Estimation of Network Effects without Network Data.

Pedro CL Souza.

November 2014.

Job Market Paper.

Abstract

Social and economic networks have far-reaching implications for the understanding of individual decisions, choices, and their economic outcomes. Yet, prevalence of networks is not usually matched with availability of network data, which often lay unobserved or are imperfectly observed, for example when data is self-reported. In this paper, I present a novel estimation method for network effects in absence of network data, backing them out only from covariates and outcomes. More specifically, I estimate the network spillovers and the parameters that capture exogenous, endogenous, and correlated effects, in addition to those that underpin network formation, such as density or reciprocity. When network data is imperfectly observed, I propose a test for whether it is the actual conveyor of spillovers. In order to achieve this goal, I couple a model of network formation with a spatial econometric model with unobserved and stochastic networks, allowing me to integrate networks away in a setting where observation of many groups is available in one period of time, such as classrooms in a school or households in a village. In this way, I achieve a maximum likelihood estimator unconditioned on network observation. In spirit, I deal with networks as a source of unobserved heterogeneity.

An application is provided for a randomized controlled trial successful in changing the occupational choice of poor villagers in Bangladesh, where treatment consisted on the provision of livestock and training to low-income households (Bandiera et al. (2013)). Without making use of network data, I show that, while treated individuals specialized in livestock rearing, connected peers followed an opposite pattern, and both groups increased food consumption two and four years after the programme implementation. These results are largely consistent with the interpretation that treated gained local comparative advantages in livestock rearing, changing also the economic lives of their peers.

Keywords: networks, spillovers, spatial econometrics.

*PhD Candidate, London School of Economics.
1 Introduction.

Personal interconnectedness is an ubiquitous presence in people’s lives. Through social and economic networks, individuals learn (Manski (1993), Angrist and Lang (2004), Ammermuller and Pischke (2009)), take decisions (Foster and Rosenzweig (1995)) and insure themselves (Bramoullé and Kranton (2007)), to cite only few examples. Understanding individuals as social entities, rather than acting in isolation, has deep consequences for the mechanisms that determine economic activity. For instance, to the extent that one own’s decisions influences others, social networks leverage the effect of covariates on outcomes, often to a very relevant degree (Miguel and Kremer (2004)). Accounting for these social mechanisms, however, remains a challenge in most cases: prevalence of social and economic networks is not usually matched with availability of network data. Networks are often imperfectly observed – for example, when data is self-reported –, or altogether unobserved by the researcher.

The main contribution of this paper is to offer an estimation method for network effects when network data is unobserved or imperfectly observed, opening up applications for datasets that were previously unsuitable for such purpose. More specifically, I propose an estimator that accomplishes three goals. Without network data, I first estimate network spillovers, which capture the amplification of the exogenous variation as, owing to mutual dependence, social networks leverage the effect of covariates on outcomes. It is also the difference between expected outcomes when networks are relevant and entirely irrelevant.

Second, also in absence of network data, I estimate the network structure, separately identifying Manski (1993)’s endogenous, exogenous and correlated effects and, in addition, parameters that capture the characteristics of the underlying networks, such as its density, reciprocity and polarity. Third, when networks are available, I provide a test for whether reported connections are actual conveyors of interaction capable of influencing economic outcomes of others.

In order to accomplish these goals, I propose a spatial econometric model with unobserved and stochastic networks, coupled with a model for random network formation where the probability of link formation may depend on exogenous variables, such as race, gender, household distances or network reports. In a setting where there are many groups in one period of time, I integrate networks away, dealing with networks as a source of unobserved heterogeneity. I then derive a maximum likelihood estimator unconditioned on network observation; in this way, I’m able to estimate effects in absence of networks, along parameters that underpin network formation. If panel data is available, I also allow for time and fixed effects if networks are invariant over time.

Under minimal assumptions, I show the estimator is consistent for a set of the structural parameters and that network spillovers are constant within the identified set – and, hence, point-identified. Confidence regions are provided within the partial identification framework. In order to establish point identification for the structural parameters, I use the fact that the presence of social interactions implies dispersion on the outcome variable which cannot be explained by covariates or peer group heterogeneity alone. Such "excess" variance in the outcome allows me to overcome the difficulty of separately identifying few strong links from a large number of weak links, along a specific path that yields constant spillover. I explore a moment condition based on outcome dispersion and solve a Generalized Method of Moments within the identified set. I also provide a test

---

1 In a setting where networks are unaccounted for but prevalent, I will show that OLS estimates are often inconsistent for individual elasticities and the size of inconsistency depends on the unobserved network.

2 Chernozhukov et al. (2007), Bugni (2010) and Romano and Shaikh (2010).

3 Graham (2008) uses a similar idea in the context of linear-in-means model, when networks are known.
for the null hypothesis that network link reports, when available, do not actually increase the probability of link formation. A rejection of the null demonstrates network data validity.

I apply the methods in this paper to investigate treatment effects in a setting conducive to spillovers. The programme of Bandiera et al. (2013) assessed whether lack of capital and skills determined the occupational choice of the poor by providing livestock and training to low-income households in Bangladesh. Due to the structure of the randomization, a large proportion of individuals in selected villagers was treated, raising the possibility that network spillovers are of considerable importance in determining outcomes. Without making use of network data, I show that, while treatment increased significantly the amount of hours dedicated to self-employment, decreased wage hours, and increased livestock value beyond treatment two and four years after implementation, connected peers followed the opposite pattern, and both groups increased food consumption and food security. This result is consistent with the takeover of vacancies left by those who started their businesses, and suggests a specialization at the village level where treated individuals gained comparative advantage in rearing, as both groups shared better off after treatment. Estimation of the network structure also shows that, in the majority of cases, densities are fairly low, suggesting local interactions via personal contacts, as opposed to interaction via village-level markets which are characterized by high-density networks. I also show that network spillovers were significant in the determination of some outcomes, particularly on food expenditures and security, amounting to around half of original treatment. These predictions are then confirmed with direct use of network data.

The methods developed in the paper are a major contribution to the spatial econometrics literature which, so far, considered estimation when networks are observed, non-stochastic, and without measurement errors. The role of randomness in network formation has received scant attention in spatial models, while an important attribute in social networks (Diestel (2010)). In this way, existing methods require the knowledge of true networks are are conditioned by its choice, which had been regarded as a limitation in the literature (Anselin (2010), Plümper and Neumayer (2010)). Representative papers are Anselin (1988), Kelejian and Prucha (1998, 1999, 2001, 2010) and especially Lee (2004, 2007) and Lee et al. (2010) who also consider a maximum likelihood estimator. Exceptions are Lam and Souza (2013a,b, 2014) and Manresa (2013) who show properties of an Adaptive Lasso estimator for the collection of pairwise links in a network, under the assumption that one group is observed for many periods of time and, consequently, clearly suit different applications. Identification results of Manski (1993), Graham (2008), Bramoullé et al. (2009) and De Giorgi et al. (2010), likewise, are derived under the assumption of network observation. In another strand of the literature, stochastic network formation models, such as Holland and Leinhardt (1981), Frank and Strauss (1986) and Strauss and Ikeda (1990), also consider estimation of network structure only when network observations are available.

Beyond the contribution to the spatial econometric theory, this paper provides a method for a systematic investigation of network effects, finding application in many fields. Examples may include peer effects in education (Sacerdote (2001), Bramoullé et al. (2009), De Giorgi et al. (2010), Fruehwirth (2013), Brakefield et al. (2014)),

---

4 Exceptions are self-employment and wage hours.
5 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 own' endogenous variable, such as Kelejian and Prucha (1998, 1999); Bramoullé et al. (2009); De Giorgi et al. (2010). To the extent that network data suffers from measurement errors, one risks violating relevance or validity assumptions without awareness.
information diffusion and technology adoption (Bandiera and Rasul (2006), Conley and Udry (2010)), social networks and labour outcomes (Rees (1966), Granovetter (1973), Conley and Topa (2002), Munshi (2003), Pellizzari (2004), Calvó-Armengol and Jackson (2004)) and crime and delinquent behaviour (Glaeser et al. (1996), Alexander et al. (2001), Cleveland and Wiebe (2003), Haynie and Osgood (2005), Dell (2012)). It is particularly suitable when eliciting network of acquaintances is difficult, time consuming or expensive. This is often the case when social networks are concerned, where link reports are often of subjective nature and prone to behavioural biases.

I proceed as follows. In section 2, I introduce the model, define network spillovers and show inconsistency that arises when networks are unaccounted for. In section 3, I present the estimator for network effects in the absence of networks and explore its asymptotic properties. Section 4 provides a simulation to confirm the performance of the estimator in small samples. Section 5 contains the application for the context of treatment spillovers. Finally, section 6 concludes.

2 Model.

The model considered in this paper is composed of two parts. I first develop a model for stochastic network generation\(^6\) allowing network links to depend on commonality of an individual characteristic, such as sharing a race or gender. Given a network, a spatial econometric model connects outcomes and covariates, model extensively considered in the literature, such as Anselin (1988), Lee (2004), Bramoullé et al. (2009), Lee et al. (2010) and De Giorgi et al. (2010). In contrast to previously-mentioned papers, I consider the estimation of network effects in the absence of network data.

I assume there are groups \(j = 1, \ldots, v\) and individuals only interact within groups. Group boundaries are observed, but networks within groups are not. An example is a randomized control trial at the household level conducted in many locations: geographical information is available and, considering households as the individual unit of analysis, the researcher may be interested in treatment spillovers through the household-level network, absent of information on the pattern of interaction between households in any given village.

Groups have individuals \(i = 1, \ldots, n_j\) and sizes are allowed to vary across \(j\). For each group \(j\), a network is described with a directed graph \(G_j\), a unordered collection of ordered pairs of individuals among \(n_j\) individuals. This set lists links along with the associated direction: \(\{i, k\} \in G_j\) implies individual \(i\) affects individual \(k\) in group \(j\). 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 very general. For example, in Figure 1 portrays estimated links between United States senators from Lam and Souza (2013b) based on their 2013 voting record. It is also convenient to express the graph with a so-called neighboring or spatial matrix \(W_j\), of \(n_j \times n_j\) dimension, 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 and so \(\{W_j\}_{ii} = 0\), for all \(i \in \{1, \ldots, n_j\}\).\(^7\)

Network formation is stochastic with a certain probability law, indexed by parameters of interest \(\theta_g\), described

---

\(^6\)Generalizations of Exponential Random Markovian Graphs model category, introduced originally by Holland and Leinhardt (1981) and expanded by Frank and Strauss (1986) and Strauss and Ikeda (1990)

\(^7\)\(G_j\) 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 \(W_j \equiv W_{n_j,j}\).
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 outcomes to covariates. 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 own’s endogenous variable on others’ outcome variables, own’s exogenous variable and others’ exogenous variables. For a particular group $j = 1, \ldots, v$ composed of $n_j$ individuals, the model is given by

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

(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 \( \epsilon_j \) is \( n_j \times 1 \), independent and normally distributed with variance \( \sigma_{\epsilon}^2 \). As particular example, this includes group-level clustering and heteroskedasticity that arises from heterogenous exposure to others’ disturbances.

In Manski (1993)’s taxonomy, term \( W_j^0 y_j \) corresponds to the *endogenous effects*, or the dependence of own’s behavior on others, through link strength scalar parameter \( \lambda_0 \). Parameter \( \beta_1 \), of dimension \( k \times 1 \), captures the direct effect of own’s exogenous variables on own’s dependent variables. Parameter \( \beta_2 \), of same dimension, describes the effects of other’s exogenous variables into own’s dependent variable. Then \( W_j^0 x_j \) is denoted *contextual effects*. Correlated effects are represented by the error \( v_j = \rho_0 M_j^0 v_j + \epsilon_j \) and fixed effects, which I will shortly allow. This model is similar to the one considered, for example, in Bramoullé et al. (2009) and Lee et al. (2010) and is known as "mixed regressive-spatial autoregressive model" in the spatial econometric literature (Anselin (1988)). Interest resides on estimation of usual spatial parameters \( \theta_s = (\lambda_0, \beta_{10}, \beta_{20}, \rho_0, \sigma_0^2, \gamma_0)’ \) and \( \theta_g \). Hence, the complete set of structural parameters of interest is \( \theta = (\theta_s, \theta_g)’ \).

Dependence of own’s outcomes on other’s outcomes and exogenous variable often means overall response to exogenous variation exceeds \( \beta_{10} \). As a consequence, to the extent that individual network spillovers depend on own exogenous variation, estimators for \( \beta_{10} \) that do not account for networks are often inconsistent.

Observation of the reduced-form model sheds light on this issue. Using the series decomposition\(^8\) \( (I_{n_j} - \lambda_0 W_j^0)^{-1} = \sum_{s=0}^{\infty} \lambda_0^s (W_j^0)^s \), expected outcomes are separated in two components: the individual reaction or elasticity with respect to \( x_j \) and its infinite reverberations through the network:

\[
\mathbb{E} y_j = x_j \beta_{10} + W_j^0 x_j \beta_{20} + \sum_{s=1}^{\infty} (\lambda_0 W_j^0)^s (x_j \beta_{10} + W_j^0 x_j \beta_{20}) .
\]

Term \( x_j \beta_{10} \) is understood as the individual-level elasticity with respect to \( x_j \) if networks were absent, while the second and third term jointly denote network spillovers, the additional effect on the mean exclusively due to individual interconnectedness,

\[
\phi(x_j, \theta_0) = W_j^0 x_j \beta_{20} + \sum_{s=1}^{\infty} (\lambda_0 W_j^0)^s (x_j \beta_{10} + W_j^0 x_j \beta_{20}) = \sum_{s=1}^{\infty} \lambda_0^{s-1} (W_j^0)^s x_j (\lambda_0 \beta_{10} + \beta_{20}) .
\]

Clearly, if \( \lambda_0 = 0 \) and \( \beta_{20} = 0_{k \times 1} \), or \( \delta_1 = \delta_0 = 0 \), then \( \phi(x_j, \theta_0) = 0 \). Spillover \( \phi(x_j, \theta_0) \) is a \( n_j \times 1 \) vector, as each individual accrues their own spillover.

Separate identification of the individual reaction and network spillovers is relevant in at least two scenarios. Provided that the ultimate goal is the consistent estimation of \( \beta_{10} \), \( \phi(x_j, \theta_0) \) presents itself as a confounding factor. As shown in subsection 2.1, when networks are unaccounted for, consistent estimation of \( \beta_{10} \) requires that a underlying network structure is such that one’s own network spillovers are independent of one’s own exogenous variation, condition that breaks down in simple counterexamples with limited interaction pools.

Moreover, network spillovers are of interest in their own right. Examples abound in the literature. Glaeser et al. (1996) argue that social interactions explain criminal behavior, particularly in petty crimes, but also of moderate importance in more serious offenses. Hence, crime prevention policies have indirect effect through

---

\(^8\)Existence is shown in section 3.
reduction of others’ proclivity towards criminal activity, and the effect’s magnitude then shapes and informs public policy debate. In another example, Foster and Rosenzweig (1995) reason that farmers’ decision of adopting high-yielding seed varieties (HYVs) depends on other farmers’ adoption decision and their accrued profit; consequently, a single farmer’s adoption decision multiplies itself by inducing others towards similar behavior. Finally, note that parameter \( \varphi(x_j, \theta) \) can be explored in order 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 be maximized by concentrating treatment in fewer groups.

**Remark 1.** Panel or spatiotemporal models can be naturally introduced from equation (2). Index outcomes and covariates by time \( t = 1, \ldots, T \) and the complete model reads

\[
y_{jt} = \lambda_0 W_j^0 y_{jt} + x_{jt} \beta_{10} + W_j^0 x_{jt} \beta_{20} + \alpha_j + \gamma_t + v_{jt}
\]

where \( \alpha_j \) is a vector of \( n_j \times 1 \) time-invariant coefficients (but allowed to vary at group or individual levels), also denoted, after Manski (1993), correlated effects. The vector \( \gamma_t \) are time effects. Under invariance of networks with respect to time, I propose a data transformation that eliminates these nuisance parameters in subsection 3.4. Arguably, model (5) can be seen as network differences-in-differences estimator. In the absence of network effects (\( \lambda_0 = 0 \) and \( \beta_{20} = 0_{k \times 1} \)), let \( x_{jt} \) denote the vector of individual treatment status and the model then reduces to the standard differences-in-differences. In this context, terms \( \lambda_0 W_j^0 \) and \( W_j^0 x_{jt} \beta_{20} \) would measure the spillovers through the network.

### 2.1 Inconsistency when Networks are Unaccounted For.

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

\[
y_j = x_j \beta_{10} + \varphi(x_j, \theta_0) + \epsilon_j.
\]

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

Consistency for \( \beta_{10} \) requires that \( \mathbb{E}(\varphi(x_j, \theta_0)|x_j) = 0 \) for all \( i = 1, \ldots, n_j \), case in which the researcher would be oblivious to network spillovers. In the other extreme, perfect correlation suggests OLS estimates are consistent for the sum of \( \beta_{10} \) and full spillovers. In general, however, independence is not generally attained, failing particularly under reciprocated networks or correlation between \( x_{ij} \) and \( x_{kj} \) for \( i \neq k \). 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, and so are also unknown the size and presence of the bias. I now give some examples.

**Example 1.** *(Classrooms and linear-in-means model).* Manski (1993) proposes the linear-in-means network
network spillovers, violating independence condition indefinitely. This is one example where treatment is clearly correlated with social networks, and therefore with the treatment. These new individuals, in turn, also refer their peers for treatment and process repeats itself.

Example 2. (Households and local interaction). Households typically interact with few others and relations are reciprocated. For the sake of example, suppose network is composed of isolated subgroups of five households, where interaction across subgroups is negligible in comparison to interactions within. In this setting, $W^0_j$ is a block-diagonal matrix with $\frac{n}{5}$ blocks, or $W^0_j = I_{\frac{n}{5}} \otimes (\frac{1}{4} I_5' \otimes \frac{1}{4} I_5)$. Suppose a proportion $\alpha$ 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 $\beta_{10}$, which can be seen by replacing $n = 5$ in equation (7). As a consequence, OLS estimates are biased for $\beta_{10}$ and capture the portion of one own’s spillovers that correlate with one own’s treatment status.

Example 3. (Snowballing a treatment). Snowballing is a common sampling technique to target hidden or difficult populations to reach, such as drug users. It takes advantage of social network of contacts of a study’s subject. One starts with a small cohort – potentially, a single individual –, who then refers his or her peers for treatment. These new individuals, in turn, also refer their peers for treatment and process repeats itself indefinitely. This is one example where treatment is clearly correlated with social networks, and therefore with network spillovers, violating independence condition $\mathbb{E}(\varphi(x_j, \theta_0) | x_j) = 0$. This violation would occur also in the

$$W^0_j = \begin{bmatrix}
0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\
\frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0
\end{bmatrix} = \frac{1}{n-1} \mathbf{t}_n \mathbf{t}'_n - \frac{1}{n-1} \mathbf{I}_n$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{t}_n$ is the $n \times 1$ vector of ones. Suppose $x_j$ is a treatment dummy and $\alpha$ proportion of individuals in the group were treated. The expectation of response conditional on treatment is obtained via the reduced-form model

$$y_j = (S^0_j)^{-1} x_j \beta_1 + (S^0_j)^{-1} W^0_j x_j \beta_2 + (S^0_j)^{-1} (R^0_j)^{-1} \epsilon_j$$

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

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

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

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

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

\(\square\)

---

For simplicity, assume $n$ is a multiple of 5.
case where, for example, drug users are more likely to have other drug users as acquaintances and more likely to receive a treatment, irrespective of whether snowballing is employed or not. In either case, treatment targeting is confounded with spillovers. Therefore, as in the example above, OLS estimates are not consistent for $\beta_{10}$.  

Generally, OLS is only consistent for $\beta_{10}$ in particular network structures. When they lay unobserved, implementation of such strategy depends on hypotheses that rule out feedback mechanisms. In section 3, I provide a method for consistent estimation of $\varphi(x_j, \theta)$ under few identifying assumptions, addressing both motivating elements. The method is based on a maximum likelihood integrated with respect to unobserved networks, resulting in a likelihood independent of network observation. In spirit, I deal with the networks as unobserved heterogeneity. As it will be shown, although point identification of $\theta$ is not obtained, spillover $\varphi(x, \theta)$ is constant within the identified set and hence point-identified. Section (3.3) uses additional identifying information to sort through the identified set and reestablish point identification for the structural parameters.

3 Estimation of Network Effects.

Spatial econometric models dealt with the case of known $W_0$ and $M_0$. Lee (2004) and Lee et al. (2010), under certain conditions, show consistency and asymptotic normality of a quasi-maximum likelihood estimator for $\theta_s$ but are of no use if $W_0$ and $M_0$ are unobserved or observed with measurement errors\(^{10}\). If consistent estimation of $\theta_s$ is guaranteed and networks are known, accounting for network spillovers would pose no challenge.

In contrast, I deal with networks as a form of unobserved heterogeneity – and integrates it away. This is possible in the context where networks are randomly formed with certain probability law and observation of many groups is available during one period of time.

More formally, I propose an integrated likelihood approach: the likelihood evaluated at a parameter set $\theta \in \Theta$ given non-network data is the integral of the likelihood given a network and non-network data (a spatial model) with respect to the probability density function for a stochastic network model:

$$\ln L(\theta | y, x, Q) = \int \ln L(\theta | y, x, W, M) dP(W, M | Q, x, \theta)$$

(8)

where $W$ and $M$ are a random block matrix of dimension $n \times n$, $n = \sum_{j=1}^{v} n_j$, with blocks of $n_j \times n_j$ dimension along the main diagonal. The conditioning of $P$ on $\theta$ and $x$ appear as $\ln L$ is in fact a function of $y$ given $x$, $W$, $M$ and $\theta$.

Given 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

$$\ln L(\theta | y, x, Q) = \sum_{s=1}^{g_{nv}} \ln L(\theta | y, x, W^s) P(W^s | Q, x, \theta).$$

(9)

Even for relatively small numbers of $n_j$ and $v$, $g_{nv}$ is an enormous number. For example, taking $v = 5$ and $n_j = 10$ for $j = 1, \ldots, v$, the total of number of graphs $g_{nv}$ exceeds $10^{135}$.

\(^{10}\)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).
In order to make the estimator computationally feasible, I propose substituting \( W_0 \) and \( M_0 \) for their expected values\(^{11} \)
\[
W^e (Q, \theta) = \int W dP \ (W \mid Q, x, \theta) \quad \text{and} \quad M^e (Q, \theta) = \int M dP \ (M \mid Q, x, \theta)
\]
in model (2). The expected networks are dependent on exogenous characteristics \( Q \) and parameters \( \theta \). Estimation of network spillovers and structural parameters is based on the likelihood of the inherently misspecified estimated model
\[
y_j = \lambda W^e_j (Q_j, \theta) y_j + x_j \beta_1 + W^e_j (Q_j, \theta) x_j \beta_2 + v^e_j
\]
(10)
with \( v^e_j (Q, \theta) = \rho M^e_j (Q_j, \theta) v_j + \epsilon_j \). Intuitively, the estimated model is equivalent to the model if networks were observed in addition to mispecification terms that are closer\(^{12} \) to zero when \( \theta = \theta_0 \),
\[
y_j = \lambda W^0_j y_j + x_j \beta_{10} + W^0_j x_j \beta_{20} + \lambda \{ W^e_j (Q_j, \theta) - W^0_j \} y_j + \{ W^e_j (Q_j, \theta) - W^0_j \} x_j \beta_{20} + v^e_j.
\]
As a result of the introduction of expected networks, pointwise identification of parameters \( \theta \) is generally not obtained: there are multiple combinations of \( \lambda \), \( \theta \) and \( \beta_2 \) such that the model is observationally equivalent, non-identification that is consistent with the difficulty of separating a large number of weak connections from a small number of strong connections along a specific path. Yet, I will show parameters that belong to the identified set are such they yield network spillovers evaluated at the true parameter \( \varphi(x_j, \theta_0) \), and so this latter quantity is point-identified. I adapt the ideas in Chernozhukov et al. (2007), Romano and Shaikh (2010) and Bugni (2010) to provide inference. Still, point identification is derived in the particular case where \( \lambda_0 \) is known, a result that lays the basis for future results. Formal presentation of the likelihood is in order.

Define \( S^e_j (Q, \theta) \equiv I - \lambda W^e_j (Q_j, \theta) \), \( S^0_j (\lambda) \equiv I - \lambda W^0_j \), \( S^0_j (\lambda_0) \), \( R^e_j (\theta) \equiv I - \rho M^e_j (Q_j, \theta) \), \( R^0_j (\rho) \equiv I - \rho M^0_j \), \( R^0_j (\rho_0) \), \( Z^e_j (Q_j, \theta_c) = (x_j, W^e_j (Q_j, \theta_c) x_j) \) and the block matrices \( W^0_n (Q_n, \theta_c) = \text{diag}(W^0_1 (Q_1, \theta_c), \ldots, W^0_\nu (Q_\nu, \theta_c)) \), \( W^e_n (Q_n, \theta_c) = \text{diag}(W^e_1 (Q_1, \theta_c), \ldots, W^e_\nu (Q_\nu, \theta_c)) \), \( M^e_n (Q_n, \theta_c) = \text{diag}(M^e_1 (Q_1, \theta_c), \ldots, M^e_\nu (Q_\nu, \theta_c)) \), \( S^e_n (Q_n, \theta_c) = \text{diag}(S^e_1 (Q_1, \theta_c), \ldots, S^e_\nu (Q_\nu, \theta_c)) \), and \( Z^e_n (Q_n, \theta_c) = (Z^e_1 (Q_1, \theta_c), \ldots, Z^e_\nu (Q_\nu, \theta_c)) \).

In order to write the model more compactly, define \( y_n = (y_1', \ldots, y_\nu')', \ x_n = (x_1', \ldots, x_\nu')', \) and model (2) can be denoted \( y_n = \lambda_0 W^0_n y_n + x_n \beta_{10} + W^0_n x_n \beta_{20} + v_n \), where \( v_n = (v_1', \ldots, v_\nu')' \). The likelihood is
\[
\ln \mathcal{L}^e (\theta \mid y, x, Q) = -\frac{n}{2} \ln (2\pi \sigma^2) + \ln |S^e_n (Q_n, \theta)| + \ln |R^e_n (Q_n, \theta)| - \frac{1}{2\sigma^2} \epsilon^e_n (Q_n, \theta) \epsilon^e_n (Q_n, \theta)
\]
(11)
with \( \epsilon^e_n (Q_n, \theta) = R^e_n (Q_n, \theta) (S^e_n (Q_n, \theta) y_n - Z^e_n (Q_n, \theta) \beta) \) for \( \beta = (\beta_1', \beta_2')' \). Parameters \( \beta \) and \( \sigma^2 \) are concentrated out of the likelihood, simplifying derivations and implementation. Denote \( \theta_c = \theta \setminus \{ \beta, \sigma^2 \} \) the non-concentrated parameters. At each \( \theta_c \), the closed-form solutions for the concentrators are
\[
\hat{\beta} (Q, \theta_c) = (Z^e_n (Q_n, \theta_c) R^e_n (Q_n, \theta_c) R^e_n (Q_n, \theta_c) Z^e_n (Q_n, \theta_c))^{-1} Z^e_n (Q_n, \theta_c) R^e_n (Q_n, \theta_c) R^e_n (Q_n, \theta_c) S^e_n (Q_n, \theta_c) y_n
\]
\[
\hat{\theta}_c = \text{arg} \min_{\theta_c} \sum_{n=1}^{N} \ln |S^e_n (Q_n, \theta_c)| + \ln |R^e_n (Q_n, \theta_c)| - \frac{1}{2\sigma^2} \epsilon^e_n (Q_n, \theta_c) \epsilon^e_n (Q_n, \theta_c)
\]
\[11\text{For simplicity of explanation, momentarily assuming } \hat{W} \text{ and } \hat{M} \text{ are independent, which does not hold for the rest of the paper. Fixed effects are reintroduced in subsection 3.4.}
\[12\text{Comparison between likelihood computed with expected network and true networks can be found in tables 7 and 8 in the appendix.}
and
\[
\hat{\sigma}^2(Q, \theta_c) = \frac{1}{n} \left( S_n^e (Q_n, \theta_c) y_n - Z_n^e (Q_n, \theta_c) \hat{\beta} (\theta_c) \right)' R_n^e (Q_n, \theta_c) R_n^e (Q_n, \theta_c) (S_n^e (Q_n, \theta_c) y_n - Z_n^e (Q_n, \theta_c) \hat{\beta} (\theta_c)) \\
= \frac{1}{n} y_n' S_n^e (Q_n, \theta_c) R_n^e (Q_n, \theta_c) P_n^e (Q_n, \theta_c) R_n^e (Q_n, \theta_c) S_n^e (Q_n, \theta_c) y_n
\]

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 likelihood brought to maximization is

\[
\ln L_n^c (\theta_c | y_n, x_n, Q_n) = -\frac{n}{2} (\ln (2\pi) + 1) - \frac{n}{2} \ln \hat{\sigma}^2 (Q_n, \theta_c) + |S_n^e (Q_n, \theta_c)| + |R^e (Q_n, \theta_c)|. \tag{12}
\]

The final estimator is \( \hat{\theta} = (\hat{\theta}_c^0, \hat{\beta} (\theta_c^0)', \hat{\sigma}^2(\theta_c^0)' \), where \( \hat{\theta}_c \equiv \arg \max_{\theta_c \in \Theta_c} \ln L_n^c (\theta_c | y_n, x_n, Q_n) \). I now lay formal hypothesis to guarantee asymptotic properties of the estimator.

### 3.1 Pointwise identification of \( \theta \) when \( \lambda_0 \) is known.

In this subsection, I present the basic assumptions for consistent estimation and pointwise identification of the parameters in the model. Assumption 6, required for pointwise identification of \( \theta \), holds as long as \( \lambda_0 \) is known to the researcher\(^{13} \). The next section relaxes this requirement as, in most circumstances, information on \( \lambda_0 \) is unavailable. I show in subsection 3.2 that in such case the parameters are identified within a set in which network spillover parameter is constant at \( \varphi(x_n, \theta_0) \). Assumptions 1-5 are maintained throughout.

The first assumption lays the true model, properties of the neighboring matrices 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 himself. The independence of \( P \) with respect to \( \beta \) and \( \sigma^2 \) allows concentrating these parameters, as described previously, and is taken for simplicity only and 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 \( \epsilon_j \sim N (0, \sigma^2 I) \). Let \( \text{mat}_{n_j} (\{0, 1\}) \) 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 \( \sigma \)-algebra of subsets of \( \Omega \) and \( P \) as probability measure. \( \{W_j^0, M_j^0\} \) is particular realization from a random matrix\(^{14} \), a measurable map from \( (\Omega, \mathcal{F}) \) to \( \text{mat}_{n_j} (\{0, 1\}) \), with probability distribution function \( P (W_j, M_j | \theta, x_j) \) with common functional form across groups. \( P \) does not depend on \( \beta \) or \( \sigma^2 \).

In some applications, it is customary to conduct a row-sum normalization of \( W_j \), an operation consisting of replacing \( W_j \) by a \( W_j^* \) with \( W_{jik} = W_{jik} \sum_{s=1}^{n_j} W_{js} \) (Anselin (1988), Kelejian and Prucha (1998, 1999, 1999)).

\(^{13}\)In fact, assumption 6 holds in the case where one parameter among \( \lambda_0 \), \( \beta_{20} \) and \( \theta_0^0 \) is known. For simplicity, I arbitrarily focus the argument on \( \lambda_0 \).

\(^{14}\)In fact, \( \{W_j^0, M_j^0\} \) are arrays and full notation should include respective dimensions, \( \{W_{n_j,j}, M_{n_j,j}\} \). This is supressed for simplicity.
dition with proposition 1. The benchmark hypothesis reads, \( \tilde{\beta}_0 \) sufficiently conditions for assumption 6 under knowledge of Assumption 5.

Assumption 4. Feasible scenarios are a misspecification term goes to zero asymptotically and variance terms are consistently estimated in the limit.

Assumption 2. The sequence of \( n_j \)-by-\( n_j \) matrices \( \{\lambda_0 W_j^0\} \), \( \{(S_j^0)^{-1}\} \), \( \{\lambda W_j^e (Q, \theta)\} \), and \( \{(S_j^e(Q, \theta))^{-1}\} \) are uniformly bounded.

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 \( S_j^0 \) 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. Feasible scenarios are \( v \to \infty \) for fixed \( n \) and \( n \to \infty \) for fixed \( v \).

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 \( \Theta_c \).

Assumption 5. The parameter set \( \Theta_c \) is compact and the true parameter \( \theta_0 \) \( \in \Theta_c \).

Next, I lay out the identification condition required for separate point identification of parameters, followed by the provision of easy-to-interpret sufficient conditions for assumption 6 under knowledge of \( \lambda_0 \), as demonstrated in proposition 1. The benchmark hypothesis reads,

Assumption 6. (Identification). Define

\[
\gamma (Q_n, \theta_c) = \frac{1}{n} \mathbb{E}\{ \beta'_0 Z_n^0 (S_n^0)^{-1} \tilde{P}_n^c (Q_n, \theta_c) (S_n^0)^{-1} Z_n^0 \beta_0 \}
\]

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

The reduced-form evaluated at the true vector of parameter \( \theta_0 \) is

\[
y = (S_n^e (Q_n, \theta_0))^{-1} Z_n^e (Q_n, \theta_0) \beta_0 + (S_n^e (Q_n, \theta_0))^{-1} (R_n^e (Q_n, \theta_0))^{-1} \varepsilon_n.
\]
As \((S_n^e(Q_n, \theta_0^e))^{-1} = I_n + \lambda_0 G_n^e(Q_n, \theta_0^e)\), where \(G_n^e(Q_n, \theta_0^e) \equiv W_n^e(Q_n, \theta_0^e) (S_n^e(Q_n, \theta_0^e))^{-1}\), the expression above can be written as

\[
y_n = Z_n^e(Q_n, \theta_0^e) \beta_0 + \lambda_0 G_n^e(Q_n, \theta_0^e) Z_n^e(Q_n, \theta_0^e) \beta_0 + (S_n^e(Q_n, \theta_0^e))^{-1} (R_n^e(Q_n, \theta_0^e))^{-1} \epsilon_n. \tag{14}
\]

For separate identification of \(\lambda_0\) and \(\beta_0 = (\beta_{10}', \beta_{20}', \cdots)\), it is then necessary to guarantee matrices \(Z_n^e(Q_n, \theta_0^e)\) and \(G_n^e(Q_n, \theta_0^e) Z_n^e(Q_n, \theta_0^e) \beta_0 = W_n^e(Q_n, \theta_0^e) (S_n^e(Q_n, \theta_0^e))^{-1} Z_n^e(Q_n, \theta_0^e) \beta_0\) are not dependent asymptotically. In turn, asymptotic independence of the concerned matrices is a necessary and sufficient condition for assumption 6. Following Lemma 3, \(\gamma(Q_n, \theta_c)\) is well approximated by \(\gamma^e(Q_n, \theta_c)\), where

\[
\gamma^e(Q_n, \theta_c) = \frac{1}{n} \mathbb{E}(\beta_0' Z_n^e(Q_n, \theta_0^e) (S_n^e(Q_n, \theta_0^e))^{-1} \tilde{P}_n^e(Q_n, \theta_c) (S_n^e(Q_n, \theta_0^e))^{-1} Z_n^e(Q_n, \theta_0^e) \beta_0).
\]

Given that \(\tilde{P}_n^e(Q_n, \theta_c) = S_n^e(Q_n, \theta_c) R_n^e(Q_n, \theta_c) S_n^e(Q_n, \theta_c)\) is positive definite, then \(\gamma(Q_n, \theta_c) = 0\) if, and only if, \((S_n^e(Q_n, \theta_0^e))^{-1} Z_n^e(Q_n, \theta_0^e) \beta_0 = 0\), which is equivalent to \(Z_n^e(Q_n, \theta_0^e) \beta_0 = 0\) using \((S_n^e(Q_n, \theta_0^e))^{-1} = I_n + \lambda_0 G_n^e(Q_n, \theta_0^e)\) or, essentially, that \(Z_n^e(Q_n, \theta_0^e)\) and \(G_n^e(Q_n, \theta_0^e) Z_n^e(Q_n, \theta_0^e) \beta_0\) are asymptotically independent. The next proposition, which resembles similar results of Bramoulle et al. (2009) and Lee et al. (2010), states the desired conditions.

**Proposition 1.** If \(\lambda_0\) is known, \(\beta_{20} \neq \lambda_0 \beta_{10}\), \(x\), \(W_n^e(Q_n, \theta_0^e) x_n\) and \((W_n^e(Q_n, \theta_0^e))^2 x_n\) are linearly independent, then \(Z_n^e(Q_n, \theta_0^e)\) and \(G_n^e(Q_n, \theta_0^e) Z_n^e(Q_n, \theta_0^e) \beta_0\) 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 \(x_n\), \(W_n^e(Q_n, \theta_0^e) x_n\) and \((W_n^e(Q_n, \theta_0^e))^2 x_n\), as also seen in the subgroup model of Lee (2007). In particular, let the probabilistic model be the pure Bernoulli, where links are formed with probability \(\delta_0\), independent of exogenous characteristic. Then \(W_n^e(Q_j, \theta_0^e) = \delta_0(\lambda_{n_j} I_{n_j} - I_{n_j})\) and \((W_n^e(Q_j, \theta_0^e))^2 = \delta_0^2(n_j - 2)(\lambda_{n_j} I_{n_j} + I_{n_j})\). With at least three distinct values of \(n_j\), independence condition in the previous proposition is guaranteed.

Under the conditions introduced above, I present the basic theorem. Proofs can be found in the appendix.

**Theorem 1.** Under assumptions 1-6, \(\hat{\theta}\) is a consistent estimator for \(\theta_0\), i.e., \(\hat{\theta} \overset{p}{\rightarrow} \theta_0\).

Asymptotic distribution can be obtained from a Taylor expansion around the point \(\frac{\partial \ln \mathcal{L}^e(\theta|y_n,x_n,Q_n)}{\partial \theta} = 0\). For a point \(\tilde{\theta}\) between \(\hat{\theta}\) and \(\theta_0\),

\[
\sqrt{n}(\tilde{\theta} - \theta_0) = \left[ \frac{1}{n} \frac{\partial \ln \mathcal{L}^e(\tilde{\theta}|y_n,x_n,Q_n)}{\partial \theta_0} \right]^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial \theta}. \tag{15}
\]

The variance matrix of the score vector is \(\Sigma_n \equiv \mathbb{E}\left[ \frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial \theta} \right]\). In the limit, \(\hat{\theta} \overset{p}{\rightarrow} \theta_0\), which implies \(\tilde{\theta} \overset{p}{\rightarrow} \theta_0\) and so the Hessian matrix converges to \(\Omega_n = \mathbb{E}\left[ \frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}^e(\theta_0|y_n,x_n,Q_n)}{\partial \theta} \right]\). 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,

\[15\text{That is, if there are three distinct values of } n_j, \text{ the only conformable vectors } c_1, c_2 \text{ and } c_3 \text{ such that } x c_1 + \delta_0(\text{diag}(\lambda_{n_1} I_{n_1}, \ldots, \lambda_{n_j} I_{n_j} - I_{n_j}) )xtc_2 + (\text{diag}(\lambda_{n_1} I_{n_1}, \ldots, (n_j - 2)\lambda_{n_j} I_{n_j} + I_{n_j} ))xtc_3 = 0 \text{ are } c_1 = c_2 = c_3 = 0.\]
Theorem 2. Under assumptions 1-5, \( \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{p} N(0, \Sigma^{-1} \Omega^{-1}) \), where \( \Sigma = \lim_{n \to \infty} \Sigma_n \) and \( \Omega = \lim_{n \to \infty} \Omega_n \).

3.2 Set identification of \( \theta \) when \( \lambda_0 \) is unknown.

There is one simple way asymptotic independence of the matrices is not assured: any path \( \{\lambda_+, \beta_+, \theta_+\} \) such that \( W_n^c(Q_n, \theta^+_c) x_n \beta_2 = W_n^e(Q_n, \theta^+_e) x_n \beta_2 \) and \( \lambda_+ W_n^c(Q_n, \theta^+_c) = \lambda_0 W_n^c(Q_n, \theta^+_c) \) 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 results on estimation and inference on partially identified models of Chernozhukov et al. (2007), Romano and Shaikh (2010) and Bugni (2010) to establish desired results. The point of departure of classic asymptotic analysis is the recognition that the identified set \( \Theta_0 = \{\hat{\theta} \in \Theta : F_n(\hat{\theta}) = F_n(\theta_0)\} \), for \( F_n(\theta) = \mathbb{E} \ln \mathcal{L}_n(\theta) \), and the estimated set \( \hat{\Theta} = \{\hat{\theta} \in \Theta : \ln \mathcal{L}_n(\hat{\theta}) = \inf_{\theta \in \Theta} \ln \mathcal{L}_n(\theta)\} \) 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 parameter set \( \Phi(\theta|y_n, x_n) \subseteq \Theta \) as combinations of parameters yielding equal network spillovers,

\[
\Phi(\theta|y_n, x_n) = \{\theta^+ \in \Theta : \lambda_+ W_n^c(Q_n, \theta^+_c) = \lambda W_n^c(Q_n, \theta_c), W_n^e(Q_n, \theta^+_e) x_n \beta_2^+ = W_n^e(Q_n, \theta_e) x_n \beta_2\}. \tag{16}
\]

The next proposition states that any parameter in the identified set, \( \theta \in \Theta_0 \), generates network spillovers evaluated at the true parameter \( \theta_0 \).

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

The objective then is to produce a sequence of sets such that: (i) in the limit, they are consistent estimates of \( \Theta_0 \), in a sense that the Hausdorff set distance metric\(^{16}\) \( d_h \) 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 \( L_n(\theta|y_n, x_n, Q_n) = -n^{-1}[\ln \mathcal{L}_n(\theta|y_n, x_n, Q_n) - \inf_{\theta \in \Theta} \ln \mathcal{L}_n(\theta|y_n, x_n, Q_n)] \) and \( \hat{\Theta}(c_n) = \{\theta \in \Theta : L_n(\theta|y_n, x_n, Q_n) \leq c_n\} \). The next theorem proves the estimator \( \hat{\Theta} = \hat{\Theta}(0) \) is consistent for \( \Theta_0 \), 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 the alternative estimator \( \hat{\Theta}(c_n) \). For the construction of a set that covers \( \Theta_0 \) with probability \( \alpha \), it is instrumental to select \( c_n = \hat{c}_n(\alpha) \) such that \( \hat{\Theta}(\hat{c}_n(\alpha)) \) possesses the desired property.

\(^{16}\) The Hausdorff set distance metric is defined as \( d_h(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \) with \( d(b, A) = \inf_{a \in A} ||b - a|| \) and \( d_h(A, B) = \infty \) if \( A \) or \( B \) are empty.
Remark also describes a bootstrap algorithm for obtaining consistent estimates of the coverage region with asymptotically controlled error probability. Given Proposition 2, the network spillover is point-identified. (4) Point-identification for \( \beta_{10} \) and \( \sigma_0^2 \) 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: firstly, it allows the provision confidence regions for the network spillovers, i.e., one wish to obtain confidence regions for \( \Phi_0 \), the image of \( \Theta_0 \) under the known function \( \varphi(x, \theta) \) for given \( \theta \in \Theta_0 \), as explored in Remark 2. Secondly, I will show it provides a framework for validation of network data, when it is available, as developed in the remark 3.

Following Romano and Shaikh (2010), in general terms, let \( f \) be a known function with \( f : \Theta \rightarrow \mathcal{Y} \), with \( \mathcal{Y}_0^f \) being the image of \( \Theta_0 \) under \( f \), and also let \( f^{-1}(v) = \{ v \in \mathcal{Y} : f(\theta) = v \} \). This suggests a modification of the inferential test statistic in the following way: note \( v \in \mathcal{Y}_0^f \) if, and only if, there exists some \( \theta \in f^{-1}(v) \) 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 \( \mathcal{Y}_\alpha \) such that coverage probability is \( 1 - \alpha \), i.e., \( \lim_{n \to \infty} P\{ \mathcal{Y}_0 \subseteq \mathcal{Y}_\alpha \} = 1 - \alpha \) and, in analogy to the previous case, this set can be defined by selecting \( c^n(\alpha) \) such that the event \( \{ \mathcal{Y}_0 \subseteq \mathcal{Y}_\alpha \} \) is equivalent to \( \{ \sup_{v \in \mathcal{Y}_0^f} \inf_{\theta \in f^{-1}(v)} L_n(\theta) \leq c^n(\alpha) \} \).

Given the construction above, if the \( \alpha \)-quantiles of the test statistic \( \sup_{v \in \mathcal{Y}_0^f} \inf_{\theta \in f^{-1}(v)} nL_n(\theta) \) were available, coverage region with asymptotically controlled error probability \( \alpha \) would be obtained directly. Next subsection also describes a bootstrap algorithm for obtaining consistent estimates \( \hat{c}^n(\alpha) \) of \( c^n(\alpha) \).

For the moment, as mentioned above, I now describe the two important applications of this procedure for the context of inference on the social multiplier and network characteristics.

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 \( \Theta \) to \( \mathbb{R}^1 \), and given proposition 2 states the network spillovers are constant in the identified set, the image \( \mathcal{Y}_0^c \) 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 \( Q_n \), 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 \( \delta_1 \) is the the estimated probability given observation of link reports, and \( \delta_0 \) otherwise. The null hypothesis of interest is \( H_0 : \delta_0 - \delta_1 = 0 \), with alternative \( H_0 : \delta_0 - \delta_1 \neq 0 \). In the setting above, suffices to take \( \hat{f} : \Theta \rightarrow \mathbb{R}^1 \) as \( \hat{f}(\theta) = \delta_1 - \delta_0 \) and build appropriate confidence intervals.

**3.2.1 Bootstrap for consistent estimation of \( c_n(\alpha) \) and \( c_n^f(\alpha) \).**

In the case of i.i.d. data, Bugni (2010) proposes a bootstrap algorithm correction consistent for \( c_n(\alpha) \) and adaptable to \( c_n^f(\alpha) \). In the current case, spatial dependence prevents immediate application of methods described therein. Instead, I propose bootstrapping at the group-level adaptable to \( c_n(\alpha) \) for the identified set under known function \( f \). 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 \( \Theta_0 \), denoted \( \hat{\Theta}_\alpha^B \) for a bootstrapped sample of arbitrary size \( B \), follow the steps:

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

Step 2. Define the bootstrapped sample \( b = 1, \ldots, B \), sampling \( v \) groups with replacement from the data and denote bootstrapped sample \( \{y_b^n, x_b^n, Q_b^n\} \). Compute

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

Step 3. Let \( \hat{c}_n^B(\alpha) \) be the \( \alpha \) 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

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

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 \( \Theta_0 \) under known function \( f \), \( \hat{\Theta}_\alpha^f \), denoted \( \hat{\Theta}_\alpha^B \), for a bootstrapped sample of arbitrary size \( B \) is:

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

Step 2. Define the bootstrapped sample \( b = 1, \ldots, B \), sampling \( v \) groups with replacement from the data and denote bootstrapped sample \( \{y_b^n, x_b^n, Q_b^n\} \). Compute

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

Step 3. Let \( \hat{c}_n^{f,B}(\alpha) \) be the \( \alpha \) 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 $Y_0$ is

$$\hat{Y}_{f,B}^f = \left\{ v \in Y : \inf_{\theta \in f^{-1}(v)} \sqrt{nL(\theta|y_n, x_n, Q_n)} \leq c_n(1 - \alpha) \right\}.$$ 

### 3.3 Pointwise identification when $\lambda_0$ is unknown using outcome dispersion.

In the previous section, I showed that parameters of interest are identified within a set and network spillovers are constant within the identified set. For some parameters, point identification was not obtained, given the model is observationally equivalent within a combination of the parameters (for example, trading link strength for number of links). A theoretically feasible restriction is when link strength $\lambda_0$ is known: in this case, theorem 1 proves consistency. Nevertheless, $\lambda_0$ is rarely observed and so, in this section, I provide additional restriction that, in certain scenarios, yields a point in the identified set that is consistent with the higher-order observed moments.

Intuition is straightforward: when social interactions are present, variance of observed outcomes cannot be explained only by the variance of exogenous variables or group or individual heterogeneity, since individuals mirror the choice of others. As a result, the variance of observed outcomes across groups is increased, while outcomes within a group are positively correlated. As network formation depends on a model, such as described in section 2, which specifies the probability for link formation, the variance across groups provides a restriction that encompasses link strength, probability of link formation and dependence on exogenous characteristics of the others, closing the identification requirements.

It has been observed elsewhere\(^{17}\) that group outcomes are substantially dispersed across groups, even for groups which are similar along observable characteristics, therefore this phenomena has been denoted *excess variance* (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. To this date, nevertheless, theoretical models considered observed networks and identification of the various modes of social interaction within this class of models (Manski (1993)).

The main idea is that, accounting for variance originating from covariates 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,

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

$$V_{B,j}(y_n) = (\bar{y}_j - \bar{y})^2$$

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, $E[V_{W,j}(y)] = n_j^{-1} \sum_{i=1}^{n_j} [V(y_j)]_{ii}$ and $E[V_{B,j}(y)] = n_j^{-2} \sum_{i,j} V(y_{ij})n_{ij}$. From the reduced-form of model (2), the covariance matrix of outcomes for group

---

\(^{17}\)Hanushek (1971), Rivkin et al. (2005), Glaeser et al. (1996).
then Theorem 6 of Rothenberg (1971, p. 585) is applied. Proofs can be found in Appendix D.8.

formed by stacking restrictions, including those originating from reduced-form estimation, has full rank, and

\[
- \text{diagonal terms}
\]

where \( s_j = (S_j^0)^{-1} - \mathbb{E}((S_j^0)^{-1}) \) 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 different 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 \( \mathbb{E}V_{W,j}(y_n) \) will be used to distinguish between competing network models in the identified set. Given it is observed from data, we only need to generate predictions from the model (17), suggesting a final estimator that minimizes a norm of the distance between observed and expected variance of \( y_j \). Naturally, this strategy depends on the evaluation of \( \nabla(y_j) \) and its expectation, 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 4. (Bernoulli network model).** In a simple setting where link formation is independent and equal to \( \delta \), 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 (17) simplifies to

\[
\nabla \{ y_j \} = \text{diag} \left( \nabla \{ W_j \} \left( \lambda^2 \text{diag} \left( x_j^{11} \right) + 2\lambda \text{diag} \left( x_j^{12} \right) + \lambda^2 \sigma^2 I_{n_j} \right) \right) + \sigma^2 I_{n_j}
\]

(18)

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

\[
\nabla \{ y_j \} = \text{diag} \left( \delta_1 (1 - \delta_1) \epsilon_{n_j}\epsilon_{n_j}' \left( \lambda^2 \text{diag} \left( x_j^{11} \right) + 2\lambda \text{diag} \left( x_j^{12} \right) + \lambda^2 \sigma^2 I_{n_j} \right) \right) + \sigma^2 I_{n_j}
\]

\[
= \delta_1 (1 - \delta_1) \left( \lambda^2 \epsilon_{n_j}\text{diag} \left( x_j^{11} \right) + 2\lambda \epsilon_{n_j}\text{diag} \left( x_j^{12} \right) + \lambda^2 \epsilon_{n_j}\sigma^2 \right) I_{n_j} + \sigma^2 I_{n_j}
\]

and the between-group variance is

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

This provides the additional restriction required for the identification of \( \theta \). 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.

\footnote{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.}

\[\square\]
An ideal estimator would be one that solves the generalized moment conditions

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

with moments $q_j(y) = [V_{B,j}(y) - \mathbb{E}V_{B,j}(y); V_{W,j}(y) - \mathbb{E}V_{W,j}(y)]$ and $2 \times 2$ weight matrix $\Omega$ on the estimated set $\hat{\Theta}$. 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 desired asymptotic properties, including point consistency for $\hat{\theta}$. The final estimator is the solution to

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

(19)

where $q_{s,j}(y) = [V_{B,j}(y) - V_{B,j}(\hat{y}_s); V_{W,j}(y) - V_{B,j}(\hat{y}_s)]$ with $\hat{y}_s = (S_n^s)^{-1} (x_n \beta_1 + W_n^s x_n \beta_2 + e_n^s)$, $S^s = (I_n - \lambda W_n^s)^{-1}$, $W_n^s$ is sampled from the distribution of the network-generating model with parameters $\theta$ and $e_n^j$ is sampled from a normal distribution with variance $\sigma^2$. If the simulator is unbiased, one can expect that $S^{-1} \sum_{s=1}^{S} q_{s,j}(y) \overset{p}{\to} q_j(y)$, and asymptotic properties follow, although variance in the limit has to take into account the imprecision of the simulator when $S$ is fixed. In addition, given $\Theta$ is $\sqrt{n}$-consistent for $\Theta_0$ on the Hausdorff metric, one might expect minimizing on the set $\hat{\Theta}$ is asymptotically equivalent to minimizing on the identified set $\Theta_0$.

**Theorem 4.** If parameters are identified, (i) estimator (19), minimized on the estimated set $\hat{\Theta}$, as defined in section 3.2, is consistent for $\theta_0$, $\hat{\theta} \overset{p}{\to} \theta_0$, and (ii) if $S \to \infty$ sufficiently fast, $\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Sigma^*)$, where $\Sigma^* = (G' (\Omega^*)^{-1} G)^{-1}$, $G = \mathbb{E} \nabla_{\theta} q_j(y, \theta_0)$ and $\Omega^* = (\mathbb{E} q_j(y, \theta_0) q_j(y, \theta_0)' )^{-1}$ with optimal choice of weight matrix $\Omega^*$.

### 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 covariates $x_j$ may correlate with unobserved components that vary at the group or individual-level, for example an unobserved "good teacher" shock in a classroom.

Bramoulle 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. Let the spatio-temporal model be, for $t = 1, \ldots, T$,

$$y_{jt} = \lambda W_j y_{jt} + x_{jt} \beta_1 + W_j x_{jt} \beta_2 + \alpha_j + \gamma_t + v_{jt} \quad (20)$$

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

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

$$\dot{y}_{jt} = \lambda W_j \dot{y}_{jt} + \dot{x}_{jt} \beta_1 + W_j \dot{x}_{jt} \beta_2 + \dot{\gamma}_t + \dot{v}_{jt} \quad (21)$$

which is a consequence of (20) because the time-differenced $W_j y_{jt}$ is equal to $W_j \dot{y}_{jt}$, and similarly for the term with $x_j$, under the hypothesis of invariance of the network over time. Explicitly, the $k$-th line of the time-differenced $W_j y_{jt}$ is

$$\sum_{i=1}^{n_j} \{W_j\}_{ki} \{y_{jt}\}_i - T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n_j} \{W_j\}_{ki} \{y_{jt}\}_i = \sum_{i=1}^{n_j} \{W_j\}_{ki} \{y_{jt}\}_i - \sum_{i=1}^{n_j} \{W_j\}_{ki} T^{-1} \sum_{t=1}^{T} \{y_{jt}\}_i$$

$$= \sum_{i=1}^{n_j} \{W_j\}_{ki} \{y_{jt}\}_i - \{y_{jt}\}_i$$

Letting $\dot{y}_n = (\dot{y}_{11}, \ldots, \dot{y}_{1T}, \ldots, \dot{y}_{n1}, \ldots, \dot{y}_{nT})'$ and $\dot{x}_n = (\dot{x}_{11}, \ldots, \dot{x}_{1T}, \ldots, \dot{x}_{n1}, \ldots, \dot{x}_{nT})'$, and similarly for $\dot{v}$ and $\dot{\gamma}$, the full model can be rewritten $\dot{y}_n = \lambda W \dot{y}_n + \dot{x}_n \beta_1 + W \dot{x}_n \beta_2 + \dot{\gamma} + \dot{v}_n$, where $W = \text{diag}\{I_T \otimes W_1, \ldots, I_T \otimes W_n\}$. The variance-covariance matrix of $\dot{v}$ is $\mathbb{E}(\dot{v} \dot{v}') = \sigma_0^2 R_0^{-1} \Sigma R_0^{-1}$, where $\Sigma_n = \sigma_0^2 I_n - \sigma_0^2 T^{-1} \cdot \text{diag}(I_{T} \otimes I_{n_1}, \ldots, I_T' \otimes I_{n_n})$. This more complicated form recognizes the dependence in $\dot{v}$ introduced by time-average subtraction. Finally, likelihood (11) can be adjusted to

$$\ln \mathcal{L}^\theta (\theta | y_n, x_n, Q_n) = -\frac{n}{2} \ln (2\pi \sigma^2) + \ln |S_n^e (Q_n, \theta)| + \ln |R_n^e (Q_n, \theta)| - \frac{1}{2\sigma^2} e_n^e (Q_n, \theta)' \hat{\Sigma}_n e_n^e (Q_n, \theta) \quad (23)$$

where $e_n^e (Q_n, \theta) = R_n^e (Q_n, \theta) (y_n - \lambda W_n^e (Q_n, \theta) \dot{y}_n - \dot{x}_n \beta_1 + W_n^e (Q_n, \theta) \dot{x}_n \beta_2 - \dot{\gamma}) = R_n^e (Q_n, \theta) (S_n^e (Q_n, \theta) \dot{y}_n - \hat{Z}_n^e (Q_n, \theta) \hat{\beta})$ where $\hat{Z}_n^e (Q_n, \theta)$ now also incorporate time effects: $\hat{Z}_n^e (Q_n, \theta) = (x_{jt}, W_j^e (Q_n, \theta) x_{jt}, 1 \{t = 1\} t n_j, \ldots, 1 \{T \} t n_j)$ and $\hat{\beta} = (\beta', \gamma_1, \ldots, \gamma_T)'$. The concentrators are now

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

$$\hat{\gamma}_n (Q, \theta) = \frac{1}{n} (S_n^e (Q_n, \theta) y - Z_n^e (Q_n, \theta) \hat{\beta}) R_n^e (Q_n, \theta) \Sigma_n R_n^e (Q_n, \theta) (S_n^e (Q_n, \theta) y - Z_n^e (Q_n, \theta) \hat{\beta})$$
Concentrated likelihood 11 remains unchanged with $\hat{\sigma}^2(Q_n, \theta)$ substituted for $\tilde{\sigma}^2(Q_n, \theta)$. Preceding theorems are applied without further modifications.

4 Simulations.

In this section, I conduct a simulation exercise to show empirical properties of the estimator. 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. Simulations are conducted for $(n = 25, v = 250)$, $(n = 100, v = 250)$, $(n = 25, v = 1000)$ and $(n = 100, v = 1000)$. 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.

True parameters are $\theta_s = (0.0125, 1, 1, 0.04, 0.4, 1)'$ 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 sharing an exogenous characteristic is 50%. Finally, $x$ and $\epsilon$ are drawn from a normal distribution with mean 0 and variance 1. The simulation is composed of 200 repetitions. The average of the estimated standard errors, following the procedure outlined in 3.2, is shown in parenthesis, while standard deviations of the point estimates computed across replications is shown in brackets. Estimation is first conducted assuming $\lambda$ is known. Then, using second moment restrictions, as outlined in section 3.3, I estimate $\lambda$, which is reported in the first row.

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 social multiplier is correctly estimated. Introduction of time dimension and fixed effects do not change the results, despite the fact that estimates of $\sigma^2$ 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.

As a robustness check, I conduct estimation and hypothesis testing when $\lambda$ is misspecified, as shown in table 9 of appendix E.1. I assume incorrectly $\lambda = 0.0250$, twice the true value. In this case, I observe halved $\hat{\lambda}_1$ and $\hat{\lambda}_0$ and $\hat{\lambda}_2$ estimated twice the true parameter. Associated standard errors followed the same pattern. This is entirely consequence of the partial identification; it is of importance, nonetheless, that multiplication of point estimates and standard errors by the same factor allows hypothesis testing although levels are not identified in absence of knowledge of $\lambda$. 

21
<table>
<thead>
<tr>
<th></th>
<th>( T = 1 )</th>
<th></th>
<th>( T = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>( n )</td>
<td>25</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>( v )</td>
<td>250</td>
<td>250</td>
<td>1000</td>
</tr>
<tr>
<td>( \hat{\lambda} )</td>
<td>0.0117</td>
<td>0.0122</td>
<td>0.0126</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.005]</td>
<td>[0.002]</td>
</tr>
<tr>
<td>( \hat{\beta}_1 )</td>
<td>0.9999</td>
<td>1.0001</td>
<td>0.9998</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.005]</td>
<td>[0.006]</td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>0.0461</td>
<td>0.0402</td>
<td>0.0400</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.005]</td>
<td>[0.006]</td>
</tr>
<tr>
<td>( \hat{\delta}_1 )</td>
<td>0.7166</td>
<td>0.7497</td>
<td>0.7605</td>
</tr>
<tr>
<td></td>
<td>[0.164]</td>
<td>[0.024]</td>
<td>[0.097]</td>
</tr>
<tr>
<td>( \hat{\delta}_0 )</td>
<td>0.2892</td>
<td>0.2995</td>
<td>0.3015</td>
</tr>
<tr>
<td></td>
<td>[0.062]</td>
<td>[0.007]</td>
<td>[0.031]</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>1.0571</td>
<td>1.2199</td>
<td>1.0547</td>
</tr>
<tr>
<td></td>
<td>[0.018]</td>
<td>[0.003]</td>
<td>[0.008]</td>
</tr>
<tr>
<td>( \varphi(\hat{x}, \hat{\theta}) )</td>
<td>-0.0007</td>
<td>0.137</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>[0.023]</td>
<td>[0.092]</td>
<td>[0.008]</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 )</td>
<td>[0.006]</td>
<td>[0.007]</td>
<td>[0.001]</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>$T = 5$, time effects.</th>
<th></th>
<th>$T = 5$, time and fixed effects.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>$n$</td>
<td>25</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>$v$</td>
<td>250</td>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>0.0121</td>
<td>0.0119</td>
<td>0.0123</td>
</tr>
<tr>
<td></td>
<td>[0.005]</td>
<td>[0.005]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.002]</td>
<td>[0.002]</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>(0.007)</td>
<td>(0.001)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>[0.006]</td>
<td>[0.001]</td>
<td>[0.003]</td>
</tr>
<tr>
<td>$\hat{\delta}_1$</td>
<td>0.7387</td>
<td>0.7508</td>
<td>0.7588</td>
</tr>
<tr>
<td>$\hat{\delta}_0$</td>
<td>(0.070)</td>
<td>(0.009)</td>
<td>(0.034)</td>
</tr>
<tr>
<td></td>
<td>[0.071]</td>
<td>[0.011]</td>
<td>[0.036]</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>0.2958</td>
<td>0.2992</td>
<td>0.3018</td>
</tr>
<tr>
<td>$\varphi(x, \hat{\theta})$</td>
<td>0.0114</td>
<td>0.0195</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>[0.007]</td>
<td>[0.001]</td>
<td>[0.003]</td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.001]</td>
<td>[0.002]</td>
</tr>
<tr>
<td></td>
<td>[0.027]</td>
<td>[0.003]</td>
<td>[0.013]</td>
</tr>
<tr>
<td>$\varphi(x, \hat{\theta})$</td>
<td>0.0005</td>
<td>-0.0050</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.039]</td>
<td>[0.005]</td>
</tr>
<tr>
<td></td>
<td>[0.014]</td>
<td>[0.008]</td>
<td>[0.007]</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
<td>[0.001]</td>
<td>[0.000]</td>
</tr>
</tbody>
</table>

Note: True parameters are $\beta_1 = 1$, $\beta_2 = 0.04$, $\delta_1 = 0.75$, $\delta_0 = 0.30$, $\sigma^2 = 1$ and $\varphi(x, \theta) = 0$. Option "with fixed effects" means DGP contains FE and were and were eliminated; "fixed effects not eliminated" means DGP contains FE but were not eliminated.
Lastly, I implement the multivariate network model described in example 5 of subsection B, where probability of link formation is described by

\[ P \{ \{ W_{jk} \}_{ik} = 1|Q_j \} = Q_{jik}^1 \delta_1 + Q_{jik}^0 \delta_0 \]

where \( Q_{jik}^1 \) is the distance between individuals \( i \) and \( k \) who belong to group \( j \), and respectively for \( Q_{jik}^0 \). 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 10 of appendix E.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 \( \lambda \) using second moments is also satisfactory.

5 Application.

Treatment provision to a large proportion individuals in a group gives rise to the possibility that spillovers or externalities play a key role in overall program evaluation (Miguel and Kremer (2004)). The method developed in the current paper is a systematic way of investigating this issue when networks are unknown or data is unreliable and information on a large number of groups is available.

In this section, I use experimental data from Bandiera et al. (2013), who conducted a randomized control trial aimed at evaluating whether absence of capital and skills determine occupational choice of poor villagers. Treatment consisted on assignment of livestock and skills training, both relevant in terms of the outlay (at around USD 140) and duration (training lasted for two years). Results point to a dramatic change in the occupational status of targeted households: four years after treatment, poor woman dedicate 92% additional hours towards self-employment running their livestock rearing business, switch away from wage hours – often insecure and temporary jobs –, and the change in occupation is associated with higher earnings, per capita expenditure, general wellbeing and life satisfaction measures.

In total, 7953 beneficiaries were surveyed in 1409 communities and, owing to village-level randomization, all eligible individuals in selected communities were treated (around 20% of the female residents in comparison to local population), a setting conducive to treatment externalities. With recourse to the estimation method developed in this paper, and without network data, I supplement these results with three additional network-dependent findings:

1. On occupation status: while treated individuals moved towards self-employment, both in terms of hours and specialization, and reduced the amount of hours spent in wage employment, connected peers moved in the opposite direction, reduced self-employment hours, decreased specialization in self-employment and increased wage hours. Overall, the supply of working hours in the village level increased. This finding is consistent with the takeover of open vacancies left by treated individuals;

2. On assets: while treated individuals increased livestock assets beyond treatment, connected peers reduced both the number and value of livestock. This is compatible with the interpretation that treated individuals
gained comparative advantage in livestock rearing, specializing in this function at the community level, and peers were better off lending, borrowing or selling their assets and taking paid jobs instead;

3. On per capita expenditure: food expenditure increased for treated individuals and their connected peers. Accordingly, food security measures registered improvements for both groups.

The extensive and rich survey contained detailed information on several networks measures, such as family relations, acquaintances, work, borrowing and lending. I then supplement and support these three findings with recourse to network data.

5.1 Program Description.

Selection of targeted individuals proceeds in stages. In collaboration with BRAC, a local non-profit organization, most vulnerable districts are selected based on food-security measures as provided by the World Food Program. Second, BRAC employees select the poorest communities within each district, which are self-contained groups of approximately 100 households. Lastly, within each community, a combination of a participatory rural appraisal exercise and survey data is used to allocate households to one of five wealth bins. Households belonging to the poorest wealth bins are selected as a potential beneficiary if other eligibility criteria are met, for example excluding microfinance borrowers and owning no productive assets.

Randomization is conducted at the local BRAC branch level, among its 40 offices in Bangladesh, stratified at subdistrict level to ensure balance between treated and control groups: within each subdistrict, one branch was randomly allocated to treatment and another to control group, and asset transfer was conducted for all selected individuals within the communities covered by a treated BRAC branch. As a consequence, a significant fraction of village populations was treated, raising the possibility that aggregate community-level effects are substantially larger than the sum of isolated individual treatment effects, for example as a consequence of learning and informal skills reinforcement from neighbors, which in turn may or not have been treated.

If eligible and selected through the randomization process, households receive a transfer of live animals (valued at around USD 140) and subsequent skill training for two years, specifically designed for the chosen asset. Program beneficiaries could select among cows, goats or chicken, adding up to the same face value, and 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 they failed to do so.

All potential beneficiaries of the program and a sample of households across the village wealth distribution are surveyed just before the intervention in 2007 and in two additional waves in 2009 and 2011. The comprehensive survey was composed 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, woman empowerment and vulnerability, such as earnings seasonality and food security, and a health module. Importantly, network links were registered when applicable: data includes family outside the household, their business activities, land transfers (through inheritance, mortgage, rent, share, received in as dowry or gift, bought or sold), business assets transfers (same possibilities as above), finance links (loans, outstanding lending or transfers) and letting of house ownerships. Most questions were repeated independently for the main female and the household head. For a subset of the villages, the survey was conducted for the
5.2 Evaluation and Identification Strategies.

Evaluation of the treatment is first conducted with a simple triple differences-in-differences as the benchmark model. Momentarily disconsidering networks, denote $S_{ij} = 1$ if individual $i$ of village $j$ was selected as a potential beneficiary of the programme, and $T_{ij} = 1$ if village $j$ was randomly selected for treatment. The benchmark model is

\[ y_{ijt} = \sum_{s=2}^{3} \beta_{1s} S_{ij} T_{ij} \{s = t\} + \sum_{s=2}^{3} \eta_{1s} S_{ij} \{s = t\} + \sum_{s=2}^{3} \eta_{2s} T_{ij} \{s = t\} + \gamma_t + \alpha_{ij} + \epsilon_{ijt} \]  \tag{24}

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), $\alpha_{ij}$ is a fixed effect at the individual level, $\gamma_t$ is a full set of time effects, $\{\cdot\}$ is an indicator function and $\epsilon_{ijt}$ is the disturbance term. The coefficients of interest are $\beta_{12}$ and $\beta_{13}$, which are identified by comparing changes over time of the outcomes of the selected individuals against non-selected individuals in treated communities; the third difference eliminates differential trends that could occur in the absence of treatment by subtracting similar differences based on non-treated communities, who were also entirely surveyed and individuals were equally selected.

I next introduce network spillovers, taking the form of two additional network-dependent terms in equation (24). In vector notation,

\[ y_{jt} = \lambda W^0_j y_{jt} + \sum_{s=2}^{3} \beta_{1s} S_j T_j \{s = t\} + \sum_{s=2}^{3} W^0_j S_j T_j \{s = t\} \beta_{2s} + \]
\[ + \sum_{s=2}^{3} \eta_{1s} S_j \{s = t\} + \sum_{s=2}^{3} \eta_{2s} T_j \{s = t\} + \gamma_t + \alpha_j + \epsilon_{jt} \]  \tag{25}

where $W^0_j$ is the unobserved household-level network and $ST_j$ is a column vector with $i$th line indicating whether individual $i$ was selected and lives in treated village $j$. Term $\lambda W^0_j y_{jt}$ represents the endogenous effects – the fact the one own’s choice depends on others’ choices – and $W^0_j S_j T_j \{s = t\} \beta_{2s}$ is the exogenous effects – one own’s choice depends on others’ treatment status. Coefficient $\beta_{22}$ and $\beta_{23}$ are interpreted as the marginal effect if one additional peer is treated, while $\beta_{12}$ and $\beta_{13}$ are the overall effects if networks were entirely absent or irrelevant.

I provide point estimates using the dispersion of outcomes, as described in section 3.3. Intuitively, when networks are present, a high dispersion of outcomes across villages is expected, while individual outcomes within groups positively correlate. This allows me to overcome the difficulty of separating a large number of weak interactions from a low number of strong interactions. By matching observed and predicted dispersion as implied by model, it is then possible to provide the desired result, given there is also sufficient variation in group sizes\(^{19}\). Estimates of network-wide parameters, particularly the density, discriminates between the polar cases where interactions occur in a localized scale, though personal interconnectedness and without intermediation of the markets (equivalent to low-density networks) or general equilibrium effects (where one own’s choices affects

\(^{19}\) It is the case in the current setting, as the survey was conducted in 1,409 villages.
all others to a small degree, and so networks are very dense).

Finally, I average network spillovers $\varphi(x_{jt}, \hat{\theta})$ for treated individuals after two and four years (denoted $\hat{\varphi}_{T,2}$ and $\hat{\varphi}_{T,4}$) and similarly for nontreated individuals ($\hat{\varphi}_{NT,2}$ and $\hat{\varphi}_{NT,4}$). Note the overall treatment effect for treated is the sum of the program effect and spillovers. Construction of confidence intervals and standard errors is described in subsection 3.2.

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 thousand of TAKAs, share of income originating from seasonal and regular activities in table 4), livestock assets (number of cows, poultry and livestock value in thousand TAKAs in table 5) and per capita expenditure (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 proximity of household identifiers, registered as $Q_{ij} = 1$ and zero otherwise. It has been anectodally noticed that identifiers were associated as surveyors followed local streets and roads, so serving as a proxy for geographical distance. This is only a generalization from the purely naive network where probability of link formation is constant and independent of any variable\textsuperscript{20}.

Results show that, while treated individuals reduced wage hours (114 and 142 yearly hours, two and four years after treatment) and increased self-employment hours (469 and 460 yearly hours) associated to livestock rearing, connected peers followed an opposite pattern, by increasing wage hours (25 and 18 yearly hours for each treated peer) and decreasing self-working hours (13 yearly hours for each treated peer). This finding is consistent with the treated individuals specializing in self-employment, and connected peers decreasing specialization. Individuals who received treatment left wage-job open, filled with individuals located in close geographic proximity\textsuperscript{21}. For self-employment hours, the role of local labour markets is pronounced, with density of estimated networks at 98% conditional on households living in close proximity and approximately 40% otherwise. A similar figure is registred in wage hours; interaction patters of all other outcomes are much more localized.

Results show that treated individuals increased livestock assets by more than original treatment; meanwhile, nontreated individuals reduced their stock of assets (not observed for poultry, consistent with low takeover of this choice of asset transfer). Livestock value followed the same pattern for both groups. These results are consistent with specialization in the village level. Because treatment consisted also on skills training, specific for the kind of assets provided, and of long duration (2 years), it is possible that these individuals are now endowed with stronger comparative advantage in livestock rearing, while connected peers specialized in other activities.

Expenditure on food consumption increased for both groups, two and four years after treatment. In this case, network spillovers are very significant in magnitude and, depending on the specification, corresponds to 25%-50% of the original treatment effect. This is further confirmed by food security measures (defined as a household that can provide at least two meals a day in most days), which also registered improvement for both

\textsuperscript{20}Estimation with naive model for probability of link formation is conducted as a robustness in table 15 in appendix E.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.

\textsuperscript{21}The null hypothesis of no differential association is rejected at the 5% level for all specifications, not shown in tables (3)-(6).
groups. Lastly, nonfood expenditures are either constant or show slight increase for the treated group; nontreated group reduced nonfood consumption after four years. This result can be explained by, as explored above, the decrease in business assets following specialization in wage labor.

Finally, I make use of network data collected in the survey to reassess the conclusions obtained in its absence, allowing the probability of link formation to depend on link reporting. I combine network reports into two categories: family and non-family (economic) links. Tables 11 to 14 of appendix E.2, show that main conclusions stand unchanged. Moreover, the null hypothesis of no network validity was rejected at 5% level for most specifications, in particular for the outcomes concerning occupational choice, earnings and seasonality.

Table 3: Occupational Choice.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>OLS.</td>
<td>Network.</td>
<td>OLS.</td>
<td>Network.</td>
<td>OLS.</td>
<td>Network.</td>
</tr>
<tr>
<td>Program effect after 2 years ($\hat{\beta}_{11}$).</td>
<td>468.928***</td>
<td>469.774***</td>
<td>-110.799***</td>
<td>-113.531***</td>
<td>0.107***</td>
<td>0.114***</td>
</tr>
<tr>
<td></td>
<td>(28.62)</td>
<td>(23.20)</td>
<td>(31.07)</td>
<td>(10.61)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Program effect after 4 years ($\hat{\beta}_{12}$).</td>
<td>465.075***</td>
<td>460.039***</td>
<td>-137.255***</td>
<td>-141.918***</td>
<td>0.112***</td>
<td>0.120***</td>
</tr>
<tr>
<td></td>
<td>(31.32)</td>
<td>(23.21)</td>
<td>(34.10)</td>
<td>(8.63)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on T after 2 years ($\hat{\varphi}_{T,2}$).</td>
<td>—</td>
<td>-6.347</td>
<td>—</td>
<td>26.855***</td>
<td>—</td>
<td>-0.032***</td>
</tr>
<tr>
<td></td>
<td>(10.55)</td>
<td>(8.45)</td>
<td>(8.54)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Spillover on T after 4 years ($\hat{\varphi}_{T,4}$).</td>
<td>—</td>
<td>28.847***</td>
<td>—</td>
<td>19.369**</td>
<td>—</td>
<td>-0.025***</td>
</tr>
<tr>
<td></td>
<td>(9.68)</td>
<td>(8.45)</td>
<td>(5.98)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Spillover on NT after 2 years ($\hat{\varphi}_{NT,2}$).</td>
<td>—</td>
<td>-3.229</td>
<td>—</td>
<td>14.491***</td>
<td>—</td>
<td>-0.018***</td>
</tr>
<tr>
<td></td>
<td>(5.37)</td>
<td>(4.55)</td>
<td>(4.75)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Spillover on NT after 4 years ($\hat{\varphi}_{NT,4}$).</td>
<td>—</td>
<td>14.676***</td>
<td>—</td>
<td>10.452***</td>
<td>—</td>
<td>-0.013***</td>
</tr>
<tr>
<td></td>
<td>(1.99)</td>
<td>(0.75)</td>
<td>(0.75)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Link to T after 2 years ($\hat{\beta}_{21}$).</td>
<td>—</td>
<td>-24.604***</td>
<td>—</td>
<td>13.904***</td>
<td>—</td>
<td>-0.050***</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
<td>(2.52)</td>
<td>(2.52)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Link to T after 4 years ($\hat{\beta}_{22}$).</td>
<td>—</td>
<td>-17.932***</td>
<td>—</td>
<td>13.030***</td>
<td>—</td>
<td>-0.043***</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
<td>(1.59)</td>
<td>(1.59)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Link probability if $Q_{ij} = 1$ ($\hat{\delta}_1$).</td>
<td>—</td>
<td>0.983***</td>
<td>—</td>
<td>0.639***</td>
<td>—</td>
<td>0.192***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Link probability if $Q_{ij} = 0$ ($\hat{\delta}_0$).</td>
<td>—</td>
<td>0.396***</td>
<td>—</td>
<td>0.331***</td>
<td>—</td>
<td>0.106***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>—</td>
<td>0.05</td>
<td>—</td>
<td>0.05</td>
<td>—</td>
<td>0.15</td>
</tr>
<tr>
<td>p-value $H_{NV}$</td>
<td>—</td>
<td>&lt; 0.001</td>
<td>—</td>
<td>&lt; 0.001</td>
<td>—</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Avg treated outcome.</td>
<td>421.8</td>
<td>421.8</td>
<td>646.7</td>
<td>646.7</td>
<td>0.303</td>
<td>646.7</td>
</tr>
<tr>
<td>Individuals ($n$).</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
</tr>
<tr>
<td>Villages ($v$).</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
</tr>
<tr>
<td>Survey waves ($T$).</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 4: Earnings and Seasonality.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Program effect after 2 years ($\hat{\beta}_{11}$)</td>
<td>0.475</td>
<td>0.506***</td>
<td>0.012</td>
<td>-0.028***</td>
<td>0.201***</td>
</tr>
<tr>
<td>Program effect after 4 years ($\hat{\beta}_{12}$)</td>
<td>2.598***</td>
<td>2.729***</td>
<td>-0.089***</td>
<td>-0.074***</td>
<td>0.191***</td>
</tr>
<tr>
<td>Spillover on T after 2 years ($\hat{\phi}_{T,2}$)</td>
<td>-0.045</td>
<td>-0.051***</td>
<td>-</td>
<td>0.023**</td>
<td></td>
</tr>
<tr>
<td>Spillover on T after 4 years ($\hat{\phi}_{T,4}$)</td>
<td>0.008</td>
<td>-0.005</td>
<td>-</td>
<td>0.029**</td>
<td></td>
</tr>
<tr>
<td>Spillover on NT after 2 years ($\hat{\phi}_{NT,2}$)</td>
<td>-0.025</td>
<td>-0.023***</td>
<td>-</td>
<td>0.012**</td>
<td></td>
</tr>
<tr>
<td>Spillover on NT after 4 years ($\hat{\phi}_{NT,4}$)</td>
<td>0.004</td>
<td>-0.002</td>
<td>-</td>
<td>0.015***</td>
<td></td>
</tr>
</tbody>
</table>

Function of $\lambda$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Link to T after 2 years ($\hat{\beta}_{21}$)</td>
<td>-0.447</td>
<td>-0.010***</td>
<td>-</td>
<td>-0.022***</td>
<td></td>
</tr>
<tr>
<td>Link to T after 4 years ($\hat{\beta}_{22}$)</td>
<td>-0.326</td>
<td>-0.016***</td>
<td>-</td>
<td>-0.015**</td>
<td></td>
</tr>
<tr>
<td>Link probability if $Q_{ij} = 1$ ($\hat{\delta}_{1}$)</td>
<td>0.075***</td>
<td>0.272***</td>
<td>-</td>
<td>0.238***</td>
<td></td>
</tr>
<tr>
<td>Link probability if $Q_{ij} = 0$ ($\hat{\delta}_{0}$)</td>
<td>0.023***</td>
<td>0.136***</td>
<td>-</td>
<td>0.106***</td>
<td></td>
</tr>
</tbody>
</table>

| λ | - | 0.50 | - | 0.20 | - | 0.20 |

p-value $H_{NV}$

| Avg treated outcome. | 4.607 | 4.607 | 0.674 | 0.674 | 0.478 | 0.478 |
| Individuals ($n$) | 69087 | 69087 | 69087 | 69087 | 69087 | 69087 |
| Villages ($v$) | 1409 | 1409 | 1409 | 1409 | 1409 | 1409 |
| Survey waves ($T$) | 3 | 3 | 3 | 3 | 3 | 3 |
Table 5: Livestock.

<table>
<thead>
<tr>
<th>Outcome Method</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS.</td>
<td>Network.</td>
<td>OLS.</td>
<td>Network.</td>
<td>OLS.</td>
<td>Network.</td>
</tr>
<tr>
<td>Program effect</td>
<td>1.119***</td>
<td>1.131***</td>
<td>2.147***</td>
<td>2.120***</td>
<td>10.326***</td>
<td>10.417***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\beta}_{11}$)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.42)</td>
<td>(0.50)</td>
<td>(0.56)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>Program effect</td>
<td>1.078***</td>
<td>1.102***</td>
<td>1.294***</td>
<td>1.326***</td>
<td>10.984***</td>
<td>11.175***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\beta}_{12}$)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.62)</td>
<td>(0.50)</td>
<td>(0.64)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Spillover on T</td>
<td>-0.033***</td>
<td></td>
<td>0.099</td>
<td></td>
<td>-0.221***</td>
<td></td>
</tr>
<tr>
<td>after 2 years ($\hat{\phi}_{T,2}$)</td>
<td></td>
<td>(0.01)</td>
<td></td>
<td>(0.17)</td>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>Spillover on T</td>
<td>-0.057***</td>
<td></td>
<td>-0.087</td>
<td></td>
<td>-0.459***</td>
<td></td>
</tr>
<tr>
<td>after 4 years ($\hat{\phi}_{T,4}$)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.20)</td>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>-0.020***</td>
<td></td>
<td>0.059</td>
<td></td>
<td>-0.132***</td>
<td></td>
</tr>
<tr>
<td>after 2 years ($\hat{\phi}_{NT,2}$)</td>
<td></td>
<td>(0.01)</td>
<td></td>
<td>(0.10)</td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>-0.033***</td>
<td></td>
<td>-0.052</td>
<td></td>
<td>-0.274***</td>
<td></td>
</tr>
<tr>
<td>after 4 years ($\hat{\phi}_{NT,4}$)</td>
<td></td>
<td>(0.01)</td>
<td></td>
<td>(0.08)</td>
<td></td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

Not function of $\hat{\lambda}$

<table>
<thead>
<tr>
<th>Function of $\lambda$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link to T</td>
<td>-0.596***</td>
<td></td>
<td>1.277</td>
<td></td>
<td>-10.456***</td>
<td></td>
</tr>
<tr>
<td>after 2 years ($\hat{\beta}_{21}$)</td>
<td></td>
<td>(0.16)</td>
<td></td>
<td>(4.12)</td>
<td></td>
<td>(1.90)</td>
</tr>
<tr>
<td>Link to T</td>
<td>-1.285***</td>
<td></td>
<td>-2.725</td>
<td></td>
<td>-16.464***</td>
<td></td>
</tr>
<tr>
<td>after 4 years ($\hat{\beta}_{22}$)</td>
<td></td>
<td>(0.17)</td>
<td></td>
<td>(4.11)</td>
<td></td>
<td>(2.33)</td>
</tr>
<tr>
<td>Link probability</td>
<td>0.024***</td>
<td></td>
<td>0.007***</td>
<td></td>
<td>0.013***</td>
<td></td>
</tr>
<tr>
<td>if $Q_{ij} = 1$ ($\hat{\delta}_{1}$)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Link probability</td>
<td>0.012***</td>
<td></td>
<td>0.009***</td>
<td></td>
<td>0.007***</td>
<td></td>
</tr>
<tr>
<td>if $Q_{ij} = 0$ ($\hat{\delta}_{0}$)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

| $\lambda$ | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| p-value $H_{NV}$ | < 0.001 | < 0.001 | < 0.001 | < 0.001 |

Avg treated outcome. 0.083 0.083 1.79 1.79 0.940 0.940
Individuals ($n$). 69087 69087 69087 69087 69087 69087
Villages ($v$). 1409 1409 1409 1409 1409 1409
Survey waves ($T$). 3 3 3 3 3 3
Table 6: Expenditures.

<table>
<thead>
<tr>
<th>Method</th>
<th>Outcome</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OLS</td>
<td>Network</td>
<td>OLS</td>
<td>Network</td>
<td>OLS</td>
<td>Network</td>
</tr>
<tr>
<td>Not function of ( \hat{\lambda} )</td>
<td>Program effect</td>
<td>-242.239</td>
<td>-220.509</td>
<td>585.304**</td>
<td>423.929***</td>
<td>0.189***</td>
<td>0.169***</td>
</tr>
<tr>
<td></td>
<td>after 2 years (( \hat{\beta}_{11} ))</td>
<td>(293.34)</td>
<td>(164.53)</td>
<td>(247.19)</td>
<td>(134.22)</td>
<td>(0.03)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>Program effect</td>
<td>175.022</td>
<td>278.277</td>
<td>585.415***</td>
<td>445.063***</td>
<td>0.010***</td>
<td>0.076***</td>
</tr>
<tr>
<td></td>
<td>after 4 years (( \hat{\beta}_{12} ))</td>
<td>(375.16)</td>
<td>(174.72)</td>
<td>(227.38)</td>
<td>(134.27)</td>
<td>(0.03)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>Spillover on T</td>
<td>—</td>
<td>-8.526</td>
<td>—</td>
<td>380.002***</td>
<td>—</td>
<td>0.017***</td>
</tr>
<tr>
<td></td>
<td>after 2 years (( \hat{\phi}_{T,2} ))</td>
<td>(68.25)</td>
<td>(55.82)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spillover on T</td>
<td>—</td>
<td>-171.985**</td>
<td>—</td>
<td>243.172***</td>
<td>—</td>
<td>0.071***</td>
</tr>
<tr>
<td></td>
<td>after 4 years (( \hat{\phi}_{T,4} ))</td>
<td>(68.15)</td>
<td>(56.88)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spillover on NT</td>
<td>—</td>
<td>-5.039</td>
<td>—</td>
<td>206.992***</td>
<td>—</td>
<td>0.027***</td>
</tr>
<tr>
<td></td>
<td>after 2 years (( \hat{\phi}_{NT,2} ))</td>
<td>(40.34)</td>
<td>(30.14)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spillover on NT</td>
<td>—</td>
<td>-101.655*</td>
<td>—</td>
<td>132.459***</td>
<td>—</td>
<td>0.032***</td>
</tr>
<tr>
<td></td>
<td>after 4 years (( \hat{\phi}_{NT,4} ))</td>
<td>(52.65)</td>
<td>(40.73)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function of ( \hat{\lambda} )</td>
<td>Link to T</td>
<td>—</td>
<td>-14.185</td>
<td>—</td>
<td>443.619***</td>
<td>—</td>
<td>0.096***</td>
</tr>
<tr>
<td></td>
<td>after 2 years (( \hat{\beta}_{21} ))</td>
<td>(988.46)</td>
<td>(85.36)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Link to T</td>
<td>—</td>
<td>-2649.43***</td>
<td>—</td>
<td>249.126***</td>
<td>—</td>
<td>0.087***</td>
</tr>
<tr>
<td></td>
<td>after 4 years (( \hat{\beta}_{22} ))</td>
<td>(980.96)</td>
<td>(84.79)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Link probability</td>
<td>—</td>
<td>0.032***</td>
<td>—</td>
<td>0.132***</td>
<td>—</td>
<td>0.128***</td>
</tr>
<tr>
<td></td>
<td>if ( Q_{ij} = 1 ) (( \hat{\delta}_1 ))</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Link probability</td>
<td>—</td>
<td>0.009***</td>
<td>—</td>
<td>0.080***</td>
<td>—</td>
<td>0.052***</td>
</tr>
<tr>
<td></td>
<td>if ( Q_{ij} = 0 ) (( \hat{\delta}_0 ))</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda )</td>
<td>—</td>
<td>0.50</td>
<td>—</td>
<td>0.20</td>
<td>—</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>( \beta )-value ( H_{NV} )</td>
<td>—</td>
<td>&lt; 0.001</td>
<td>—</td>
<td>&lt; 0.001</td>
<td>—</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Avg treated outcome.</td>
<td>1054.5</td>
<td>1054.5</td>
<td>2953.7</td>
<td>2953.7</td>
<td>0.457</td>
<td>0.457</td>
<td></td>
</tr>
<tr>
<td>Individuals (( n )).</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td></td>
</tr>
<tr>
<td>Villages (( v )).</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td></td>
</tr>
<tr>
<td>Survey waves (( T )).</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
6 Conclusion.

This paper is concerned with estimation of network effects and its structure in absence of network data. Under the assumption that observations of many groups are available for one period of time, I suggest an estimator that, in spirit, integrates network away. In this process, it recovers network spillovers and, with additional assumptions, the underlying network structure, such as its density, reciprocity or polarity. The key feature that allows estimation in absence of network data is the large number of groups and a probabilistic model for network formation.

The procedure is useful for at least three reasons. First, network spillovers constitute a confounding element when the final objective is the estimation of individual responses and, therefore, controlling for spillovers is paramount. In second place, networks may leverage variation from covariates to outcomes and, as such, may amount to an important reinforcement mechanism. For example, providing external validity to treatment evaluations requires, in part, the understanding of treatment spillovers through networks. Finally, it may incorporate network data, when available, by allowing the probability of link formation to depend on link reporting, and test its validity.

The procedure is a substantial contribution in the literature by considering estimation when networks are unobserved or stochastic. Literature in spatial econometrics (Anselin (2010) and references therein), identification of network models (Manski (1993), Bramoulé et al. (2009), De Giorgi et al. (2010) and others), estimation of random network formation (Snijders (2011)) dealt with a number of techniques only when networks are observed, which does not in many datasets. Therefore, the current paper considerably widens prospective applications and investigation of network phenomena in data previously unsuitable for such purpose.

While the estimator is set-identified without using information provided by previous knowledge or second moments, network spillovers are constant within the identified set. This important property allows the assessment of spillovers and hypothesis testing on structural parameters under fairly weak hypothesis. For this purpose, I adapt the methods of Chernozhukov et al. (2007), Bugni (2010) and Romano and Shaikh (2010) to the current setting. In order to provide consistent structural estimation, a restriction originating from the excess variance – or the dispersion of outcomes that cannot be explained by covariates or disturbance term alone – is used to close point-identification requirements. All procedures have been shown to work well in simulated settings, including models with fixed and time effects.

I apply the estimator on a randomized control trial that provided livestock and skills training to poor households in Bangladesh. While treated individuals moved towards specialization in livestock rearing, their peers reduced livestock value and increased wage hours. This is consistent with the interpretation that treatment changed comparative advantages of treated towards rearing, hence affecting the economic choices of their peers; overall, the both groups registered improvement in food consumption and security, an evidence that benefit of the treatment spread to a wider population than directly targeted.

The methods described in this paper are instrumental in elucidating mechanisms that arise as consequence of individual interconnectedness, especially in a setting where network data is not available or contains measurement errors. Estimating network spillovers, and discriminating between endogenous, contextual and correlated effects in absence of network data provides a very useful empirical tool for further applied research.
References


A Summary of Notation.

\[ \beta = (\beta'_1, \beta'_2), \theta = \theta \setminus \{ \beta, \sigma^2 \}, n = \sum_{j=1}^{n_j} n_j. \]

\( I_n \) an identity matrix of dimensions \( n \times n \), \( \mathbf{u}_n \) is a \( n \times 1 \) vector of ones.

\[ y_n = (y'_1, \ldots, y'_i, \ldots, y'_n) , \quad y_j = y_{n,j} (y_{n,j}, \ldots, y_{n,j})', \quad j = 1, \ldots, n_j. \]

\[ x_n = (x'_1, \ldots, x'_i, \ldots, x'_n) , \quad x_j = x_{n,j} = (x'_{nj}, \ldots, x'_{nj}), \quad j = 1, \ldots, n_j. \]

\[ \epsilon_n = (\epsilon'_1, \ldots, \epsilon'_i, \ldots, \epsilon'_n), \quad \epsilon_{n,j} = (\epsilon_{nj}, \ldots, \epsilon_{nj})', \quad j = 1, \ldots, n_j. \]

\[ y_j = \lambda_0 w_j^0 + x_j \beta_{10} + w_j^0 x_j \beta_{20} + v_j, \quad v_j = \rho_0 \mathbf{M}_j^0 v_j + \epsilon_j. \]

\[ Z_j^0 = Z_{n,j}^0 = (x_j, w_j^0 x_j), \quad Z_j^0 = (z_{1,j}^0, \ldots, z_{n,j}^0)', \quad z_{1,j}^0, \ldots, z_{n,j}^0. \]

\[ y_j = \lambda w_j^0 (Q, \theta_c) + x_j \beta_1 + w_j^0 (Q, \theta_c) x_j \beta_2 + \epsilon_j. \]

\[ Z_j^+(Q, \theta_c) = Z_{n,j}^+(Q, \theta_c) = (x_j, w_j^+ (Q, \theta_c) x_j), \quad Z_j^+(Q, \theta_c) = (z_{1,j}^+, \ldots, z_{n,j}^+ (Q, \theta_c))'. \]

\[ W_n^0 = \text{diag} (W_1^0, \ldots, W_n^0), \quad W_n^+ = \text{diag} (W_1^+, \ldots, W_n^+ (Q, \theta_c)). \]

\[ S_j^0 (\lambda) = S_{n,j}^0 (\lambda) = I_n - \lambda w_j^0, \quad S_j^+ = S_j^0 (\lambda_0), \quad S_n^0 = \text{diag} (S_1^0, \ldots, S_n^0). \]

\[ (S_n^0)^{-1} = \lambda_0 G_n^0 + I_n, \quad C_n^0 = W_n^0 (S_n^0)^{-1}. \]

\[ S_j^+ (Q, \theta) = I_n - \lambda w_j^+ (Q, \theta), \quad S_n^0 (Q, \theta) = \text{diag} (S_1^+ (Q, \theta), \ldots, S_n^+ (Q, \theta)) = W_n^0 (Q, \theta_c) (S_n^0 (Q, \theta)). \]

\[ (S_n^0 (Q, \theta))^{-1} = I_n + \lambda G_n^0 (Q, \theta_c), \quad G_n^0 (Q, \theta_c) = W_n^0 (Q, \theta_c) (S_n^0 (Q, \theta_c))^{-1}. \]

\[ R_n^0 (\rho) = I_n - \rho \mathbf{M}_j^0, \quad R_n^0 = R_n^0 (\rho_0), \quad R_n^0 = \text{diag} (R_1^0, \ldots, R_n^0). \]

\[ R_j^0 (\theta) = I_n - \rho \mathbf{M}_j^0 (Q, \theta), \quad R_n^0 (Q, \theta) = \text{diag} (R_1^0 (Q, \theta), \ldots, R_n^0 (Q, \theta)). \]

\[ M_n^0 = \text{diag} (M_1^0, \ldots, M_n^0), \quad M_n^0 (Q, \theta_c) = \text{diag} (M_1^0 (Q, \theta_c), \ldots, M_n^0 (Q, \theta_c)). \]

\[ P_n^0 (Q, \theta_c) = I_n - R_n^0 (Q, \theta_c) Z_n^0 (Q, \theta_c) [Z_n^0 (Q, \theta_c) R_n^0 (Q, \theta_c) R_n^0 (Q, \theta_c) Z_n^0 (Q, \theta_c)]^{-1} Z_n^0 (Q, \theta_c) R_n^0 (Q, \theta_c). \]

\[ B_n (Q, \theta_c) = \lambda (W_n^0 - W_n^0 (Q, \theta_c)) (I_n + \lambda_0 G_n^0). \]

\( \mu [A] \) 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 5. (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 \( Q_{ji}^j \) 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, \(Q^B_{jk}\) 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 \(\theta^R_g, \theta^R_g\) and \(\theta^B_g\) of compatible dimensions.

\[
P \{ \{W_{j}\}_{ik} = 1\mid Q_j \} = Q^R_{j} \theta^R_g + Q^B_{j} \theta^B_g + Q^R_{jk} \theta^R_g + Q^B_{jk} \theta^B_g. \tag{26}
\]

Because probabilities should stay in the range \([0, 1]\), it is plausible to use, instead, \(P \{ \{W_{j}\}_{ik} = 1\mid Q_j \} = \logit(Q^R_{j} \theta^R_g + Q^R_{jk} \theta^R_g + Q^B_{jk} \theta^B_g)\) 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 \(Q^R_{j}\), \(Q^R_{jk}\) and \(Q^B_{jk}\) 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 \(H_0 : \theta^B_g = 0\), 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 (19) is provided.

Example 6. (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\(^{22}\)) and homogeneous\(^{23}\), and all isomorphic graphs have same probability,

\[
P \{ W_j = w_j \} = \frac{1}{\kappa(\theta_g)} \cdot \exp \left\{ \theta^R_g T(w_j) + \sum_{s=1}^{n-1} \theta^R_g S_s(w_j) \right\} \tag{27}
\]

where \(T(w_j) = \sum_{i,k,l} \{ w_{j} \}_{ik} \{ w_{j} \}_{kl} \) 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 \(\theta_g = (\theta^R_g, \theta^R_g, \ldots, \theta^R_g)\). 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

\[
P \{ W_j = w_j \} = \frac{1}{\kappa(\theta_g)} \cdot \exp \left\{ \theta^R_g Z(w_j) \right\}. \tag{28}
\]

Note that, as a consequence of homogeneity, edges have equal probability of being formed with expected network \(W^*_j(\theta_g) = p_{u v}^n \cdot I_{n_j} - p_{u w}^n\). 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 (12) is much larger than might initially appear.

\(^{22}\)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 \(v\).

\(^{23}\)That is, nodes are a priori indistinguishable.
C  Score Vector and Hessian Matrix.

The likelihood is \( \ln L (\theta | y, x, Q_n) = -\frac{1}{2} \ln (2\pi \sigma^2) + \ln |S_n (Q_n, \theta)| + \ln |R_n (Q_n, \theta)| - \frac{1}{2\sigma^2} e_n^\theta (Q_n, \theta) e_n (Q_n, \theta) \) where \( e_n^\theta (Q_n, \theta) = R_n^\theta (Q_n, \theta) (S_n^\theta (Q_n, \theta) y_n - x_n^\theta - W_n^\theta (Q_n, \theta) x_n \beta) \). First-order derivatives are

\[
\begin{align*}
\frac{\partial \ln L (\theta)}{\partial \lambda} &= -\left[ (S_n (Q_n, \theta))^{-1} W_n (Q_n, \theta) \right] - \frac{1}{2\sigma^2} y_n W_n^\sigma (Q_n, \theta) R_n^\sigma (Q_n, \theta) R_n (Q_n, \theta) W_n (Q_n, \theta) y_n
\end{align*}
\]

and second-order derivatives

\[
\begin{align*}
\frac{\partial^2 \ln L (\theta)}{\partial \lambda \partial \lambda^*} &= -\left[ (S_n (Q_n, \theta))^{-1} W_n (Q_n, \theta) (S_n (Q_n, \theta))^{-1} W_n (Q_n, \theta) \right] - \frac{1}{2\sigma^2} y_n W_n^\sigma (Q_n, \theta) R_n^\sigma (Q_n, \theta) R_n (Q_n, \theta) W_n (Q_n, \theta) y_n
\end{align*}
\]
Proof. if link formation is independent, $D.1$ Useful Lemmas.

Derivatives

$$\frac{\partial^2 \ln \mathcal{L}^\varepsilon(\theta)}{\partial \mu_{i}^2 \partial \mu_{j}^2} = \rho \text{tr} \left[ \left( R_{n}^{\varepsilon}(Q_{n}, \theta) \right)^{-1} \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta) \left( R_{n}^{\varepsilon}(Q_{n}, \theta) \right)^{-1} M_{n}^{\varepsilon}(Q_{n}, \theta) \right] - \text{tr} \left[ \left( R_{n}^{\varepsilon}(Q_{n}, \theta) \right)^{-1} \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta) \right]
$$

$$- \frac{1}{n} \lambda \left( \begin{array}{c} \tau_{0} \tau_{1} \\
\end{array} \right) W_{n}^{\varepsilon}(Q_{n}, \theta) y_{n} - x_{n} \beta_{1} - W_{n}^{\varepsilon}(Q_{n}, \theta) x_{n} \beta_{2} \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta) c_{n}(Q_{n}, \theta)
$$

$$+ \frac{1}{n} \lambda \left( \begin{array}{c} \tau_{0} \tau_{1} \\
\end{array} \right) W_{n}^{\varepsilon}(Q_{n}, \theta) y_{n} - x_{n} \beta_{1} - W_{n}^{\varepsilon}(Q_{n}, \theta) x_{n} \beta_{2} \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta) c_{n}(Q_{n}, \theta)
$$

$$- \rho \frac{1}{n} \lambda \left( \begin{array}{c} \tau_{0} \tau_{1} \\
\end{array} \right) S_{n}^{\varepsilon}(Q_{n}, \theta) y_{n} - x_{n} \beta_{1} - W_{n}^{\varepsilon}(Q_{n}, \theta) x_{n} \beta_{2} \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta) c_{n}(Q_{n}, \theta)
$$

$$- \frac{1}{n} \lambda \left( \begin{array}{c} \tau_{0} \tau_{1} \\
\end{array} \right) \left( S_{n}^{\varepsilon}(Q_{n}, \theta) y_{n} - x_{n} \beta_{1} - W_{n}^{\varepsilon}(Q_{n}, \theta) x_{n} \beta_{2} \right) \nabla_{\theta_{ij}} M_{n}^{\varepsilon}(Q_{n}, \theta)
$$

Similarly for derivatives of $M_{j}^{\varepsilon}(Q_{j}, \theta)$ and model-dependent and so are omitted here.

\section*{D Proofs.}

\subsection*{D.1 Useful Lemmas.}

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

\begin{lemma}
For any $n \times n$ matrix $A_{n}$ with uniformly bounded column sums in absolute value, uniformly bounded $n \times k$ matrix $Z_{n}$, and if $u_{n} \sim N(0, \sigma^{2}I)$ of dimension $n \times 1$, then $\frac{1}{\sqrt{n}} Z_{n}^{\top} A_{n} u_{n} = o_{p}(1)$.
\end{lemma}

\begin{lemma}
The $u_{i}^{n} A_{n} u_{n}$ is $\sigma^{2} \text{tr}(A_{n})$ and $\text{Var}(u_{i}^{n} A_{n} u_{n}) = (\mu_{i} - 3\sigma^{2}) \text{vec}_{D}(A_{n}) \text{vec}_{D}(A_{n}) + \sigma^{4} \left( \text{tr}(A_{n} A_{n}^{\top}) + \text{tr}(A_{n} A_{n}^{\top}) \right)$.
\end{lemma}

\begin{lemma}
Define $A_{n}^{-1} = \left( S_{n}^{0} \right)^{-1} A_{n}^{-1} \left( S_{n}^{0} \right)^{-1}$, $A_{n}^{-1} = \left( S_{n}^{0} \right)^{-1} A_{n}^{-1} \left( S_{n}^{0} \right)^{-1}$, and $A_{n} = \left( S_{n}^{0} \right)^{-1} R_{n}^{c}(Q_{n}, \theta) R_{n}^{c}(Q_{n}, \theta) P_{n}^{c}(Q_{n}, \theta) R_{n}^{c}(Q_{n}, \theta) S_{n}^{0}(Q_{n}, \theta))^{-1}$. Then, for any randomly distributed vector $\epsilon_{n}$ of dimension $n \times 1$ such that $E_{i} \epsilon_{j} = 0$ for $i \neq j$ with $E\epsilon_{i}^{2} < \infty$ and if link formation is independent, $\frac{1}{\sqrt{n}} E(\epsilon_{n} A_{n}^{-1} \epsilon_{n}) = \frac{1}{\sqrt{n}} E(\epsilon_{n} A_{n}^{-1} \epsilon_{n}) + o_{p}(1)$.
\end{lemma}

\begin{proof}
For simplicity, consider $R_{n}^{c}(Q_{n}, \theta_{0}) = R_{n}^{c} = I_{n}$. Proof generalizes immediately otherwise. Then $\frac{1}{n} E \left( \epsilon_{n}^{2} A_{n}^{-1} - (E A_{n}^{-1})^{2} \right) = \frac{1}{n} \sum_{i,j=1}^{n} E \left( \epsilon_{i}^{2} A_{n}^{-1} \right) = \frac{1}{n} \sum_{i,j=1}^{n} E \left( \epsilon_{i}^{2} \right) E \left( A_{n}^{-1} \right) = o_{p}(1)$.
\end{proof}

It follows that $E A_{n} = \lambda_{n} - \lambda_{n} W_{n}(\theta_{0}) \lambda_{n} - \lambda_{n} W_{n}(\theta_{0}) + \lambda_{n}^{2} E W_{n}^{2} A_{n} W_{n}^{2} = \lambda_{n} - \lambda_{n} W_{n}(\theta_{0}) \lambda_{n} - \lambda_{n} W_{n}(\theta_{0}) + \lambda_{n}^{2} W_{n}(\theta_{0}) \lambda_{n} W_{n}(\theta_{0}) = \lambda_{n}$ where the second equality holds only if link formation is independent, i.e., if $E \{ W_{n}^{2} \}_{i,k} \{ W_{n}^{2} \}_{i',k'} = \{ W_{n}^{2} \}_{i,k} \{ W_{n}^{2} \}_{i',k'}$ if either $i \neq i'$ or $k \neq k'$.
\hfill \qed
Lemma 4. Let \( \varepsilon_n \) be a \( n \times 1 \) stationary, ergodic process with \( \mathbb{E}\varepsilon_n = 0 \). Then \( \frac{1}{n} \mathbb{E}(\varepsilon_n' \Lambda_n^{-1} \varepsilon_n) = \frac{1}{n} \mathbb{E}(\varepsilon_n' \Lambda_n^{-1} \varepsilon_n) + o_p(1) \).

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

\[ \square \]

Lemma 5. \( \frac{1}{n} \mathbb{E}\{ g(Z_n) - \mathbb{E}\{ g(Z_n) \}\} \to 0 \).

Proof. Apply Lemma 4 with minor modifications twice. First, note that \( \frac{1}{n} \mathbb{E}(\varepsilon_n' Z_n^{0}) (S_n^{0})^{-1} S_n' (Q, \theta_0) P_n' (Q, \theta_0) R_n (Q, \theta_0) S_n (Q, \theta_0) (S_n^{0})^{-1} Z_n^{0} \beta_0) = o_p(1) \). Second, \( \frac{1}{n} \mathbb{E}(\varepsilon_n' Z_n^{0} R_n (Q, \theta_0) P_n' (Q, \theta_0) R_n (Q, \theta_0) Z_n (Q, \theta_0) \beta_0) = o_p(1) \)

Lemma 6. \( \frac{1}{n} \mathbb{E}(\varepsilon_n' (R_n^{0})^{-1} (S_n^{0})^{-1} S_n' (Q, \theta_0) R_n' (Q, \theta_0) P_n' (Q, \theta_0) R_n' (Q, \theta_0) S_n (Q, \theta_0) (S_n^{0})^{-1} (R_n^{0})^{-1} \varepsilon_n) = \sigma^2 + o_p(1) \).

Proof. Direct consequence of Lemma 4 taken with \( \theta_n = \theta_0^n \).

\[ \square \]

D.2 Derivation of pdf of networks.

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

\[
P(W = w) = \prod_{i<j} \delta_{g} w_{ij} \prod_{i<j} \delta_{A} w_{ij} (1-w_{ij}) + (1-w_{ij}) w_{ji} \prod_{i<j} \delta_{N} (1-w_{ij}) (1-w_{ji})
\]

\[
= \exp \left\{ \ln \delta_{g} \sum_{i<j} w_{ij} w_{ji} + \ln \delta_{A} \sum_{i<j} w_{ij} (1-w_{ij}) + (1-w_{ij}) w_{ji} + \ln \delta_{N} \sum_{i<j} (1-w_{ij}) (1-w_{ji}) \right\}
\]

\[
= \frac{1}{\kappa} \exp \left\{ \theta_{g}^{1} \sum_{i<j} w_{ij} + \theta_{g}^{2} \sum_{i<j} w_{ij} w_{ji} \right\}
\]

where \( \theta_{g}^{1} = \ln \frac{\delta_{g}}{\delta_{A}} \) and \( \theta_{g}^{2} = \delta_{g} / \delta_{A} \) and \( \kappa = \left( \prod_{i<j} \delta_{N} \right)^{-1} \). Introducing dependence on sharing exogenous characteristics, the pdf is

\[
P(W = w | Q = q) = \prod_{i<j} \left( \delta_{g}^{1-q_{ij}} \delta_{A}^{1-q_{ij}} \right)^{w_{ij} w_{ji}} \prod_{i<j} \left( \delta_{N}^{1-q_{ij}} \delta_{A}^{1-q_{ij}} \right)^{(1-w_{ij}) w_{ji} + \ln \delta_{A} (1-w_{ij}) w_{ji} + \ln \delta_{N} \sum_{i<j} (1-w_{ij}) (1-w_{ji})}
\]

\[
= \exp \left\{ \ln \left( \prod_{i<j} \left( \delta_{g}^{1-q_{ij}} \delta_{A}^{1-q_{ij}} \right)^{w_{ij} w_{ji}} \prod_{i<j} \left( \delta_{N}^{1-q_{ij}} \delta_{A}^{1-q_{ij}} \right)^{(1-w_{ij}) w_{ji} + \ln \delta_{A} (1-w_{ij}) w_{ji} + \ln \delta_{N} \sum_{i<j} (1-w_{ij}) (1-w_{ji})} \right) \right\}
\]

\[
= \exp \left\{ \sum_{i<j} w_{ij} w_{ji} (q_{ij} \ln \delta_{1} + (1-q_{ij}) \ln \delta_{2}) + \ln \delta_{A} \sum_{i<j} (1-w_{ij}) w_{ji} (q_{ij} \ln \delta_{1} + (1-q_{ij}) \ln \delta_{2}) \right\}
\]

\[
= \frac{1}{\kappa} \exp \left\{ \theta_{g}^{1} \sum_{i<j} w_{ij} + \theta_{g}^{2} \sum_{i<j} w_{ij} w_{ji} + \theta_{g}^{3} \sum_{i<j} w_{ij} w_{ji} + \theta_{g}^{4} \sum_{i<j} w_{ij} w_{ji} q_{ij} \right\}
\]

40
where \( \theta_1 = \ln \frac{\delta_{1A}}{\delta_{1N}}, \theta_2 = \ln \frac{\delta_{2A}}{\delta_{1N}} \), 
\( \theta_3 = \ln \frac{\delta_{3A}}{\delta_{1N}}, \theta_4 = \ln \frac{\delta_{4A}}{\delta_{1N}} \) 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 \( x_n c_1 + W_n^e (Q_n, \theta_n^e) x_n c_2 + G_n^e (Q_n, \theta_n^e) x_n b_{10} c_3 + G_n^o (Q_n, \theta_n^o) W_n^e (Q_n, \theta_n^e) x_n b_{20} c_3 = 0 \) are \( c_1 = c_2 = c_3 = 0 \). Under the assumption that \( G_n^e (Q_n, \theta_n^e) \equiv W_n^e (Q_n, \theta_n^e) \left( S_n^e (Q_n, \theta_n^e) \right)^{-1} = \left( S_n^e (Q_n, \theta_n^e) \right)^{-1} W_n^e (Q_n, \theta_n^e) \), i.e., assuming symmetry of \( W_n^e (Q_n, \theta_n^e) \), expression is equal to \( x_n c_1 + W_n^e (Q_n, \theta_n^e) x_n c_2 + \left( S_n^e (Q_n, \theta_n^e) \right)^{-1} W_n^e (Q_n, \theta_n^e) x_n b_{10} c_3 + \left( S_n^e (Q_n, \theta_n^e) \right)^{-1} W_n^e (Q_n, \theta_n^e) x_n b_{20} c_3 \), then equivalent to assessing

\[
S_n^e (Q_n, \theta_n^e) x_n c_1 + S_n^e (Q_n, \theta_n^e) W_n^e (Q_n, \theta_n^e) x_n c_2 + W_n^e (Q_n, \theta_n^e) x_n b_{10} c_3 + \left( W_n^e (Q_n, \theta_n^e) \right)^2 x_n b_{20} c_3
\]

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

**D.4 Theorem 1.**

**Proof.** (Uniform Convergence). The goal is to show that the concentrated log-likelihood \( 2 (n)^{-1} \ln L_n (\theta_c) - Q_n (\theta_c) \) converges uniformly to zero on \( \Theta_c \), where \( F_n (\theta_c) = \max_{\theta \in \Theta_c} E \ln L_n^e (\theta_c) \), that is,

\[
\sup_{\theta_c \in \Theta_c} \left| \frac{1}{n} \ln L_n (\theta_c) - \frac{1}{n} F_n (\theta_c) \right| = \sup_{\theta_c \in \Theta_c} \left| \ln \tilde{\sigma}^2 (\theta_c) - \ln \sigma^2 (\theta_c) \right| = o_p (1).
\]

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

\[
P_n^e (Q_n, \theta_n^e) R_n^e (Q_n, \theta_n^e) S_n^e (Q_n, \theta_n^e) y_n = P_n^e (Q_n, \theta_n^e) R_n^e (Q_n, \theta_n^e) Z_n^o \beta_0 + (\lambda_0 - \lambda) P_n^e (Q_n, \theta_n^e) R_n^e (Q_n, \theta_n^e) G_n^0 Z_n^o \beta_0
\]

++ \( P_n^e (Q_n, \theta_n^e) R_n^e (Q_n, \theta_n^e) B_n (Q_n, \theta_n^e) Z_n^o \beta_0 + P_n^e (Q_n, \theta_n^e) R_n^e (Q_n, \theta_n^e) S_n^e (Q_n, \theta_n^e) (S_n^o)^{-1} (R_n^o)^{-1} \epsilon_n \).
Given that \( \sigma^2(Q_n, \theta_c) = \frac{1}{n} \sum_{i=1}^{10} K_i(Q, \theta_g) \), where

\[
K_1(Q_n, \theta_g) = \frac{1}{n} \left[ R_n^\top (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) Z_n \beta_0 \right]
\]

\[
K_2(Q_n, \theta_g) = \frac{2}{n} (\lambda_0 - \lambda) \left[ R_n^\top (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) G_n^\top Z_n \beta_0 \right]
\]

\[
K_3(Q_n, \theta_g) = \frac{2}{n} \left[ R_n^\top (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) B_n (Q_n, \theta_c) Z_n \beta_0 \right]
\]

\[
K_4(Q_n, \theta_g) = \frac{2}{n} \left[ R_n^\top (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) (S_n^\top)^{-1} (R_n^\top)^{-1} \epsilon_n \right]
\]

\[
K_5(Q_n, \theta_g) = \frac{1}{n} (\lambda_0 - \lambda)^2 \left[ R_n^\top (Q_n, \theta_c) G_n^\top Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) G_n^\top Z_n \beta_0 \right]
\]

\[
K_6(Q_n, \theta_g) = \frac{2}{n} (\lambda_0 - \lambda) \left[ R_n^\top (Q_n, \theta_c) G_n^\top Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) B_n (Q_n, \theta_c) Z_n \beta_0 \right]
\]

\[
K_7(Q_n, \theta_g) = \frac{2}{n} (\lambda_0 - \lambda) \left[ R_n^\top (Q_n, \theta_c) G_n^\top Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) (S_n^\top)^{-1} (R_n^\top)^{-1} \epsilon_n \right]
\]

\[
K_8(Q_n, \theta_g) = \frac{1}{n} \left[ R_n^\top (Q_n, \theta_c) B_n (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) B_n (Q_n, \theta_c) Z_n \beta_0 \right]
\]

\[
K_9(Q_n, \theta_g) = \frac{2}{n} \left[ R_n^\top (Q_n, \theta_c) B_n (Q_n, \theta_c) Z_n^\top \beta_0 \right]^\top P_n^\top (Q_n, \theta_c) \left[ R_n (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) (S_n^\top)^{-1} (R_n^\top)^{-1} \epsilon_n \right]
\]

\[
K_{10}(Q_n, \theta_g) = \frac{1}{n} \left[ R_n^\top (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) (S_n^\top)^{-1} (R_n^\top)^{-1} \epsilon_n \right] \left[ R_n^\top (Q_n, \theta_c) P_n^\top (Q_n, \theta_c) R_n (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) (S_n^\top)^{-1} (R_n^\top)^{-1} \epsilon_n \right]
\]

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

\[
\tilde{\beta}(Q_n, \theta_c) = \left[ Z_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) Z_n^\top (Q_n, \theta_c) \right]^{-1} Z_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) R_n (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) \beta_0
\]

\[
\tilde{\sigma}^2(Q_n, \theta_c) = \frac{1}{n} \mathbb{E} \left\{ \left[ S_n^\top (Q_n, \theta_c) y_n - Z_n^\top (Q_n, \theta_c) \tilde{\beta}(Q_n, \theta_c) \right] \left[ R_n^\top (Q_n, \theta_c) P_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) y_n - Z_n^\top (Q_n, \theta_c) \tilde{\beta}(Q_n, \theta_c) \right] \right\}
\]

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

\[
\tilde{\sigma}^2(Q_n, \theta_c) = \frac{1}{n} \mathbb{E} \left\{ y_n S_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) P_n^\top (Q_n, \theta_c) R_n^\top (Q_n, \theta_c) S_n^\top (Q_n, \theta_c) y_n \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \left[ (S_n^\top)^{-1} (R_n)^{-1} \epsilon_n \right]^\top \left[ (S_n^\top)^{-1} Z_n^\top \beta_0 \right] \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \left[ (S_n^\top)^{-1} Z_n^\top \beta_0 \right]^\top \left[ (S_n^\top)^{-1} Z_n^\top \beta_0 \right] \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \left[ (S_n^\top)^{-1} (R_n)^{-1} \epsilon_n \right]^\top \left[ (S_n^\top)^{-1} Z_n^\top \beta_0 \right] \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \left[ (S_n^\top)^{-1} (R_n)^{-1} \epsilon_n \right]^\top \left[ (S_n^\top)^{-1} Z_n^\top \beta_0 \right] \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \beta_0^\top Z_n^\top \left[ (\lambda_0 - \lambda) G_n^\top + I_n + B(Q_n, \theta_c) \right] \right\}
\]

\[
= \frac{1}{n} \mathbb{E} \left\{ \beta_0^\top Z_n^\top \left[ (\lambda_0 - \lambda) G_n^\top + I_n + B(Q_n, \theta_c) \right] \right\}
\]

42
and so $\hat{\sigma}^2(Q, \theta_c) = \sum_{i=1}^{\tau} \hat{K}_i(Q, \theta_c)$ with

$$
\hat{K}_1(Q, \theta_c) = \frac{1}{n} \left\{ \epsilon^j_n \left( R^c_n \right)^{-1} \left( S^q_n \right)^{-1} S^q_n (Q, \theta_c) R^c_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_2(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_3(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_4(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_5(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_6(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

$$
\hat{K}_7(Q, \theta_c) = \frac{1}{n} \left\{ \left( \lambda - \lambda^2 \right) \beta^q_n \gamma^q_n \gamma^q_n (Q, \theta_c) P^c_n (Q, \theta_c) R^c_n (Q, \theta_c) \right\}
$$

By Lemma 2, $\hat{K}_1(Q, \theta_c) = K_{10}(Q, \theta_c) + o_p(1)$. Also, $\hat{K}_2(Q, \theta_c) = K_2(Q, \theta_c) + o_p(1)$, $\hat{K}_3(Q, \theta_c) = K_3(Q, \theta_c) + o_p(1)$, $\hat{K}_4(Q, \theta_c) = K_4(Q, \theta_c) + o_p(1)$, $\hat{K}_5(Q, \theta_c) = K_5(Q, \theta_c) + o_p(1)$, $\hat{K}_6(Q, \theta_c) = K_6(Q, \theta_c) + o_p(1)$, $\hat{K}_7(Q, \theta_c) = K_7(Q, \theta_c) + o_p(1)$ and $\hat{K}_8(Q, \theta_c) = K_8(Q, \theta_c) + o_p(1)$. As a consequence, $\hat{\sigma}^2(Q, \theta_c) = \hat{\sigma}^2(Q, \theta_c) + o_p(1)$ uniformly on $\theta_c$. Convergence is uniform on the parameter space as $\lambda$, $\rho$ and $\theta_c$ appear as polynomial factors.

*Identification for $\lambda = \lambda_0$. Consider the non-stochastic auxiliary model $y_j = \lambda_0 W_j^y (Q, \theta^0_c) y_j + x_j \beta_1 + W_j^y (Q, \theta^0_c) x_j \beta_2 + v_j$ where true neighboring matrices are given by expected network at true parameter values, $W^0_j = W_j^y (Q, \theta^0_c)$ and $M^0_j = M_j^y (Q, \theta^0_c)$. Its likelihood is

$$
\ln \mathcal{L}_n^y (\theta) = -\frac{n}{2} \ln (2\pi \sigma^2) + \ln |S^q_n(Q, \theta)| + \ln |R^c_n(Q, \theta)| - \frac{1}{2\sigma^2} \sum_{j=1}^{\tau} \epsilon^y_j(Q_j, \theta) \epsilon^y_j(Q_j, \theta)
$$

where $\epsilon^y_j(Q_j, \theta) = R^c_j(Q_j, \theta) \left( S^q_j(Q_j, \theta) y_j - x_j \beta_1 - W^y_j(Q_j, \theta) x_j \beta_2 \right)$. As usual, parameters $\beta$ and $\sigma^2$ can be concentrated out of the likelihood.

The concentrators are given by

$$
\hat{\beta}^* (Q, \theta_c) = \left[ Z^q_n(Q, \theta_c) R^c_n(Q, \theta_c) R^c_n(Q, \theta_c) Z^q_n(Q, \theta_c) \right]^{-1} Z^q_n(Q, \theta_c) R^c_n(Q, \theta_c) R^c_n(Q, \theta_c) S^q_n(Q, \theta_c) y_n
$$

$$
\hat{\sigma}^{*2} (Q, \theta_c) = \frac{1}{n} \left[ S^q_n(Q, \theta_c) y_n - Z^q_n(Q, \theta_c) \hat{\beta}(\theta_c) \right] \left[ R^c_n(Q, \theta_c) R^c_n(Q, \theta_c) S^q_n(Q, \theta_c) y_n - Z^q_n(Q, \theta_c) \hat{\beta}(\theta_c) \right]^{-1}
$$

The final form for the concentrated likelihood is $\ln \mathcal{L}_n^y(\theta_c) = -\frac{n}{2} \ln (2\pi) + \frac{1}{2} \ln |S^q_n(Q, \theta)| + \ln |R^c_n(Q, \theta)|$. The problem in expectation $\mathcal{L}_n^{**}(\theta) = \max_{\beta, \sigma^2} \mathbb{E} [\ln \mathcal{L}_n^y(\theta)]$ is $F^*_n(\theta) = -\frac{n}{2} \ln (2\pi) + \ln |S^q_n(Q, \theta)| + \ln |R^c_n(Q, \theta)| - \frac{1}{2} \hat{\sigma}^{*2}(\theta)$, where $\hat{\sigma}^{*2}(Q, \theta_c)$
is given by

\[
\frac{1}{n} \mathbb{E} \left\{ y_n^2 s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) y_n \right\} \\
= \frac{1}{n} \mathbb{E} \left\{ c_n \left( R_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} (r_n^\epsilon (Q_n, \theta_c^0))^{-1} \epsilon_n \right\} \\
+ \frac{1}{n} \mathbb{E} \left\{ \beta_0^e z_n^\epsilon (Q_n, \theta_c^0) \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} z_n^\epsilon (Q_n, \theta_c^0) \beta_0 \right\} \\
= \frac{1}{n} \mathbb{E} \left\{ c_n \left( R_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} (r_n^\epsilon (Q_n, \theta_c^0))^{-1} \epsilon_n \right\} \\
+ \frac{1}{n} \mathbb{E} \left\{ \beta_0^e z_n^\epsilon (Q_n, \theta_c^0) \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} z_n^\epsilon (Q_n, \theta_c^0) \right\}. 
\]

By Jensen’s Inequality, \( Q_n^\ast (\theta) \leq Q_n^\ast (\theta_0) \). Identification in the original model follows from

\[
\frac{1}{n} F_n (\theta_c) - \frac{1}{n} F_n (\theta_0^0) = \frac{1}{n} \left[ F_n^\ast (\theta_c) - F_n^\ast (\theta_0^0) \right] + \frac{1}{2} \left[ \ln \sigma^\ast^2 (\theta_c) - \ln \sigma^\ast^2 (\theta_0^0) \right]. 
\]

It is immediate that \( \sigma^\ast^2 (\theta_0^0) = \sigma_0^2 \). Lemmas 5 and 6 imply that \( \bar{\sigma}^2 (\theta_0^0) = \sigma_0^2 \). Notice also

\[
\bar{\sigma}^2 (\theta_c) = \frac{1}{n} \mathbb{E} \left\{ c_n \left( R_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} (r_n^\epsilon (Q_n, \theta_c^0))^{-1} \epsilon_n \right\} \\
+ \frac{1}{n} \mathbb{E} \left\{ \beta_0^e z_n^\epsilon (Q_n, \theta_c^0) \left( s_n^\epsilon (Q_n, \theta_c^0) \right)^{-1} s_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) p_n^\epsilon (Q_n, \theta_c) r_n^\epsilon (Q_n, \theta_c) s_n^\epsilon (Q_n, \theta_c) (s_n^\epsilon (Q_n, \theta_c^0))^{-1} z_n^\epsilon (Q_n, \theta_c^0) \right\}. 
\]

Finally, Lemma 3 and Assumption 6 imply \( \ln \sigma^\ast^2 (\theta_c) - \ln \bar{\sigma}^2 (\theta_c) < 0 \). This completes the proof. \( \square \)

\section{D.5 Theorem 2.}

\textbf{Proof.} Jacobian and Hessian matrices are given in Appendix C. The asymptotic distribution can be obtained from a Taylor expansion around the point \( \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta} | y_n, x_n, Q_n)}{\partial \hat{\theta}} = 0 \). For a point \( \hat{\theta} \) between \( \hat{\theta} \) and \( \theta_0 \),

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

(Showing \( \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta} | y_n, x_n, Q_n)}{\partial \theta} \right) \xrightarrow{p} \frac{\partial \ln \mathcal{L}^\epsilon (\theta_0 | y_n, x_n, Q_n)}{\partial \theta} \). Convergence is shown explicitly for three terms: \( \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta})}{\partial \lambda_0^3} \), \( \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta})}{\partial \lambda_0^2 \lambda_1} \) and \( \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta})}{\partial \lambda_1^3} \); other terms can be shown with little or no modifications. For

\[
\frac{1}{n} \left\{ \frac{\partial \ln \mathcal{L}^\epsilon (\hat{\theta})}{\partial \lambda_0^3} - \frac{\partial \ln \mathcal{L}^\epsilon (\theta_0)}{\partial \lambda_0^3} \right\} = \frac{1}{n \sigma_0^2} y_n^2 w_n^\epsilon (Q_n, \theta_0) r_n^\epsilon (Q_n, \theta_0) x_n - \frac{1}{n \sigma_0^2} y_n^2 w_n^\epsilon (Q_n, \hat{\theta}) r_n^\epsilon (Q_n, \hat{\theta}) x_n \\
= \frac{1}{n} \left[ \frac{1}{n \sigma_0^2} - \frac{1}{n \sigma_0^2} \right] y_n^2 w_n^\epsilon (Q_n, \theta_0) r_n^\epsilon (Q_n, \theta_0) x_n + \frac{1}{n \sigma_0^2} y_n^2 \left[ w_n^\epsilon (Q_n, \theta_0) r_n^\epsilon (Q_n, \theta_0) - w_n^\epsilon (Q_n, \hat{\theta}) r_n^\epsilon (Q_n, \hat{\theta}) \right] x_n. 
\]
The argument follows by noticing \( W_n^* (Q_n, \theta_0) \) and \( R_n^* (Q_n, \theta_0) \) are row and column-sum bounded, so \( \frac{1}{n} y_n^* W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) x = O_p (1) \), while by continuity of the inverse, \( \left[ \frac{\partial \ln \mathcal{L}^* (\hat{\theta})}{\partial \lambda} - \frac{\partial \ln \mathcal{L}^* (\theta_0)}{\partial \lambda} \right] = o_p (1) \). The second term converges in probability as \( \frac{1}{n \sigma^2} y_n^* [W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) - W_n^* (Q_n, \hat{\theta}) R_n^* (Q_n, \hat{\theta})] x_n = \frac{1}{n \sigma^2} \beta_n^* Z_n^* \left( S_n^* \right)^{-1} \left[ W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) - W_n^* (Q_n, \hat{\theta}) R_n^* (Q_n, \hat{\theta}) \right] x + o_p (1) \). Given that \( Z_n^* = [x_n; W_n^* x_n] \), \( x_n \) is non-stochastic, \( W_n^* \) is row and column-sum bounded, and \( \left[ W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) - W_n^* (Q_n, \hat{\theta}) R_n^* (Q_n, \hat{\theta}) \right] = o_p (1) \), it has been shown that \( \frac{1}{n} \left\{ \frac{\partial \ln \mathcal{L}^* (\hat{\theta})}{\partial \lambda} - \frac{\partial \ln \mathcal{L}^* (\theta_0)}{\partial \lambda} \right\} = o_p (1) \). The next term is

\[
\frac{1}{n} \left\{ \frac{\partial \ln \mathcal{L}^* (\hat{\theta})}{\partial \lambda^2} - \frac{\partial \ln \mathcal{L}^* (\theta_0)}{\partial \lambda^2} \right\} = \frac{1}{n \sigma^2} y_n^* W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) \epsilon_n^* (Q_n, \theta_0) - \frac{1}{n \sigma^2} y_n^* W_n^* (Q_n, \hat{\theta}) R_n^* (Q_n, \hat{\theta}) \epsilon_n^* (Q_n, \hat{\theta})
\]

\[
= \frac{1}{n \sigma^2} y_n^* W_n^* (Q_n, \theta_0) R_n^* (Q_n, \theta_0) \epsilon_n^* (Q_n, \theta_0) + \frac{1}{n \sigma^2} y_n^* W_n^* (Q_n, \hat{\theta}) R_n^* (Q_n, \hat{\theta}) \epsilon_n^* (Q_n, \hat{\theta})
\]

as \( \epsilon_n^* (Q_n, \hat{\theta}) = R_n^* (Q_n, \hat{\theta}) [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2] - R_n^* (Q_n, \theta_0) [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \theta_0) x_n \tilde{b}_2] + \epsilon_n^* (Q_n, \theta_0)
\]

\[
= R_n^* (Q_n, \hat{\theta}) [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2] - R_n^* (Q_n, \theta_0) [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \theta_0) x_n \tilde{b}_2] + [R_n^* (Q_n, \hat{\theta}) - R_n^* (Q_n, \theta_0)] [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2] + [R_n^* (Q_n, \hat{\theta}) - R_n^* (Q_n, \theta_0)] [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2] + [R_n^* (Q_n, \hat{\theta}) - R_n^* (Q_n, \theta_0)] [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2]
\]

\[
= [R_n^* (Q_n, \hat{\theta}) - R_n^* (Q_n, \theta_0)] [S_n^* (Q_n, \theta_0) y_n - x_n \tilde{b}_1 - W_n^* (Q_n, \hat{\theta}) x_n \tilde{b}_2] + \epsilon_n^* (Q_n, \theta_0) - \epsilon_n^* (Q_n, \hat{\theta}) + \epsilon_n^* (Q_n, \theta_0) - \epsilon_n^* (Q_n, \hat{\theta}) + \epsilon_n^* (Q_n, \theta_0) - \epsilon_n^* (Q_n, \hat{\theta}) + \epsilon_n^* (Q_n, \theta_0) - \epsilon_n^* (Q_n, \hat{\theta}) + \epsilon_n^* (Q_n, \theta_0) - \epsilon_n^* (Q_n, \hat{\theta})
\]

By Mean Value Theorem, defining \( G_j (\lambda, \theta_0) = (S_j^* (Q_j, \theta_0))^{-1} W_j^* (Q_j, \theta_0) \), \( G_n (\hat{\lambda}, \hat{\theta}_0) \) then

\[
\frac{1}{n} \left\{ \frac{\partial \ln \mathcal{L}^* (\hat{\theta})}{\partial \lambda^2} - \frac{\partial \ln \mathcal{L}^* (\theta_0)}{\partial \lambda^2} \right\} = 2 \left\{ G_n (\hat{\lambda}, \hat{\theta}_0) \right\} \left\{ \frac{1}{n \sigma^2} y_n^* W_n^* (\hat{\lambda}, \hat{\theta}_0) \epsilon_n^* (\hat{\lambda}, \hat{\theta}_0) \right\} \left\{ \frac{1}{n \sigma^2} y_n^* W_n^* (\hat{\lambda}, \hat{\theta}_0) \epsilon_n^* (\hat{\lambda}, \hat{\theta}_0) \right\} = 0_p (1) \]

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

(Showing \( \frac{1}{n} \frac{\partial \ln \mathcal{L}^* (\hat{\theta})}{\partial \lambda^2} \left|_{\theta_0} \right| \left| x_n ; Q_n \right| \left. \right| E \left( \frac{1}{n} \frac{\partial \ln \mathcal{L}^* (\theta_0)}{\partial \lambda^2} \left|_{\theta_0} \right| x_n ; Q_n \right) \right) \). Terms that generically fit into the format \( \omega_\lambda (\theta) = \frac{1}{n} \varphi^* \Delta (\varphi) \), where \( \varphi \) is non-stochastic vector of dimension \( n \) and \( \Delta \) is a stochastic matrix of conformable dimension can be shown to \( \mathbb{V} \{ \omega (\theta) \} \rightarrow 0 \). For example, \( -\frac{1}{n} \frac{\partial^2 \ln \mathcal{L}^* (\theta_0)}{\partial \lambda \partial \alpha} \left|_{\theta_0} \right| = \frac{1}{n} x_n^* R_n^* (Q_n, \theta_0) W_n^* (Q_n, \theta_0) y = \frac{1}{n} x_n^* R_n^* (Q_n, \theta_0) W_n^* (Q_n, \theta_0) \left( S_n^* \right)^{-1} Z_n^* \beta_0 + \left( S_n^* \right)^{-1} (R_n^* \beta_n^*) x_n^* R_n^* (Q_n, \theta_0) \)
\[ W_n^e (Q_n, \theta_0) (S_n^0)^{-1} \tilde{x}_n, \gamma_1 + \frac{1}{2} x_i^\top R_n^e (Q_n, \theta_0) W_n^e (Q_n, \theta_0) (S_n^0)^{-1} W_n^0 \tilde{x}_n, \gamma_2 + o_p (1) \]

Defining \( x_n^{(l)} \) as the \( l \)-th column of \( x_n \),

\[
\omega_2 (\theta) \equiv \frac{1}{\sigma_n^2} x_n^{(l)\top} R_n^e (Q_n, \theta_0) W_n^e (Q_n, \theta_0) (S_n^0)^{-1} x_n^{(l)} = \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_n^{(l)i} x_n^{(l)j} \left( R_n^e (Q_n, \theta_0) W_n^e (Q_n, \theta_0) (S_n^0)^{-1} \right)_{ij}.
\]

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

\[
\mathcal{V} (\gamma_1) = \left[ \frac{1}{\sigma_n^2} \right] \sum_{i=1}^{n} \sum_{j=1}^{n} \left( x_n^{(l)i} x_n^{(l)j} \right)^2 \mathcal{V} \left\{ (R_n^e (Q_n, \theta_0) W_n^e (Q_n, \theta_0) (S_n^0)^{-1} \right\}.
\]

Noticing \( (W_n^0) \) is a matrix of constants, \( R_n^e (Q_n, \theta_0) W_n^e (Q_n, \theta_0) \) is column and row-sum bounded, then \( \mathcal{V} \{ \cdot \} \) goes to zero and so does \( \mathcal{V} (\gamma_1) \).

An equivalent argument goes through if terms in the middle contains matrix of derivatives. Terms that generically fit into \( \omega_2 (\theta) = \frac{1}{\sigma_n^2} \epsilon_n \Delta (\theta) \epsilon_n \), for example,

\[
- \frac{\sigma_n^2}{n} \frac{\partial^2 \ln L (\theta)}{\partial \epsilon \partial \sigma^2} = \frac{1}{\sigma_n^2} \epsilon_n^\top W_n^e (Q_n, \theta) R_n^e (Q_n, \theta) \epsilon_n (Q_n, \theta) = \frac{1}{\sigma_n^2} \epsilon_n^\top (S_n^0)^{-1} W_n^e (Q_n, \theta) R_n^e (Q_n, \theta) R_n^e (Q_n, \theta) \left( (S_n^0 (Q_n, \theta))^{-1} \right) y_n - Z_n^e (Q_n, \theta) \beta = \frac{1}{\sigma_n^2} \epsilon_n^\top (S_n^0)^{-1} W_n^e (Q_n, \theta) R_n^e (Q_n, \theta) R_n^e (Q_n, \theta) \left( (S_n^0 (Q_n, \theta))^{-1} \right) y_n + o_p (1) = \frac{1}{\sigma_n^2} \epsilon_n^\top (S_n^0)^{-1} W_n^e (Q_n, \theta) R_n^e (Q_n, \theta) R_n^e (Q_n, \theta) \left( (S_n^0 (Q_n, \theta))^{-1} \right) \epsilon_n + o_p (1)
\]

by Lemma 1, and straightforward adaptation of Lemma 3, converges to

\[
\mathbb{E} \left\{ \frac{\sigma_n^2}{n} \frac{\partial^2 \ln L (\theta)}{\partial \epsilon \partial \sigma^2} \right\} = \frac{1}{n} \text{tr} \left\{ \mathbb{E} \left\{ (S_n^0)^{-1} W_n^e (Q_n, \theta) R_n^e (Q_n, \theta) R_n^e (Q_n, \theta) (S_n^0 (Q_n, \theta))^{-1} (S_n^0)^{-1} \right\} \right\}.
\]

(Asymptotic distribution). Given existence of higher order moments of \( \epsilon_n \), the Central Limit Theorem in Kelejian and Prucha (2001) can be applied to show that \( \frac{1}{\sqrt{n}} \frac{\partial \ln L (\theta_0)}{\partial \theta} \xrightarrow{d} N (0, \Sigma_{\theta}^{-1}) \). 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,

\[
\varphi (x_n; W_n^e (Q_n, \theta_0^l), \beta_10, \beta_2) = \sum_{j=1}^{\infty} \lambda_{j+1}^{-1} (W_n^e (Q_n, \theta_0^l))^j x_n (\lambda_+ \beta_10 + \beta_2) = \sum_{j=1}^{\infty} \lambda_0 \lambda_{j+1}^{-1} (W_n^e (Q_n, \theta_0^l))^j x_n (\lambda_0 \beta_10 + \lambda_0 \lambda_{j+1}^{-1} \beta_2) = \sum_{j=1}^{\infty} \lambda_0^{-1} (W_n^e (Q_n, \theta_0^l))^j x_n (\lambda_0 \beta_10 + \beta_2) = \varphi (x_n; W_n^e (Q_n, \theta_0^l), \lambda_0, \beta_10, \beta_2) \tag{29}
\]

46
where the last equality follows by $W_n^c(Q_n, \theta_n^0) x_n | \beta_{21} - \beta_{22} = 0$. (ii). Define $\Phi^* (\theta | y_n, x_n, Q_n) = \{ \tilde{\theta} \in \Theta : Q_n(\tilde{\theta}) = Q_n(\theta) \}$. Sets $\Phi (\theta^0 | y_n, x_n) = \Phi^* (\theta^0 | y_n, x_n)$, as I now show. Inclusion $\Phi (\theta^0 | y_n, x_n) \subseteq \Phi^* (\theta^0 | y_n, x_n)$ is immediate from the first part. The reverse $\Phi^* (\theta^0 | y_n, x_n) \subseteq \Phi (\theta^0 | y_n, x_n)$ follows from a contradiction: suppose there is a $\theta^* \in \Phi^* (\theta^0 | y_n, x_n)$ and $\theta^* \notin \Phi^* (\theta^0 | y_n, x_n)$ by construction and Jensen’s inequality, $Q_n(\theta^*) < Q_n(\theta^0)$. Observation of the reduced-form implies $S_n^e(\theta_n^0) = e^0(\theta_n^0)$, $|S_n^e(\theta_n^0)| = |S_n^e(\theta^0)|$ and $|R_n^e(\theta_n^0)| = |R_n^e(\theta^0)|$, and so $Q_n(\theta^*) = Q_n(\theta^0)$, a contradiction. Therefore, given that $\Phi (\theta^0 | y_n, x_n) = \Phi^* (\theta^0 | y_n, x_n)$, for any $\theta_e \in \Phi^* (\theta^0 | y_n, x_n)$, and by definition, $\Phi^* (\theta^0 | y_n, x_n) = \Theta_0$, the result is proven.

\section{Theorem 3.}

\begin{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 $\Theta$. Parts (3) and (4) are immediate corollaries.
\end{proof}

\section{Example 4.}

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

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

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

$$\nabla y_j = \mathbb{E} \left\{ \left. \left( (S_j^0)^{-1} - \mathbb{E}(S_j^0)^{-1} \right) x_j \beta_{10} \right| y_j \right\} + \mathbb{E} \left\{ \left. \left( (S_j^0)^{-1} W_j^0 - \mathbb{E}(S_j^0)^{-1} W_j^0 \right) x_j \beta_{20} \right| y_j \right\} + \mathbb{E} \left\{ \left. \left( (S_j^0)^{-1} \epsilon_j \right) \right| y_j \right\}.$$
element of $W_j^0$. This simplifies term $A_j$ to

$$A_j = \lambda^2 \begin{bmatrix} \sum_i V \{w_{1i}\} x_{i1}^{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i V \{w_{ni}\} x_{ni}^{11} \end{bmatrix}$$

which then implies $A_j = \text{diag} \left( \lambda^2 V \{W_j\} \text{diag}(x_j^{11}) \right)$. Proceeding in a similar fashion, $B_j = s_j x_j^{12} s_j^*$ with $x_j^{12} = x_j \beta_{10} \beta_{20} x_j'$ and $s_j^* = W_j^0 + \lambda_0 (W_j^0)^2 + \lambda_0^2 (W_j^0)^3 + \cdots$. The second equality uses independence between Bernoulli trials. For $C_j$,

$$C_j = \begin{bmatrix} \sum_i E \{s_i^2 \} x_{i1}^{22} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i E \{s_i^2 \} x_{ni}^{22} \end{bmatrix} = \text{diag} (V \{W_j\} \text{diag}(x_j^{22}))$$

Lastly,

$$D_j = \begin{bmatrix} \sum_{i,j} E \{s_i s_j\} E \{e_{ij}\} & \cdots & \sum_{i,j} E \{s_i s_j\} E \{e_{ij}\} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} E \{s_i s_j\} E \{e_{ij}\} & \cdots & \sum_{i,j} E \{s_i s_j\} E \{e_{ij}\} \end{bmatrix} = \begin{bmatrix} \sum_i E \{s_i^2 \} \sigma^2 & \cdots & \sum_i E \{s_i s_i\} \sigma^2 \\ \vdots & \ddots & \vdots \\ \sum_i E \{s_i s_i\} \sigma^2 & \cdots & \sum_i E \{s_i^2 \} \sigma^2 \end{bmatrix} = \lambda^2 \sigma^2 \begin{bmatrix} \sum_i V \{w_{1i}\} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i V \{w_{ni}\} \end{bmatrix} + \sigma^2 I_{n_j}$$

The entire expression reads $\forall y_j = \text{diag} \left( V \{W_j\} \left( \lambda^2 \text{diag}(x_j^{11}) + 2\lambda \text{diag}(x_j^{12}) + \lambda^2 \sigma^2 I_{n_j} \right) \right)$. 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
\( \beta_{10} \) and \( \sigma_0^2 \) 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

\[
\begin{align*}
J_{K1}(\theta) &= 2n_j^{-1} \delta_1 (1 - \delta_1) \lambda \left( \lambda^2 \lambda' \frac{\partial \text{diag}(x_{ij}^1)}{\partial \beta_1} + 2 \lambda \lambda' \frac{\partial \text{diag}(x_{ij}^2)}{\partial \beta_2} \right) \\
J_{K2}(\theta) &= n_j^{-1} \delta_1 (1 - \delta_1) \left( \lambda^2 \lambda' \frac{\partial \text{diag}(x_{ij}^1)}{\partial \beta_2} + 2 \lambda \lambda' \frac{\partial \text{diag}(x_{ij}^2)}{\partial \beta_2} \right) \\
J_{K3}(\theta) &= n_j^{-1} \delta_1 (1 - \delta_1) \left( 2 \lambda \lambda' \frac{\partial \text{diag}(x_{ij}^1)}{\partial \beta_1} + \lambda \lambda' \frac{\partial \text{diag}(x_{ij}^2)}{\partial \beta_1} \right) \\
J_{K4}(\theta) &= n_j^{-1} (1 - 2\delta_1) \left( \lambda^2 \lambda' \text{diag}(x_{ij}^1) + 2 \lambda \lambda' \text{diag}(x_{ij}^2) + \lambda \lambda' \text{diag}(x_{ij}^2) + n_j \lambda^2 \sigma^2 \right) \\
J_{K5}(\theta) &= \delta_1 (1 - \delta_1) - n_j \lambda^2 + 1.
\end{align*}
\]

Identification is guaranteed with rank \((\mathcal{J}(\theta)) = K\), where \(K\) is the number of parameters in the structural model. Given \(\sigma_0^2\) is identified, the last equation gives a solution for \(\delta_1\) and \(\lambda\). 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 \(c_1 J_{K1}(\theta) + c_3 J_{K3}(\theta) + c_4 J_{K4}(\theta) = 0\). Substituting out \(c_1\) and \(c_3\) in the third equation, one obtains the condition that \(c_4 [-\lambda \delta^{-1} J_{K1}(\theta) - \lambda \delta^{-1} \beta J_{K3}(\theta) + J_{K4}(\theta)] = 0\). If \(\lambda \neq 0\), it is equivalent to \(-\lambda \delta^{-1} J_{K1}(\theta) - \lambda \delta^{-1} \beta J_{K3}(\theta) + J_{K4}(\theta) \neq 0\) at \(\theta_0\). 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 \(\Theta_0\) in the Hausdorff metric, \(\hat{\Theta} \subseteq \Theta_0^c\) for \(\Theta_0^c = \{\theta \in \Theta : d(\theta, \Theta_0) \leq \epsilon\}\) with \(\epsilon = o(1)\) and \(\epsilon \geq 0\). It follows that

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

When \(S\) and \(v\) are going to infinity,

\[
v^{-2} \left( \sum_{j=1}^{v} S^{-1} \sum_{s=1}^{S} q_{s,j}(y, \theta) \right)^{\prime} \Omega \left( \sum_{j=1}^{v} S^{-1} \sum_{s=1}^{S} q_{s,j}(y, \theta) \right) \overset{a.s.}{\rightarrow} \left( E_y^{\theta} W_c q_{s,j}(y, \theta) \right)^{\prime} \Omega \left( E_y^{\theta} W_c q_{s,j}(y, \theta) \right)
\]

49
where $\mathbb{E}_{W,e}$ is the conditional expectation taken with respect to the distribution of $W$ and $e$, given $y$ and $x$ and $\mathbb{E}_y^n$ is the expectation with respect to the true distribution of $y$, given $x$. Given that $(\mathbb{E}_y^n\mathbb{E}_{W,e} q_{s,j} (y, \theta))^t \Omega \left( \mathbb{E}_y^n\mathbb{E}_{W,e} q_{s,j} (y, \theta) \right) = (\mathbb{E}_y^n q_j (y, \theta))^t \Omega \left( \mathbb{E}_y^n q_j (y, \theta) \right)$ and $\mathbb{E}_y^n q_j (y, \theta) = 0$ only at $\theta_0$, 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). $\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Sigma^*)$, where $\Sigma_n = (G_n \Omega_n G_n)^{-1} G_n \Omega_n O_n \gamma_n G_n (G_n \Omega_n G_n)^{-1}$, $G_n = \mathbb{E}q_j (y_n, \theta_0)$, $O_n = \mathbb{E}q_j (y_n, \theta_0)q_j (y_n, \theta_0)^t$ and $\Sigma = \lim_{n \to \infty} \Sigma_n$. Optimal weight matrix is $\Omega_n^* = O_n^{-1}$ and, in this case, $\Sigma_n^* = (G_n (\Omega_n^*)^{-1} G_n)^{-1}$ 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

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

where $q_{s,j}^* (y, \theta) = [\nabla \theta \ln \mathcal{L}^c (\theta) \ q_{s,j} (y, \theta)]^t$ and $\Omega^*$ is a weight matrix of conformable 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, $\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Sigma^{**})$, where $\Sigma_n^* = (G_n^* \Omega_n^* G_n^*)^{-1} G_n^* \Omega_n^* O_n^* \gamma_n^* G_n^* (G_n^* \Omega_n^* G_n^*)^{-1}$, $G^* = \mathbb{E}q_j^* (y, \theta_0)$, $O^* = \mathbb{E}q_j^* (y, \theta_0)q_j^* (y, \theta_0)^t$ and $q_j^* (y, \theta_0) = \lim_{S \to \infty} S^{-1} \sum_{s=1}^{S} q_{s,j}^* (y, \theta_0)$ and $\Sigma^* = \lim_{n \to \infty} \Sigma_n^{**}$. Using optimal matrix $\Omega_n^{**} = (O_n^*)^{-1}$, $\Sigma_n^{**} = (G_n^* (\Omega_n^{**})^{-1} G_n^*)^{-1}$, $\lim_{n \to \infty} \Sigma_n^{**}$. \hfill $\Box$
E Additional figures and tables.

E.1 Estimator and simulations.

Table 7: Likelihood as a function of $\beta_1$.

<table>
<thead>
<tr>
<th>N = 250</th>
<th>N = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.25$</td>
<td>$0.5$</td>
</tr>
<tr>
<td>$0.75$</td>
<td>$1$</td>
</tr>
<tr>
<td>$1.25$</td>
<td>$1.5$</td>
</tr>
<tr>
<td>$1.75$</td>
<td>$W_0$</td>
</tr>
</tbody>
</table>

Note: Rescaled additive inverse of likelihood as a function of $\beta_1$, 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 $\delta_1$.

<table>
<thead>
<tr>
<th>N = 250</th>
<th>N = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.05$</td>
<td>$0.15$</td>
</tr>
<tr>
<td>$0.25$</td>
<td>$0.35$</td>
</tr>
<tr>
<td>$0.45$</td>
<td>$0.55$</td>
</tr>
<tr>
<td>$0.65$</td>
<td>$0.75$</td>
</tr>
<tr>
<td>$0.85$</td>
<td>$0.95$</td>
</tr>
</tbody>
</table>

Note: Rescaled additive inverse of likelihood as a function of $\delta_1$, 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 $W^0 = W^0 (\theta_0)$ is the likelihood where true network is equal to expected network.
Table 9: Simulations: Bernoulli model under misspecified $\lambda_{ref}$.

<table>
<thead>
<tr>
<th></th>
<th>$T = 1$</th>
<th>$T = 5$, FE</th>
<th>$T = 5$, TE</th>
<th>$T = 5$, FE and TE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$n$</td>
<td>25</td>
<td>100</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>$v$</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>0.1002</td>
<td>0.0094</td>
<td>0.0092</td>
<td>0.0094</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.000</td>
<td>0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>1.0016</td>
<td>1.0009</td>
<td>0.9996</td>
<td>1.0008</td>
</tr>
<tr>
<td></td>
<td>0.009</td>
<td>0.005</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.0892</td>
<td>0.0817</td>
<td>0.0793</td>
<td>0.0826</td>
</tr>
<tr>
<td></td>
<td>0.009</td>
<td>0.005</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>$\hat{\delta}_1$</td>
<td>0.3766</td>
<td>0.3722</td>
<td>0.3606</td>
<td>0.3764</td>
</tr>
<tr>
<td></td>
<td>0.095</td>
<td>0.014</td>
<td>0.050</td>
<td>0.005</td>
</tr>
<tr>
<td>$\hat{\delta}_0$</td>
<td>0.1427</td>
<td>0.1496</td>
<td>0.1499</td>
<td>0.1475</td>
</tr>
<tr>
<td></td>
<td>0.033</td>
<td>0.007</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>1.0550</td>
<td>1.2193</td>
<td>0.8458</td>
<td>0.9767</td>
</tr>
<tr>
<td></td>
<td>0.037</td>
<td>0.007</td>
<td>0.019</td>
<td>0.003</td>
</tr>
<tr>
<td>$\varphi(x, \hat{\theta})$</td>
<td>(0.018)</td>
<td>(0.085)</td>
<td>(0.019)</td>
<td>(0.030)</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
<td>0.001</td>
</tr>
</tbody>
</table>

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 10: Simulations: multivariate network model.

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

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


### E.2 Application.

Table 11: Occupational Choice.

<table>
<thead>
<tr>
<th>Method</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program effect</td>
<td>473.219***</td>
<td>473.581***</td>
<td>–113.002***</td>
<td>–113.146***</td>
<td>0.113***</td>
<td>0.114***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\beta}_{11}$).</td>
<td>(12.99)</td>
<td>(13.89)</td>
<td>(8.33)</td>
<td>(8.33)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Program effect</td>
<td>464.069***</td>
<td>463.441***</td>
<td>–142.755***</td>
<td>–143.009***</td>
<td>0.120***</td>
<td>0.121***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\beta}_{12}$).</td>
<td>(13.07)</td>
<td>(5.10)</td>
<td>(8.53)</td>
<td>(8.25)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Link probability</td>
<td>–20.438***</td>
<td>–23.097***</td>
<td>24.394***</td>
<td>26.933***</td>
<td>–0.029***</td>
<td>–0.034***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\varphi}_{T,2}$).</td>
<td>(7.01)</td>
<td>(6.95)</td>
<td>(8.50)</td>
<td>(9.21)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on T</td>
<td>17.396***</td>
<td>14.734**</td>
<td>19.805**</td>
<td>22.105**</td>
<td>–0.023***</td>
<td>–0.027**</td>
</tr>
<tr>
<td>after 4 years ($\hat{\varphi}_{T,4}$).</td>
<td>(6.41)</td>
<td>(7.04)</td>
<td>(8.37)</td>
<td>(10.30)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>–9.771***</td>
<td>–11.346***</td>
<td>12.692***</td>
<td>14.259***</td>
<td>–0.015***</td>
<td>–0.018***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\varphi}_{NT,2}$).</td>
<td>(3.35)</td>
<td>(3.42)</td>
<td>(4.41)</td>
<td>(4.87)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>8.317**</td>
<td>7.237*****</td>
<td>10.304**</td>
<td>11.703</td>
<td>–0.012***</td>
<td>–0.014***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\varphi}_{NT,4}$).</td>
<td>(3.28)</td>
<td>(1.88)</td>
<td>(5.21)</td>
<td>(13.28)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Family effect</td>
<td>0.113***</td>
<td>0.114***</td>
<td>0.120***</td>
<td>0.121***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.115***</td>
<td>0.116***</td>
</tr>
<tr>
<td>if $Q_{ij} = 1$ ($\hat{\delta}_1$).</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Link probability</td>
<td>0.317***</td>
<td>0.464***</td>
<td>0.364***</td>
<td>0.362***</td>
<td>0.115***</td>
<td>0.116***</td>
</tr>
<tr>
<td>if $Q_{ij} = 0$ ($\hat{\delta}_0$).</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.075</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>p-value $H_{NV}$</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Avg treated outcome.</td>
<td>421.8</td>
<td>421.8</td>
<td>646.7</td>
<td>646.7</td>
<td>0.303</td>
<td>646.7</td>
</tr>
<tr>
<td>Individuals ($n$).</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
</tr>
<tr>
<td>Villages ($v$).</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
</tr>
<tr>
<td>Survey waves ($T$).</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 12: Earnings and Seasonality.

<table>
<thead>
<tr>
<th>Outcome Method</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program effect after 2 years ((\hat{\beta}_{11}))</td>
<td>0.562***</td>
<td>0.556***</td>
<td>-0.029***</td>
<td>-0.029***</td>
<td>0.181***</td>
<td>0.182***</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.148)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Program effect after 4 years ((\hat{\beta}_{12}))</td>
<td>2.726***</td>
<td>2.806***</td>
<td>-0.075***</td>
<td>-0.075***</td>
<td>0.166***</td>
<td>0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.196)</td>
<td>(0.108)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on T after 2 years ((\hat{\varphi}_{T,2}))</td>
<td>-0.258**</td>
<td>-0.187*</td>
<td>-0.063***</td>
<td>-0.062***</td>
<td>0.037***</td>
<td>0.030***</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.113)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on T after 4 years ((\hat{\varphi}_{T,4}))</td>
<td>-0.098</td>
<td>-0.188*</td>
<td>-0.016*</td>
<td>-0.016</td>
<td>0.044***</td>
<td>0.035***</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.112)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on NT after 2 years ((\hat{\varphi}_{NT,2}))</td>
<td>-0.133**</td>
<td>-0.102*</td>
<td>-0.026***</td>
<td>-0.026***</td>
<td>0.017***</td>
<td>0.014***</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.062)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Spillover on NT after 4 years ((\hat{\varphi}_{NT,4}))</td>
<td>-0.051</td>
<td>-0.103</td>
<td>-0.002</td>
<td>-0.007</td>
<td>0.020**</td>
<td>0.017**</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.78)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

Function of \(\lambda\):

| Link to T after 2 years (\(\hat{\beta}_{21}\)) | -0.236 | -0.245 | -0.023*** | -0.017*** | -0.051*** | -0.053*** |
|                | (0.456) | (0.478) | (0.02) | (0.00) | (0.01) | (0.01) |
| Link to T after 4 years (\(\hat{\beta}_{22}\)) | -0.740 | -0.375 | -0.019*** | -0.014*** | -0.037*** | -0.039*** |
|                | (0.541) | (0.596) | (0.01) | (0.00) | (0.01) | (0.01) |
| Link probability if \(Q_{ij} = 1\) (\(\hat{\delta}_1\)) | 0.155*** | 0.064*** | 0.234*** | 0.236*** | 0.100*** | 0.077*** |
|                | (0.01) | (0.00) | (0.02) | (0.01) | (0.01) | (0.00) |
| Link probability if \(Q_{ij} = 0\) (\(\hat{\delta}_0\)) | 0.030*** | 0.030*** | 0.152*** | 0.203*** | 0.054*** | 0.051*** |
|                | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) |
| \(p\)-value \(H_{NV}\) | < 0.001 | < 0.001 | < 0.001 | 0.022 | < 0.001 | < 0.001 |

<p>| Avg treated outcome. | 4.607 | 4.607 | 0.674 | 0.674 | 0.478 | 0.478 |
| Individuals ((n)). | 69087 | 69087 | 69087 | 69087 | 69087 | 69087 |
| Villages ((v)). | 1409 | 1409 | 1409 | 1409 | 1409 | 1409 |</p>
<table>
<thead>
<tr>
<th>Survey waves ((T)).</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program effect</td>
<td>1.132**</td>
<td>1.132**</td>
<td>2.116**</td>
<td>2.117**</td>
<td>10.412***</td>
<td>10.420***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\beta}_{11}$).</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(365.41)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Program effect</td>
<td>1.103***</td>
<td>1.101***</td>
<td>1.296***</td>
<td>1.330***</td>
<td>11.175***</td>
<td>11.173***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\beta}_{12}$).</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(459.21)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Spillover on T</td>
<td>-0.032***</td>
<td>-0.033***</td>
<td>0.039</td>
<td>0.107</td>
<td>-0.184***</td>
<td>-0.230***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\phi}_{T,2}$).</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.11)</td>
<td>(0.18)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Spillover on T</td>
<td>-0.055***</td>
<td>-0.055***</td>
<td>0.029</td>
<td>-0.095</td>
<td>-0.407***</td>
<td>-0.456***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\phi}_{T,4}$).</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.12)</td>
<td>(0.21)</td>
<td>(0.11)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>-0.018***</td>
<td>-0.020***</td>
<td>0.014</td>
<td>0.064</td>
<td>-0.106***</td>
<td>-0.137***</td>
</tr>
<tr>
<td>after 2 years ($\hat{\phi}_{NT,2}$).</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.06)</td>
<td>(0.11)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Spillover on NT</td>
<td>-0.031***</td>
<td>-0.032***</td>
<td>0.011</td>
<td>-0.056</td>
<td>-0.234***</td>
<td>-0.272***</td>
</tr>
<tr>
<td>after 4 years ($\hat{\phi}_{NT,4}$).</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function of $\lambda$.</th>
<th>Link to T</th>
<th>Link to T</th>
<th>Link probability</th>
<th>Link probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>after 2 years ($\hat{\beta}_{21}$).</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(19.65)</td>
</tr>
<tr>
<td></td>
<td>after 4 years ($\hat{\beta}_{22}$).</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(21.05)</td>
</tr>
<tr>
<td></td>
<td>Link probability if $Q_{ij} = 1$ ($\hat{\delta}_1$).</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>Link probability if $Q_{ij} = 0$ ($\hat{\delta}_0$).</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>p-value $H_{NV}$.</td>
<td>0.003</td>
<td>0.300</td>
<td>0.045</td>
<td>1.000</td>
</tr>
<tr>
<td>Avg treated outcome.</td>
<td>0.083</td>
<td>0.083</td>
<td>1.79</td>
<td>1.79</td>
</tr>
<tr>
<td>Individuals ($n$).</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
</tr>
<tr>
<td>Villages ($v$).</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
</tr>
<tr>
<td>Survey waves ($T$).</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 14: Expenditures.

<table>
<thead>
<tr>
<th>Outcome Method</th>
<th>(1) Nonfood PCE.</th>
<th>(2) Food PCE.</th>
<th>(3) Food Security.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program effect after 2 years ((\bar{\beta}_{11}))</td>
<td>-208.803 (160.98)</td>
<td>-208.049 (160.05)</td>
<td>421.741** (133.67)</td>
</tr>
<tr>
<td>Program effect after 4 years ((\bar{\beta}_{12}))</td>
<td>280.309* (145.11)</td>
<td>279.158 (178.65)</td>
<td>444.980** (133.66)</td>
</tr>
<tr>
<td>Spillover on T after 2 years ((\bar{\phi}_{T,2}))</td>
<td>-29.966 (69.80)</td>
<td>-32.452 (56.88)</td>
<td>401.713*** (56.47)</td>
</tr>
<tr>
<td>Spillover on T after 4 years ((\bar{\phi}_{T,4}))</td>
<td>-161.955** (40.98)</td>
<td>-161.161** (69.72)</td>
<td>253.726*** (59.58)</td>
</tr>
<tr>
<td>Spillover on NT after 2 years ((\bar{\phi}_{NT,2}))</td>
<td>-17.507 (49.98)</td>
<td>-19.103 (41.09)</td>
<td>215.298*** (30.18)</td>
</tr>
<tr>
<td>Spillover on NT after 4 years ((\bar{\phi}_{NT,4}))</td>
<td>-94.620*** (26.64)</td>
<td>-94.869*** (39.98)</td>
<td>135.984*** (51.07)</td>
</tr>
<tr>
<td>Link to T after 2 years ((\bar{\beta}_{21}))</td>
<td>-311.329 (966.77)</td>
<td>-349.080 (968.78)</td>
<td>343.343*** (62.93)</td>
</tr>
<tr>
<td>Link to T after 4 years ((\bar{\beta}_{22}))</td>
<td>-2386.991** (959.21)</td>
<td>-2389.737** (962.22)</td>
<td>190.068*** (62.48)</td>
</tr>
<tr>
<td>Link probability if (Q_{ij} = 1 (\bar{\delta}_{1}))</td>
<td>0.020** (0.01)</td>
<td>0.014** (0.01)</td>
<td>0.158*** (0.03)</td>
</tr>
<tr>
<td>Link probability if (Q_{ij} = 0 (\bar{\delta}_{0}))</td>
<td>0.013*** (0.00)</td>
<td>0.013*** (0.00)</td>
<td>0.118*** (0.00)</td>
</tr>
<tr>
<td>(\lambda) p-value (H_{NV})</td>
<td>0.50 0.389</td>
<td>0.50 0.835</td>
<td>0.15 0.159</td>
</tr>
<tr>
<td>Avg treated outcome. Individuals ((n))</td>
<td>1054.5 6907</td>
<td>2953.7 6907</td>
<td>2953.7 6907</td>
</tr>
<tr>
<td>Villages ((v))</td>
<td>1409 6907</td>
<td>1409 6907</td>
<td>1409 6907</td>
</tr>
<tr>
<td>Survey waves ((T))</td>
<td>3 3</td>
<td>3 3</td>
<td>3 3</td>
</tr>
</tbody>
</table>
Table 15: Occupational Choice, Bernoulli model.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self hours.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network.</td>
<td>474.153***</td>
<td>−112.859***</td>
<td>0.114***</td>
</tr>
<tr>
<td>Program effect after 2 years (β_{11})</td>
<td>(14.55)</td>
<td>(8.34)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Program effect after 4 years (β_{12})</td>
<td>464.304***</td>
<td>−143.481***</td>
<td>0.121***</td>
</tr>
<tr>
<td>Spillover on T after 2 years (ϕ_{T,t})</td>
<td>−26.577***</td>
<td>25.865***</td>
<td>−0.033***</td>
</tr>
<tr>
<td>Spillover on T after 4 years (ϕ_{T,t})</td>
<td>13.148</td>
<td>22.082***</td>
<td>−0.027***</td>
</tr>
<tr>
<td>Spillover on NT after 2 years (ϕ_{NT,t})</td>
<td>−12.862**</td>
<td>13.714***</td>
<td>−0.018***</td>
</tr>
<tr>
<td>Spillover on NT after 4 years (ϕ_{NT,t})</td>
<td>6.363</td>
<td>11.708***</td>
<td>−0.015***</td>
</tr>
<tr>
<td>Link to T after 2 years (β_{21})</td>
<td>−27.891***</td>
<td>13.355***</td>
<td>−0.050***</td>
</tr>
<tr>
<td>Link to T after 4 years (β_{22})</td>
<td>−12.862***</td>
<td>13.758***</td>
<td>−0.045***</td>
</tr>
<tr>
<td>Link probability (δ_{l})</td>
<td>0.492***</td>
<td>0.380***</td>
<td>0.120***</td>
</tr>
<tr>
<td>Avg treated outcome.</td>
<td>421.8</td>
<td>646.7</td>
<td>646.7</td>
</tr>
<tr>
<td>Individuals (n).</td>
<td>69087</td>
<td>69087</td>
<td>69087</td>
</tr>
<tr>
<td>Villages (v).</td>
<td>1409</td>
<td>1409</td>
<td>1409</td>
</tr>
<tr>
<td>Survey waves (T).</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>