Social Networks with Mismeasured Links

Arthur Lewbel
Boston College

Xi Qu
Shanghai Jiao Tong University

Xun Tang
Rice University

April 2021

Abstract

We consider estimation of peer effects in social network models where some network links are incorrectly measured. We show that if the number of mismeasured links does not grow too quickly with the sample size, then standard instrumental variables estimators that ignore the measurement error remain consistent, and standard asymptotic inference methods remain valid. These results hold even when measurement errors in the links are correlated with regressors, or with the model errors. Monte Carlo simulations and real data experiments confirm our results in finite samples. These findings imply that researchers can ignore small amounts of measurement errors in networks.

JEL classification: C31, C51

Keywords: Social networks, Peer effects, Misclassified links, Missing links, Mismeasured network.

1 Introduction

In many social and economic environments, an individual's behavior or outcome (such as a consumption choice or a test score) depends not only on his or her own characteristics, but also on the behavior and characteristics of other individuals. Call such dependence between two individuals a *link,* and call individuals with such links *friends*. A *social network* consists of a group of linked individuals. Each individual may have a different set of friends in the network, and each individual may assign heterogenous weights to his or her links. The structure of a social network is fully characterized by a square *adjacency matrix*, which lists all links (with possibly heterogenous weights) among the individuals in the network.

Much of the econometric literature on social networks focuses on disentangling and estimating various social or network effects, based on observed outcomes and characteristics of network members. These structural parameters include the effects on each individual's outcome by (i) the individual's own characteristics (*direct effects*) and possibly group characteristics (*correlated effects*), (ii) the characteristics of the individual's friends (*contextual effects*) and (iii) the outcomes of the individual's friends (*peer effects*). Standard methods of identifying and estimating these

structural network effect parameters assume that the adjacency matrix of links among individuals in the sample is perfectly observed.

1.1. Our contribution. We consider the case where network links are misclassified, or generally measured with errors. Here we provide good news for empirical researchers, by showing that relatively small amounts of measurement error in the network can be safely ignored in estimation. More precisely, we show that instrumental variable estimators like Bramoullé, Djebbari and Fortin (2009), and their standard errors, remain consistent and valid, despite the presence of misclassified, unreported, or mismeasured links, as long as the number and size of these measurement errors grows sufficiently slowly with the sample size. Moreover, these results hold even when the measurement errors are correlated with the regressors, or with the model errors. Below in subsection 1.3 we give examples of applications where measurement errors grow at these slow rates.

It may not be surprising that measurement errors growing at sufficiently slow rates are asymptotically negligible, but it is also not automatic. slow measurement error rates could still blow up an estimator if the stochastic order of quadratic terms in the parameter estimator errors isn't bounded. What we essentially show is that, in the case of two stage least squares (2SLS) estimators of network models, minimal and standard regularity conditions suffice to bound these terms.

1.2. Motivation. There are many reasons why network links can be mismeasured in practice. In some data sets, links are imputed from measures of proximity or similarity of individuals (e.g., use of distance as a link in gravity models of trade). Such imputations are generally imperfect. Mismeasurement may also arise because links that are observed in one context may be irrelevant for outcomes under study in another. For example, two people who are observed as linked on a social media platform may be connected there for business or political reasons, but have no effect on each other's personal outcomes (or vice versa). Or in a school setting, some reported friends and not others may be study partners who affect academic performance.

Even in data sets where observed links are directly relevant for observed outcomes, link data may suffer from a variety of reporting or recording errors. For example, many surveys limit the number of links (such as the number of friends) one can report, leading to missing links for popular individuals. Studies that focus on links within groups, such as within classrooms or villages, may not report links across these groups, (e.g., friendships with people in other schools). Also, in some surveys an individual A could claim to be a friend with B, but B does not report being a friend with A. This leaves the status of their link uncertain.

Yet another potential source of network measurement errors is that the adjacency matrix that determines peer effects could differ from the adjacency matrix that determines contextual effects. For example, in student achievement, many students may benefit from a contextual effect like parent volunteers, while peer effects may come only, or primarily, from a student's immediate friends. Typically, only a single adjacency matrix is observed, and used for estimating both peer and contextual effects. This results in another potential source of network measurement errors.

1.3. Examples. For a sample of n individuals, let H_n^* be the matrix of reported links between these n individuals, and suppose the actual adjacency matrix is G_n^*. Our asymptotic framework is

one in which n grows to infinity. Our main results focus on situations where the sum of network measurement errors (the differences between H_n^* and G_n^*) grows at a rate less than $\sqrt{n}$. Here we list a range of empirical situations in which network measurement errors would be expected to grow at these slow rates.

Consider first the common modeling environment in which data are collected from many groups of individuals, like villages or schools. Data are often collected on links within these groups, such as friendships within class rooms, or kinship relationships within villages. Models using such data often assume no links between individuals in different groups, either for theoretical convenience, or because data are not collected on links between groups. This is equivalent to misclassifying as zero all links that exist outside of diagonal blocks of G_n^*. In other words, this means using a block-diagonal H_n^* in place of the actual G_n^* in the data-generating process. The measurement errors will grow at a rate slower than $\sqrt{n}$ if the number of sampled groups grows at rate slower than $\sqrt{n}$, and the number of links between groups is relatively small.

Another example comes from panel data. Suppose the sample consists of L individuals, each of which is observed for T time periods, so the sample size is $n = LT$. For example, the data could be weekly test scores, for T weeks, by L students in a school. Suppose the friendship network is only observed occasionally, and is assumed to be fixed between observations of the network. Then friendships that are created or dissolved between observations of the network will be misreported. The resulting misclassification rates will be of an order less than $\sqrt{n}$ if the number of times the network is observed is sufficiently frequent relative to T, or if L grows sufficiently quickly relative to T.[1]

A third example is data collection that limits the number of observed non-zero links per row, such as surveys that limit the number of friends any one person can report. If this maximum allowable number of reported friends grows moderately with sample size (which is plausible as a model of survey design), then the resulting misclassification rates will be of order less than$\sqrt{n}$.

A fourth example is ordinary recording errors on surveys. These errors may grow at a slow rate relative to sample size if quality control in the data collection increases the bigger (and hence the more expensive) the survey design is.

More generally, as long as the number and size of measurement errors in an observed adjacency matrix is relatively small, asymptotics that assume these measurement errors grow slowly with n should provide a good approximation for inference.

1.4. The Model. With a sample size n, let $Y_n = (y_1, ..., y_n)' \in \mathbb{R}^n$ be a vector of individual outcomes, let $\iota = (1, ..., 1)'$ and $\epsilon_n = (\varepsilon_1, ..., \varepsilon_n)'$ be n-dimensional column vectors, and let $X_n = (x_1, ..., x_n)'$ be an n-by-K matrix that consists of n vectors of exogenous regressors $x_i \in \mathbb{R}^K$ for $i \leq n$. Let G_n^* and C_n^* be n-by-n adjacency matrices that list the actual links for peer effects

[1] Our base case analysis assumes i.i.d. errors, which is restrictive for panel data. However, our results readily generalize to allow some degree of error dependence in the usual way, since the estimator takes the form of linear two-stage least squares.

and contextual effects respectively.[2] Let G^*_{ij} (C^*_{ij}) denote the element in row i and column j of G^*_n (C^*_n). We have $G^*_{ij} > 0$ if i and j are linked for peer effects and $G^*_{ij} = 0$ otherwise. Similarly, $C^*_{ij} > 0$ if i and j are linked for contextual effects and $C^*_{ij} = 0$ otherwise. For each individual i, let $G^*_{ii} = 0$ and $C^*_{ii} = 0$ by convention in the literature. Note that G^*_{ij} can be binary (with $G^*_{ij} \in \{0, 1\}$ indicating the absence or presence of a link), or continuous and non-negative with $G^*_{ij} \in \mathbb{R}_+$ signifying the strength of the link. The same applies for C^*_n. Throughout the paper, we maintain that $\min_i \sum_{j=1}^n G^*_{ij} > 0$ and $\min_i \sum_{j=1}^n C^*_{ij} > 0$ with probability one. This means there are no isolated individuals in the network, or equivalently no rows of zeros in G^*_n or C^*_n, almost surely. This condition is standard in the literature.

We assume a linear social network model:

$$Y_n = \alpha_0 \iota + \lambda_0 G_n Y_n + X_n \beta_0 + C_n X_n \gamma_0 + \epsilon_n, \tag{1}$$

where G_n and C_n can be either the original adjacency matrices G^*_n and C^*_n, or normalized versions of G^*_n and C^*_n. For example, a row-normalized G_n is defined by $G_{ij} = G^*_{ij} / \left(\sum_{j'=1}^n G^*_{ij'} \right)$. Row normalization is common; we will show our results hold with or without such normalization.

The parameters in equation (1) are as follows: $\lambda_0 \in \mathbb{R}$ is a scalar peer effect, $\beta_0 \in \mathbb{R}^K$ is a vector of direct effects, $\gamma_0 \in \mathbb{R}^K$ is a vector of contextual effects, and $\alpha_0 \in \mathbb{R}$ is the structural intercept. If individuals are divided into groups (such as villages or classrooms), then what are known as correlated effects are group-level fixed effects, i.e., elements of β_0 where the corresponding element of X_n is a group membership indicator.

Our goal is estimation of $\theta_0 \equiv (\alpha_0, \lambda_0, \beta_0', \gamma_0')'$. If Y_n, X_n, G^*_n and C^*_n (and hence G_n and C_n) were perfectly observed, the structural model would take the form of a linear regression of Y_n on a constant and the sets of regressors $G_n Y_n$, X_n, and $C_n X_n$. However, even if X_i is uncorrelated with ε_j for all i and j, making X_n and $C_n X_n$ strictly exogenous, this regression could not be consistently estimated by ordinary least squares, because of the endogeneity of $G_n Y_n$. Instead, one can use an instrument-based, 2SLS estimator using friends of friends of i to construct instruments for $G_n Y_n$ (see, e.g., Lee (2007) and Bramoullé, Djebbari and Fortin (2009)). For example, $G^2_n X_n$ can be instruments for $G_n Y_n$. To implement this 2SLS estimator, one is assumed to have perfect measures of G^*_n and C^*_n so that the regressors $G_n Y_n$ and $C_n X_n$, and instruments such as $G^2_n X_n$, can all be constructed without errors.

1.5. Estimation with misclassified links.

Instead of observing Y_n, X_n, and the true adjacency matrices G^*_n and C^*_n, we assume that what is observed is Y_n, X_n, and a single mismeasured adjacency matrix H^*_n. The differences $H^*_n - G^*_n$ and $H^*_n - C^*_n$ are measurement errors in links. Our analysis will assume G^*_n and C^*_n are very similar, so that a single mismeasured H^*_n can approximate both.[3] We could instead allow G^*_n and C^*_n

[2] Most studies assume a single adjacency matrix, i.e., $G^*_n = C^*_n$, since only a single matrix is usually observed in practice. We allow G^*_n and C^*_n to differ, but require the differences be quite small, unless two different matrices can be observed in practice. See, e.g., Blume, et al. (2015).

[3] In particular, we will assume that measurement errors $H^*_n - G^*_n$ and $H^*_n - C^*_n$ both grow a slow rate relative to

to be completely different, by assuming that two different adjacency matrices are observed, one a mismeasured version of G_n^* and the other a mismeasure of C_n^*. We do not do so to save on notation, and because it is extremely rare in practice to observe two different adjacency matrices, where one is known to measure peer effects and the other is known to measure contextual effects.

Like G_n^* and C_n^*, the matrix H_n^* by convention has zeros on the diagonal. When $G_{n,ij}^*$ equals zero or one, misclassification of that link corresponds to $H_{ij}^* = 1 - G_{ij}^*$, and similarly for C_{ij}^*. More generally, a measurement error in a link occurs whenever $H_{ij}^* \neq G_{ij}^*$ or $H_{ij}^* \neq C_{ij}^*$. The differences $H_n^* - G_n^*$ and $H_n^* - C_n^*$ summarize the measurement errors in the network. These measurement errors can be any combination of misclassified links or incorrectly weighted links.

We investigate the asymptotic properties of 2SLS estimation of (1) when the mismeasured adjacency matrix H_n^* is observed instead of the true unknown matrices G_n^* and C_n^*. So instead of a 2SLS regression of Y_n on G_nY_n, X_n, and C_nX_n, using as instruments $G_n^2X_n$, X_n, and C_nX_n, we consider 2SLS regression of Y_n on H_nY_n, X_n, and H_nX_n, using as instruments $H_n^2X_n$, X_n, and H_nX_n. Note this means that both some regressors and some instruments are mismeasured, and that the measurement errors in regressors and instruments are correlated. Moreover, we do not impose any of the uncorrelatedness or conditional independence conditions on measurement errors that are usually assumed in measurement error models. For example, we allow the measurement errors $H_n^* - G_n^*$ and $H_n^* - C_n^*$ to be arbitrarily correlated with X_n, Y_n, and ϵ_n.

We find that if the magnitude of measurement errors grows at a rate slower than $\sqrt{n}$, then the 2SLS estimator remains $\sqrt{n}$-consistent and asymptotically normal, and the usual formulas for inference and standard errors remain valid. As a result, under these conditions researchers can safely ignore the presence of misclassified or mismeasured links, because the estimator and inference based on H_n^* instead of G_n^* and C_n^* remains consistent and valid.

We also find that if the magnitude of measurement errors in the observed adjacency matrix grows at a rate faster than $\sqrt{n}$ but slower than n, then the 2SLS estimator is still consistent. However, in this case the rate of convergence of the coefficients is less than $\sqrt{n}$ (due to a bias term that shrinks at a slower rate than $\sqrt{n}$), so the usual standard error formulas would no longer apply.

1.6 Outline. The next section is a short literature review. This is followed by our formal model. We then present our results for 2SLS estimation of mismeasured networks. This is followed by some simulation results and an empirical illustration. Proofs are in the appendix.

2 Literature Review

Typical social network models may allow an individual's outcome to depend on his or her own characteristics, contextual influences from peers' characteristics, and peer effects from peer outcomes. The traditional linear-in-means model (which assumes everyone is linked with everyone else equally, either within groups or in the whole network) suffers from the "reflection problem" as

n, which requires that the difference $G_n^* - C_n^*$ also grows at a slow rate, and so requires that the differences between these true matrices be small.

pointed out by Manski (1993). This identification problem can be overcome in models with more complicated social interaction structures. Lee (2007) uses conditional maximum likelihood and instrumental variable methods to estimate peer and contextual effects in a spatial autoregressive social interaction model, assuming links are perfectly observed in the data. Bramoullé, Djebbari and Fortin (2009) and Lin (2010) provide specific conditions on observed network structure in order to identify peer effects in social interaction models, using characteristics of friends of friends as instruments.

Given results like these, the model described in the introduction has been widely used to estimate peer effects in a variety of settings (usually assuming either $C_n^* = G_n^*$ or $C_n^* = 0$, though see Blume et. al. 2015). Examples are studies of peer influence on students' academic performance, sport and club activities, and delinquent behaviors (Hauser et al., 2009; Calvó-Armengol et al., 2009; Lin, 2010; Lee et al., 2010; Liu et al., 2014; Boucher et al., 2014; Patacchini and Zenou, 2012). These models all assume that the network structure is correctly measured in the data.

Regarding selection and comparison of adjacency matrices, LeSage and Pace (2009) use the Bayesian posterior distribution to choose among models with different adjacency matrices. Empirical research may also report estimates using different link weights as robustness checks. These practices are feasible in, e.g., spatial econometric models, where link weights are assumed to be a function of observable geographic information, as in gravity models of trade. Errors in constructing such links would fit in our framework. There is also a small literature on identification and estimation of peer effects when networks are unobserved. Examples include de Paula et al. (2018) and Lewbel et al. (2021).

The issue of potentially misclassified links is acknowledged and discussed in Patacchini and Venanzoni (2014), Liu et al. (2014), and Lin (2015) among others. But these papers do not provide a formal analysis of the asymptotic impact of mismeasured links on the performance of standard estimators. Griffith (2021) studied the impact on inference when misclassification in the adjacency matrix occurs because of binding caps on the number of self-reported links. Our results fill a void in the literature by analyzing how ignoring small amounts of general measurement errors in the adjacency matrix affects the consistency of standard estimators and the validity of inference.[4]

[4] Referring to potential omission of friends, Patacchini and Venanzoni (2014) say that, "in the large majority of cases (more than 94%), students tend to nominate best friends who are students in the same school and thus are systematically included in the network (and in the neighborhood patterns of social interactions)". Liu et al. (2014) report that "less than 1% of the students in our sample show a list of ten best friends, less than 3% a list of five males and roughly 4% a list of five females. On average, they declare that they have 4.35 friends with a small dispersion around this mean value (standard deviation equal to 1.41), and in the large majority of cases (more than 90%) the nominated best friends are in the same school." Lin (2015) says, "this nomination constraint only affects a small portion of our sample, as less than 10% of the sample have listed five male or female friends. Therefore, this restriction should not have a significant impact on the results." This last speculation is precisely what our first set of results establishes: that consistency of estimates will not be effected if the number of omitted (and hence misclassified) links is sufficiently small.

3 2SLS Estimation With Mismeasured Links

In this section we derive asymptotic properties of the 2SLS estimator for the model in (1) when the mismeasured adjacency matrix H_n^* is used in place of the actual, unknown G_n^* and C_n^*. This means the regressors G_nY_n, C_nX_n and instruments $G_n^2X_n$ are replaced by H_nY_n, H_nX_n and $H_n^2X_n$ in the estimator.

Write equation (1) as

$$Y_n = R_n\theta_0 + \epsilon_n = \widetilde{R}_n\theta_0 + \widetilde{\epsilon}_n,$$

where $R_n \equiv (\iota, G_nY_n, X_n, C_nX_n)$ is the true matrix of regressors, $\widetilde{R}_n \equiv (\iota, H_nY_n, X_n, H_nX_n)$ is its observed proxy, θ_0 is the true value of θ, and $\widetilde{\epsilon}_n \equiv \epsilon_n - \lambda_0\Delta_{1n}Y_n - \Delta_{2n}X_n\gamma_0$.

Let $\widetilde{V}_n \equiv (\iota, H_n^2X_n, X_n, H_nX_n)$ denote an n-by-$(3K+1)$ matrix of instruments. This $\widetilde{V}_n$ is an observable proxy for the unobservable desired instrument matrix $V_n \equiv (\iota, G_n^2X_n, X_n, C_nX_n)$. The 2SLS estimator is:

$$\widehat{\theta} = \left[\widetilde{R}_n'\widetilde{V}_n(\widetilde{V}_n'\widetilde{V}_n)^{-1}\widetilde{V}_n'\widetilde{R}_n\right]^{-1}\widetilde{R}_n'\widetilde{V}_n(\widetilde{V}_n'\widetilde{V}_n)^{-1}\widetilde{V}_n'Y_n. \tag{2}$$

We show that this estimator is consistent when the measurement errors in the adjacency matrices are small in the following sense.

Assumption 1 *$\sum_i\sum_j E\left(\left|H_{ij}^* - G_{ij}^*\right|\right) = O(n^s)$ and $\sum_i\sum_j E\left(\left|H_{ij}^* - C_{ij}^*\right|\right) = O(n^s)$ for some $s < 1$.*

Assumption 1 requires the expected sum of absolute measurement errors in G_n^* and C_n^* increase at a rate slower than the sample size n. This condition holds, for example, if mismeasurement occurs only for a subset of individuals of order $O(n^s)$ with $s < 1$, and if the magnitude and expected number of mismeasured links for each individual in the subset is bounded. See the introduction for more examples under a variety of contexts.

As discussed earlier, this assumption implies that the sum of the absolute differences between G_n^* and C_n^* also increases at a rate slower than n. We can eliminate this constraint if there are two different H_n^* matrices reported in the sample, with one being a mismeasurement of G_n^* and the other a mismeasurement of C_n^*. However, this is very rarely the case in practice.

Denote $S_n \equiv I_n - \lambda_0G_n$. When S_n is non-singular, the reduced form for outcomes is:

$$Y_n = S_n^{-1}(\alpha_0\iota + X_n\beta_0 + C_nX_n\gamma_0 + \epsilon_n).$$

We maintain the following regularity conditions.

Assumption 2 *(i) ϵ_n is independent from X_n; individual errors ε_i are independent across i, with $E(\varepsilon_i) = 0$ and there exists a constant $M_0 < \infty$ such that $\Pr\{\sup_{i\leq n} E(|\varepsilon_i| \mid H_n) \leq M_0\} = 1$ for all n. (ii) G_n^* and C_n^* are sequences of pre-determined, non-stochastic matrices, and S_n is non-singular for all n. The sequences $\{G_n^*\}$, $\{C_n^*\}$, and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums. The row and column sums in the sequence $\{H_n^*\}$ is uniformly bounded in probability. (iii) The elements of X_n are uniformly bounded for all n; and $V_n'V_n/n$ converges in probability to a non-singular matrix as $n \to \infty$.*

Part (i) of Assumption 2 states that X_n are exogenous. Notice that we do not impose exogeneity of H_n^*, i.e., the measurement errors $H_n^* - G_n^*$ and $H_n^* - C_n^*$ can be correlated with both ϵ_n and X_n. This is in sharp contrast to most measurement error models, which typically require measurement errors to be independent of some observed or unobserved variables for point identification and estimation. Part (ii) requires the row and column sums of G_n^* and H_n^* to be uniformly bounded, and that the reduced form of outcomes is well-defined. Invertibility of S_n holds if $\sum_j |\lambda G_{ij}| < 1$ for all i. In the special case of non-negative elements and row-normalization in G_n^*, $|\lambda| < 1$ is sufficient for non-singular S_n. Part (iii) requires the matrix of actual instruments to have full column rank. All these assumptions regarding the true adjacency matrices are standard for linear social interactions models.

Proposition 1 *Under Assumptions 1 and 2,*

$$\widehat{\theta} - \theta_0 = O_p(n^{-1/2} \vee n^{s-1}).$$

This proposition holds because we can establish the following relationship between the mismeasured proxies and their actual counterparts:

$$\begin{aligned} \widehat{\theta} - \theta_0 &= \left[\frac{\widetilde{R}_n'\widetilde{V}_n}{n}\left(\frac{\widetilde{V}_n'\widetilde{V}_n}{n}\right)^{-1}\frac{\widetilde{V}_n'\widetilde{R}_n}{n}\right]^{-1}\frac{\widetilde{R}_n'\widetilde{V}_n}{n}\left(\frac{\widetilde{V}_n'\widetilde{V}_n}{n}\right)^{-1}\frac{\widetilde{V}_n'\widetilde{\epsilon}_n}{n} \\ &= \left[\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'R_n}{n}\right]^{-1}\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'\epsilon_n}{n} + O_p(n^{s-1}). \end{aligned} \tag{3}$$

Under regularity conditions in Assumption 2, $(R_n'V_n)/n$ and $(V_n'V_n)/n$ both converge in probability to constant matrices with full rank $(2K+2)$. Under exogeneity of X_n, the term $V_n'\epsilon_n/n$ is $O_p(n^{-1/2})$ by Chebyshev's Inequality. Combining these results, we conclude that the estimation errors in (3) is $O_p(n^{-1/2} \vee n^{s-1})$. Thus the 2SLS estimator using $H_n^2 X_n$ as an instruments for $H_n Y_n$ is consistent when $s < 1$.

It follows from this proposition that with $s < 1$ the 2SLS estimator $\widehat{\theta}$ using instruments $H_n^2 X_n$ is consistent. Furthermore, if $s < 1/2$, the effect of measurement errors vanishes fast enough so that it does not affect the $\sqrt{n}$-rate of convergence or the asymptotic distribution of the 2SLS estimator. That is, we have

Proposition 2 *Under Assumptions 1 and 2, if $s < 1/2$ then*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega),$$

where Ω is the asymptotic variance of the 2SLS estimator using the actual adjacency matrix G_n; and Ω can be consistently estimated by $\widehat{A}^{-1}\widehat{B}\widehat{A}^{-1}$, where $\widehat{A} \equiv \frac{1}{n}\widetilde{R}_n' P_{\widetilde{V},n} \widetilde{R}_n$ and $\hat{B} \equiv \frac{1}{n}\widetilde{R}_n' P_{\widetilde{V},n} \widehat{\Sigma}_n P_{\widetilde{V},n} \widetilde{R}_n$, with $P_{\widetilde{V},n} \equiv \widetilde{V}_n \left(\widetilde{V}_n'\widetilde{V}_n\right)^{-1} \widetilde{V}_n'$ and $\widehat{\Sigma}_n$ being a diagonal n-by-n matrix whose i-th diagonal entry is the square of the i-th residual in $Y_n - \widetilde{R}_n\widehat{\theta}$.

As noted in the introduction, even slowly growing measurement errors could asymptotically corrupt $\widehat{\theta}$ if the stochastic order of quadratic terms in $\widehat{\theta} - \theta_0$ isn't bounded. The closed form of the 2SLS estimator plays a key role in deriving our results. In our proofs, this closed form allows us to use Cauchy-Schwartz inequalities to bound the stochastic order of these errors. Key conditions we use for this are boundedness of S_n^{-1} and X_n. Without those, the estimation errors might not obey the stochastic orders we derive.

4 Simulations

We investigate the performance of the 2SLS estimator with mismeasured links using simulated data. The structural equation in our data-generating process (DGP) is (1) where X_n consists of two regressors: the first is independently drawn from $\{-1, 1, 2\}$ with equal probability, and the second is from $N(0, 1)$. The error terms ε_i are i.i.d. from $N(0, 1)$. Links in G_n^* are independent draws from a Bernoulli distribution with success probability $p_n = \mu/n$ for some constant $\mu < \infty$. By this construction, the expected number of friends for each individual is equal to μ. Let G_n be a row-normalization of G_n^* and $C_n^* = G_n^*$.

We generate misclassified links using $H_{ij}^* = G_{ij}^* \cdot e_{1i} + (1 - G_{ij}^*) \cdot e_{2i}$ for $i \neq j$, where e_{1i} and e_{2i} are Bernoulli random variables with success probabilities $1 - \tau_{1i}$ and τ_{2i} respectively. Therefore, τ_{1i} is the misclassification probability that $H_{ij}^* = 0$ when the true $G_{ij}^* = 1$, and τ_{2i} is the misclassification probability that $H_{ij}^* = 1$ when $G_{ij}^* = 0$. We set $\tau_{1i} = \rho_i n^{s-1}$ and $\tau_{2i} = 100\rho_i n^{s-2}$, where $\rho_i = (\sum_{j=1}^n G_{ij}^*/\mu + |\varepsilon_i|)/3$. For each individual i, the probability of misclassification increases in the number of individual i's friends $\sum_{j=1}^n G_{ij}^*$, and in the magnitude of i's unobserved error $|\varepsilon_i|$. This construction makes the measurement errors both endogenous (correlated with the model errors) and correlated with the actual row-normalized G_n.

We set the model parameters to be $\alpha = 1$; $\lambda = 0.4$; $\beta = (1.5, 2)'$ and $\gamma = (0.9, 0.6)'$; choose $\mu = 20$; and experiment with the rates in measurement errors $s = 0.1$, 0.3, 0.5, and 0.7. We experiment with sample sizes $n = 200$, 500, and 1000. For each value of s and n, we simulate $T = 200$ samples and calculate the mean squared error, the bias, the sample standard deviation, and the asymptotic standard error of the estimator. We also report the total number of misclassified links based on the average over $T = 200$ simulated samples.

Results are summarized in Table 1. We observe several patterns:

1. The 2SLS estimates of all parameters appear to converge at $\sqrt{n}$ rate. The mean-squared errors decrease appropriately as the sample size increases.

2. Consistent with our asymptotic theory, the 2SLS estimator using the misclassified adjacency matrix H_n works almost as well as its infeasible analog based on the actual G_n when the measurement error rate is $s < 0.5$. This suggests that the sample sizes we consider are large enough for the asymptotic approximations to apply. Note that with our data generating process, the estimates in Table 1 where $s < 0.5$ have error rates where the expected number of misclassified links is less than n.

3. The average standard errors do a good job of estimating the standard deviations for all values of s. This is as expected, because the problem with inference for larger values of s is that the bias in the estimator shrinks at rate n^{s-1}. Similarly, with $s \geq 0.5$, the parameter estimates deteriorate primarily due to bias rather than variance.

4. With both the true and mismeasured adjacency matrices, the mean-squared errors are much smaller for the direct effects β than for the peer and contextual effects λ and γ, and the mean squared errors are much lower for the discrete regressor effects β_1 and γ_1 than for the continuous regressor effects β_2 and γ_2

Table 1. 2SLS Estimators with Misclassified Links

	$n = 200$				$n = 500$				$n = 1000$			
	m.s.e.	bias	std	a.s.e.	m.s.e.	bias	std	a.s.e.	m.s.e.	bias	std	a.s.e.
True		Mis.#	0				0				0	
α	3.880	-0.114	1.971	2.197	1.519	0.031	1.235	1.310	0.762	0.065	0.873	0.887
λ	0.336	0.025	0.581	0.654	0.131	-0.010	0.362	0.386	0.068	-0.019	0.260	0.264
β_1	0.003	0.005	0.058	0.058	0.001	-0.003	0.036	0.036	0.001	-0.000	0.027	0.026
β_2	0.005	0.008	0.072	0.073	0.002	0.001	0.048	0.045	0.001	-0.000	0.032	0.032
γ_1	0.802	-0.029	0.898	1.006	0.301	0.019	0.549	0.597	0.165	0.030	0.406	0.410
γ_2	1.571	-0.040	1.256	1.348	0.561	0.020	0.750	0.796	0.278	0.032	0.528	0.545
$s = 0.1$		Mis.#	66				81				88	
α	4.100	-0.058	2.029	2.254	1.576	0.033	1.258	1.325	0.780	0.070	0.883	0.894
λ	0.365	0.008	0.605	0.672	0.135	0.010	0.368	0.391	0.070	-0.020	0.263	0.266
β_1	0.003	0.004	0.058	0.058	0.001	-0.003	0.036	0.036	0.001	-0.000	0.027	0.026
β_2	0.005	0.008	0.072	0.073	0.002	0.001	0.048	0.045	0.001	-0.000	0.032	0.032
γ_1	0.877	-0.015	0.938	1.033	0.307	0.015	0.556	0.604	0.168	0.030	0.410	0.413
γ_2	1.610	-0.012	1.272	1.382	0.574	0.019	0.760	0.805	0.284	0.033	0.533	0.549
$s = 0.3$		Mis.#	193				278				351	
α	4.599	0.058	2.149	2.388	1.678	0.014	1.2985	1.367	0.833	0.083	0.911	0.912
λ	0.405	-0.023	0.638	0.712	0.144	-0.002	0.380	0.403	0.074	-0.023	0.271	0.272
β_1	0.004	0.003	0.059	0.059	0.001	-0.003	0.035	0.037	0.001	-0.000	0.027	0.026
β_2	0.005	0.009	0.073	0.074	0.002	0.001	0.048	0.046	0.001	-0.000	0.032	0.032
γ_1	0.949	0.018	0.977	1.094	0.334	-0.005	0.579	0.622	0.179	0.031	0.423	0.421
γ_2	1.756	0.041	1.328	1.461	0.598	-0.005	0.775	0.830	0.305	0.035	0.552	0.560
$s = 0.5$		Mis.#	556				968				1401	
α	5.620	0.136	2.373	2.773	1.995	0.0670	1.414	1.519	1.060	0.133	1.023	0.982
λ	0.498	-0.032	0.706	0.828	0.172	-0.012	0.416	0.449	0.093	-0.035	0.303	0.293
β_1	0.004	0.001	0.062	0.060	0.001	-0.003	0.036	0.037	0.001	-0.000	0.028	0.026
β_2	0.005	0.011	0.073	0.075	0.002	0.001	0.049	0.046	0.001	-0.000	0.033	0.032
γ_1	1.174	-0.022	1.086	1.272	0.408	-0.015	0.640	0.691	0.218	0.032	0.467	0.453
γ_2	2.157	0.021	1.472	1.691	0.732	-0.021	0.857	0.921	0.376	0.041	0.614	0.602
$s = 0.7$		Mis.#	1605				3346				5572	
α	17.93	0.433	4.223	4.212	4.581	0.253	2.131	2.075	1.812	0.157	1.340	1.291
λ	1.549	-0.095	1.244	1.252	0.395	-0.0470	0.628	0.613	0.158	-0.025	0.398	0.385
β_1	0.004	0.002	0.066	0.064	0.002	-0.004	0.038	0.039	0.001	-0.001	0.028	0.027
β_2	0.006	0.009	0.076	0.081	0.003	-0.001	0.052	0.048	0.001	0.001	0.033	0.033
γ_1	3.643	-0.050	1.913	1.898	0.894	-0.058	0.946	0.934	0.374	-0.069	0.610	0.589
γ_2	6.452	0.011	2.547	2.533	1.545	-0.047	1.245	1.250	0.649	-0.056	0.806	0.786

Note: m.s.e (mean squared error), bias, std (standard deviation) are calculated using the empirical distribution of 200 estimates. "a.s.e." is the average of standard errors in 200 samples.

5 Application

Lin and Lee (2010) model teenage pregnancy rates, using the model

$$Teen_i = \lambda \sum_{j=1}^{n} G_{ij} Teen_j + \alpha + Edu_i\beta_1 + Inco_i\beta_2 + FHH_i\beta_3 + Black_i\beta_4 + Phy_i\beta_5 + \varepsilon_i,$$

where $Teen_i$ is the teenage pregnancy rate in county i, which is the percentage of pregnancies occurring to females 12-17 years old, and G_{ij} is the row-normalized entry of the original link matrix G_n^*, where $G_{ij}^* = 1$ if counties i and j are neighboring counties. Edu_i is the education service expenditure (in units of \$100), $Inco_i$ is median household income (divided by 1000), FHH_i is the percentage of female-headed households, $Black_i$ is the proportion of black population and Phy_i is the number of physicians per 1000 population, all in county i.[5]

The sample size is $n = 761$. Among all the $761 \times 760 = 578,360$ entries (diagonal are zero) in the original network G_n^*, there are $4,606$ non-zero links. We treat the adjacency matrix they report as the true network, artificially introduce misclassified links, and then evaluate how this affect the 2SLS estimates. We generate misclassified links using $H_{ij}^* = G_{ij}^* \cdot e_{1i} + (1 - G_{ij}^*) \cdot e_{2i}$, where e_{1i} and e_{2i} are binary variables with probabilities $\tau_{1i} = \rho_i n^{s-1}$ and $\tau_{2i} = 100\rho_i n^{s-2}$ of equaling 1. We set $\rho_i = (y_i/\overline{y})^2$, so for each individual i misclassification is more likely to happen the larger is the magnitude of the observed outcome y_i.

We report 2SLS estimates using H_nX_n and $H_n^2X_n$ as instruments. Unlike our structural model, Lin and Lee (2010) assume contextual effects (the γ coefficients) are zero, so G_nX_n does not appear as regressors. It would therefore have been possible to just use H_nX_n as instruments for estimation. Nonetheless, to illustrate our proposition, we use both H_nX_n and $H_n^2X_n$ as instruments here.

Table 2 reports results based on 1000 Monte Carlo replications for each value of s. Results are reported where the model is estimated both with and without row normalization.

Consistent with our propositions, when the misclassification rate is low ($s < 0.5$), the 2SLS parameter and standard error estimates using the mis-measured H_n are very similar to those based on G_n. The same is true for estimation based on matrices H_n^* and G_n^* that are not row-normalized. When s increases, the bias and inaccuracy of the estimators increases, as expected. In particular, the parameter estimates (especially λ) become quite biased when $s \geq 0.5$ (which, by our theory, is when bias shrinks at a slower rate than variance).

[5] The data are collected from 761 counties in Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, Nebraska, North Dakota, South Dakota, and Wyoming. Data details are in Lin and Lee (2010).

Table 2. Estimation Results with Different Misclassification Rates

	λ	α	$100\beta_1$	β_2	β_3	β_4	β_5	Mis. #
Row-normalized adjacency matrices $G_{ij} = G^*_{ij} / \left(\sum_j G^*_{ij}\right)$ and $H_{ij} = H^*_{ij} / \left(\sum_j H^*_{ij}\right)$								
True	0.4813	6.1911	-0.9824	-0.1871	0.7347	0.1267	-0.4956	0
	(0.079)	(1.469)	(0.651)	(0.040)	(0.063)	(0.057)	(0.188)	
$s = 0.1$	0.4897	6.1085	-0.9910	-0.1878	0.7355	0.1289	-0.4980	111
	(0.081)	(1.480)	(0.651)	(0.040)	(0.063)	(0.057)	(0.188)	
$s = 0.3$	0.5132	5.8759	-1.0086	-0.1895	0.7375	0.1341	-0.5049	418
	(0.085)	(1.512)	(0.652)	(0.040)	(0.063)	(0.057)	(0.188)	
$s = 0.5$	0.6017	4.9578	-1.0542	-0.1943	0.7422	0.1465	-0.5227	1578
	(0.099)	(1.626)	(0.654)	(0.040)	(0.063)	(0.057)	(0.189)	
$s = 0.7$	0.8138	2.7629	-1.1726	-0.2092	0.7589	0.1683	-0.5535	5948
	(0.139)	(1.985)	(0.660)	(0.040)	(0.064)	(0.057)	(0.191)	
Original adjacency matrices G^*_{ij} and H^*_{ij} without normalization								
True	0.0239	10.840	-1.5244	-0.2348	0.8151	0.2061	-0.5731	0
	(0.009)	(1.261)	(0.669)	(0.041)	(0.064)	(0.058)	(0.194)	
$s = 0.1$	0.0275	10.491	-1.5290	-0.2317	0.8087	0.2069	-0.5658	111
	(0.009)	(1.248)	(0.666)	(0.040)	(0.064)	(0.057)	(0.193)	
$s = 0.3$	0.0356	9.6492	-1.5361	-0.2239	0.7916	0.2079	-0.5463	418
	(0.008)	(1.216)	(0.659)	(0.040)	(0.063)	(0.057)	(0.191)	
$s = 0.5$	0.0486	7.5887	-1.5473	-0.2039	0.7351	0.2058	-0.4813	1578
	(0.005)	(1.130)	(0.633)	(0.038)	(0.061)	(0.055)	(0.184)	
$s = 0.7$	0.0442	4.9575	-1.5211	-0.1749	0.6170	0.1858	-0.3396	5948
	(0.003)	(0.984)	(0.571)	(0.034)	(0.055)	(0.049)	(0.166)	

Note: The table reports average estimates and average standard errors (in parentheses) from 1000 simulated samples.

6 Conclusions

We show that in 2SLS estimation of linear social network models, measurement errors in the network can be safely ignored by the researcher if the number and magnitude of measurement errors in the adjacency matrix grows sufficiently slowly with the sample size. Moreover, these results hold even if the measurement errors are correlated with model errors, covariates, and outcomes. A useful agenda for future work would be to see if similar results can be obtained for more general network models.

Appendix

For a generic matrix A, let $A_{(i)}$, $A_{[k]}$ denote its i-th row and k-th column respectively; and A_{ij} denote its (i,j)-th component, so that $A_{(i)}\iota$ is the sum of the i-th row in A. Let $\Delta^*_{1n} \equiv H^*_n - G^*_n$ and $\Delta^*_{2n} \equiv H^*_n - C^*_n$ with $H^*_{ii} = 0$ by construction. With row normalization,

$$\Delta_{1n} \equiv H_n - G_n = diag\left\{\left(\tfrac{1}{G^*_{(1)}\iota}, ..., \tfrac{1}{G^*_{(n)}\iota}\right)\right\}\Delta^*_{1n} + diag\left\{\left(\tfrac{1}{H^*_{(1)}\iota} - \tfrac{1}{G^*_{(1)}\iota}, ..., \tfrac{1}{H^*_{(n)}\iota} - \tfrac{1}{G^*_{(n)}\iota}\right)\right\}H^*_n.$$

Similarly for Δ_{2n}. The following two lemmas are useful for the proofs.

Lemma A1. *Let a_n, b_n be random vectors in $\mathbb{R}^n$. Let the matrix Δ_n* be either Δ_{1n} or Δ_{2n}. *Suppose there exist constants $M_1, M_2 < \infty$ such that* $\Pr\{\sup_{i\le n}|a_i| \le M_1\} = 1$ *and* $\Pr\{\sup_{j\le n} E\left(|b_j||\Delta_n\right) \le M_2\} = 1$ *for all n. Then* $\frac{1}{n}a_n'\Delta_n b_n = O_p(n^{s-1})$ *and* $\frac{1}{n}a_n'\Delta^*_n b_n = O_p(n^{s-1})$ *under Assumption 1.*

Proof of Lemma A1. The following proof uses $\Delta_n = \Delta_{1n}$ (and $\Delta^*_n = \Delta^*_{1n}$) as an example. Similarly for $\Delta_n = \Delta_{2n}$. From the triangle inequality

$$\begin{aligned} & E\left(\textstyle\sum_i\sum_j|\Delta_{ij}|\right) = E\left(\sum_i\sum_j\left|\frac{1}{G^*_{(i)}\iota}\Delta^*_{ij} + \frac{\left(G^*_{(i)} - H^*_{(i)}\right)\iota}{\left(G^*_{(i)}\iota\right)\left(H^*_{(i)}\iota\right)}H^*_{ij}\right|\right) \\ \le\ & E\left[\sum_i\sum_j\left(\frac{1}{G^*_{(i)}\iota}\left|\Delta^*_{ij}\right| + \frac{1}{\left(G^*_{(i)}\iota\right)\left(H^*_{(i)}\iota\right)}\left|\left(G^*_{(i)} - H^*_{(i)}\right)\iota\right| \times H^*_{ij}\right)\right] \\ \le\ & E\left[\sum_i\left(\frac{1}{G^*_{(i)}\iota}\sum_j\left|\Delta^*_{ij}\right| + \frac{1}{G^*_{(i)}\iota}\sum_j\left|\Delta^*_{ij}\right|\right)\right] = O(n^s), \end{aligned}$$

Furthermore,

$$E\left(\left|\tfrac{1}{n}a_n'\Delta_n b_n\right|\right) \le \tfrac{1}{n}E\left[\sup_{i,j}E\left(|a_ib_j||\Delta_n\right)\cdot\left(\textstyle\sum_i\sum_j|\Delta_{ij}|\right)\right] = O(n^{s-1}).$$

Similar arguments can be applied to show $\frac{1}{n}a_n'\Delta^*_n b_n = O_p(n^{s-1})$. □

Lemma A2. *Under Assumption 2,* $\sup_{i\le n}|V_{iq}| = O(1)$ and $\sup_{i\le n}V^2_{iq} = O(1)$ *for $q = 1, ..., K$, and there exists constant $M^* < \infty$ such that* $\Pr\{\sup_i E(|y_i||\Delta_n) \le M^*\} = 1$ *for all n.*

Proof of Lemma A2. Note

$$\sup_{i\le n}\left(\left[G^2_{(i)}X_{[q]}\right]^2\right) \le \left(\sup_{i\le n}\textstyle\sum_k|G_{ik}|\right)^2\left(\sup_{k\le n}\sum_j|G_{kj}|\right)^2\left(\sup_{j\le n}x^2_{jq}\right) = O(1).$$

It follows that $\sup_i V^2_{iq} = O(1)$. By Liapounov's Inequality, $\sup_i V^2_{iq} = O(1)$ implies $\sup_i|V_{iq}| = O(1)$ for all $q = 1, ..., K$.

It then follows from reduced form for Y_n that

$$\begin{aligned} & \sup_i E(|y_i||\Delta_n) = \sup_i E\left(\left|\textstyle\sum_j(S_n^{-1})_{ij}\left(\alpha_0 + x_j'\beta_0 + \sum_k C_{jk}x_k'\gamma_0 + \varepsilon_j\right)\right|\Big|\,\Delta_n\right) \\ \le\ & \sup_i\left[\textstyle\sum_j(S_n^{-1})_{ij}\right] \times \sup_j E\left(|\alpha_0| + |x_j'\beta_0| + \textstyle\sum_k|C_{jk}| \times |x_k'\gamma_0| + |\varepsilon_j|\,\Big|\,\Delta_n\right). \end{aligned}$$

Hence, there exists some constant $M^* < \infty$ with $\Pr\{\sup_i E(|y_i||\Delta_n) \leq M^*\} = 1$. $\square$

Proof of Proposition 1 . Recall

$$\widehat{\theta} - \theta_0 = \left[\frac{\widetilde{R}_n'\widetilde{V}_n}{n}\left(\frac{\widetilde{V}_n'\widetilde{V}_n}{n}\right)^{-1}\frac{\widetilde{V}_n'\widetilde{R}_n}{n}\right]^{-1}\frac{\widetilde{R}_n'\widetilde{V}_n}{n}\left(\frac{\widetilde{V}_n'\widetilde{V}_n}{n}\right)^{-1}\frac{\widetilde{V}_n'\widetilde{\epsilon}_n}{n} \tag{4}$$

where

$$\begin{aligned}\frac{1}{n}\widetilde{V}_n'\widetilde{R}_n &= \frac{1}{n}V_n'R_n + \frac{1}{n}V_n'(0,\Delta_{1n}Y_n,0,\Delta_{2n}X_n)\\ &\quad+\frac{1}{n}(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)'R_n\\ &\quad+\frac{1}{n}(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)'(0,\Delta_{1n}Y_n,0,\Delta_{2n}X_n).\end{aligned}$$

$$\begin{aligned}\frac{1}{n}\widetilde{V}_n'\widetilde{V}_n &= \frac{1}{n}V_n'V_n + \frac{1}{n}V_n'(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)\\ &\quad+\frac{1}{n}(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)'V_n\\ &\quad+\frac{1}{n}(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)'(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)\end{aligned}$$

$$\begin{aligned}\frac{1}{n}\widetilde{V}_n'\widetilde{\epsilon}_n &= \frac{1}{n}V_n'\epsilon_n - \frac{1}{n}\lambda_0 V_n'\Delta_{1n}Y_n - \frac{1}{n}V_n'\Delta_{2n}X_n\gamma_0\\ &\quad+\frac{1}{n}(0,(G_n\Delta_{1n}+\Delta_{1n}G_n+\Delta_{1n}^2)X_n,0,\Delta_{2n}X_n)'(\epsilon_n-\lambda_0\Delta_{1n}Y_n-\Delta_{2n}X_n\gamma_0).\end{aligned} \tag{5}$$

Due to Assumption 2 and Lemma A2, $\sup_i V_iV_i' = O(1)$. Lemma A2 also suggests that V_n and $X_n\gamma_0$ satisfy the dominance conditions on the vectors a_n; Y_n and ϵ_n satisfy the dominance conditions b_n in Lemma A1. The second to the fourth terms on the RHS of (5) can all be expressed as $\frac{1}{n}a_n'\Delta_n b_n$ in Lemma A1, and hence are $O_p(n^{s-1})$. Because $\frac{1}{n}V_n'\epsilon_n = O_p(n^{-1/2})$, it then follows that $\frac{1}{n}\widetilde{V}_n'\widetilde{\epsilon}_n = O_p(n^{-1/2}\vee n^{s-1})$.

Proof of Proposition 2. As

$$\sqrt{n}(\widehat{\theta}-\theta_0) = \left[\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'R_n}{n}\right]^{-1}\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'\epsilon_n}{\sqrt{n}} + O_p(n^{s-1/2}),$$

when $s < 1/2$, $\sqrt{n}(\widehat{\theta}-\theta_0)$ has the same asymptotic distribution as the 2SLS estimator using true network links.

Consider the asymptotic variance Ω. Let Σ_n be the diagonal matrix of the error variance, i.e., $\Sigma_{ii} = E(\varepsilon_i^2)$. We have $\Omega = A^{-1}BA^{-1}$, where

$$\begin{aligned}A &= p\lim\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'R_n}{n};\\ B &= p\lim\frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\left(\frac{1}{n}V_n'\Sigma_nV_n\right)\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'R_n}{n}.\end{aligned}$$

Using Lemma A1, we can show that

$$\widehat{A} = A + O_p(n^{s-1})$$

and

$$\widehat{B} = B + \frac{R_n'V_n}{n}\left(\frac{V_n'V_n}{n}\right)^{-1}\left(\frac{1}{n}\widetilde{V}_n'\widehat{\Sigma}_n\widetilde{V}_n - \frac{1}{n}V_n'\Sigma_nV_n\right)\left(\frac{V_n'V_n}{n}\right)^{-1}\frac{V_n'R_n}{n} + O_p(n^{s-1}).$$

Then, what left is to show that from the fact that $\frac{1}{n}\widetilde{V}_n'\widehat{\Sigma}_n\widetilde{V}_n - \frac{1}{n}V_n'\Sigma_nV_n$ is $o_p(1)$. As

$$\frac{1}{n}\widetilde{V}_n'\widehat{\Sigma}_n\widetilde{V}_n - \frac{1}{n}V_n'\Sigma_nV_n = \frac{1}{n}V_n'\left(\widehat{\Sigma}_n - \Sigma_n\right)V_n + O_p(n^{s-1}),$$

and the first term on the RHS is $O_p(n^{-1/2} \vee n^{s-1})$ because

$$\begin{aligned}
\frac{1}{n}V_n'\left(\widehat{\Sigma}_n - \Sigma_n\right)V_n &= \frac{1}{n}\sum_{i=1}^{n}\left((Y_n - \widetilde{R}_n\widehat{\theta})_{(i)}]^2 - E(\varepsilon_i^2)\right)v_iv_i' \\
&= \frac{1}{n}\sum_{i=1}^{n}v_iv_i'[\varepsilon_i^2 - E(\varepsilon_i^2)] + \frac{1}{n}\sum_{i=1}^{n}v_iv_i'\left([\widetilde{R}_i(\theta_0 - \widehat{\theta})]^2 + [(\lambda_0\Delta_{1n}Y_n + \Delta_{2n}X_n\gamma_0)_{(i)}]^2\right) \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}v_iv_i'\widetilde{R}_i(\theta_0 - \widehat{\theta})\varepsilon_i - \frac{2}{n}\sum_{i=1}^{n}v_iv_i'[\widetilde{R}_i(\theta_0 - \widehat{\theta}) + \varepsilon_i](\lambda_0\Delta_{1n}Y_n + \Delta_{2n}X_n\gamma_0)_{(i)} \\
&= O_p(n^{-1/2}) + O_p(\theta_0 - \widehat{\theta}) + O_p(n^{s-1}) = O_p(n^{-1/2} \vee n^{s-1})
\end{aligned}$$

Together, we have $\widehat{A}^{-1}\widehat{B}\widehat{A}^{-1} - A^{-1}BA^{-1} = O_p(n^{-1/2} \vee n^{s-1}) = o_p(1)$.

References

Blume, L., W. Brock, S. Durlauf, R. Tayaraman, 2015. Linear social interactions models. Journal of Political Economy 123 (2): 444-496.

Boucher, V., Y. Bramoullé, H. Djebbari, B. Fortin, 2014. Do peers affect student achievement? Evidence from Canada using group size variation. Journal of Applied Econometrics 29: 91–109.

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

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

De Paula Á., I. Rasul, P. CL. Souza, 2018. Recovering social networks from panel data: identification, simulations and an application, CeMMAP working papers CWP58/18.

Griffith, A. 2021. Name your friends, but only five? The importance of censoring in peer effects estimates using social network data. Working paper, The University of Washington.

Hauser, C., M. Pfaffermayr, G. Tappeiner, J. Walde, 2009. Social capital formation and intra familial correlation: A social panel perspective. Singapore Economic Review 54: 473–488.

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

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

Lewbel, A., X. Qu, X. Tang, 2021 Social Networks with Unobserved Links. Boston College Working Papers in Economics 1004.

LeSage, J. P., R. K. Pace, 2009. Introduction to spatial econometrics. Taylor Francis/CRC Press, Boca Raton.

Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. Journal of Labor Economics 28: 825–860.

Lin, X., L. Lee, 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34-52.

Lin, X., 2015. Utilizing spatial autoregressive models to identify peer effects among adolescents. Empirical Economics, 49(3), 929-960.

Liu X, E. Patacchini, Y. Zenou, 2014. Endogenous peer effects: local aggregate or local average? Journal of Economic Behavior and Organization 103: 39–59.

Manski, C F., 1993. Identification of Endogenous Social Effects: The Reflection Problem. Review of Economic Studies 60 (3): 531–42.

Patacchini E., Y. Zenou, 2012. Juvenile delinquency and conformism. Journal of Law, Economics, and Organization 28: 1–31.

Patacchini, E., G. Venanzoni, 2014. Peer effects in the demand for housing quality. Journal of Urban Economics, 83(0), 6-17.