Heterogeneous Endogenous Effects in Networks *

Sida Peng†

November 16, 2017

Abstract

This paper proposes a new method to identify leaders and followers in a network. Prior works use spatial autoregression models (SARs) which implicitly assume that each individual in the network has the same peer effects on others. Mechanically, they conclude the key player in the network to be the one with the highest centrality. However, when some individuals are more influential than others, centrality may fail to be a good measure. I develop a model that allows for individual-specific endogenous effects and propose a two-stage LASSO (2SLSS) procedure to identify influential individuals in a network. My method allows me to identify leaders and followers through their observed behaviors on the network. I develop robust inference for my estimators including uniformly valid confidence intervals. I extend the analysis to allow for multiple types of connections (multiple networks), and I show how to use the square-root sparse group LASSO to detect which of the multiple connection types is more influential. Simulation evidence shows that my estimator has good finite sample performance. I further apply my method to the data in Banerjee et al. (2013) and my proposed procedure is able to identify leaders and effective networks.

key words: key players, network, endogenous effects, spillovers, high-dimensional models, LASSO, model selection, robust inference

*I would like to thank Francesca Molinari, Matthew Backus, David Easley, Marten Wegkamp, Donald Kenkel, Zhuan Pei, Douglas Miller, Joris Pinkse, Yanlei Ma and participants in seminars at which this paper were presented. All remaining errors are mine.

†Microsoft Research, sidpeng@microsoft.com

1 Introduction

How an individual's behavior is affected by the behavior of her neighbors in an exogenously given network is an important research question in applied economics. With the increasing availability of detailed data documenting connections among individuals, spatial autoregression models (SARs) have been widely applied in the empirical networks literature to estimate endogenous effects.

In SARs, an individual's behavior depends on the weighted average of other individuals' behaviors (see Anselin, 1988; Kelejian and Prucha, 1998). Standard SARs assume that the peer effects/endogenous effects are the same across individuals in a network. Each individual influences her neighbors *at the same rate* regardless of who she is. However, in many contexts, some individuals are clearly more influential than others. For example, Mas and Moretti (2009) finds that the magnitude of spillovers varies dramatically among workers with different skill levels. Clark and Loheac (2007) also notes that popular teenagers in a school have much stronger influence on their classmates' smoking decisions than their less popular peers.

I propose a novel SAR model which allows for *heterogeneous* endogenous effects. Each individual in a network simultaneously generates an outcome that takes into account all her neighbors' behaviors. Unlike standard SARs, each individual has an individual-specific effect on her neighbors. As a result, there are as many coefficients for individual-specific endogenous effects as there are individuals in the network. To achieve identification, I assume that "truly-influential" individuals only constitute a small fraction of the total population. In other words, individual-specific coefficients are assumed to be sparse. This assumption allows me to estimate the model via the least absolute shrinkage and selection operator (LASSO). The LASSO procedure penalizes the l_1 norm for the coefficients of heterogeneous endogenous effects. The geometry of the l_1 norm enforces the sparsity in the LASSO estimators. If a coefficient is selected by LASSO (i.e. the estimated coefficient is non-zero), the individual associated with this coefficient can influence all her neighbors at her specific rate. Otherwise the LASSO estimator will indicate that the individual has no influence on her neighbors. With some restrictions on the network structures, I show that the LASSO estimates for heterogeneous endogenous effects have near oracle performance (see Bühlmann and van de Geer, 2011). In other words, the selection of influential individuals is consistent and the convergence rate of non-zero LASSO estimates is the same as the convergence rate that would have been achieved if the truly influential individuals were known.

One challenge in my estimation process is the presence of endogeneity in the spatial lag and error term. As with standard SARs, the dependent variable in my model is used to construct spatial lags as an independent variable. As a result, the regressors are correlated with the error term and estimates would be biased if we were to apply LASSO directly.

First I propose a set of novel instruments to address the endogeneity. Following Kelejian and Prucha (1998), I express the dependent variable as an infinite sum of functions consisting of exogenous characteristics and an adjacency matrix. I show that the exogenous characteristics of influential individuals can be used as instruments for their neighbors. Then I design a two-stage estimation process for heterogeneous endogenous effects using LASSO at each stage. *In the first stage*, I use LASSO to estimate the coefficients for the instruments. These estimated coefficients and instruments are then used to create a synthetic dependent variable. *In the second stage*, I replace the dependent variable in the spatial lags with the synthetic variable to perform the LASSO estimation. Unlike in the standard two stage least square estimation process, the synthetic dependent variable in the first stage suffers from a shrinkage bias due to the LASSO fitting. However, I show that with certain restrictions on the network structure, the shrinkage bias is negligible (i.e. $o(1/\sqrt{n})$).

The next challenge is to construct robust confidence intervals for my LASSO type two-stage estimator. As pointed out in Leeb and Potscher (2008), it is impossible to construct uniformly valid confidence intervals for estimates based on model selection. Consistent model selection by LASSO is only guaranteed when all non-zero coefficients are large enough to be distinguished from zero in a finite sample (i.e. usually called the "beta-min" condition). LASSO may fail to select regressors with very small coefficients, resulting in omitted variable bias in the post LASSO inference.

I propose a bias correction for my two-stage estimator following the recent LASSO inference literature (see Chernozhukov et al., 2017; Belloni et al., 2015; van de Geer et al., 2014; Javanmard and Montanari, 2014; Zhang and Zhang, 2011). The idea is to correct the first order bias and make the estimators independent from the model selection. Heuristically, shrinkage bias due to the l_1 penalty in LASSO can be expressed as a function of the LASSO estimators. Normality can still be achieved after adding back this bias. I show that this strategy also works in a two-stage estimation process. I derive the asymptotic normality for my "de-sparse" two-stage LASSO estimator and conduct robust inference including confidence intervals.

My model can be extended to allow for more flexible network structures. One real world scenario is a network which consists of multiple cliques. Each clique has its local leaders, who only influence individuals within their own cliques but have no influence on individuals outside their cliques. One identification difficulty in this setting is that the number of leaders increases with the number of cliques. Hence, the sparsity assumption can potentially be violated.

To solve the problem in this scenario, I modify my model by bringing back the classical SAR model. I assume that there are both local leaders and global leaders in the network. In contrast to local leaders, global leaders can influence individuals across different cliques. I assume global leaders are sparse and show that identification can be achieved for this modified model. The endogenous effects of local leaders will be captured by the classical SAR model, which becomes an average endogenous effects in the network. The endogenous effects of global leaders, whose influence remains individual-specific, can be identified in the same way as it was in the previous model. If there is no global leader in the network, the model is effectively just the standard SAR model.

Another real world scenario is the existence of multiple types of connections among individuals. For example, connections among individuals can be classified as social (e.g. friendship, kinship) or economic (e.g. lending, employment). In epidemiology, infectious disease can be spread through air, insects, or direct contact. It is important to identify which networks are more efficient at transmitting the endogenous effects.

I model different types of connections as multiple networks. I propose the use of square-root sparse group LASSO to estimate a heterogeneous endogenous effects model with multiple networks. The standard sparse group LASSO penalizes both the l_1 norm and the l_2 norm for each coefficient in each type of connection. I modify the sparse group LASSO by taking the square-root of the mean square error and thus make the estimation process pivotal. I derive the convergence rate and prove the consistency of selection. To the best of my knowledge, my paper is the first to show statistical properties for square-root sparse group LASSO.

I provide simulation evidence for networks of different sizes and different data generating algorithms. The empirical coverage of my proposed estimators is close to the nominal level in all scenarios. Similar results are also found in models with multiple networks and with cliques.

I apply my method to study villagers' decisions to participate in micro-finance programs in rural areas of India as in Banerjee et al. (2013). Among different social and economic networks, my method shows that some networks such as "visit go-come" and "borrow money", are much more effective at influencing villagers' decisions than other networks such as "temple company" and "medical help". I further show that individuals in certain careers such as agricultural workers, Anganwadi teachers and small business owners are more likely to influence villagers.

My proposed methodology can be applied to detect influential individuals in empirical work when there are both leaders and followers. It is important to identify such individuals because we can then study why certain people are more influential than others. On the one hand, we can examine individuals exogenous characteristics and see if any of them contribute to an individual's influence. On the other hand, we can study how the position of an individual within a network may impact her influence by further introducing network formation into the model.

Being able to identify influential individuals could also lead to more effective policy outcomes. If individuals with certain characteristics are found to be more influential than others, policy makers could potentially implement policies solely targeting influential individuals rather than the entire population. Since more resources are directed to the small group of highly influential individuals, one would expect much more effective policies. For example, online opinion leaders have influence on what people tweet and share on the Internet. In an election, instead of advertising on television and trying to influence every voter, a candidate could invest in these online opinion leaders and let them influence the public in a more efficient way. This technique could also work in employment contexts. Union leaders are often those workers who have the strongest influence on their fellow workers' opinions. Instead of reading through complaints from every worker, employers could identify those union leaders and make sure their complaints were addressed to prevent any ongoing strike. When studying peer effects in smoking behavior, my method can identity a group of teenagers who have a strong influence on their peers. A policy can target this group of students and encourage them to quit smoking.

1.1 Literature Review

This paper brings together literature on spatial autoregression model, LASSO and networks.

SARs:
SARs have been widely applied in empirical studies. For instance, they have been used to study peer effects in labor productivity (see Mas and Moretti, 2009; Guryan et al., 2009; Bandiera et al., 2009), smoking behavior among teenagers (see Krauth, 2005; Clark and Loheac, 2007; Nakajima, 2007), educational achievements among different student groups (see Sacerdote, 2001; Neidell and Waldfogel, 2010), systemic risk in finance (see Bonaldi et al., 2015; Denbee et al., 2015), and the adoption of new agricultural technologies (see Coelli et al., 2002; Conley and Udry, 2010). My paper proposes a novel extension of standard SARs that could be used to identify influential individuals in a given network. My methodology for estimating such a model could easily be adopted in existing empirical SARs analyses to identify influential individuals who influence their peers productivity, smoking decisions, or financial holdings.

More specifically, my model extends existing SARs literature by introducing *heterogeneous* endogenous effects. Until very recently, SARs always assume a constant rate of dependence for endogenous effects across different individuals (see Cliff and Ord (1973), the first monograph on the topic, and the later studies, Upton and Fingleton (1985); Anselin (1988); Cressie (1993); Lee and Liu (2010); Lee and Yu (2010); Jin and Lee (2016)). Recent developments in social interaction literature incorporate individual characteristics into SARs, essentially modeling the heterogeneity through a linear combination of exogenous effects (see Manski, 1993; Bramoullé et al., 2009). In contrast, my model considers the heterogeneity in the endogenous effects. Heterogeneous endogenous effects can be identified from individuals' outcomes instead of being pre-specified through individuals' characteristics. To my knowledge, my proposed model is the first to capture the direct impact of an individuals neighbors' decisions on her own decision.

To estimate the heterogeneous endogenous effects in my model, I propose a methodology that is different from standard SARs literature. In classic SARs, there is only one endogenous variable and hence it is sufficient to identify the model through only one instrument. In my model, the number of potentially endogenous variables increases as the number of observations increases. As a result, I propose a set of instruments that contain the same number of instruments as the total number of individuals. Moreover, each instrument is different from the standard SARs instrument as in Kelejian and Prucha (1998), Lee (2002), Lee (2003) and Lee (2004).

This paper also contributes to literature that models multiple networks through SARs. In standard SARs, multiple networks are modeled as higher order spatial lags (see Lee and Liu, 2010). Even though different networks are assumed to have different constant rates for endogenous effects in these models each individual in a given network faces the same constant rate. In contrast, my model allows for the a more realistic scenario where each individual has her own specific endogenous effects in each network. Moreover, my methodology allows some networks to be classified as completely irrelevant to decision-making *ex ante* and these networks can be consistently identified.

LASSO:
My paper extends LASSO literature by deriving statistical bounds and consistency of selection for the square-root *sparse group* LASSO estimator. This estimator builds on the group LASSO, square-root LASSO, square-root group LASSO, and sparse group LASSO. Belloni et al. (2011) introduced the square-root LASSO, which does not require a pre-estimation of an unknown standard deviation σ. Yuan and Lin (2006) proposes the group LASSO, in which explanatory variables are represented by different groups. The group LASSO assumes that sparsity exists only among groups, i.e. some groups of variables are relevant while other groups are not. Simon et al. (2013) proposes the sparse group LASSO, which further allows sparsity within each group, i.e. some regressors within the relevant groups can also be irrelevant. Bunea et al. (2014) derives statistical properties for the square-root group LASSO, which combines group LASSO and square-root LASSO. When estimating a heterogeneous endogenous effects model with multiple networks, I provide proof for both statistical bounds and consistency of selection for the square-root *sparse group* LASSO estimator. To the best of my knowledge, this paper is the first to show asymptotic statistical properties

for the square-root *sparse group* LASSO estimator.

This paper also contributes to the growing literature on endogenous regressors in LASSO estimators. For instance, Belloni et al. (2014a) proposes the double selection mechanism to study confounded treatment effects. Fan and Liao (2014) proposes a GMM type estimator to deal with many endogenous regressors. Gautier and TsyBakov (2014) proposes a Self Tuning Instrumental Variables (STIV) estimator. The paper that is closest to mine is Zhu (2016), which studies the statistical properties of two-stage least square procedure with high-dimensional endogenous regressors. She studied a case when there exists p endogenous regressors. For each regressor j, she assumed that one can find d_j instruments. Both p and d_j may grow as n increases. I consider a case that is tailored to my spatial autoregression model. There are n endogenous regressors and each regressor shares the same n instruments. I show that a modified "de-sparse" LASSO estimator can be constructed for my estimator in a manner similar to Zhang and Zhang (2011), Bühlmann (2013), van de Geer et al. (2014), and Zhu (2016). I derive its asymptotic distributions and show how to perform inference.

Networks:
My paper shares similar microfoundations with SARs as discussed in Blume et al. (2015), where the individual utility function can be written as a linear summation of the private and social components. The private component is a quadratic loss function on individual's efforts. The social components depend on the network structure as well as the efforts of one's neighbors. While the marginal rate of substitution between the private and social components of utility is assumed fixed in SARs, I assume this rate is individual-specific and depends on one's neighbors. My paper applies and extends LASSO approaches to deal with a high-dimensional problem in networks. The total number of possible edges in a network is n^2, however, the social interaction networks we often observe are far more sparse. This is an ideal setting where LASSO could be applied. Manresa (2013) studies the heterogeneous exogenous effects in a network using LASSO. de Paula et al. (2015) explore the use of LASSO to recover network structures. Both these two papers consider panel data and rely on repeated observations of the same network to identify their models. My model considers cross-sectional data. To identify an individual's endogenous effects, I rely on the variations in her neighbors' outcome.

My paper also relates to the literature on identifying the key players in the network following Ballester et al. (2006), Calvó-Armengol et al. (2009), and Horracea et al. (2016). Under the framework of SARs, every individual is assumed to have the same endogenous effects. As a result, individuals who are well-connected in the network (with high centrality measure) become the key players in the network. However, this is not necessarily the case in my model, as well connected individuals can have zero endogenous effects on her neighbors. Indeed, as shown in the empirical application, well connected villagers such as tailors, hotel workers, veterans, and barbers are not influential in other villagers' decisions to join the micro-finance program.

The rest of this paper is organized as follows: in Section 2, I introduce the model; in Section 3, I discuss identification assumptions; in Section 4, I design estimation procedures; in Section 5, I derive consistency and asymptotic properties; in Section 6, I show finite sample performance using Monte Carlo simulations; in Section 7, I apply my proposed model to study influential individuals and effective networks in promoting micro-finance programs in rural India; and in Section 8, I conclude.

2 Models

In this section, I first lay out the benchmark endogenous effects model and introduce the central model of this paper - the heterogeneous endogenous effects model. Then I discuss two extensions of the heterogeneous endogenous effects model: a model for networks consisting of multiple cliques and a model for multiple networks. Finally, I provide two examples and illustrate how my model fits into these settings.

2.1 Benchmark Endogenous Effects Model

I first introduce the standard spatial autoregression model as the benchmark endogenous effects model. Let n denotes the total number of observed individuals in a network. The outcome of individual i is denoted as d_i and is the variable of interest. Here d_i can represent any outcome associated with individual i, such as whether to smoke, whether to join a program, or whether to tweet a message from a friend. It is assumed that the outcome of each neighbor of individual i impacts her outcome homogeneously through a constant rate λ_0:

$$d_i = \lambda_0 \sum_{j \in N_i} d_j + x_i \beta_0 + \epsilon_i, \tag{1}$$

where the set N_i is defined as individual i's neighbors. The matrix form of this model is expressed as follows:

$$D_n = \lambda_0 M_n D_n + X_n \beta_0 + \epsilon_n, \tag{2}$$

where $D_n = (d_1, d_2, \cdots, d_n)'$ is the n-dimensional vector of observable outcomes. The n by k matrix X_n represents the observable exogenous characteristics of individuals. When ϵ_n is specified as an

n-dimensional vector of independent and identically distributed disturbances with zero mean and a constant variance σ^2, equation (2) is also called a mixed regression model.

The spatial weight matrix M_n is of size n by n, where the (i,j)th entry represents the connection between individual i and individual j. In empirical studies, the spatial weight matrix is often replaced by the adjacency matrix (see Ammermuller and Pischke, 2009; Acemoglu et al., 2012; Banerjee et al., 2013): the (i,j)-th entry of the matrix M_n takes value 1 if individual i and individual j are connected and takes value 0 otherwise.

In this model, endogenous effects (see Manski, 1993) or network effects (see Bramoullé et al., 2009) are captured by the scalar λ_0. An implicit assumption in equation (2) is that λ_0, the rate of endogenous effects, is identical across all individuals in the network. Every individual affects her neighbors at this same rate λ_0 no matter who she is, how many neighbors she has and where is she in the network. This limitation has been noted in various studies (see Ammermuller and Pischke, 2009; de Paula et al., 2015). I relax this assumption by proposing a more flexible model that allows individual-specific endogenous effects as discussed below.

2.2 Heterogeneous Endogenous Effects Model

I propose the following model to allow for heterogeneous endogenous effects:

$$d_i = \sum_{j \in N_i} d_j \eta_j + x_i \beta + \epsilon_i \tag{3}$$

where N_i represents the set of individual i's neighbors and η_j represents the endogenous effects of individual j on the outcome of all her neighbors $i \in N_j$. the model can be rewritten in matrix form as:

$$D_n = \Big(M_n \circ D_n\Big)\eta_0 + X_n \beta_0 + \epsilon_n, \tag{4}$$

where $\eta_0 = (\eta_1, \eta_2, \cdots, \eta_n)'$ is a vector of parameter of size n by 1. The ith entry in η_0 represents the endogenous effects of individual i on her neighbors. This model allows for individual heterogeneity to interact with endogenous effects so that every individual is allowed to have her own coefficient η_i. My model allows some $\eta_j = 0$. In other word, there are individuals that have no endogenous effects on their neighbors. I define those individuals with $\eta_j \neq 0$ as influential.

The operator $\circ$ is defined between a n by n matrix M_n and a n by 1 vector D_n as

$$M_n \circ D_n = M_n \cdot \ \text{diag}(D_n) = C,$$

where $diag(\cdot)$ is the diagonalization operator and $C_{i,j} = M_{i,j} d_j$.

Note that in contrast to fixed rate λ_0 specified in equation (2), even though each neighbor of individual j is assumed to receive the same influence $d_j\eta_j$ from her[1], each individual is allowed to influence her neighbors at her own rate η_j.

Similar to equation (2), equation (4) can be derived from a bayesian Nash Equilibrium. Let (x_i, ϵ_i) denotes an individual's type, where x_i is publicly observed characteristics and ϵ_i is private characteristics only observable by i. Individual i's utility depends on her own action and characteristics as well as her neighbors' actions. Individual i chooses action d_i to maximize the following utility:

$$U_i(d_i, d_{-i}) = (x_i\beta + \epsilon_i)d_i - \frac{1}{2}d_i^2 + \sum_{j\in N_i} d_j d_i \eta_j$$

The first order condition yields equation (4). The micro-foundations derived above is similar to the one for SARs as discussed in Blume et al. (2015).

2.3 Examples

To help readers conceptualize the heterogeneous endogenous effects model, here I apply the model to two specific contexts one invloving labor productivity and the other online opinion leaders.

Peer Effects in Labor Productivity:

Understanding the mechanism and magnitude of the dependence of labor productivity on coworkers is an important question for economists and policy makers. As found in Mas and Moretti (2009), workers respond more to the presence of coworkers with whom they frequently interact. In this case, the influence level of each individual to hers coworkers is not necessarily the same. Equation (4) can be used to incorporate such differences.

$$y_i = \sum_{j\in N_i} y_j\eta_j + x_i\beta + \epsilon_i,$$

where y_i is individual i's productivity, x_i represents individual i's characteristics (education levels, ages, etc) and N_i is the set of coworkers that works directly with i. η_j represents the size of influence of coworker j – all else being equal, the additional effect on individual i's productivity if individual j becomes her coworker

Note that if we restrict the parameters η_j to be the same across different workers, then we are back to the classical SARs setting as laid out in equation (2). Thus, $\lambda = \frac{1}{n}\sum_{j=1}^{n}\eta_j$ can be interpreted as the averaged spillover effects in the canonical sense.

[1]Further relaxation of the model considering different individual j's influence on each of her neighbors requires panel data.

Define

$$\lambda^* = \frac{1}{\sum \mathbf{1}_{\eta_j \neq 0}} \sum_{j=1}^{n} \eta_j \mathbf{1}_{\eta_j \neq 0}$$

as the averaged endogenous effects for influential workers. λ^* does not include non-influential individuals in the calculation. It is a more precise measure of endogenous effects compared with λ from equation (2).

Online Opinion Leaders:

A decision can represent whether to "tweet" a news story seen online. When individuals make such decisions, they are often influenced by several online opinion leaders – whether those people "tweet" the news or not. There are also many types of online opinion leaders, including political figures and some are celebrities. For certain types of news, some opinion leaders may be very influential while the rest may have no influence on the public. Opinion leaders may also influence each other when deciding whether to "tweet" the news or not. Assume a binary decision $(0, 1)$ is made from a bayesian Nash Equilibrium, such that

$$d_i^* = \sum_{j \in N_i} d_j^* \eta_j + x_i \beta + \epsilon_i,$$

where d_i^* is the probability of individual i playing action 1, and $\sum_{j \in N_j} d_j^* \eta_j$ is the expected endogenous effects from i's neighbors N_i . X_i is the individual i's characteristics such as political views, age, career, etc.

Similarly, we can define $\lambda = \frac{1}{n} \sum_{j=1}^{n} \eta_j$ as the averaged endogenous effects. Since the number of opinion leaders is very small compared with total online users, λ can be very close to 0. A more precise measure would be

$$\lambda^* = \frac{1}{\sum \mathbf{1}_{\eta_j \neq 0}} \sum_{j=1}^{n} \eta_j \mathbf{1}_{\eta_j \neq 0}$$

λ^* will be the average endogenous effects for online opinion leaders. On the other hand, it is also important to identify the set:

$$S = \{j : \eta_j \neq 0\}$$

as truly influential opinion leaders. If a similar type of news story needs to be spread the next time, contacting those leaders and obtaining their endorsement would be a good starting strategy.

2.4 Heterogeneous Endogenous Effects Model with Cliques

I propose an extension to my heterogeneous endogenous effects model which could address such challenges. Consider a network composed of many cliques (small groups of connected individuals).

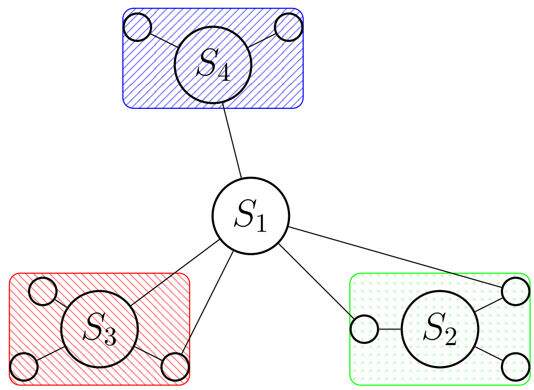


Figure 1: Local Leader

Each clique has its local leader who only influences individuals within her own clique. Figure 1 provides an example of such a network structure. Note that in Figure 1, node S_2, S_3 and S_4 represent local leaders who only influence individuals within their own cliques. On the contrary, node S_1 represents a global leader who can influence individuals across different cliques. For example, one can think about the local leaders S_2, S_3 and S_4 as local news channels while S_1 is the national news channel. I assume that all local news will influence the public at a small but similar rate while different national channels can have different effects on their audience.

In the above network structure, if the number of local leaders is increasing with the number of cliques but the number of individuals in each clique stays fixed, it is impossible to identify the individual-specific influence of all those local leaders. To address this problem, I assume a homogeneous effect γ_0 among all individuals. This rate will capture all influence from local leaders. However, I allow global leaders to heterogeneously influence their neighbors at rates that differ from γ_0 and show that γ_0 and the heterogeneous effects can be consistently estimated.

More specifically, I consider the following model:

$$d_i = \sum_{j \in N_i} d_j \eta_j + \gamma_0 \sum_{j \in N_i} d_j + x_i \beta_0 + \epsilon_i, \tag{5}$$

which can be represented in matrix form as:

$$D_n = \Big(M_n \circ D_n \Big) \eta_0 + M_n D_n \gamma_0 + X_n \beta_0 + \epsilon_n, \tag{6}$$

where $\eta_0' = (\eta_1, \eta_2, \cdots, \eta_n)'$. The new term $\gamma_0 \sum_{j \in N_i} d_j$ captures influence from the local level. Note that this is the same term as the spatial lag in the benchmark spatial autoregression model. The vector η_0 captures the heterogeneous endogenous effects of global leaders.

If no global leader exists, i.e. $\eta_j = 0, \forall j$, the model collapses back to the classical SARs as in equation (2). If there is no local level influence, i.e. $\gamma_0 = 0$, then the model coincides with the heterogeneous endogenous effects model in section 2.2. I expand the discussion of the identification assumptions for this model in the next section.

2.5 Heterogeneous Endogenous Effects Model with Multiple Networks

In reality, individuals are often connected with each other through more than one type of network. For example, ones colleague (connection in an employment network) could also be her friend (connection in a friendship network), and ones uncle (connection in a relative network) could also be the person she lends money to (connection in a borrowing/lending network). In such scenarios, an individuals outcome could potentially be influenced by the outcomes of her neighbors from more than one type of network.

To capture different types of connections among the same set of individuals, we can incorporate multiple networks in my heterogeneous endogenous model. More specifically, a separate adjacency matrix can be constructed for each type of network. For instance, the (i, j)-th entry of the adjacency matrix representing friendship takes value 1 if individual i and individual j are friends and takes value 0 otherwise; that representing the borrowing/lending network takes value 1 if individual i and individual j lend money to each other and takes value 0 otherwise.

Let q be the total number of different types of network. Define M_n^l as the adjacency matrix for the lth network. The heterogeneous endogenous effects model with multiple networks is defined as

$$d_i = \sum_{l=1}^{q} \sum_{k \in N_i} d_k^l \eta_k^l + x_i \beta_0 + \epsilon_i \tag{7}$$

Note that in this model, different network could potentially bear different endogenous effects for the same individual. In equation (7), coefficient η_k^l represents the rate of endogenous effect of individual k through network l. As a result, we have $nq + k$ coefficients for endogenous effects. In addition, I assume endogenous effects from different types of networks are linearly additive. The model can also be rewritten in matrix form as:

$$D_n = \sum_{l=1}^{q} \left(M_n^l \circ D_n \right) \eta_0^l + X_n \beta_0 + \epsilon_n, \tag{8}$$

where M_n^l is the adjacency matrix for network l. $\eta^l = (\eta_1^l, \eta_2^l, \cdots, \eta_n^l)'$ is an n by 1 vector for $l = 1, 2, \cdots, q$. Define a network l as efficient network if $\eta_i^l \neq 0$ for at least one individual $i = 1, 2, \cdots, n$.

3 Identification

In this section, I discuss the conditions under which the heterogeneous endogenous effects model is identified. My assumptions combine both standard SARs type assumptions and LASSO type assumptions. SARs type assumptions ensure the existence of valid instruments to identify the model. LASSO type assumptions guarantee consistent model selection and estimation using a LASSO estimator. In what follows, I will first present the assumptions needed for a standard heterogeneous endogenous effects model to be identified. Then I will discuss identification assumptions for two model extensions laid out in the previous section – one heterogeneous endogenous effects model for networks consisting of multiple cliques and one with multiple types of networks.

Before discussing identification assumptions for the heterogeneous endogenous effects model, lets first recall the benchmark SAR model:

$$D_n = \lambda_0 M_n D_n + X_n \beta_0 + \epsilon_n, \tag{9}$$

Note that by rearranging the above equation, we can express endogenous variable $M_n D_n$ solely as a function of X_n and M_n, since:

$$D_n = J_n^{-1} X_n \beta_0 + J_n^{-1} \epsilon_n$$

where I_n is the n by n identity matrix and $J_n = I_n - \lambda_0 M_n$. It is straightforward that $J_n^{-1} X_n$ can serve as valid instruments for $M_n D_n$. As a result, the identification and estimation of equation (9) can be achieved through either 2SLS or GMM as proposed in papers such as Kelejian and Prucha (1995), Kelejian and Prucha (1998) Lee (2002), Lee (2003), and Lee (2004).

As will be explained in detail in subsequent sections, to estimate the individual specific effects in the heterogeneous endogenous effects model, I derive a set of instruments in a similar way by solving D_n as a function of exogenous variables and an adjacency matrix. The assumptions listed below essentially guarantee the existence and consistency of the 2SLS estimates.

3.1 Identification Assumptions for the Heterogeneous Endogenous Effects Model

Recall that the heterogeneous endogenous effects model is specified as:

$$D_n = \Big(M_n \circ D_n\Big)\eta_0 + X_n \beta_0 + \epsilon_n,$$

First note that without additional restrictions, this model could not be point identified through canonical method as the number of parameters $n + k$ is greater than the number of observations

n. To achieve identification, the key assumption that I maintain is that only a small number of individuals in the network are influential (i.e. $\eta_j \neq 0$).

Assumption 1. *Let $S_n \subset \{1, 2, \cdots, n\}$ denote the set of influential individuals (i.e. $\eta_j \neq 0$). Let $s_n = |S_n|$ be the number of elements in S_n.*

$$s_n = o\left(\frac{\sqrt{n}}{\log n}\right), \qquad \text{as } n \to \infty$$

Assumption 1 is usually referred to as "sparsity" assumption. The assumption that most individuals in a network are not influential is plausible under many circumstances. For example, opinion leaders on social media only constitute a very small fraction of internet users; there are only a couple of "cool" kids at school that might influence their friends' smoking decisions; passionate workers that can boost the productivity of their coworkers are also relatively rare. On the other hand, assumption 1 might not be plausible under some circumstances. For example, the peer effects in obesity among school children might not be sparse. When many local leaders exist within a network, the sparsity assumption could also be violated. In section 3.2, I propose an extension of the current model to address the problem of many local leaders.

Assumption 2.

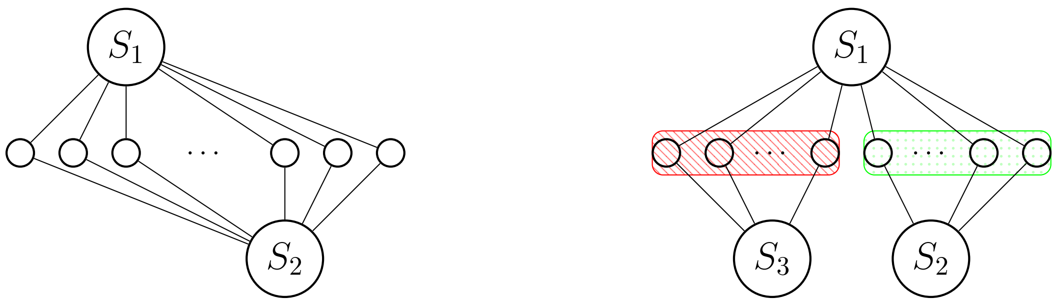
- *There exists an $\eta_{\max} < 1$ such that $\|\eta_0\|_\infty \leq \eta_{max}$*
- *The ϵ_j are i.i.d with 0 mean and variance σ^2*
- *The regressors x_i in X_n are non-stochastic and uniformly bounded for all n. $\lim_{n\to\infty} X_n' X_n / n$ exists and is nonsingular*

Assumption 2 guarantees the invertibility of $\left(I_n - M_n \circ \eta_0\right)$. The restriction on η_0 excludes the unit root process and ensures the uniqueness of equilibrium. The assumptions on the error term and the assumption that X_n is a fixed design matrix are the same as those imposed in the mixed regression model[2] (see Lee, 2002). I focus on the case where X_n is an n by 1 vector and study identification as in Bramoullé et al. (2009). It is straightforward to generalize the algebra when X_n is n by k. More instruments can be constructed in this scenario.

To proceed, recall the definition of the operator "$\circ$" as $M_n \circ D_n = M_n \cdot \text{diag}(D_n)$, where $\text{diag}(\cdot)$ is the diagonalization operator. Note the following property of the "$\circ$":

$$\left(M_n \circ D_n\right)\eta_0 = \left(M_n \circ \eta_0\right)D_n$$

[2]The assumption on error terms exclude exogenous effects and correlated effects from my model. An identification problem similar to the "reflection problem" arises when including exogenous effects. More instruments need to be constructed, which requires better data. These are interesting directions for future research.



(1) (2)

Figure 2: Examples of networks which violate assumption 3

If the invertibility of $\Big(I_n - M_n \circ \eta_0\Big)$ is guaranteed, then

$$D_n = \Big(M_n \circ D_n\Big)\eta_0 + X_n\beta_0 + \epsilon_n \Leftrightarrow D_n = \sum_{i=0}^{\infty}\Big(M_n \circ \eta_0\Big)^i (X_n\beta_0 + \epsilon_n) \tag{10}$$

This is formally shown in Appendix.

Since $\Big(M_n \circ D_n\Big)\eta_0$ is correlated with ϵ_n and η_0 is sparse (i.e. having at most s_n non-zero elements), we need at least s_n instruments to deal with the endogeneity in the model. Using equation (10), we can express the expectation of D_n as follows:

$$E(D_n) = X_n\beta_0 + \Big(M_n \circ X_n\Big)(\beta\eta_0) + \sum_{i=2}^{\infty}\Big(M_n \circ \eta_0\Big)^i \beta_0 X_n, \tag{11}$$

Let $(\cdot)_S$ denote the operator such that $(M_n)_S$ is a sub matrix of M_n with its columns restricted to columns corresponding to the elements of S. The first and second terms of equation (11) suggest that X_n and $(M_n \circ X_n)_S$ can serve as valid instruments to point identify β_0 and η_0.

Assumption 3. $\big[X_n, (M_n \circ X_n)_S\big]$ *is full rank.*

Assumption 3 is the key assumption that leads to identification. The linear independence among $(M_n \circ X_n)_S$ requires the assumption that any two influential individuals may not necessarily connect with identical neighbors. Moreover, assumption 3 also requires that neighbors of an influential individual cannot be a linear combination of neighbors of several other influential individuals, which rules out network structures as depicted in Figure 2.

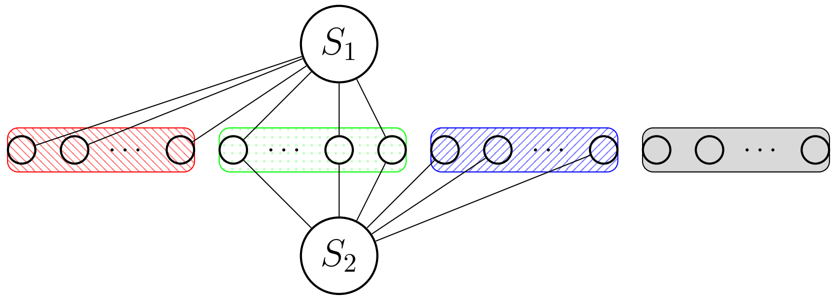


Figure 3: Examples of networks which satisfies assumption 3

In other words, as long as each influential individual has a neighbor that is not connected with any other influential individuals, assumption 3 is satisfied. One can think of the identification here as estimating fixed effects from influential individuals. Collinearity arises when the fixed effects of two influential individuals are imposed on exactly the same observations. As shown in Figure 3, the influence of S_1 can be identified by comparing red (right shaded) and grey groups (plain), while the influence of S_2 can be identified by comparing blue (left shaded) and grey (plain) groups. Or the influence of S_1 can be identified by comparing green (dotted) and blue (left shaded) groups, while the influence of S_2 can be identified by comparing red (right shaded) and green (dotted) groups.

Further, as shown in Appendix, one can rewrite equation (11) as:

$$E(D_n) = X_n\beta_0 + \Big(M_n \circ X_n\Big)\tilde{\eta}, \tag{12}$$

where $\tilde{\eta}_j = \eta_j f(\beta_0, X_n, M_n)$ for some function f depends on β_0, X_n, and M_n. Note that $\tilde{\eta}_j = 0$ as long as $\eta_j = 0$. As a result, the sparsity assumption is also satisfied in equation (14), and I can thus estimate equation (14) as the first stage in using a LASSO type estimator.

At this point, if the truly influential individuals set S_n were available to us, we would be able to estimate the model using 2SLS method or GMM. However, in most cases, S_n is not known beforehand. I propose to use a LASSO type estimator to both recover the set of influential individuals and estimate the model. For LASSO to achieve correct recovery, I need the following assumptions:

Assumption 4.
(Irrepresentable Condition) *There exists* $N \in \mathbb{N}$*:* $\forall n \geq N$*, there is a* $\vartheta \in (0,1)$ *such that*

$$P\left(\left\|diag((\hat{D}_n)_{S^c})\Sigma_n diag((\hat{D}_n)_S)^{-1} sign(\eta_0)\right\|_\infty \leq \vartheta\right) = 1;$$

where

$$\Sigma_n = (M_n)'_{S^c}(M_n)_S\left((M_n)'_S(M_n)_S\right)^{-1},$$

(Beta Min Condition) *There exists $N \in \mathbb{N}$: $\forall n \geq N$, there is a $m > 0$ such that*

$$\min(|\eta_0|)_S \geq m/\sqrt{n},$$

Here $(M_n)_S$ represents the sub-matrix of M_n given by the columns corresponding to influential individuals. Similarly, $(M_n)_{S^c}$ represents the sub-matrix of M_n given by the columns corresponding to non-influential individuals.

Assumption 4 is required for the LASSO estimator to achieve a consistent selection for the set S_n in the second stage. The Irrepresentable Condition imposes restrictions on non-influential individuals such that the neighbors of a non-influential individual will not be exactly the same as those of any influential individual. This is because when two individuals connect with exactly the same neighbors, we cannot distinguish which individual is the true source of influence. This assumption rules out identification in complete networks (i.e. all individuals are connected). The Beta Min Condition requires the magnitude of the endogenous effects to be sufficiently strong in order to be detected by LASSO. For example, there does not exist a sequence of individuals whose influence decay to 0 faster than $m/\sqrt{n}$.

Irrepresentable Condition implies the compatibility condition as in van de Geer et al. (2014) and restricted eigenvalues condition (Condition RE) as in Belloni et al. (2011). The stronger version is assumed to ensure selection consistency. As shown in Zhao and Yu (2006), the Irrepresentable Condition together with the Beta Min Condition are necessary and sufficient conditions for LASSO to achieve consistent model selection. If assumption 5 is violated, inference on the de-sparse coefficients is still valid when compatibility condition is satisfied. Details will be discussed in the next section.

Assumption 5.
(Maximum Neighbors Condition) *There exists $N \in \mathbb{N}$: $\forall n \geq N$,*

$$\|M_n'\mathbf{1}_n\|_\infty = O([\log n]^\epsilon), \quad \epsilon \in (0,1]$$

(Variance Condition)

$$\frac{1}{n}M_n'W_n(I - M_n \circ \eta_0)^{-1}(I - M_n \circ \eta_0)^{-1'}W_nM_n \to \Omega,$$

where $W_n = \Big(I - X_n(X_n'X_n)^{-}X_n'\Big)$

The Maximum Neighbors Condition requires the network structure (edges) to be sparse. More specifically, it requires that the number of direct neighbors not increase faster than $O(\log n)$ when the number of influential individuals increases at speed $o\left(\frac{\sqrt{n}}{\log n}\right)$. This rate can be improved when the number of influential individuals is fixed. The Maximum Neighbors Condition is an asymptotic bound on the number of neighbors for each individual as the network increases. This condition is required to prevent shrinkage bias carried from the first stage LASSO estimation from growing faster than $o(1/\sqrt{n})$ in the second stage.

The Variance Condition requires the variance-covariance matrix to converge to a limit. In classical SARs, the spatial weight matrix is assumed to be uniformly row sum bounded. This assumption implies the Variance Condition but imposes restrictions on the network structure. Each individual may only connect with a finite number of neighbors. In my case, the identification of an influential individual comes from the difference in responses between neighbors that solely connect with her and individuals who connect with no influential individuals. For example, consider two groups of individuals that have the same characteristic X where one group all connects with individual j and the other does not. If the mean response of the two groups is significantly different, we can conclude that j is influential. To identify the influence of individual j as a fixed effect, the number of individuals affected by individual j must grow as the sample size increases. As a result, the row sum for influential individuals cannot be bounded by a fixed number.

The heterogeneous endogenous effects model is identified under assumptions 1-5 as a linear system with a unique solution. I discuss the identification of my model with cliques and with multiple networks in the following two sections.

3.2 Identification Assumptions with Cliques

Recall the heterogeneous endogenous effects model with cliques, represented as follows:

$$D_n = \Big(M_n \circ D_n\Big)\eta_0 + M_n D_n \gamma_0 + X_n \beta_0 + \epsilon_n$$

Define global leaders as those influential individuals who influence multiple cliques and whose neighborhoods increase as n increases. Define local leaders as influential individuals who are not global leaders.

Assumption 1'. *Among n individuals in the network, let $S_n \subset \{1, 2, \cdots, n\}$ be the set of global leaders. Let $s_n = |S_n|$ be the number of elements in S_n. Assume:*

$$s_n = o\left(\frac{\sqrt{n}}{\log n}\right), \qquad \text{as } n \to \infty$$

Assumption 1' only requires the number of global leaders to be sparse. My model does not impose any restriction on the number of local leaders. As a result, it does not rule out situations where everyone is (locally) influential. Local leaders' influence will be captured by the γ_0, coefficient of classical spatial lag.

To ensure invertibility of the matrix $\Big(I_n - M_n \circ \eta_0 - M_n\gamma_0\Big)$, I modify the first part of assumption 2 as:

Assumption 2'. *There exists an* $\eta_{\max} < 1$ *such that* $\|\eta_0 + \gamma_0\|_\infty \leq \eta_{max}$

Similar to assumption 2, this assumption excludes unit root processes. Since there exists a local level influence γ_0 in the network, global level influence η_0 needs to be further bounded above by 1. As a result, equation (6) can be transformed into the following:

$$E(D_n) = X_n\beta_0 + \Big(M_n \circ X_n\Big)(\beta_0\eta_0) + M_nX_n(\beta_0\gamma) + \sum_{i=2}^{\infty}\Big(M_n \circ \eta_0 + \gamma M_n\Big)^i \beta_0 X_n \tag{13}$$

The details of the algebra can be found in the appendix. Equation (6) introduces one more coefficient γ_0 compared with equation (4). As a result, assumption 3 is modified to include an extra instrument M_nX_n, which is also the classic instrument used in equation (2):

Assumption 3'. $\big[X_n, (M_n \circ X_n)_S, M_nX_n\big]$ *is full rank.*

Assumption 3' is similar to assumption 3 and requires the additional instrument M_nX_n to be linearly independent with $\big[X_n, (M_n \circ X_n)_S\big]$. Assumptions 4' and 5' are the same as assumption 4 and 5.

3.3 Identification Assumptions with Multiple Networks

Recall the heterogeneous endogenous effects model with multiple networks, represented as follows:

$$D_n = \sum_{j=1}^{q}\left(M_n^j \circ D_n\right)\eta_0^j + X_n\beta_0 + \epsilon_n$$

First notice that the number of coefficients in this model becomes $nq + k$. The number of observed networks q is also allowed to increase as the number of observations increases. As a result, the sparsity assumption will be imposed on both the influential individuals and the effective networks. I assume that some of the networks are completely irrelevant (i.e. $\eta_0^j = 0$) and that relevant networks are not necessarily passing influence for everyone (i.e. $\eta_0^j \neq 0$ but $\eta_{0,i}^j = 0$ for some i).

Second, to ensure invertibility, for any matrix norm $\|.\|$:

$$\left\|\sum_{j=1}^{q}\Big(M_n^j \circ \eta_0^j\Big)\right\| \leq \sum_{j=1}^{q}\Big\|\Big(M_n^j \circ \eta_0^j\Big)\Big\| \leq \sum_{j=1}^{q}\|\eta_0^j\|_\infty\Big\|\left(M_n^j\right)\Big\|$$

Because M_n^j is the adjacency matrix such that each entry is 0 or 1, $\sum_{j=1}^q \|\eta_0^j\|_\infty < 1$ guarantees the invertibility of $I - \sum_{j=1}^q \left(M_n^j \circ \eta_0^j\right)$.

Third, I require $\left[X_n, \left(M_n^1 \circ X_n\right)_S, \left(M_n^2 \circ X_n\right)_S, \cdots, \left(M_n^q \circ X_n\right)_S\right]$ to be full rank. Compared with the standard model, this assumption requires the independence condition to hold across different networks. Again, we cannot identify the source of influence if two influential individuals connect to the same neighbors.

Fourth, I assume conditions that guarantee a consistent selection of square-root sparse group LASSO. The difference between sparse group LASSO and standard LASSO estimator is we have additional information on the groups of the regressors. As a result, a set of weaker conditions can be applied to guarantee consistent selection. And, finally, the Maximum Neighbor Condition needs to be satisfied in all q adjacency matrices. Since the five conditions for multiple networks are very similar to assumption 1-5, I list them formally in the appendix as assumption 1*-5*.

4 Estimation

I propose an estimator similar to the two-stage least square method but use LASSO in both stages. The estimator proposed here differs from the double selection or double machine learning (DML) estimator as proposed in Belloni et al. (2014a) and Chernozhukov et al. (2017). Double selection or DML is targeting the treatment effect and applys "orthogonalization" to ignore those penalized regressors. My procedure is targeting those penalized regressors and trying to provide valid inference for them. My estimator is a type of two-stage LASSO (2SLSS) estimator as in Zhu (2016). In this section, I define this 2SLSS procedure and propose a bias corrected version of the estimator. I show how this procedure can be extended to estimate my model for networks consisting of multiple cliques and my model for multiple networks.

4.1 Two-Stage LASSO Estimator

I propose to estimate equation (4) using the following estimator:

Two-Stage LASSO Estimator:

- First Stage:

$$(\tilde{\beta}, \tilde{\eta}) = \arg\min_{\beta,\eta} \|D_n - X_n\beta - \Big(M_n \circ X_n\Big)\eta\|_2 + \lambda|\eta|_1 \tag{14}$$

Obtain a LASSO fitting $\hat{D}_n$

$$\hat{D}_n = X_n\tilde{\beta} + \Big(M_n \circ X_n\Big)\tilde{\eta}$$

- Second Stage:

$$(\hat{\beta}, \hat{\eta}) = \arg\min_{\beta,\eta} \|D_n - \Big(M_n \circ \hat{D}_n\Big)\eta - X_n\beta\|_2 + \lambda|\eta|_1 \tag{15}$$

As shown in section 3, $\Big(M_n \circ D_n\Big)$ is correlated with ϵ_n. Thus equation (4), equation (6) and equation (8) cannot be estimated directly using LASSO or sparse group LASSO. The instruments proposed in section 3 are $[X_n, (M_n \circ X_n)_S]$. We do not observe the set S but note that $[X_n, (M_n \circ X_n)]$ is a set of regressors that contains the valid instruments.

The two-stage least square method can be used to address endogeneity in SARs as in Lee (2003). In the first stage, $M_n X_n$ are used as instruments to estimate D_n. In the second stage, $M_n\hat{D}_n$ is used to replace $M_n D_n$ to avoid endogeneity.

Following the same idea, I estimate a first stage using $[X_n, (M_n \circ X_n)]$. Since there are $n + k$ regressors, I use the square-root LASSO to select those instruments in set S. I choose the square-root LASSO over standard LASSO to avoid a pre-estimation of the unknown variance of the error term σ^2. I construct a synthetic $\hat{D}_n$ variable using square-root LASSO estimates. In the second stage, I replace D_n with $\hat{D}_n$ in the regressors and estimate the coefficients $\hat{\eta}$ using the square-root LASSO again.

The statistical properties of two-stage estimators using LASSO have been studied in Zhu (2016), where she derives bounds for the estimator and proves consistency of model selection in a general setting. Zhu (2016) studied the over identified case where the number of endogenous regressors goes to infinity while the number of instruments for each regressor also goes to infinity. Although infinitely many instruments can be constructed by considering $\{(M_n \circ X_n), (M_n \circ X_n)^2, \cdots\}$ in my case, I studied the just identified situation using the instruments proposed in section 3, where the number of endogenous regressors is the same as the number of instruments and both go to infinity.

4.2 De-sparse 2SLSS Estimator

The estimator $\hat{\beta}$ and $\hat{\eta}$ suffer from LASSO shrinkage bias. Moreover, post model selection inference conditioning on the selected model $\hat{S}_n = \{i | \hat{\eta} \neq 0\}$ suffers from the omitted variable bias and thus is not uniformly valid (see Leeb and Potscher, 2005, 2008, 2009). I construct a "de-sparse" estimator under my setting and derive the asymptotic distribution for it. I propose the following de-sparse LASSO estimator:

De-sparse 2SLSS Estimator:

$$\hat{e} = \hat{\eta} + \hat{\Theta}(M_n \circ \hat{D}_n)'(D_n - X_n\hat{\beta} - (M_n \circ \hat{D}_n)\hat{\eta})/n$$

$$\hat{b} = \hat{\beta} - (X_n'X_n)^{-}X_n'(M_n \circ \hat{D}_n)'(D_n - X_n\hat{\beta} - (M_n \circ \hat{D}_n)\hat{\eta})/n$$

$\hat{\beta}$ and $\hat{\eta}$ are estimators from the 2SLSS. $\hat{\Theta}$ is defined by the nodewise regression as in Meinshausen and Bühlmann (2006). Nodewise regression explores the correlation between the columns of the design matrix $W_n(M_n \circ \hat{D}_n)$ by regressing each column on all the rest of the columns while penalizing the coefficients. An approximation of the inverse of the matrix $\frac{1}{n}(M_n \circ \hat{D}_n)'W_n(M_n \circ \hat{D}_n)$ can be constructed based on nodewise regression. Further, define $\hat{S}_n = \{i | \hat{\eta} \neq 0\}$, which represents the LASSO selected active set. The estimators $(\hat{e}, \hat{b})$ are adjusted for the LASSO shrinkage bias and are a consistent estimator for β and η. They are similar to the estimators proposed in van de Geer et al. (2014), but are constructed through a two-stage process as well as using square-root LASSO.

The de-sparse LASSO estimator does not depend on the selected active set. Thus, it does not suffer from the non-uniformity problem. Notice that the double selection or DML method proposed in Belloni et al. (2014a) and Chernozhukov et al. (2017) could also be applied to conduct inference on $\hat{\beta}$. Belloni et al. (2014b) shows the first order equivalence of the double selection method and the de-sparse method. On the other hand, the main interest of this paper is the coefficients $\hat{\eta}$. The double selection method does not provide a way to conduct inference on the model selection.

As the instruments are known in my case, I can derive the asymptotics for my estimator explicitly. By considering a sparse network structure (e.g. Maximum Neighbors Condition), shrinkage bias from the first stage is negligible ($o(1/\sqrt{n})$). I will defer the proof of consistency for the LASSO selected set $\hat{S}_n$ as well as consistency and asymptotic distribution for my estimator $(\hat{e}, \hat{b})$ to section 5. In the remainder of this subsection, I will define the estimators for the two extended models.

4.3 2SLSS with Cliques

To estimate equation (6), I propose the following 2SLSS:

Two-Stage LASSO Estimator with Homogenous Effects:

- First Stage:

$$(\tilde{\beta}, \tilde{\gamma}, \tilde{\eta}) = \arg\min_{\beta,\gamma,\eta} \|D_n - X_n\beta - M_nX_n\gamma - \Big(M_n \circ X_n\Big)\eta\|_2 + \lambda(|\eta|_1 + |\gamma|)$$

  Obtain a LASSO fitting $\hat{D}_n$

$$\hat{D}_n = X_n\tilde{\beta} + M_nX_n\tilde{\gamma} + \Big(M_n \circ X_n\Big)\tilde{\eta}$$

- Second Stage:

$$(\hat{\beta}, \hat{\gamma}, \hat{\eta}) = \arg\min_{\beta,\gamma,\eta} \|D_n - M_n\hat{D}_n\gamma - \Big(M_n \circ \hat{D}_n\Big)\eta - X_n\beta\|_2 + \lambda(|\eta|_1 + |\gamma|)$$

The estimator is similar to that for the previous model except that the classical spatial lag M_nX_n is now included in the estimation. In the above estimator, I penalize ηs and γ at the same rate because I assume no prior knowledge of these two effects. One can penalize them at a different rate or not penalize γ when influence from local leaders is more likely than that from global leaders or vice versa. Since γ and ηs are both penalized coefficients, a similar de-sparse LASSO estimator can be constructed for γ:

De-sparse 2SLSS Estimator with Cliques:

$$\hat{r} = \hat{\gamma} + \hat{\Theta}(M_n\hat{D}_n)'(D_n - X_n\hat{\beta} - M_n\hat{D}_n\tilde{\gamma} - (M_n \circ \hat{D}_n)\hat{\eta})/n$$

With an abuse of notation here, $\hat{\Theta}$ is changed compare to the one in section 4.2. In this scenario, it is an approximation for the inverse of the matrix $\frac{1}{n}[M_n\hat{D}_n, (M_n \circ \hat{D}_n)]'W_n[M_n\hat{D}_n, (M_n \circ \hat{D}_n)]$ instead of $\frac{1}{n}(M_n \circ \hat{D}_n)'W_n(M_n \circ \hat{D}_n)$.

4.4 Multiple Networks

When multiple networks exist, each individual will have network-specific endogenous effects. The number of unknown coefficients increases from $n + k$ to $nq + k$ compared with the standard case. These coefficients can also be classified into q different groups based on networks. By applying the sparsity assumption to the relevant networks, we can estimate the model using the square-root sparse group LASSO instead of the square-root LASSO and propose the following estimator. The square-root sparse group LASSO penalizes both the l_1 and l_2 norm in each group. It can identify all the relevant groups under weaker assumptions compared with the square-root LASSO estimator.

Two-Stage LASSO Estimator with Multiple Networks:

- First Stage:

$$(\tilde{\beta}, \tilde{\eta}) = \arg\min_{\beta,\eta} \left\{ \left\| D_n - X_n\beta - \sum_{j=1}^{q} (M_n^j \circ X_n)\eta^j \right\|_2 + \left(\sum_{j=1}^{q} \left(\lambda_1 \|\eta^j\|_2 + \lambda_2 \|\eta^j\|_1 \right) \right) \right\}$$

Obtain a LASSO fitting $\hat{D}_n$

$$\hat{D}_n = X_n\tilde{\beta} + \sum_{j=1}^{q} (M_n^j \circ X_n)\tilde{\eta}^j$$

- Second Stage:

$$(\hat{\beta}, \hat{\eta}) = \arg\min_{\beta,\eta} \left\{ \left\| D_n - X_n\beta - \sum_{j=1}^{q} (M_n^j \circ \hat{D}_n)\eta^j \right\|_2 + \left(\sum_{j=1}^{q} \left(\lambda_1 \|\eta^j\|_2 + \lambda_2 \|\eta^j\|_1 \right) \right) \right\}$$

The square-root sparse group LASSO introduces two tuning parameters, λ_1 and λ_2, to penalize both the l_1 and the l_2 norm in each network. Similar to the LASSO estimator, the geometric shape of the penalties allows the square-root sparse group LASSO to identify sparsity not only within each network (group) but also among networks (groups). In other words, some networks could be completely irrelevant (i.e. $\eta^j = 0$) and within relevant networks, some individuals can have no influence on their neighbors (i.e. $\eta^j \neq 0$ but $\eta_i^j = 0$ for some i). The sparse group Lasso was first proposed by Simon et al. (2013). They provide an algorithm to solve this problem without deriving any statistical properties. I modify the estimator by taking the square-root of the mean square error term in the minimization problem. Similar to the square-root LASSO proposed in Belloni et al. (2011), the method becomes pivotal since it does not require a pre-estimation of the standard deviation σ. I will prove the statistical properties of square-root sparse group LASSO in appendix.

The de-sparse LASSO estimator for square-root sparse group LASSO is proposed as follows:

De-sparse 2SLSS Estimator for Square-root Sparse Group LASSO:

$$\hat{e}_m = \hat{\eta} + \hat{\Theta}_Z \hat{Z}_n'(D_n - \hat{Z}_n\hat{\eta} - X_n\hat{\beta})/n$$
$$\hat{b}_m = \hat{\beta} - (X_n'X_n)^{-1}X_n'\hat{Z}_n\hat{\Theta}_Z X_n'(D_n - \hat{Z}_n\hat{\eta} - X_n\hat{\beta})/n$$

---

where $\hat{Z}_n = \Big[(M_n^1 \circ \hat{D}_n), (M_n^2 \circ \hat{D}_n), \cdots (M_n^q \circ \hat{D}_n)\Big]$ and $\hat{\Theta}_Z$ is the approximation of the inverse of the matrix $\frac{1}{n}\hat{Z}_n'W_n\hat{Z}_n$.

5 Statistical Properties

In this section, I consider the statistical properties for the de-sparse 2SLSS estimators $(\hat{e}, \hat{b}, \hat{S}_n)$ proposed in section 4. I show consistency and derive asymptotic normality for my de-sparse estimators. In order to show consistency and asymptotic normality for the de-sparse 2SLSS estimator with multiple networks, I derive the statistical properties for square-root sparse group LASSO, which have not been previously explored in statistics literature.

5.1 Consistency

The proof of consistency has two parts. 1) I show that the selected active set converges to the true non-zero parameter set. 2) I show that the de-sparse estimators converge to the true parameters.

Theorem 1. *In heterogeneous endogenous effects model and with assumption 1-5, if* $\lambda \propto \sqrt{\frac{\log n}{n}}$,

1) $\lim_{n\to\infty} \mathbb{P}(\hat{S}_n = S) = 1$; *2)* $\hat{e} \to \eta_0$; *3)* $\hat{b} \to \beta_0$.

The consistency of the LASSO active set $\hat{S}_n$ follows from assumption 4 as is shown in Zhao and Yu (2006). The consistency of $\hat{e}$ and $\hat{b}$ can be shown by taking the Karush-Kuhn-Tucker conditions of the LASSO minimization problem in the second stage. The shrinkage bias carried from the first stage: $\frac{1}{n}(M_n \circ \hat{D}_n)'\Big(M_n \circ (\hat{D}_n - D_n)\Big)\eta_0$ can be shown of order $o(1/\sqrt{n})$. The details of this proof are provided in the appendix.

In the presence of cliques, if γ is penalized, it can be treated as one of the components in η. On the other hand, if it is not penalized, it can be treated as one of the components in β. The consistency follows directly from Theorem 1:

Corollary 1. *In the heterogeneous endogenous effects model with cliques and under assumptions 1'-3', assumptions 4-5, if* $\lambda \propto \sqrt{\frac{\log n}{n}}$,

1) $\lim_{n\to\infty} \mathbb{P}(\hat{S}_n = S) = 1$; *2)* $\hat{e} \to \eta_0$; *3)* $\hat{r} \to \gamma_0$; *4)* $\hat{b} \to \beta_0$.

In the presence of multiple networks, theorem 2 summarizes the consistency results.

Theorem 2. *In the heterogeneous endogenous effects model with multiple networks and under assumptions 1*-5*, if* $\lambda_1 \propto \sqrt{\frac{\log n}{n}}$ *and* $\lambda_2 \propto \sqrt{\frac{\log n}{n}}$,

1) $\lim_{n\to\infty} \mathbb{P}(\hat{S}_n = S) = 1$; *2)*$\hat{e}_m^j \to \eta_0^j$ *for* $j = 1, \cdots, q$; *3)* $\hat{b}_m \to \beta_0$.

The derivation of theorem 2 is similar to that of theorem 1 except that the square-root LASSO is replaced with the square-root sparse group LASSO. Theorem 1, corollary 1 and theorem 2 establish the consistency for my de-sparse 2SLSS estimators.

5.2 Asymptotics

Post inference or inference after model selection are not uniformly valid. Define the set:

$$B(s) = \{\eta \in \mathbb{R}^n | \{j, \eta_j \neq 0\} \leq s\}$$

As shown in Leeb and Potscher (2005), Leeb and Potscher (2008), and Kasy (2015)

$$\sup_{\eta_0 \in B(s)} \left| P\left(\frac{\sqrt{n}(\hat{\eta}_j - \eta_0)}{\hat{V}_j} < t \right) - \Phi(t) \right| \nrightarrow 0 \tag{16}$$

where $\hat{\eta}_j$ can be any estimator based on a selected model, $\hat{V}_j$ is the associated standard deviation and $\Phi(t)$ is the normal CDF function. When η_j is of order $O(1/\sqrt{n})$, the probability that LASSO fails to select this regressor into the active set can be non-zero. The resulting post model selection estimator will carry the omitted variable bias because of the exclusion of regressor j from the model. Thus, post inference conditioning on the selected model cannot converge to the true parameters uniformly over the models defined by sparsity.

On the other hand, the de-sparse LASSO estimator is uniformly valid since the inference is not conditioned on the selected model (see van de Geer et al., 2014). I follow the same idea and show that my de-sparse 2SLSS estimators achieve asymptotic normality with square-root LASSO and square-root sparse group LASSO.

Theorem 3. *In the heterogeneous endogenous effects model and under assumption 1-5, if* $\lambda \propto \sqrt{\log n/n}$

$$\sqrt{n}(\hat{e} - \eta_0) = E_1 + \Delta_1, \qquad E_1 \sim N(0, \sigma^2 \Theta_1 diag(\Gamma)\Omega diag(\Gamma)\Theta_1')$$
$$\sqrt{n}(\hat{b} - \beta_0) = E_2 + \Delta_2, \qquad E_2 \sim N(0, \sigma^2 \Theta_2 diag(\Gamma)\Omega diag(\Gamma)\Theta_2')$$

and

$$\|\Delta_1\|_\infty = o_p(1), \quad \|\Delta_2\|_\infty = o_p(1), \quad \Gamma = \lim_{n\to\infty}(I - M_n \circ \eta_0)^{-}X_n\beta_0,$$
$$\Theta_1 = \lim_{n\to\infty}\hat{\Theta}, \quad Z_n = (M_n \circ \hat{D}_n), \quad \tilde{Z}_n = X_n(X_n'X_n)^{-1}X_n'Z,$$
$$\Theta_2 = \lim_{n\to\infty}\frac{1}{n}\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)'X_n(X_n'X_n)^{-1}X_n'\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)$$

Theorem 3 shows that the 2SLSS estimator achieves normality at the standard rate $\sqrt{n}$. The shifts Δ_1 and Δ_2 represent the bias from using nodewise regression and they are shown to be $o_p(1)$ with the proper choice of tuning parameters.

Corollary 2. *In the heterogeneous endogenous effects model with cliques and under assumption 1'-3', and assumptions 4-5, if* $\lambda \propto \sqrt{\log n/n}$

$$\sqrt{n}\begin{pmatrix}(\hat{e} - \eta_0)\\(\hat{r} - \gamma_0)\end{pmatrix} = E_1 + \Delta_1, \qquad E_1 \sim N(0, \sigma^2 \Theta_1 diag(\Gamma)\Omega diag(\Gamma)\Theta_1'),$$
$$\sqrt{n}(\hat{b} - \beta_0) = E_2 + \Delta_2, \qquad E_2 \sim N(0, \sigma^2 \Theta_2 diag(\Gamma)\Omega diag(\Gamma)\Theta_2'),$$

and

$$\|\Delta_1\|_\infty = o_p(1), \quad \|\Delta_2\|_\infty = o_p(1), \quad \Gamma = \lim_{n\to\infty}(I - M_n \circ \eta_0)^{-}X_n\beta_0,$$
$$\Theta_1 = \lim_{n\to\infty}\hat{\Theta}, \quad Z_n = [(M_n \circ \hat{D}_n), M_n\hat{D}_n], \quad \tilde{Z}_n = X_n(X_n'X_n)^{-1}X_n'Z,$$
$$\Theta_2 = \lim_{n\to\infty}\frac{1}{n}\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)'X_n(X_n'X_n)^{-1}X_n'\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)$$

For my setting with multiple networks, I derive the following results:

Theorem 4. *In the heterogeneous endogenous effects model with multiple networks and under* *assumptions 1*-5*, if* $\lambda_1 \propto \sqrt{\frac{\log n}{n}}$ *and* $\lambda_2 \propto \sqrt{\frac{\log n}{n}}$

$$\sqrt{n}(\hat{e}_m - \eta_0) = E_{m1} + \Delta_{m1}, \qquad E_{m1} \sim N(0, \sigma^2 \Theta_{Z1} diag(\Gamma) \Omega_m diag(\Gamma) \Theta'_{Z2})$$
$$\sqrt{n}(\hat{b}_m - \beta_0) = E_{m2} + \Delta_{m2}, \qquad E_{m2} \sim N(0, \sigma^2 \Theta_{Z2} diag(\Gamma) \Omega_m diag(\Gamma) \Theta'_{Z2}),$$

and

$$\|\Delta_{m1}\|_\infty = o_p(1), \quad \|\Delta_{m2}\|_\infty = o_p(1),$$
$$\Theta_{Z1} = \lim_{n\to\infty} \hat{\Theta}_Z, \quad Z_n = (M_n \circ \hat{D}_n), \quad \tilde{Z}_n = X_n(X_n'X_n)^{-1}X_n'Z,$$
$$\Theta_{Z2} = \lim_{n\to\infty} \frac{1}{n}\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)' X_n(X_n'X_n)^{-1}X_n'\Big(I - Z_n\hat{\Theta}\tilde{Z}_n'/n\Big)$$

The proof of Theorem 2 and Theorem 4 requires the following results from the square-root sparse group LASSO: 1) Bounds on the prediction, i.e.

$$\left\|\sum_{j=1}^{q}(M^j \circ X_n)(\hat{\eta}^j - \eta_0^j) + X_n(\hat{\beta} - \beta_0)\right\|_2 \lesssim \lambda$$

and 2) Consistency of selection i.e. $\hat{S}_n = S$. I prove these two statistical properties in the appendix.

6 Simulations

In this section, I report Monte Carlo simulation results for my heterogeneous endogenous effects model and its extension with cliques and with multiple networks. My results are robust when applied to networks generated by different algorithms and to networks of different sizes.

6.1 Heterogeneous Endogenous Effects Model

To assess the finite sample performance of my estimator for the heterogeneous endogenous effects model, I use the Erdos-Renyi algorithm to simulate a network of size n. Individuals are added into the graph one at a time. When one individual is added to the network, she has probability p of generating a link with all existing individuals independently. I choose $p = 0.1$ and $p = 0.2$ in the simulation. I avoid a large p because collinearity among regressors may arise when links become very dense, violating assumption 5.

I set the first 5 individuals to be influential by letting their coefficients η_j be non-zero. To guarantee the existence of endogenous effects, I arbitrarily specify the connections among these five individuals. The adjacency matrix M_n for the five influential individuals is given in the appendix. If the connections among these five individuals are not fixed, there is a possibility that no connections are formed among these five and thus there is no endogeneity in the network. In this case, the results will be too good in such a case.

The true parameters are fixed as $\beta_0 = 3$, $\eta_{0,1} = \eta_{0,2} = \eta_{0,3} = \eta_{0,4} = \eta_{0,5} = 0.5$, and $\eta_{0,j} = 0$ for $j > 5$. Individual characteristics X_n are generated from a standard normal distribution.

Individual outcomes Y_n are then generated as $Y_n = (I - M_n \circ \eta_0)^{-1}(X_n\beta_0 + \epsilon_n)$ where ϵ_n is drawn independently from a standard normal distribution.

I use (M_n, X_n, Y_n) as observations and apply my two-stage LASSO estimator. I construct the desparse 2SLSS estimator and repeat the above process 200 times in a manner similar to van de Geer et al. (2014).

I report the average coverage probability (Avgcov) and average length (Avglength) of confidence intervals for the coefficients for influential individuals, $\{\eta_1, \cdots, \eta_5\}$, the coefficient for individual characteristics, β_0, and the coefficients for non-influential individuals, the η_js $(j > 5)$. For example:

$$\text{Avgcov } S_0 = s_0^{-1} \sum_{j \in S_0} \mathbb{P}[\eta_{0,j} \in CI_j] \tag{17}$$

$$\text{Avglength } S_0 = s_0^{-1} \sum_{j \in S_0} length(CI_j) \tag{18}$$

I separately report the average coverage and average length for each of the five influential individuals. As shown in appendix table A1, the coverage is around the nominal 95% level and the length of the confidence intervals decreases as the sample size grows.

Since we can construct confidence intervals for all n coefficients, joint inference can be performed under the control of False Discover Rate (FDR). As shown in equation (19), the power reported in appendix table A1 represents the average percentage in the active set (i.e. $\{1, 2, 3, 4, 5\}$) that is significant after controlling for the False Discover Rate (FDR) at 5% using the Benjamini-Hochberg method. The FDR reported in appendix table A1 represents the average percentage of the non-active set (i.e. $\{6, 7, \cdots, n\}$) that is significant after controlling the FDR at 5% using the Benjamini-Hochberg method. The exact definition is as in equation (20).

$$\text{Power } = s_0^{-1} \sum_{j \in S_0} \mathbb{P}[H_{0,j} \text{ is rejected}] \tag{19}$$

$$\text{FDR} = \sum_{j \in S_0^c} \mathbb{P}[H_{0,j} \text{ is rejected}] / \sum_{j=1}^{n} \mathbb{P}[H_{0,j} \text{ is rejected}] \tag{20}$$

The power varies because the networks change when the sample size increases. It is strictly increasing when the network is sparse (i.e. $p = 0.1$). The power decreases in the $p = 0.2$ case as the problem of endogeneity increases when the network is dense. The empirical FDR is controlled well, which all under the 5% rate. Notice that the confidence interval's length is large when the sample size equals 50. This is because when the number of individuals is small, some individuals might only connect to 1 or 2 other individuals. This means that the regressors that represent this individual are all 0s except for a small numbers of non-zero terms, which leads to a large standard error.

The two-stage LASSO estimator requires the choice of two tuning parameters (i.e. the two λs from both stages as in section 4.1). Moreover, when calculating $\hat{\Theta}$ in the De-sparse 2SLSS estimator (section 4.2) and using the nodewise regression, one also need to choose a tuning parameter. Following the suggestion in Belloni et al. (2011) and Zhu (2016), I use a benchmark choice of λ for the nodewise regression (i.e. $\lambda \propto \sqrt{\log(n)/n}$). For the first stage, I choose $\lambda \propto (\log(n)/n)^{1/4}$ to "over-penalize", and for the second stage, I use cross-validation to pick λ to enhance finite sample performance.

I further increase the number of influential individuals to 10 and report the results in appendix table A2. Again, to guarantee the existence of endogeneity, the adjacency matrix for these ten individuals is set as shown in the appendix. All average coverages and average confidence interval lengths are separately reported for these ten individuals.

The choice of the tuning parameters is similar to those used to generate appendix table A1 for networks with 50 and 200 individuals. For networks with 500 individuals, I use benchmark λ to replace cross validation in the second stage. The idea is to show the converge of the process, such that valid coverage can still be generated under theory guide tuning parameters (see Belloni et al., 2011).

As shown in appendix table A2, all coverages are very close to the nominal levels. The average lengths of confidence intervals is slightly larger compared with appendix table A1. This is due to the increase in influential individuals; it is more difficult to differentiate them from those irrelevant individuals.

Appendix table A3 presents the result when a network is generated using the Watts-Strogatz mechanism or the "small world" network. Define the pN (even number) as the mean degree for each node and a special parameter $\omega = 0.4$. The WattsStrogatz mechanism works as follows:

- construct a graph with N nodes each connected to pN neighbors, which $\frac{pN}{2}$ on each side.
- For each node n_i, take every edge (n_i, n_j) with $i < j$ and rewrite it with probability ω. Rewrite means replace (n_i, n_j) with (n_i, n_k) where k is choosing uniformly among all nodes that are not currently connected with n_i

The influential individuals are chosen as the 1st, 5th, 15th, 40th and 50th individuals in the network. As shown in appendix table A3, my estimator is robust under a "small world" algorithm. Nominal level is reached as the size of the network grows and the length of confidence intervals is slightly smaller than in the standard case.

6.2 Heterogeneous Endogenous Effects Model with Cliques

Appendix table A5 presents results for the heterogeneous endogenous effects model with cliques. The outcome variable Y_n is now generated as $Y_n = (I - M_n \circ \eta_0 - M_n\gamma_0)^{-1}(X_n\beta_0 + \epsilon_n)$. The coefficient of the homogeneity effects γ_0 is set at 0.05.

The choice of the tuning parameters is similar to that used to generate appendix table A1 for networks with 50 and 200 individuals. For networks with 500 individuals, I use benchmark λ (i.e. $\lambda \propto \sqrt{\log(n)/n}$) to replace cross validation in the second stage.

The coverage is above the 95% nominal level in all cases. I also report the mean coverage and average length of the confidence interval for the coefficient of the homogeneous effects. My model gives above 95% coverage in all cases. I also report the empirical probability of rejecting a null hypothesis of zeros effects at 95% nominal level. The probability of rejecting the test converges to 1 when the sample size grows to 500.

6.3 Heterogeneous Endogenous Effects Model with Multiple Networks

In this Monte Carlo exercise, I include two different networks generated by the Erdos-Renyi algorithm, where one is influential and the other is not. I use the two-stage LASSO estimator with multiple networks to estimate the parameters. The square-root sparse group LASSO requires two tuning parameters, one for the l_2 norm and the other for the l_1 norm. I set the two parameters to be equal to each other as the correlations among the columns of the adjacency matrices are very small. The choice of tuning parameters is similar to that used to generate table 1 for networks with 50 and 200 individuals. For networks with 500 individuals, I use benchmark λ instead of cross-validation in the second stage. Appendix table A4 summarizes the results. As in previous results, all coverages exceed the nominal 95% level.

I report the empirical probabilities such that at least one individual is detected in a given network controlling for the FDR at 5% using the Benjamini-Hochberg method. I also report the average number of detections conditioning on at least one individual who is detected in a given network. Appendix table A4 shows that network 1, which is the relevant network, is more likely to be detected in all cases than network 2, the irrelevant network. The average number of identified individuals for network 1 is also more than that of network 2.

7 Empirical Application

I use the proposed estimator to study the importance of different networks in spreading the participation in a micro finance program within rural Indian villages. I show that different kinds of networks have different effects on individuals decisions. I identify the influential individuals in each village. My analysis shows that leaders among agricultural laborers, Anganavadi teachers, construction workers, small business owners and mechanics are very likely to be influential in the villages.

7.1 Background

A non-profit organization named Bharatha Swamukti Samsthe (BSS) has been running micro finance programs in rural southern Karnataka, India since 2007. It provides small loan products to poor women and, through them, to their families. The villages covered by the program are geographically isolated and heterogeneous in terms of caste.

When BSS initially introduces a micro finance program to a village, the credit officers of BSS first approached a number of "predefined leaders", such as teachers, shopkeepers and village elders. BSS held a private meeting with these leaders and explained the program. Then these predefined leaders passed the information onto other villagers. Those who were interested in the program and contacted BSS were trained and assigned to groups to receive credit. Each group consisted of 5 borrowers and group members were jointly liable for loans. Loans were around 10,000 rupees (approximately $200) at an annualized rate of approximately 28%. Note that 74.5 percent of the households in rural area said the monthly income of their highest earning member is less than 5,000 rupees (source: Socio-Economic Caste Census-2011). This loan had to be repaid within 50 weeks.

In 2006, 75 villages in Karnataka were surveyed 6 months before the initiation of the BSS micro finance program. This survey consisted of a village questionnaire and a detailed follow-up survey conducted among a subsample of villagers. The village questionnaire gathered demographic information on all households in a village including GPS coordinates, age, gender, number of rooms, whether the house had electricity, and whether the house had a latrine. The data set also contains information on the "pre-defined leaders" set who helped spread the information to the entire village. The follow-up survey collected data from a villager sample stratified according to age, education level, caste, occupancy, etc. It also asked questions about social network structures along 12 dimensions, including:

- Friends: Name the 4 non-relatives whom you speak to the most.
- Visit-go: In your free time, whose house do you visit?
- Visit-come: Who visits your house in his or her free time?
- Borrow-kerorice: If you needed to borrow kerosene or rice, to whom would you go?
- Lend-kerorice: Who would come to you if he/she needed to borrow kerosene or rice?
- Borrow-money: If you suddenly needed to borrow Rs. 50 for a day, whom would you ask?
- Lend-money: Who do you trust enough that if he/she needed to borrow Rs. 50 for a day you would lend it to him/her?
- Advice-come: Who comes to you for advice?
- Advice-go: If you had to make a difficult personal decision, whom would you ask for advice?
- Medical-help: If you had a medical emergency and were alone at home whom would you ask for help in getting to a hospital?
- Relatives: Name any close relatives, aside from those in this household, who also live in this village.
- Temple-company: Do you visit a temple/mosque/church? Do you go with anyone else? What are the names of these people?

For the 43 villages where micro finance was introduced by the time of 2011, BSS also collects information on which villagers have joined the program. These survey questions reveal the underlying structures for connections among any two individuals in the network. Figure 4 presents all those connections at the household-level in a graph. Each node in the graph represents a household. A black node indicates that the household joined the micro finance program, while a white node indicates that it did not. Bigger nodes represent those households in which at least one family member has been chosen as being among the "pre-defined leaders". An edge between two nodes signifies that the two nodes are connected in at least one of the 12 networks. The darker the color of the edge, the more connections it represents.

This dataset provides an ideal framework for application of the heterogeneous endogenous effects model. First, it allows me to model endogenous effects. An individual may decide to join the micro finance program if her neighbors or friends plan to join. Second, the endogenous effects are individual specific. Given the diversity of the villagers, it is possible that some villagers are more influential than others. Third, it allows me to implement the heterogeneous endogenous effects model with multiple networks. The questions asked regarding multiple dimensions of the network structure allow me to explore which network is most influential.

7.2 Data

In this empirical study, I focus on the 38 villages that have been introduced to the micro finance programs by BSS and have data publicly available [3]. For each village, I can observe both its social network structure and the villagers' decisions about joining the program. I drop the data for one village (Village 46) that contains incorrect entries on the index of households. Appendix table A6 summarizes the descriptive statistics for each village.

Among the 12 questions about the social network structure, 4 pairs essentially capture the same connections among the villagers [4]. Therefore, I consolidate each pair of questions into one dimension:

[3] The dataset can be downloaded from http://web.stanford.edu/ jacksonm/Data.html

[4] Assuming every villager truthfully answers a pair of questions, the adjacency matrices associated with each question are the same. It is also plausible to treat villagers' answers to each question as a separate directed graph. However, these questions do not allow for clear determination of the directions. For example, if villager A visits villager B's house, it is not clear whether villager A influences villager B or vice versa

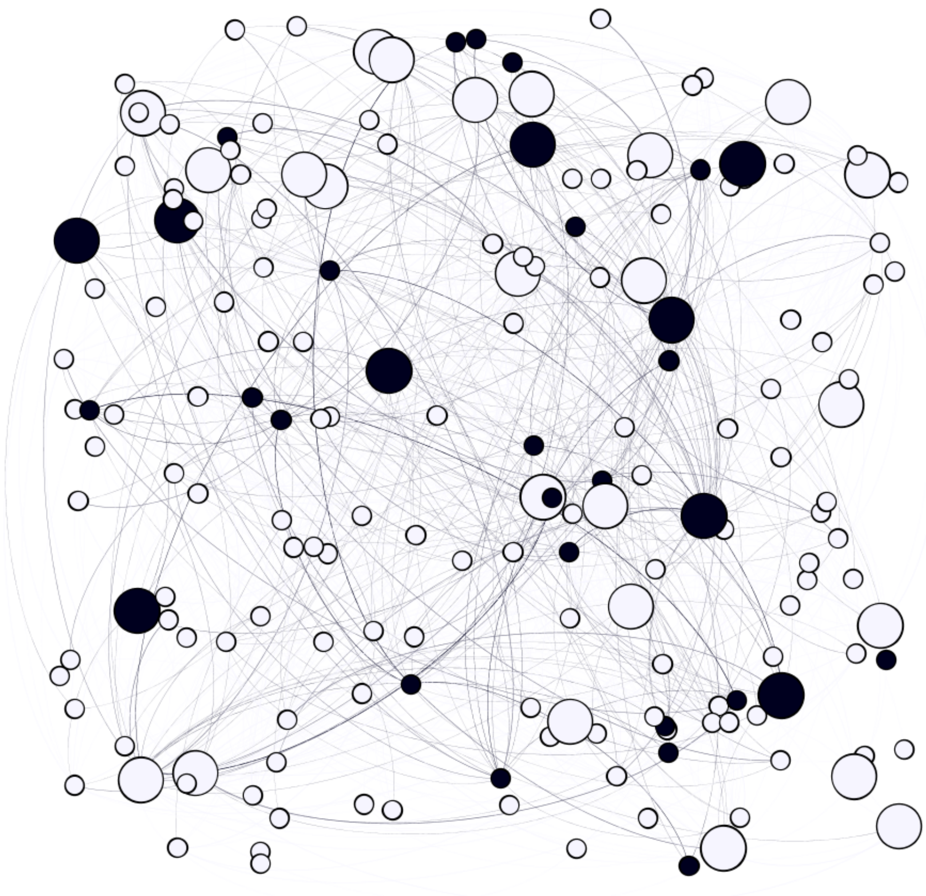


Figure 4: Network in Village 1

$$
\text{Visit-go-come} \begin{cases} \text{In your free time, whose house do you visit?} \\ \text{Who visits your house in his or her free time?} \end{cases}
$$

$$
\text{Borrow-Lend-kerorice} \begin{cases} \text{If you needed to borrow kerosene or rice, to whom would you go?} \\ \text{Who would come to you if he/she needed to borrow kerosene or rice?} \end{cases}
$$

$$
\text{Borrow-Lend-money} \begin{cases} \text{If you needed to borrow Rs.50 for a day, whom would you ask?} \\ \text{Who do you trust enough that if he/she needed to borrow Rs.50 for a} \\ \quad \text{day you would lend it to him/her?} \end{cases}
$$

$$
\text{Help decision} \begin{cases} \text{Who comes to you for advice?} \\ \text{If you had to make a difficult personal decision, whom would you ask} \\ \quad \text{for advice?} \end{cases}
$$

I restructure all the data at the household level as only women are allowed to apply for the micro finance program because the goal of BSS is to support families through the women in them. As a result, a woman's decision to join or not join the micro finance program becomes her family's decision. A connection between two villagers becomes a connection between two families. A "predefined leader" is a villager selected by BSS to help spread information about the micro finance program to the other villagers. At the household level, I use the term "predefined leader" for a household that contains at least one such villager.

7.3 Sparsity and Equilibrium

To demonstrate how my method identifies influential households, I model families' decisions regarding joining the micro finance program as a network game with Bayesian Nash Equilibrium. For household i, let d_i^* be the expected probability that i chooses to join the micro finance program. The decision of household i depends on its neighbors' decisions as well as the types of connections between them. The decision also depends on its characteristics X_i and on unobserved information ϵ_i. Formally, it can be written as:

$$
d_i^* = \sum_{l \in N_i} d_l^* (\sum_{j=1}^{q} \eta_l^j) + x_i \beta + \epsilon_i
$$

Rewritten in matrix form:

$$
D_n^* = \sum_{j=1}^{q} \left(M_n^j \circ D_n^* \right) \eta^j + X_n \beta + \epsilon_n
$$

I assume that only a small number of households are influential over their neighbors. Leaders and followers are usually observed in those rural villages. Big decisions are often made by the village elders or by the more educated among the villagers. BSS recognized the importance of leaders and gathered a group of predefined leaders, asking them to inform the rest of the villagers about their program. I do not consider the local level influence in these villages given the size and how complicated the network structures are. Households are closely connected by these 8 networks as shown in Figure 4 and there is no form of clique visible.

Because the villages are considered geographically isolated, I apply my estimator separately to each of the 38 villages. I use the number of rooms per person in a houshold as the independent variable X_n. Number of rooms per person is a proxy for the wealth in the family. As shown in table 1, it is negatively correlated with the decision to join the micro finance program. The richer the family, the less likely the family is to participate in the micro-finance program. I further check the robustness of my independent variable by including additional controls. The adjacency matrix M_n^j is constructed from the questions in the survey. Households i and k are connected in network j if either i or k reported the other in question j. Finally, d_i^* is replaced with the household's choice.

Table 1: Predictive Power of Characteristics X_n

	(1) Participate	(2) Participate	(3) Participate
Average Num. rooms x100	-8.19***	-7.12***	-3.36***
	(1.30)	(1.30)	(1.12)
Household Size x100		0.45**	0.40**
		(0.21)	(0.20)
Electricity x100			1.29
			(1.26)
Latrine x100			-5.66***
			(1.38)
Average Num. workers x100			0.64*
			(0.50)
Average age x100			-0.27***
			(0.03)
Village Fixed Effects	Y	Y	Y
n	8,375	8,375	8,375
# of villages	37	37	37
R^2	0.05	0.05	0.06

Table 1 provides a robustness check for variable X_n: Average Num. rooms, which used to construct instruments. Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard deviation clustered at village level. Dependent variable is households' decision on whether to join the micro finance program or not. All design control village fixed effects.

The instruments are constructed as $\left(M_n^j \circ X_n\right)$ for $j = 1, 2, \cdots, 8$. I use the heterogeneous endogenous effects model with multiple networks to: 1) Identify the effective networks affecting a household's decision and 2) Identify that households that are leaders in the village and study the association between observable characteristics and leader status. If a new program is going to try to recruit these households, the organizers can target those influential households and try to persuade them to join first.

7.4 Results

7.4.1 Identifying Effective Networks

First, I study how LASSO selects networks. I define a coefficient for a household's endogenous effect in a network as significant according to two different criteria. The first criterion, "Cross-Validation", determines a coefficient to be significant if LASSO predicts the coefficient to be non-zero after cross-validation. The second criterion, "De-sparse", first constructs a bias-adjusted coefficient and calculates its standard error. It then determines a coefficient to be significant if the Benjamini-Hochberg method rejects the null hypothesis of zero effect at the 5% false discovery rate. A network is defined as significant if at least one coefficient for heterogeneous endogenous effects in this network is significant.

Table 2 presents the empirical probability of the 8 networks being significant among the 37 villages. Note that certain types of networks (such as visit go-come) are more likely to pass influence then others (such as temple company). For example, by cross-validation criterion, the visit go-come network is detected as significant in 19 out of the 37 villages (i.e. 51%) whereas temple company is detected as significant in only 5 out of the 37 villages (i.e. 14%). I also present the average number of households associated with significant endogenous effects in each significant network. For example, according to the cross-validation criterion, 342 households in 19 villages have significant coefficients associated with the visit go-come network, which averages to 18 households per detection. On the other hand, 32 households in 5 villages have significant coefficients associated with the temple company network, which averages to 6 households per detection.

In terms of variable selection, if Assumption 4 holds, the cross-validation criterion may consistently select the truly influential households with high probability even in a finite sample. On the other hand, the de-sparse criterion is likely to be conservative because of its use of the false discovery control process. In terms of coefficients estimated, de-sparse estimators are asymptotically consistent. On the other hand, estimates based on the LASSO estimator suffer from shrinkage bias and are not consistent.

Table 3 reports the average absolute heterogeneous endogenous effects within significant networks using the de-sparse estimators. For example, if all else is equal, an additional influential neighbor in the visit go-come network will, on average, increase the probability of joining the micro-finance program by 16%; moreover, an additional influential neighbor in both the visit go-come network and the friendship network will increase the probability of joining the micro-finance program by 16% on average. The magnitude of those coefficients should not be over interpreted as exogenous effects and correlated effects are not considered in the model. Similar to Table 2, certain types of networks (such as visit go-come) pass stronger influence than others (such as temple company). Note that, in most of the cases, networks that are more likely to pass influence also pass stronger influence. The relative network is an exception. Even though the relative network is less likely to pass influence compared to the friendship network, the borrow-lend-money network and the help decision network, it passes stronger influence once it is significant. Table 3 also presents the percentage of positive effects detected among different networks. For networks such as visit-go-come and friendship, more than 70% of influential villagers are "true leaders" – if they decide to join the micro-finance program, their neighbors will follow them and join the program. On the contrary, for the temple company network, it is almost equally likely for neighbors of influential households to either follow the same decision or choose the opposite.

Table 2: Second Stage: influential networks

		visit go-come	borrow-lend keroric	borrow-lend money	friendship	medical help	help decision	relatives	temple company
Cross [1] Validation	probability [3]	51%	43%	41%	41%	30%	32%	30%	14%
	identified [4]	18	12	13	14	9	14	9	6
De-sparse [2]	probability [3]	51%	46%	51%	43%	32%	41%	43%	19%
	identified [4]	3	3	3	2	3	3	3	2

Table 2 reports the probability of detection for different networks among the 38 villages. A network is detected as influential if at least one leader is detected within this network. 1. Cross Validation represents those networks detected from lasso using cross validation. 2. De-sparse represents those networks detected from De-sparse criterion using FDR control. 3. Probability reports the empirical probability that at least one coefficient $\hat{e}_i^j$ is significant in network j. 4. Identified reports the averaged number of significant $\hat{e}_i^j$ in the network j conditioning on the network being detected.

The results in Table 2 and Table 3 suggest villagers are more likely to discuss the micro-finance program when they visit each other, chat with friends, and meet with people to whom they are economically connected. Villagers are not likely to talk about the micro finance program when they go to the temple.

Table 3: Second Stage: estimated $\hat{e}$ for each network

	visit go-come	borrow-lend keroric	borrow-lend money	friendship	medical help	help decision	relatives	temple company
Absolute [1] magnitudes	0.1543	0.1443	0.1245	0.1194	0.1214	0.1217	0.1404	0.0555
Percentage of positive effects [2]	77%	67%	69%	70%	68%	77%	67%	55%

Table 3 summarize the estimated size of influential: $\hat{e}_i^j$. 1. The absolute magnitudes is the mean of $|\hat{e}_i^j|$ and represents the average endogenous effects through network j. 2. Percentage of positive effects is the percentage of $\hat{e}_i^j > 0$ among all leaders in network j and represents the direction of the influence through network j.

To verify my findings above, I provide exogenous evidence using centrality measures. Intuitively, the more a villager is exposed to a network, the more likely she is to be connected to influential villagers, and hence she is more likely to join the program. Following Banerjee et al. (2013), I measure the centrality of each villager in each network through "degree", "closeness", "betweenness" and "eigenvector" (see Appendix for definitions). Then I regress households' decisions on whether to join the micro finance program on each centrality measure separately while controlling for village fixed effects:

$$d_j = C_j^q \beta + \gamma_j + \epsilon_j \tag{21}$$

where d_j is household j's decision; C_j^q is household j's centrality in the network q; γ_j is the village fixed effect; and ϵ_j is the error term.

Table 4 presents the regression results for equation (21). Visit go-come and borrow-lend kerorice are positively correlated with degree, closeness and eigenvector centrality. Friendship, borrow-lend money and medical help are positively correlated with degree and closeness centrality. This is consistent with my findings that these four networks are more effective in passing influence. Meanwhile, neither help decision, relative nor temple company are found to be correlated with any of the centrality measures defined above. This is also consistent with the lower probability of passing influence as found in table 2. Note that none of the networks are found to correlate with betweenness centrality. This is because betweenness centrality is based on the shortest paths in a network, which is not a direct measure of the exposure of an individual to a network.

Table 4: Centrality Measure

	visit go-come	borrow-lend kerorice	borrow-lend money	friendship	medical help	help decision	relatives	temple company
degree	0.0025**	0.0032**	0.0020**	0.0022**	0.0032**	0.0013	0.0035	0.0061
	(0.0009)	(0.0011)	(0.0010)	(0.0010)	(0.0014)	(0.0011)	(0.0019)	(0.0032)
closeness	32.9116***	40.2695***	29.7981**	31.0882***	32.5602**	18.1944	7.9509	231.0770
	(9.5639)	(10.7603)	(9.3901)	(9.0446)	(11.1383)	(9.8242)	(16.5557)	(134.3147)
betweenness	1.3565	0.1751	0.2940	1.6713	1.1662	-0.5728	0.3055	-0.2093
	(1.0101)	(0.8240)	(0.9504)	(1.0207)	(0.8634)	(0.8283)	(0.7736)	(0.2178)
eigenvector	3.6201***	1.5239**	0.1161	1.3927	-0.7338	-0.2387	0.7753	3.3015
	(0.8888)	(0.6250)	(0.8245)	(0.8271)	(0.7709)	(0.7603)	(0.5672)	(3.5585)

Table 4 presents the regression results for equation (21). High centrality implies more likely to be influenced. Table 4 provides reduce formed evidence on which network is influential. Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Definition and description for centrality measures are provided in appendix.

7.4.2 Identifying Influential Households

Second, I focus on how LASSO selects households. I compare the LASSO selected influential households with the BSS selected "predefined leaders". It is important to point out that these "predefined leaders" are *not* necessarily influential villagers in a network. Recall that predefined leaders are a set of villagers that BSS select to help spread the information about the micro finance program. The fact that a villager is selected as a "predefined leader" to *pass information* about the micro finance program does not a priori guarantee her or her family's *influence* – her decision to join the micro finance program may not lead to her neighbors' decisions to join. In the analyses below, I will examine how influential villagers are associated with "predefined leaders" and explore their potential differences.

1. Influential Predefined Households

In table 5, I report results indicating that influential households selected by LASSO partly overlap with "predefined leaders". This is intuitive because some "predefined leaders" such as school headmasters and village elders are highly respected figures in a village. Therefore, their decisions are likely to be followed by others in the village. On average, BSS selected 27 villagers as "predefined leaders" in each village. In comparison, Cross-Validation criterion selects around 22 villagers and De-sparse criterion selects around 6. Furthermore, on average, 4 out of 22 influential villagers (i.e. 19%) selected by Cross-Validation criterion are also BSS "predefined leaders"; 1 out of 6 influential villagers (i.e. 13%) selected by De-sparse criterion are also BSS "predefined leaders". In Table 8 below, I show that small business owners are more likely to be both influential and selected as "predefined leaders".

Table 5: Second Stage: coverage of predefined leaders

	coverage [1]	total number of discovery [2]
Cross Validation [3]	19%	22
De-sparse [4]	13%	6

Table 5 depicts the overlapping between influential households selected by LASSO and "predefined leaders". Predefined leaders are a set of villagers defined by BSS, who helped spread the information about the micro-finance program. 1. Coverage reports the percentage of individuals detected by LASSO and also selected as "predefined leaders" in total detection. 2. Total number of discovery reports the total number of individuals discovered by lasso using each method. 3. Cross Validation represents those individuals identified from lasso using cross validation. 4. De-sparse represents those individuals identified from De-sparse criterion controlling FDR. 6. The average number of predefined leaders in one village is 27.

2. Influential Non-Predefined Households

In this and the following section, I focus on understanding the differences between the influential households selected by LASSO and the "predefined households" selected by BSS. I investigate the likelihood that a household being selected by LASSO or by BSS, as associated with the careers of its family members. More specifically, I regress whether a household is selected as "predefined leader" (Column 1 in table 6), whether a household is selected by LASSO as influential (Column 3 in table 6), and whether a household joins the micro finance program (Column 2 in table 6), separately on dummy variables based on the full set of careers as reported in the survey data controlling for other household characteristics and village fixed effects. The full results of these regressions are reported in appendix table A7.

Table 6 summarizes all careers that have a significant impact on the likelihood of a household being selected by LASSO as influential. Note that except for small business owners, all the other careers in this table are not significantly associated with the likelihood of a household being selected by BSS as being among the "predefined leaders". Over 67% of the villagers are agricultural laborers and 75% of the LASSO selected influential households have agricultural laborers in the family. Anganwadi Teacher is a set of groups that provides pre-school education to the children. They are part of the government's health care system in the rural areas. There are 31 Anganwadi Teachers in all villages, and LASSO detects 7 of their families to be influential. BSS also selects 7 of them as "predefined leaders" but only 2 of the 7 are selected by LASSO as influential. Other careers that are correlated with LASSO selection include police officer, mechanic, and skilled laborers. These are more educated individuals and it seems compelling that they are selected as influential individuals.

Table 7 summarizes all careers that have a significant impact on the likelihood of a household being selected by BSS as being among the "predefined leaders". Poojari are Indian priests in those villages and they are very likely to be included as "predefined leaders". However, they are not likely to influence people to join the micro finance program. Other careers as tailor, hotel workers, veteran, and barber are included as "predefined leaders" because individuals doing these jobs can spread information quickly in the village. However, LASSO does not find these individuals to be influence.

Table 8 reports the counter factual study when selected leaders all decide to join the micro-finance program. The participation rate for non-leaders in the data is 16%. When all "predefined leaders" decide to join, the participation rate for non-leaders will increase to 20%. And when all LASSO selected leaders decide to join, the participation rate for non-leaders will further increase to 33%.

Table 6: Second Stage: LASSO selected leaders' careers

	Predefined leaders	Participate	Selected by LASSO
Agriculture labour	-0.0141	0.0476*	0.0672***
	(0.0136)	(0.0286)	(0.0134)
Anganwadi Teacher	0.0386	0.0664	0.1248**
	(0.0602)	(0.1269)	(0.0593)
Blacksmith	-0.0752	-0.2279	0.1606*
	(0.0927)	(0.1954)	(0.0913)
Construction/mud work	0.0050	0.2199***	0.0562**
	(0.0258)	(0.0544)	(0.0254)
Small business	0.2006***	0.1287***	0.0606***
	(0.0227)	(0.0479)	(0.0224)
Police officer	-0.1459	-0.0374	0.3282*
	(0.1917)	(0.4044)	(0.1890)
Mechanic	0.0106	-0.1237	0.1274**
	(0.0634)	(0.1337)	(0.0625)
Skilled labour/work for company	0.0469	0.0252	0.0809*
	(0.0491)	(0.1036)	(0.0484)
Control other careers	Y	Y	Y
Control village fix effect	Y	Y	Y

Table 6 summarizes all careers that have a significant impact on the likelihood of a household being selected by LASSO from appendix table A7. The first column uses whether one is predefined leaders as response variable, the second column uses whether one joins the micro-finance program as response variable and the third column uses whether one is selected by lasso as response variable. Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Second Stage: predefined leaders' careers

	Predefined leaders	Participate	Selected by LASSO
Small business	0.2006***	0.1287***	0.0606***
	(0.0227)	(0.0479)	(0.0224)
Tailor Garment worker	0.0903***	0.1169*	0.0309
	(0.0304)	(0.0642)	(0.0300)
Hotel worker	0.3299***	0.4257***	0.0759
	(0.0750)	(0.1581)	(0.0739)
Poojari	0.3697***	-0.1542	0.1501
	(0.1369)	(0.2887)	(0.1349)
Veterinary clinic	0.8649***	1.9114***	0.0377
	(0.3314)	(0.6990)	(0.3266)
Barber/saloon	0.4883***	-0.0036	0.0443
	(0.1005)	(0.2119)	(0.0990)
Doctor/Health assistant	0.2691**	0.2703	0.0874
	(0.1053)	(0.2222)	(0.1038)
Control other careers	Y	Y	Y
Control village fix effect	Y	Y	Y

Table 7 summarizes all careers that have a significant impact on the likelihood of a household being selected by BSS as being among the "predefined leaders" from appendix table A7. The first column uses whether one is predefined leaders as response variable, the second column uses whether one joins the micro-finance program as response variable and the third column uses whether one is selected by lasso as response variable. Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Participation Rate when Targeting Different Leaders

	In data	Predefined Leaders	LASSO Leaders
Participation Rate (non-leaders)	16%	20%	33%

Table 8 reports the participation rate of non-leaders when all targeted leaders decided to join. The true participation rate in data is 16%. If all predefined leaders decide to join, the participation rate will increase to 20%. If all LASSO detected leaders decide to join, the participation rate will increase to 33%.

8 Conclusions

In this paper, I propose a novel spatial autoregression model which allows for *heterogeneous* endogenous effects. Specifically, each individual has an individual-specific endogenous effect on her neighbors. My approach is useful for modeling a network with leaders and followers.

I propose a set of instruments as well as a two stage LASSO (2SLSS) method to estimate my model. The instruments are constructed as a function of the independent variables and an adjacency matrix. I use a LASSO type estimator to select the valid instruments in the first stage and the influential individuals in the second stage. I propose a bias correction for my two-stage estimator following van de Geer et al. (2014). I derive the asymptotic normality for my "de-sparse" two-stage LASSO estimator and conduct robust inference including confidence intervals.

My model can be extended to allow for more flexible structures. To apply LASSO, I assume that the number of influential individuals is sparse. I propose heterogeneous endogenous effects model with cliques to incorporate locally influential individuals, where the sparsity assumption is only applied to globally influential individuals. My model can also be extended to situations where there are multiple networks. I propose the use of the square-root sparse group LASSO in my 2SLSS process. I derive the convergence rate and prove the consistency of selection for the square-root sparse group LASSO estimator.

I apply my method to study villagers' decisions to participate in micro-finance programs in rural areas of Indian. I show that leaders in those villages have significant influence over their neighbors' decision to join the micro-finance program, and I provide rankings for the different social and economic networks among villagers. Based on how effectively each network spreads the impact of influential individuals' decisions, my method shows that some networks such as "visit go-come" and "borrow money" are much more effective in influencing villagers' decisions than other networks such as "temple company" and "medical help". I further show that individuals from certain careers such as agricultural workers, Anganwadi teachers and small business owners are more likely to influence other villagers.

References

Acemoglu, D., García-Jimeno, C., and Robinson, J. A. (2012). Finding eldorado: Slavery and long-run development in colombia. NBER WORKING PAPER SERIES.

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

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, (74):1403–1417.

Bandiera, O., Barankay, I., and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094.

Banerjee, A., Chandrasekhar, A., Duflo, E., and Jackson, M. (2013). The diffusion of microfinance. *Science*, 341(6144).

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). Inference on treatment effects after selection amongst high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Belloni, A., Chernozhukov, V., and Kato, K. (2014b). Uniform post selection inference for lad regression and other z-estimation problems.

Belloni, A., Chernozhukov, V., and Kato, K. (2015). Uniform post selection inference for lad regression and other z-estimation problems.

Belloni, A., Chernozhukov, V., and Wang, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, pages 1–18.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2):444–496.

Bonaldi, P., Hortacsu, A., and Kastl, J. (2015). An empirical analysis of funding costs spillovers in the euro-zone with application to systemic risk.

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

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 41(2):802–837.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data*. Springer.

Bunea, F., Lederer, J., and She, Y. (2014). The square root group lasso: theoretical properties and fast algorithms. *IEEE-Information Theory*, 60:1313–1325.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, (76):1239–1267.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2017). Double/debiased machine learning for treatment and causal parameters.

Clark, A. E. and Loheac, Y. (2007). "it wasn't me, it was them!" social influence in risky behavior by adolescents. *Journal of Health Economics*, 26:763–784.

Cliff, A. and Ord, J. K. (1973). *Spatial autocorrelation.* London: Pion.

Coelli, T., Rahman, S., and Thirtle, C. (2002). Technical, allocative, cost and scale efficiencies in bangladesh rice cultivation: A nonparametric approach. *Journal of Agricultural Economics*, 53(3):607–626.

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

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* John Wiley & Sons, Inc.

de Paula, A., Rasul, I., and Souza, P. C. (2015). Estimating and identifying social interactions.

Denbee, E., Julliard, C., Li, Y., and Yuan, K. (2015). Network risk and key players: A structural analysis of interbank liquidity.

Fan, J. and Liao, Y. (2014). Endogeneity in high dimensions. *The Annals of Statistics*, 42(3):872–917.

Gautier, E. and TsyBakov, A. B. (2014). High-dimensional instrumental variables regression and confidence sets.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Horracea, W. C., Liu, X., and Patacchini, E. (2016). Endogenous network production functions with selectivity. *Journal of Econometrics*, 190(2):222–232.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(1):2869–2909.

Jin, F. and Lee, L.-F. (2016). Lasso maximum likelihood estimation of parametric models with singular information matrices.

Kasy, M. (2015). Uniformity and the delta method. *arXiv preprint arXiv:1507.05731*.

Kelejian, H. H. and Prucha, I. R. (1995). A generalized moments estimator for the autoregressive parameter in a spatial model. *INTERNATIONAL ECONOMIC REVIEW*, 40.

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

Krauth, B. V. (2005). Peer effects and selection effects on smoking among canadian youth. *Canadian Journal of Economics*, 38(3):735–757.

Lee, L. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial. *Econometric Theory*, 18(2):252–277.

Lee, L. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive. *Econometric Reviews*, 22(4):305–335.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica*, 72:1899–1926.

Lee, L. and Liu, X. (2010). Efficient gmm estimation of high order spatial autoregrssive models with autoregressive disturbances. *Econometric Theory*, 26:187–230.

Lee, L.-f. and Yu, J. (2010). A spatial dynamic panel data model with both time and individual effects. *Econometric Theory*, 26:564–597.

Leeb, H. and Potscher, B. M. (2005). Model selection and inference: facts and fiction. *Econometric Theory*, 21(1):21–59.

Leeb, H. and Potscher, B. M. (2008). Can one estimate the unconditional distribution of post-modelselection estimators? *Econometric Theory*, 24(2):38–376.

Leeb, H. and Potscher, B. M. (2009). Model selection. *Handbook of Financial Time Series*, pages 889–925.

Manresa, E. (2013). Estimating the structure of social interactions using panel data.

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

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(1436-1462).

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74(3):897–935.

Neidell, M. and Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *Review of Economics and Statistics*, 92(3):562–576.

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

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). The sparse group lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245.

Upton, G. and Fingleton, B. (1985). *Spatial data analysis by example. Volume 1: Point pattern and quantitative data.* John Wiley and Sons Ltd.

van de Geer, S., Buhlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society*, B(68):49–67.

Zhang, C.-H. and Zhang, S. S. (2011). Confidence intervals for low-dimensional parameters in high-dimensional linear models. *Journal of the Royal Statistical Society*, 76(1):217–242.

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541–2563.

Zhu, Y. (2016). Sparse linear models and l1regularized 2sls with high-dimensional endogenous regressors and instruments.