e}q_{s,j}(y,\theta)\right)$$

where $\mathbb{E}_{W,e}$ is the conditional expectation taken with respect to the distribution of W and e, given y and x and $\mathbb{E}_y^0$ is the expectation with respect to the true distribution of y, given x. Given that $\left(\mathbb{E}_y^0\mathbb{E}_{W,e}q_{s,j}\left(y,\theta\right)\right)'\Omega\left(\mathbb{E}_y^0\mathbb{E}_{W,e}q_{s,j}\left(y,\theta\right)\right)=\left(\mathbb{E}_y^0q_j\left(y,\theta\right)\right)'\Omega\left(\mathbb{E}_y^0q_j\left(y,\theta\right)\right)$ and $\mathbb{E}_y^0q_j\left(y,\theta_0\right)=0$ only at θ_0, consistency follows.

(Asymptotic normality). In the cases where $S\to\infty$ fast enough, results follow from standard asymptotic theory and Gouriéroux and Monfort (1997, Ch. 2). $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N\left(0,\Sigma^*\right)$, where $\Sigma_n=\left(G_n'\Omega_nG_n\right)^{-1}G_n'\Omega_nO_n\Omega_nG_n\left(G_n'\Omega_nG_n\right)^{-1}$, $G_n=\mathbb{E}\nabla_\theta q_j(y_n,\theta_0)$, $O_n=\mathbb{E}q_j(y_n,\theta_0)q_j(y_n,\theta_0)'$ and $\Sigma=\lim_{n\to\infty}\Sigma_n$. Optimal weight matrix is $\Omega_n^*=O_n^{-1}$ and, in this case, $\Sigma_n^*=\left(G_n'(\Omega_n^*)^{-1}G_n\right)^{-1}$ and $\Sigma^*=\lim_{n\to\infty}\Sigma_n^*$. When it can be shown that the local maximum is unique, the estimator can also be seen as the solution to

$$\hat{\theta}^\star = \arg\min_{\theta\in\Theta}\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}^\star\left(y,\theta\right)\right)'\Omega^\star\left(\sum_{j=1}^{v}S^{-1}\sum_{s=1}^{S}q_{s,j}^\star\left(y,\theta\right)\right)$$

where $q_{s,j}^\star\left(y,\theta\right)=\left[\nabla_\theta\ln\mathcal{L}^e\left(\theta\right)\quad q_{s,j}\left(y,\theta\right)\right]'$ and $\Omega^\star$ is a weight matrix of comformable dimensions with possibly arbitrary large weights for the first-order conditions, so that the restriction $\theta\in\hat{\Theta}$ is implemented. In the case where $S\to\infty$ fast enough, given identification, $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N\left(0,\Sigma^{\star*}\right)$, where $\Sigma_n^\star=(G_n^{\star'}\Omega_n^\star G_n^\star)^{-1}G_n^{\star'}\Omega_n^\star O_n^\star\Omega_n^\star G_n^\star(G_n^{\star'}\Omega_n^\star G_n^\star)^{-1}$, $G^\star=\mathbb{E}\nabla_\theta q_j^\star(y_j,\theta_0)$, $O^\star=\mathbb{E}q_j^\star(y,\theta_0)q_j^\star(y,\theta_0)'$ and $q_j^\star(y,\theta_0)=\lim_{S\to\infty}S^{-1}\sum_{s=1}^{S}q_{s,j}^\star(y,\theta_0)$ and $\Sigma^\star=\lim_{n\to\infty}\Sigma_n^\star$. Using optimal matrix $\Omega_n^{\star*}=\left(O_n^\star\right)^{-1}$, $\Sigma_n^{\star*}=\left(G_n^{\star'}\left(\Omega_n^{\star*}\right)^{-1}G_n^\star\right)^{-1}$, $\lim_{n\to\infty}\Sigma_n^{\star*}$. □

E Algorithms.

E.1 Bootstrap for $c_n\left(\alpha\right)$ and $c_n^f\left(\alpha\right)$

In the case of i.i.d. data, Bugni (2010) proposes a bootstrap algorithm correction consistent for $c_n\left(\alpha\right)$ and adaptable to $c_n^f\left(\alpha\right)$. In the current case, spatial dependence or social interactions in groups prevents immediate application of methods described therein. Instead, I propose bootstrapping at the group-level j, while maintaining within-group observations $i=1,\ldots,n_j$. In this way, dependence of observed data is preserved. Apart from the straightforward modification proposed here, proofs can be found in the aforementioned paper.

Algorithm 1. *(Bugni (2010) bootstrap). In order to produce confidence regions with coverage probability $1-\alpha$, $\alpha\in(0,1)$, for Θ_0, denoted $\hat{\Theta}_\alpha^B$ for a bootstrapped sample of arbitrary size B, follow the steps:*

Step 1. Estimate the identified set $\hat{\Theta}=\{\theta\in\Theta:L_n\left(\theta|\,y_n,x_n,Q_n\right)=0\}$.

Step 2. Define the bootstrapped sample $b=1,\ldots,B$, sampling v groups with replacement from the data and denote bootstrapped sample $\{y_n^b,x_n^b,Q_n^b\}$. Compute

$$\hat{c}_n^b = \sup_{\theta\in\hat{\Theta}}\sqrt{n}\left(L_n(\theta|\,y_n^b,x_n^b,Q_n^b)-L_n\left(\theta|\,y_n,x_n,Q_n\right)\right).$$

Step 3. Let $\hat{c}_n^B\left(\alpha\right)$ be the α quantile of the empirical distribution of $\{\hat{c}_n^1,\ldots,\hat{c}_n^B\}$. The $(1-\alpha)$ confidence set for the identified set is

$$\hat{\Theta}_\alpha^B = \left\{\theta\in\Theta:\sqrt{n}L\left(\theta|\,y_n,x_n,Q_n\right)\leq\hat{c}_n^B\left(1-\alpha\right)\right\}$$

Next, I produce an adaptation of the algorithm to be able to generate confidence regions for the image of the identified set under known function f, hence completing the statistical toolkit necessary for implementation of remarks 2 and 3.

Algorithm 2. *(Adaptation of Bugni (2010) bootstrap for projection under f). The modified algorithm to produce confidence regions with probability $1-\alpha$, $\alpha \in (0,1)$, for the projection of Θ_0 under known function f, Υ_0^f, denoted $\hat{\Upsilon}_\alpha^B$, for a bootstrapped sample of arbitrary size B is:*

Step 1. Estimate the projection of the identified set $\hat{\Upsilon} = \left\{ \upsilon \in \Upsilon : \inf_{\theta \in f^{-1}(\upsilon)} L_n\left(\theta | y_n, x_n, Q_n\right) = 0 \right\}$.

Step 2. Define the bootstrapped sample $b = 1, \ldots, B$, sampling v groups with replacement from the data and denote bootstrapped sample $\{y_n^b, x_n^b, Q_n^b\}$. Compute

$$\hat{c}_n^{f,b} = \sup_{\upsilon \in \hat{\Upsilon}} \inf_{\theta \in f^{-1}(\upsilon)} \sqrt{n} \left(L_n(\theta | y_n^b, x_n^b, Q_n^b) - L_n\left(\theta | y_n, x_n, Q_n\right) \right).$$

Step 3. Let $\hat{c}_n^{f,B}(\alpha)$ be the α quantile of the empirical distribution of $\{\hat{c}_n^{f,1}, \ldots, \hat{c}_n^{f,B}\}$. The $(1-\alpha)$ confidence set for the projected identified set Υ_0 is

$$\hat{\Upsilon}_\alpha^{f,B} = \left\{ \upsilon \in \Upsilon : \inf_{\theta \in f^{-1}(\upsilon)} \sqrt{n} L\left(\theta | y_n, x_n, Q_n\right) \leq \hat{c}_n^{f,B}(1-\alpha) \right\}.$$

E.2 Main algorithms

Algorithm 3. *If λ_0 is known and there are at least three distinct group sizes n_j, follow the steps:*

Step 1. Maximize the concentrated pseudo-likelihood

$$\ln \mathcal{L}_n^c\left(\theta_c | y_n, x_n, Q_n\right) = -\frac{n}{2}\left(\ln(2\pi)+1\right) - \frac{n}{2} \ln \hat{\sigma}^2\left(Q_n, \theta_c\right) + \left|S_n^e\left(Q_n, \theta_c\right)\right| + \left|R^e\left(Q_n, \theta_c\right)\right|$$

with respect to θ_g, where

$$\hat{\sigma}^2(Q_n, \theta_c) = \frac{1}{n} y_n' S_n^{e'}(Q_n, \theta_c) R_n^{e'}(Q_n, \theta_c) P_n^e(Q_n, \theta_c) R_n^e(Q_n, \theta_c) S_n^e(Q_n, \theta_c) y_n$$

and $P_n^e(Q_n, \theta_c) = I_n - R_n^e(Q_n, \theta_c) Z_n^e(Q_n, \theta_c) (Z_n^{e'}(Q_n, \theta_c) R_n^{e'}(Q_n, \theta_c) R_n^e(Q_n, \theta_c) Z_n^e(Q_n, \theta_c))^{-1} Z_n^{e'}(Q_n, \theta_c) R_n^{e'}(Q_n, \theta_c)$. Obtain the full solution $\hat{\theta} = (\hat{\theta}_c', \hat{\beta}(\hat{\theta}_c)', \hat{\sigma}^2(\hat{\theta}_c))'$, where $\hat{\theta}_c \equiv \arg\max_{\theta \in \Theta_c} \ln \mathcal{L}_n^c\left(\theta_c | y_n, x_n, Q_n\right)$ and

$$\hat{\beta}(\hat{\theta}_c) = (Z_n^{e'}(Q_n, \theta_c) R_n^{e'}(Q_n, \theta_c) R_n^e(Q_n, \theta_c) Z_n^e(Q_n, \theta_c))^{-1} Z_n^{e'}(Q_n, \theta_c) R_n^{e'}(Q_n, \theta_c) R_n^e(Q_n, \theta_c) S_n^e(Q_n, \theta_c) y_n.$$

Calculate and store the expected network $\hat{W}_n^e = W_n^e(Q_n, \hat{\theta})$.

Step 2. (C.I. of structural parameters). Calculate the asymptotic variance given by Theorem 1. The full expressions of the Jacobian and Hessian are given in Appendix C or can be numerically approximated.

Step 3. (Network spillovers). Network spillovers are calculated as

$$\varphi(x_n, \hat{\theta}) = \hat{W}_n^e x_n \hat{\beta}_2 + \sum_{s=1}^{s_{max}} (\lambda_0 \hat{W}_n)^s (x_n \hat{\beta}_1 + \hat{W}_n x_j \hat{\beta}_2).$$

In practice, $s_{max} = 25$ has been shown to provide a good approximation to the case where $s_{max} \to \infty$. Confidence intervals follow from a simple Delta Method, $\sqrt{n^*}(\varphi(x_n, \hat{\theta}) - \varphi(x_n, \theta_0)) \xrightarrow{d} N(0, \nabla\varphi(x_n, \theta_0) \Sigma^{-1}(\lambda_0) \Omega(\lambda_0) \Sigma^{-1}(\lambda_0) \nabla\varphi(x_n, \theta_0))$.

Step 4. (Network data validity). When network data are available, a Delta Method also is employed to provide confidence intervals for the null hypothesis $\mathcal{H}_0 : \delta_1 - \delta_0 = 0$.

Algorithm 4. *The following algorithm generalizes for the case in which λ_0 is unknown. If there are at least three distinct group sizes n_j, follow the steps:*

Step 1. Select a candidate λ_0.

Step 2. Maximize the concentrated pseudo-likelihood

$$\ln \mathcal{L}_n^c\left(\theta_c | y_n, x_n, Q_n\right) = -\frac{n}{2}\left(\ln(2\pi)+1\right) - \frac{n}{2}\ln \hat{\sigma}^2\left(Q_n, \theta_c\right) + \left|S_n^e\left(Q_n, \theta_c\right)\right| + \left|R^e\left(Q_n, \theta_c\right)\right|$$

with respect to θ_g *and obtain the set of solution* $\hat{\theta} = (\hat{\theta}_c', \hat{\beta}(\hat{\theta}_c)', \hat{\sigma}^2(\hat{\theta}_c))'$ *such that* $\hat{\theta}_c \equiv \arg\max_{\theta \in \Theta_c} \ln \mathcal{L}_n^c\left(\theta_c | y_n, x_n, Q_n\right)$. *Denote this set* $\hat{\Theta}$. *Full expressions for the concentrated parameters* $\hat{\beta}(\hat{\theta}_c)$ *and* $\hat{\sigma}^2(\hat{\theta}_c)$ *are given in Step 1 of Algorithm 3.*

Step 3. Check if probability of peers forming link is in the $[0, 1]$ *range. Otherwise, go back to Step 1 and adjust* λ_0 *accordingly.*

Step 4. (C.I. of structural parameters). Obtain confidence regions for θ_g *following the bootstrap Algorithm 1.*

Step 5. (Network spillovers). Take any point $\hat{\theta}^*$ *in the identified* $\hat{\Theta}$. *Network spillovers are calculated as*

$$\varphi(x_n, \hat{\theta}^*) = \hat{W}_n^e x_n \hat{\beta}_2^* + \sum_{s=1}^{s_{max}} (\lambda_0 \hat{W}_n^*)^s (x_n \hat{\beta}_1^* + \hat{W}_n^* x_j \hat{\beta}_2^*).$$

Where $\hat{W}_n^{e*} = W_n^e(Q_n, \hat{\theta}^*)$. *Again,* $s_{max} = 25$ *has been shown to provide a good approximation to the case where* $s_{max} \to \infty$. *Confidence intervals are calculated following Algorithm 2.*

Step 6. (Network data validity). When network data are available, Algorithm 2 is reemployed to provide confidence intervals for the null hypothesis $\mathcal{H}_0 : \delta_1 - \delta_0 = 0$.

Step 7. (Identifying λ*). Solve the GMM problem*

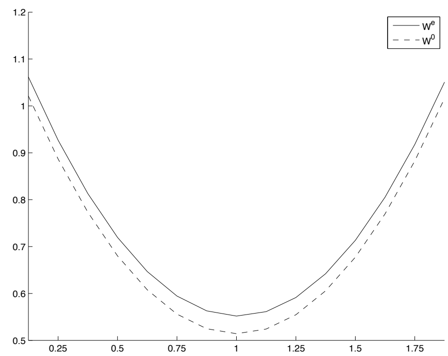
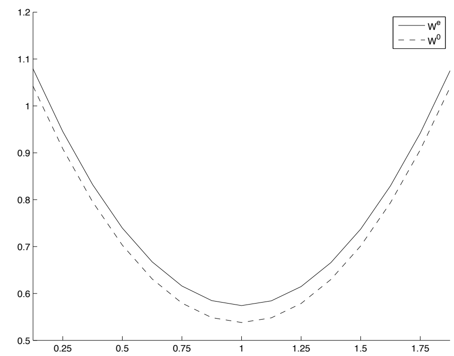
$$\hat{\theta} = \underset{\theta \in \hat{\Theta}}{arg\,min} \left(\sum_{j=1}^{v} S^{-1} \sum_{s=1}^{S} q_{s,j}(y_j, x_j, \theta) \right)' \Omega \left(\sum_{j=1}^{v} S^{-1} \sum_{s=1}^{S} q_{s,j}(y_j, x_j, \theta) \right)$$

where $q_{s,j}(y_j, x_j, \theta) = [V_{B,j}(y_j, x_j, \theta) - V_{B,j}(\hat{y}_j, x_j, \theta); V_{W,j}(y_j, x_j, \theta) - V_{W,j}(\hat{y}_j, x_j, \theta)]'$ *with* $\hat{y}_{j,s} = (S_j^s)^{-1}(x_j\beta_1 + W_j^s x_j \beta_2 + e_j^s)$ *and* $S^s = (I_{n_j} - \lambda W_j^s)^{-1}$. W_j^s *is sampled from the distribution of the network-generating model and* e_j^s *is sampled from a normal distribution with variance* σ^2. *Confidence intervals are given in Theorem 4.*

F Additional figures and tables.

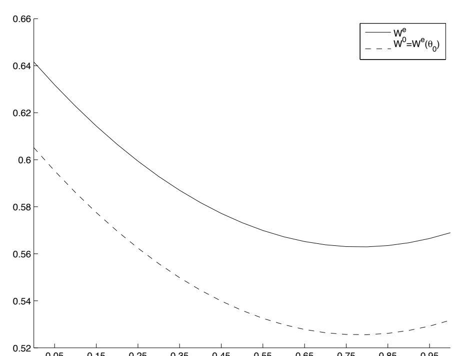
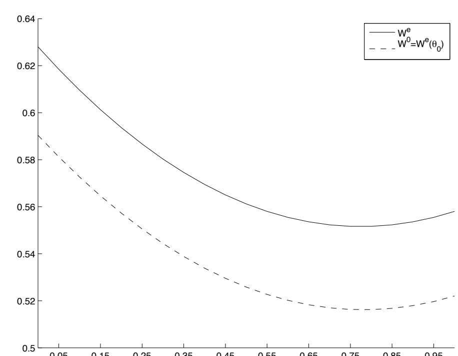
F.1 Estimator and simulations.

Table 7: Likelihood as a function of β_1.

$N = 250$	$N = 500$
	

Note: Rescaled additive inverse of likelihood as a function of β_1, with all other parameters at the true value. True $\beta_{10} = 1$. Solid line represents likelihood computed with expected network $W^e = W^e(Q, \theta_0)$, and dashed with real network W^0. True networks are realizations from the stochastic generating process.

Table 8: Likelihood as a function of δ_1.

$N = 250$	$N = 500$
	

Note: Rescaled additive inverse of likelihood as a function of δ_1, with all other parameters at the true value. True $\delta_{10} = 0.75$. Solid line represents likelihood computed with expected network $W^e = W^e(Q, \theta_0)$ and underlying networks are realization from the stochastic generating process. Dashed line $W^0 = W^e(\theta_0)$ is the likelihood where true network is equal to expected network.

Table 9: Simulations: bias in $\hat{\beta}_{OLS}$.

	$T = 1.$				$T = 5$, fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	25	100	25	100	25	100	25	100
v	250	250	1000	1000	250	250	1000	1000
$\hat{\beta}_1$	1.0670	1.1127	1.0670	1.1120	1.0667	1.1116	1.0669	1.1118
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

	$T = 5$, time effects.				$T = 5$, time and fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	25	100	25	100	25	100	25	100
v	250	250	1000	1000	250	250	1000	1000
$\hat{\beta}_1$	1.0665	1.1110	1.0670	1.1120	1.0660	1.1110	1.0671	1.1120
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: True parameters is $\beta_1 = 1$.

Table 10: Simulations: baseline model with small n and v.

	$T=1$.				$T=5$, fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	15	25	15	25	15	25	15	25
v	25	25	50	50	25	25	50	50
$\hat{\lambda}$	0.0270	0.0219	0.0213	0.0178	0.0199	0.0172	0.0168	0.0141
	[0.022]	[0.019]	[0.011]	[0.014]	[0.015]	[0.011]	[0.011]	[0.005]
$\hat{\beta}_1$	1.0023	0.9972	1.0035	1.0002	1.0023	1.0003	0.9997	1.0003
	(0.036)	(0.030)	(0.025)	(0.021)	(0.017)	(0.014)	(0.013)	(0.010)
	[0.036]	[0.029]	[0.025]	[0.021]	[0.016]	[0.013]	[0.011]	[0.009]
$\hat{\beta}_2$	0.0631	0.0175	0.0654	0.0190	0.1263	0.0501	0.0718	0.0469
	(0.904)	(0.651)	(0.405)	(0.527)	(0.648)	(0.071)	(0.120)	(0.020)
	[0.818]	[0.933]	[0.582]	[0.627]	[0.793]	[0.055]	[0.145]	[0.018]
$\hat{\delta}_1$	0.3637	0.5942	0.6637	0.6264	0.6971	0.7605	0.6823	0.7201
	(1.032)	(0.558)	(0.607)	(0.359)	(0.451)	(0.247)	(0.306)	(0.175)
	[0.804]	[0.498]	[0.556]	[0.374]	[0.360]	[0.228]	[0.258]	[0.163]
$\hat{\delta}_0$	0.0606	0.2101	0.2458	0.2235	0.2551	0.2956	0.2635	0.2901
	(0.539)	(0.234)	(0.276)	(0.139)	(0.161)	(0.103)	(0.114)	(0.070)
	[0.394]	[0.218]	[0.247]	[0.143]	[0.145]	[0.086]	[0.103]	[0.062]
$\hat{\sigma}^2$	1.0248	1.0448	1.0204	1.0475	0.8229	0.8436	0.8286	0.8455
	(1.424)	(0.598)	(0.461)	(0.449)	(0.641)	(0.068)	(0.126)	(0.021)
	[0.071]	[0.059]	[0.050]	[0.042]	[0.026]	[0.021]	[0.018]	[0.015]
$\varphi(x,\hat{\theta})$	0.0062	0.0013	-0.0022	0.0006	0.0008	0.0005	-0.0023	0.0004
	(0.051)	(0.067)	(0.036)	(0.046)	(0.020)	(0.030)	(0.015)	(0.018)
	[0.167]	[0.147]	[0.067]	[0.069]	[0.033]	[0.013]	[0.010]	[0.006]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x,\theta) = 0$.

Table 11: Simulations: baseline model with small n and v.

	$T=5$, time effects.				$T=5$, time and fixed effects.			
	(1)	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
n	15	25	15	25	15	25	15	25
v	25	25	50	50	25	25	50	50
$\hat{\lambda}$	0.0178	0.0141	0.0156	0.0137	0.0197	0.0164	0.0166	0.0142
	[0.014]	[0.007]	[0.001]	[0.004]	[0.017]	[0.013]	[0.012]	[0.006]
	1.0025	1.0015	1.0003	1.0000	0.9997	0.9995	1.0004	1.0000
$\hat{\beta}_1$	(0.016)	(0.014)	(0.011)	(0.009)	(0.017)	(0.015)	(0.012)	(0.010)
	[0.016]	[0.013]	[0.011]	[0.009]	[0.016]	[0.013]	[0.011]	[0.009]
	0.0021	0.0664	0.0632	0.0444	0.0501	0.0808	0.0225	0.0508
$\hat{\beta}_2$	(0.396)	(0.241)	(0.054)	(0.013)	(0.318)	(0.178)	(0.525)	(0.025)
	[0.379]	[0.237]	[0.056]	[0.016]	[0.304]	[0.238]	[0.625]	[0.021]
	0.5643	0.6124	0.6611	0.7255	0.5651	0.6160	0.6345	0.6956
$\hat{\delta}_1$	(0.406)	(0.241)	(0.255)	(0.137)	(0.434)	(0.268)	(0.285)	(0.183)
	[0.360]	[0.237]	[0.259]	[0.161]	[0.353]	[0.224]	[0.253]	[0.162]
	0.2225	0.2479	0.2648	0.2883	0.2102	0.2480	0.2537	0.2732
$\hat{\delta}_0$	(0.164)	(0.091)	(0.106)	(0.052)	(0.178)	(0.108)	(0.110)	(0.070)
	[0.147]	[0.092]	[0.103]	[0.061]	[0.144]	[0.087]	[0.103]	[0.061]
	0.0139	0.0082	0.0123	0.0041	0.0049	0.0110	0.0372	0.0138
$\hat{\sigma}^2$	(0.380)	(0.032)	(0.055)	(0.013)	(0.052)	(0.171)	(0.054)	(0.025)
	[0.053]	[0.018]	[0.037]	[0.031]	[0.023]	[0.027]	[0.027]	[0.021]
	-0.0023	-0.0082	-0.0000	0.0011	-0.0004	-0.0006	-0.0000	-0.0008
$\varphi(x,\hat{\theta})$	(0.022)	(0.032)	(0.016)	(0.021)	(0.022)	(0.028)	(0.015)	(0.019)
	[0.034]	[0.018]	[0.010]	[0.006]	[0.046]	[0.016]	[0.015]	[0.006]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

Table 12: Simulations: baseline model with across-group connections.

$T = 1$

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(6)*
n	100	100	100	100	100	100	100
v	250	250	250	250	250	250	250
δ_A	0.00	0.01	0.025	0.05	0.075	0.10	0.25
$\hat{\lambda}$	0.0120	0.0123	0.0132	0.0132	0.0140	0.0151	0.0278
	[0.001]	[0.002]	[0.003]	[0.002]	[0.002]	[0.006]	[0.001]
$\hat{\beta}_1$	0.9999	1.0000	1.0000	0.9999	0.9995	0.9992	0.9958
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
$\hat{\beta}_2$	0.0403	0.0396	0.0399	0.0379	0.0355	0.0313	0.0124
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.002)
	[0.007]	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]	[0.002]
$\hat{\delta}_1$	0.7569	0.7578	0.7605	0.7857	0.8243	0.8855	1.0000
	(0.081)	(0.074)	(0.081)	(0.079)	(0.086)	(0.076)	(0.071)
	[0.080]	[0.080]	[0.080]	[0.081]	[0.079]	[0.079]	[0.069]
$\hat{\delta}_0$	0.3010	0.3036	0.3035	0.3131	0.3264	0.3484	0.5397
	(0.028)	(0.029)	(0.032)	(0.030)	(0.032)	(0.031)	(0.036)
	[0.031]	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]	[0.035]
$\hat{\sigma}^2$	1.0558	1.0582	1.0629	1.0696	1.0778	1.0849	1.1255
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.002)
	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
$\varphi(x, \hat{\theta})$	0.0008	-0.0010	-0.0015	0.0014	-0.0000	-0.0011	-0.0001
	(0.010)	(0.010)	(0.011)	(0.011)	(0.009)	(0.010)	(0.010)
	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

Table 13: Simulations: Bernoulli model under misspecified λ_{ref}.

	$T = 1$.		$T = 5$, FE.		$T = 5$, TE.		$T = 5$, FE and TE.	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
n	25	100	25	100	25	100	25	100
v	250	250	250	250	250	250	250	250
$\hat{\lambda}$	0.1002	0.0094	0.0092	0.0094	0.0099	0.0105	0.0977	0.0113
	[0.002]	[0.000]	[0.006]	[0.009]	[0.001]	[0.000]	[0.006]	[100.00%]
$\hat{\beta}_1$	1.0016	1.0009	0.9996	1.0008	1.0000	1.0002	1.0005	0.9999
	(0.008)	(0.005)	(0.004)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)
	[0.009]	[0.005]	[0.004]	[0.002]	[0.004]	[0.002]	[0.004]	[0.002]
$\hat{\beta}_2$	0.0892	0.0817	0.0878	0.0793	0.0826	0.0801	0.0825	0.0813
	(0.037)	(0.007)	(0.019)	(0.003)	(0.010)	(0.002)	(0.011)	(0.003)
	[0.033]	[0.007]	[0.014]	[0.003]	[0.012]	[0.003]	[0.012]	[0.003]
$\hat{\delta}_1$	0.3766	0.3722	0.3606	0.3764	0.3702	0.3745	0.3700	0.3729
	(0.095)	(0.014)	(0.050)	(0.005)	(0.033)	(0.005)	(0.032)	(0.006)
	[0.082]	[0.013]	[0.037]	[0.006]	[0.036]	[0.006]	[0.036]	[0.006]
$\hat{\delta}_0$	0.1427	0.1496	0.1449	0.1499	0.1475	0.1501	0.1488	0.1498
	(0.038)	(0.005)	(0.018)	(0.002)	(0.012)	(0.001)	(0.013)	(0.002)
	[0.029]	[0.004]	[0.014]	[0.002]	[0.013]	[0.002]	[0.013]	[0.002]
$\hat{\sigma}^2$	1.0550	1.2193	0.8458	0.9767	0.0678	0.0495	0.0376	0.0312
	(0.037)	(0.007)	(0.019)	(0.003)	(0.010)	(0.002)	(0.011)	(0.003)
	[0.019]	[0.011]	[0.007]	[0.004]	[0.014]	[0.009]	[0.012]	[0.003]
$\varphi(x, \hat{\theta})$	-0.0032	-0.0095	0.0029	-0.0037	-0.0026	-0.0061	-0.0013	-0.0007
	(0.018)	(0.085)	(0.019)	(0.030)	(0.009)	(0.053)	(0.011)	(0.027)
	[0.006]	[0.006]	[0.007]	[0.001]	[0.001]	[0.009]	[0.001]	[0.001]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

Table 14: Simulations: multivariate network model.

	$T = 1.$		$T = 5$, FE.		$T = 5$, TE.		$T = 5$, FE and TE.	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
n	25	100	25	100	25	100	25	100
v	250	250	250	250	250	250	250	250
$\hat{\lambda}$	0.0132	0.0124	0.0122	0.0125	0.0124	0.0125	0.0128	0.0125
	[0.002]	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]	[0.001]	[0.000]
	0.9979	0.9997	0.9999	1.0001	0.9999	1.0000	1.0000	0.9999
$\hat{\beta}_1$	(0.009)	(0.005)	(0.004)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)
	[0.009]	[0.005]	[0.004]	[0.002]	[0.004]	[0.002]	[0.004]	[0.002]
	0.0432	0.0400	0.0398	0.0398	0.0394	0.0399	0.0403	0.0401
$\hat{\beta}_2$	(0.018)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
	[0.014]	[0.001]	[0.002]	[0.001]	[0.002]	[0.001]	[0.002]	[0.001]
	0.2395	0.2500	0.2509	0.2515	0.2574	0.2509	0.2452	0.2503
$\hat{\delta}_1$	(0.048)	(0.009)	(0.022)	(0.005)	(0.017)	(0.004)	(0.020)	(0.005)
	[0.016]	[0.007]	[0.011]	[0.005]	[0.015]	[0.071]	[0.010]	[0.004]
	0.4949	0.5013	0.5029	0.5021	0.5077	0.5010	0.4969	0.4990
$\hat{\delta}_0$	(0.077)	(0.011)	(0.025)	(0.006)	(0.024)	(0.005)	(0.025)	(0.008)
	[0.018]	[0.010]	[0.011]	[0.007]	[0.015]	[0.068]	[0.012]	[0.005]
	1.0193	1.0697	0.8159	0.8546	0.1854	0.1956	0.1062	0.2845
$\hat{\sigma}^2$	(0.016)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)
	[0.018]	[0.010]	[0.007]	[0.003]	[0.013]	[0.007]	[0.005]	[0.002]
	0.0088	0.0027	0.0045	0.0009	0.0053	0.0009	0.0037	0.0021
$\varphi(x, \hat{\theta})$	(0.007)	(0.002)	(0.003)	(0.008)	(0.004)	(0.009)	(0.003)	(0.010)
	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.25$, $\delta_0 = 0.50$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$.

F.2 Application.

Table 15: Occupational Choice.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Self hours.		Wage hours.		Self emp. only.	
	Method	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.
Not function of $\hat{\lambda}$.	Program effect	473.219***	473.581***	−113.002***	−113.146***	0.113***	0.114***
	after 2 years ($\hat{\beta}_{11}$).	(12.99)	(13.89)	(8.33)	(8.33)	(0.01)	(0.01)
	Program effect	464.069***	463.441***	−142.755***	−143.009***	0.120***	0.121***
	after 4 years ($\hat{\beta}_{12}$).	(13.07)	(5.10)	(8.53)	(8.25)	(0.01)	(0.01)
	Spillover on T	−20.438***	−23.097***	24.394***	26.933***	−0.029***	−0.034***
	after 2 years ($\hat{\varphi}_{T,2}$).	(7.01)	(6.95)	(8.50)	(9.21)	(0.01)	(0.01)
	Spillover on T	17.396***	14.734**	19.805**	22.105**	−0.023***	−0.027**
	after 4 years ($\hat{\varphi}_{T,4}$).	(6.41)	(7.04)	(8.37)	(10.30)	(0.00)	(0.01)
	Spillover on NT	−9.771***	−11.346***	12.692***	14.259***	−0.015***	−0.018***
	after 2 years ($\hat{\varphi}_{NT,2}$).	(3.35)	(3.42)	(4.41)	(4.87)	(0.00)	(0.00)
	Spillover on NT	8.317**	7.237***	10.304**	11.703	−0.012***	−0.014***
	after 4 years ($\hat{\varphi}_{NT,4}$).	(3.28)	(1.88)	(5.21)	(13.28)	(0.01)	(0.00)
Function of $\hat{\lambda}$.	Link to T	−40.247***	−27.635***	12.794***	13.663***	−0.045***	−0.051***
	after 2 years ($\hat{\beta}_{21}$).	(1.99)	(1.42)	(2.48)	(2.72)	(0.01)	(0.01)
	Link to T	−30.758***	−20.648***	12.938***	13.721***	−0.040***	−0.045***
	after 4 years ($\hat{\beta}_{22}$).	(1.53)	(1.77)	(1.57)	(2.73)	(0.01)	(0.01)
	Link probability	0.776***	0.759***	0.985***	0.726***	0.336***	0.196***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).	(0.05)	(0.05)	(0.08)	(0.05)	(0.03)	(0.02)
	Link probability	0.317***	0.464***	0.364***	0.362***	0.115***	0.116***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
	$\hat{\lambda}$	0.075	0.05	0.05	0.05	0.15	0.15
	p-value $\mathcal{H}_{NV}$.	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Avg treated outcome.	421.8	421.8	646.7	646.7	0.303	646.7
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes as in Table 3.

Table 16: Earnings and Seasonality.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Earnings.		Share Seas.		Share Reg.	
	Method	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.
Not function of $\hat{\lambda}$.	Program effect	0.562***	0.556***	−0.029***	−0.029***	0.181***	0.182***
	after 2 years ($\hat{\beta}_{11}$).	(0.207)	(0.148)	(0.01)	(0.01)	(0.01)	(0.01)
	Program effect	2.726***	2.806***	−0.075***	−0.075***	0.166***	0.166***
	after 4 years ($\hat{\beta}_{12}$).	(0.196)	(0.108)	(0.01)	(0.01)	(0.01)	(0.01)
	Spillover on T	−0.258**	−0.187*	−0.063***	−0.062***	0.037***	0.030***
	after 2 years ($\hat{\varphi}_{T,2}$).	(0.116)	(0.113)	(0.02)	(0.02)	(0.01)	(0.01)
	Spillover on T	−0.098	−0.188*	−0.016*	−0.016	0.044***	0.035***
	after 4 years ($\hat{\varphi}_{T,4}$).	(0.117)	(0.112)	(0.01)	(0.02)	(0.02)	(0.01)
	Spillover on NT	−0.133**	−0.102*	−0.026***	−0.026***	0.017***	0.014***
	after 2 years ($\hat{\varphi}_{NT,2}$).	(0.060)	(0.062)	(0.01)	(0.01)	(0.01)	(0.01)
	Spillover on NT	−0.051	−0.103	−0.002	−0.007	0.020**	0.017**
	after 4 years ($\hat{\varphi}_{NT,4}$).	(0.057)	(0.78)	(0.01)	(0.01)	(0.01)	(0.00)
Function of $\hat{\lambda}$.	Link to T	−0.236	−0.245	−0.023***	−0.017***	−0.051***	−0.053***
	after 2 years ($\hat{\beta}_{21}$).	(0.456)	(0.478)	(0.02)	(0.00)	(0.01)	(0.01)
	Link to T	−0.740	−0.375	−0.019***	−0.014***	−0.037***	−0.039***
	after 4 years ($\hat{\beta}_{22}$).	(0.541)	(0.596)	(0.01)	(0.00)	(0.01)	(0.01)
	Link probability	0.155***	0.064***	0.234***	0.236***	0.100***	0.077***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)
	Link probability	0.030***	0.030***	0.152***	0.203***	0.054***	0.051***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	$\hat{\lambda}$	0.50	0.50	0.20	0.15	0.50	0.50
	p-value $\mathcal{H}_{NV}$.	< 0.001	< 0.001	< 0.001	0.022	< 0.001	< 0.001
	Avg treated outcome.	4.607	4.607	0.674	0.674	0.478	0.478
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes as in Table 3.

Table 17: Livestock.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Cows.		Poultry.		Livestock Value.	
	Method	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.
Not function of $\hat{\lambda}$.	Program effect	1.132***	1.132***	2.116***	2.117***	10.412***	10.420***
	after 2 years ($\hat{\beta}_{11}$).	(0.03)	(0.03)	(0.50)	(0.50)	(365.41)	(0.45)
	Program effect	1.103***	1.101***	1.296***	1.330***	11.175***	11.173***
	after 4 years ($\hat{\beta}_{12}$).	(0.03)	(0.03)	(0.50)	(0.50)	(459.21)	(0.44)
	Spillover on T	−0.032***	−0.033***	0.039	0.107	−0.184***	−0.230***
	after 2 years ($\hat{\varphi}_{T,2}$).	(0.01)	(0.01)	(0.11)	(0.18)	(0.07)	(0.06)
	Spillover on T	−0.055***	−0.055***	0.029	−0.095	−0.407***	−0.456***
	after 4 years ($\hat{\varphi}_{T,4}$).	(0.02)	(0.02)	(0.12)	(0.21)	(0.11)	(0.06)
	Spillover on NT	−0.018***	−0.020***	0.014	0.064	−0.106***	−0.137***
	after 2 years ($\hat{\varphi}_{NT,2}$).	(0.01)	(0.01)	(0.06)	(0.11)	(0.04)	(0.04)
	Spillover on NT	−0.031***	−0.032***	0.011	−0.056	−0.234***	−0.272***
	after 4 years ($\hat{\varphi}_{NT,4}$).	(0.01)	(0.01)	(0.10)	(0.08)	(0.08)	(0.03)
Function of $\hat{\lambda}$.	Link to T	−0.965***	−0.996***	9.169	1.495	−9.251***	−10.634***
	after 2 years ($\hat{\beta}_{21}$).	(0.15)	(0.15)	(19.65)	(4.22)	(2.64)	(1.22)
	Link to T	−1.227***	−1.256***	6.975	−2.914	−14.504***	−16.332***
	after 4 years ($\hat{\beta}_{22}$).	(0.16)	(0.16)	(21.05)	(4.21)	(2.30)	(2.07)
	Link probability	0.039***	0.019***	0.020**	0.008	0.029***	0.010**
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Link probability	0.014***	0.014***	0.011	0.008***	0.008***	0.008***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
	$\hat{\lambda}$	0.50	0.50	0.50	0.50	0.50	0.50
	p-value $\mathcal{H}_{NV}$.	0.003	0.300	0.045	1.000	0.024	0.764
	Avg treated outcome.	0.083	0.083	1.79	1.79	0.940	0.940
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes as in Table 3.

Table 18: Expenditures.

		(1)	(2)	(3)	(4)	(5)	(6)
	Outcome	Nonfood PCE.		Food PCE.		Food Security.	
	Method	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.	Network. Family.	Network. Economic.
Not function of $\hat{\lambda}$.	Program effect	−208.803	−208.049	421.741***	424.602***	0.169***	0.169***
	after 2 years ($\hat{\beta}_{11}$).	(160.98)	(160.05)	(133.67)	(133.61)	(0.01)	(0.01)
	Program effect	280.309*	279.158	444.980***	447.736***	0.075***	0.076***
	after 4 years ($\hat{\beta}_{12}$).	(145.11)	(178.65)	(133.66)	(133.61)	(0.01)	(0.01)
	Spillover on T	−29.966	−32.452	401.713***	387.106***	0.028***	0.083***
	after 2 years ($\hat{\varphi}_{T,2}$).	(70.13)	(69.80)	(56.88)	(56.47)	(0.01)	(0.03)
	Spillover on T	−161.955**	−161.161**	253.726***	242.561***	0.080**	0.163***
	after 4 years ($\hat{\varphi}_{T,4}$).	(71.28)	(69.72)	(59.58)	(55.82)	(0.03)	(0.05)
	Spillover on NT	−17.507	−19.103	215.298***	208.075***	0.012***	0.033***
	after 2 years ($\hat{\varphi}_{NT,2}$).	(40.98)	(41.09)	(30.18)	(30.97)	(0.00)	(0.01)
	Spillover on NT	−94.620***	−94.869**	135.984***	130.380***	0.033***	0.065***
	after 4 years ($\hat{\varphi}_{NT,4}$).	(26.64)	(39.08)	(51.07)	(29.85)	(0.00)	(0.02)
Function of $\hat{\lambda}$.	Link to T	−311.329	−349.080	343.343***	438.309***	0.102***	0.123***
	after 2 years ($\hat{\beta}_{21}$).	(966.77)	(968.78)	(62.93)	(83.73)	(0.01)	(0.01)
	Link to T	−2386.991**	−2389.737**	190.068***	238.308***	0.088***	0.113***
	after 4 years ($\hat{\beta}_{22}$).	(959.21)	(962.22)	(62.48)	(83.19)	(0.01)	(0.01)
	Link probability	0.020**	0.014**	0.158***	0.132***	0.184***	0.092***
	if $Q_{ij} = 1$ ($\hat{\delta}_1$).	(0.01)	(0.01)	(0.03)	(0.01)	(0.00)	(0.00)
	Link probability	0.013***	0.013***	0.118***	0.087***	0.059***	0.065***
	if $Q_{ij} = 0$ ($\hat{\delta}_0$).	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	$\hat{\lambda}$	0.50	0.50	0.15	0.20	0.50	0.50
	p-value $\mathcal{H}_{NV}$.	0.389	0.835	0.002	0.159	< 0.001	< 0.001
	Avg treated outcome.	1054.5	1054.5	2953.7	2953.7	0.457	0.457
	Individuals (n).	23029	23029	23029	23029	23029	23029
	Villages (v).	1409	1409	1409	1409	1409	1409
	Survey waves (T).	3	3	3	3	3	3

Notes as in Table 3.

Table 19: Occupational Choice, Bernoulli model.

		(1)	*(2)*	*(3)*
	Outcome	Self hours.	Wage hours.	Self emp. only.
	Method	Network.	Network.	Network.
Not function of $\hat{\lambda}$.	Program effect	474.153***	−112.859***	0.114***
	after 2 years ($\hat{\beta}_{11}$).	(14.55)	(8.34)	(0.01)
	Program effect	464.304***	−143.481***	0.121***
	after 4 years ($\hat{\beta}_{12}$).	(9.50)	(8.47)	(0.01)
	Spillover on T	−26.577***	25.865***	−0.033***
	after 2 years ($\hat{\varphi}_{T,2}$).	(7.92)	(6.55)	(0.01)
	Spillover on T	13.148	22.082***	−0.027***
	after 4 years ($\hat{\varphi}_{T,4}$).	(9.59)	(7.06)	(0.01)
	Spillover on NT	−12.862**	13.714***	−0.018***
	after 2 years ($\hat{\varphi}_{NT,2}$).	(6.56)	(3.77)	(0.00)
	Spillover on NT	6.363	11.708***	−0.015***
	after 4 years ($\hat{\varphi}_{NT,4}$).	(4.58)	(1.97)	(0.00)
Function of $\hat{\lambda}$.	Link to T	−27.891***	13.355***	−0.050***
	after 2 years ($\hat{\beta}_{21}$).	(1.38)	(2.50)	(0.01)
	Link to T	−12.862***	13.758***	−0.045***
	after 4 years ($\hat{\beta}_{22}$).	(1.63)	(1.59)	(0.01)
	Link probability	0.492***	0.380***	0.120***
	($\hat{\delta}_1$).	(0.03)	(0.03)	(0.00)
	$\hat{\lambda}$	0.05	0.05	0.15
	Avg treated outcome.	421.8	646.7	646.7
	Individuals (n).	23029	23029	23029
	Villages (v).	1409	1409	1409
	Survey waves (T).	3	3	3

Notes as in Table 3.