Improved Estimation of Peer Effects using Network Data

Richard Startz

Alex Wood-Doughty*

October 2016

Preliminary: Please do not cite or distribute

Abstract

The ability to estimate peer effects in network models has been advanced considerably by the IV model of Bramoullé *et al.* (2009). While such IV estimates work well for very sparse networks, they exhibit very weak power for networks of even modest densities. We review and extend the findings of Bramoullé *et al.* (2009) and then propose an alternative estimator. We show that our new estimator works approximately as well as IV in very sparse networks and performs much better in networks of moderate density. To highlight the benefits of our proposed estimator, we provide an empirical application where we estimate peer effects in individual schools.

JEL codes: D85, C3, C36, L14

1   Introduction

Estimation of peer effects in networks has undergone enormous growth. Much of this growth is due to the breakthrough work of Bramoullé *et al.* (2009) (hereafter BDF) who show that if we know the network structure–who interacts with whom–and there is enough sparsity of the network, peer effects are identified. Much of the peer-effect literature is framed in the Manski (1993) linear-in-means framework,i.e.,my outcome depends linearly on the average outcomes of my friends. Manski shows that, without knowing anything about the group structure, the linear-in-means model is not identified. BDF make two contributions. First, they show that the additional information available from knowing the structure of network interactions often permits identification. Second, they offer an instrumental variable strategy for estimating such models.

While BDF prove that identification holds for almost all networks, their simulations show that meaningful inference is only possible in very sparse networks. For networks with density above 5%, (the average person is connected with 5% of the network), their instrumental variables estimation offers very limited power. Our primary contribution is to offer a nonlinear estimator that we show has enormously improved power in denser networks with only a small departure from nominal size. We also extend and confirm BDF's Monte Carlo results on bias and precision, and further frame the issues in the weak instrument framework. (See Nelson & Startz (1990a,b) and Staiger & Stock (1997))

Much of the practical concern about network density arises in the study of friendship networks in school, which are typically relatively dense within a classroom or a single school. BDF use their technique to identify peer effects of recreational activities by middle and high-schoolers. Their technique has also been used in a number of other papers: De Giorgi *et al.* (2010) look at peer effects in college major choice, Lin (2010) looks at academic achievement, and Fortin & Yazbeck (2015) look at obesity. Calvo-Armengol *et al.* (2009) use a similar method to investigate peer effects in school performance. All of these studies use data on the friendship networks in schools to identify the peer effects.

While some schools are sufficiently large to ensure a low density friendship network, there are many with densities above 5%. To get around this problem, the previous studies generate a sparse network by stacking multiple school networks and assuming there is no interaction across schools. To illustrate the increased power of our proposed method, we give an empirical example in which we identify peer effects without stacking across schools.

The rest of the paper is organized as follows. Section 2 describes the linear-in-means model and explains the common method of estimating peer effects. Section 3 provides some Monte Carlo simulations that highlight the problems with the common method for

higher density networks. Section 4 proposes an alternative estimation strategy and Section 5 provides simulations that highlight its benefits. Section 6 applies the new method to the Add Health data and estimates school level peer effects.

2   Linear-in-Means Model

Consider the following model, taken from BDF:

$$y_i = \alpha + \beta \frac{1}{n_i} \sum_{j \in g_i} y_j + \gamma x_i + \delta \frac{1}{n_i} \sum_{j \in g_i} x_j + \epsilon_i \tag{1}$$

$$\epsilon_i | X \sim iidN(0, \sigma_{\mathcal{E}}^2)$$

where g_i is the set of friends of person i and n_i is the size of g_i. We model person i's outcome, y_i, as a simple average of her friends' outcomes, y_j, an observable characteristic, x_i, and an average of her friends' observable characteristics, x_j. Following Manski (1993), we will call β the endogenous social effect and δ the exogenous social effect. The endogenous social effect, or peer effect, can be interpreted as the relationship between an average friend's outcome and your outcome. In a classroom setting, we might be interested in the peer effect of test scores: how your average friend's test score affects your own score. We assume $|\beta| < 1$, as is standard. The exogenous social effect measures the influence of your friends' characteristics on your outcome. If we believe that the education level of a student's parents affects their achievement in the classroom, we might also believe the education level of their friends' parents has an effect. Importantly, we want to separately identify the endogenous social effect from the exogenous effect. They are both capturing the influence of peers, however they have very different policy implications. In particular, in many situations the endogenous social effect creates a multiplier on changes in the exogenous variables.

Following BDF, we can stack the observations and write the model as:

$$Y = \alpha \iota + \beta GY + \gamma X + \delta GX + \mathcal{E} \tag{2}$$

$$\mathcal{E} | X \sim N(0, \sigma_{\mathcal{E}}^2 I_n)$$

where G is the adjacency matrix, a row-normalized $n \times n$ matrix that describes the network and ι is an $n \times 1$ vector of ones. The adjacency matrix starts as a matrix of ones and zeros where $G_{ij} = 1$ if i is friends with j. Each row is then normalized by its sum so both the endogenous and exogenous social effects can be interpreted as the effect of an average friend. Since this is a simultaneous system, in order to separately identify both β and δ we need an

instrument. To see this, we can rewrite Equation 2 to give the reduced form by collecting terms on Y:

$$Y = \alpha\iota + \beta GY + \gamma X + \delta GX + \mathcal{E}$$
$$= (I_n - \beta G)^{-1}(\alpha\iota + \gamma X + \delta GX + \mathcal{E}) \tag{3}$$

We can write $(I_n - \beta G)^{-1}$ as a series expansion, $\sum_{k=0}^{\infty} \beta^k G^k$, so:

$$Y = \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \gamma X + \delta GX + \mathcal{E}) \tag{4}$$

$$= \sum_{k=0}^{\infty} \beta^k G^k (\gamma X + \delta GX) + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) \tag{5}$$

$$= \gamma X + \sum_{k=1}^{\infty} \beta^k G^k \gamma X + \sum_{k=0}^{\infty} \beta^k G^k \delta GX + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) \tag{6}$$

$$= \gamma X + \sum_{k=1}^{\infty} \beta^k G^k \gamma X + \sum_{k=1}^{\infty} \beta^{k-1} G^k \delta X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) \tag{7}$$

$$= \gamma X + \sum_{k=1}^{\infty} (\beta^k \gamma + \beta^{k-1}\delta) G^k X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) \tag{8}$$

$$= \gamma X + \sum_{k=1}^{\infty} \beta^{k-1}(\beta\gamma + \delta) G^k X + \sum_{k=0}^{\infty} \beta^k G^k (\alpha\iota + \mathcal{E}) \tag{9}$$

$$= \gamma X + (\beta\gamma + \delta) GX + \beta(\beta\gamma + \delta) GGX + \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta) G^k X + \frac{\alpha}{1-\beta}\iota + \sum_{k=0}^{\infty} \beta^k G^k \mathcal{E} \tag{10}$$

If we only use information on X and GX, we cannot separately identify $[\alpha, \beta, \gamma, \delta]$. However, since GGX does not show up directly in Equation 2 but it does show up in the reduced form of GY, BDF propose that the characteristics of the friends of friends (GGX) can be used to instrument for the friends' outcomes, GY, and Equation 2 can be estimated via 2SLS. The structural parameters are identified as long as GGX is not perfectly collinear with ι, X, and GX.[1] Another way of saying this is that there must be some intransitive triads: second-order friends who are not also first-order friends. The instrument must be providing some information about GY. While this is true for most networks, the amount of information decreases as the network gets more dense. As everyone has more friends,

---

[1]Generally, for any $p > 1$, $G^p X$ is a valid instrument for GY as long as $G^p X$ is not perfectly collinear with $[\iota, X, GX]$.

everyone's second-order friends start to look more and more similar. We show below that "weak identification" in this context results in biased standard errors but only modestly biased coefficients. In other words, the classical IV asymptotic distributions are about right although the estimated distributions are not. The more important issue, correctly identified by the asymptotic distributions, is that the estimator is of very low power.

2.1 Estimation

Once we have an instrument, estimation proceeds via 2SLS.[2] Following BDF, we assume there are network-specific unobservables, α. To control for these unobservable characteristics, we apply a within-transformation by multiplying each term by $I_n - \frac{1}{n}(\iota \iota')$. Note, that this transformation changes the identification requirement slightly. As proven in BDF, now the structural parameters are identified as long as $GGGX$ is not perfectly collinear with $[X, GX, GGX]$.

Define the following:

$$R = I_n - \frac{1}{n}(\iota \iota') \tag{11}$$

$$W = [RGY, RX, RGX] \tag{12}$$

$$Z = [RX, RGX, RGGX] \tag{13}$$

$$P = Z(Z'Z)^{-1}Z' \tag{14}$$

$$\theta = [\beta, \gamma, \delta]' \tag{15}$$

Then the 2SLS estimator is:

$$\hat{\theta}_{2SLS} = (W'PW)^{-1}W'PRY \tag{16}$$

and the asymptotic variance is:

$$Var(\hat{\theta}_{2SLS}) = \sigma_{\mathcal{E}}^2 (W'PW)^{-1} \tag{17}$$

where σ^2 is estimated by

$$\hat{\sigma_{\mathcal{E}}}^2 = \frac{1}{n}((RY - W\hat{\theta})'(RY - W\hat{\theta})) \tag{18}$$

---

[2]Lee *et al.* (2010) adopted spatial models to allow estimation via maximum likelihood. They show their method is more efficient than 2SLS, however they are still relying on variation in GGX for identification.

3   2SLS Simulation Results

To examine the properties of this estimator, we run Monte Carlo simulations. We generate data from known parameters and then estimate the model in an attempt to recover those parameters. We will vary the density of the network to see how that affects the estimation. We follow BDF by setting the network size, $N = 240$, $\alpha = 0.7683$, $\beta = 0.4666$, $\gamma = 0.0834$ and $\delta = 0.1507$. We draw X from a log-normal distribution with mean 1 and variance 3. For each simulation, we generate a random reciprocal network with the probability of any link equal to the density. We then row normalize so the weight on each individual's friends sums to 1. We also draw the error vector, $\mathcal{E} \sim N(0, 0.1)$. From there, we can generate the outcome Y according to Equation 3. We simulate 10,000 networks with each of the following densities: 0.01, 0.025, 0.05, 0.1, 0.2, 0.3.

Figure 1 plots the distribution of the estimated $\hat{\beta}$s. As in the BDF findings, the estimates become imprecise around a density of 5% and there is no meaningful inference for dense networks, in that the distributions become essentially flat. Note also some increase in bias as the density increases.

We are mainly interested in inference about $\hat{\beta}$, so we construct the following test statistic: $t_{\hat{\beta}} = \frac{\hat{\beta} - 0.4666}{SE_{\hat{\beta}}}$ and compare the statistic to a 5 percent critical value. Figure 2 plots the empirical size of this test. At very low densities, the test achieves approximately the correct size. However, in more dense networks the test is notably undersized.
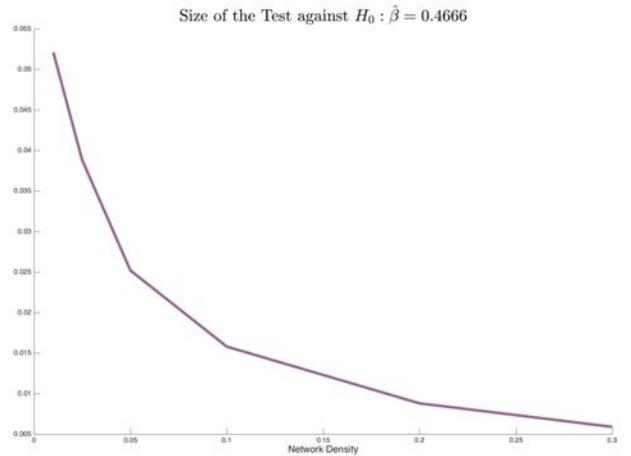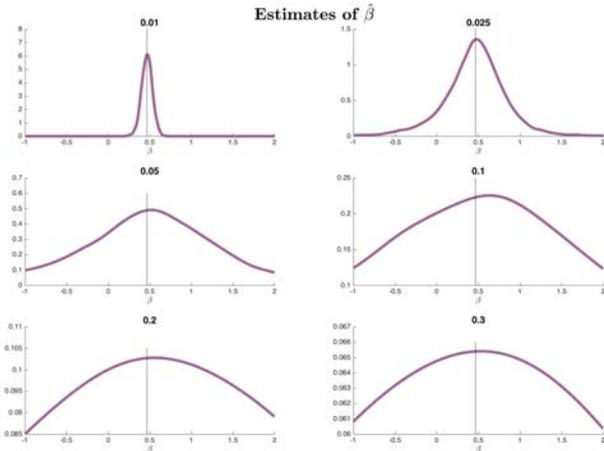

Figure 1: Distribution of $\hat{\beta}$ for varying network densities

Figure 2: Size of $\hat{\beta}$ for varying network densities

To test the power, we let the true β vary uniformly between $[0, 1]$ and then test the hypothesis that $\hat{\beta} = 0.4666$. Figure 3 plots the power of the test at specified densities.

Even for the 5% network there is limited power. For densities of 10 percent or larger power
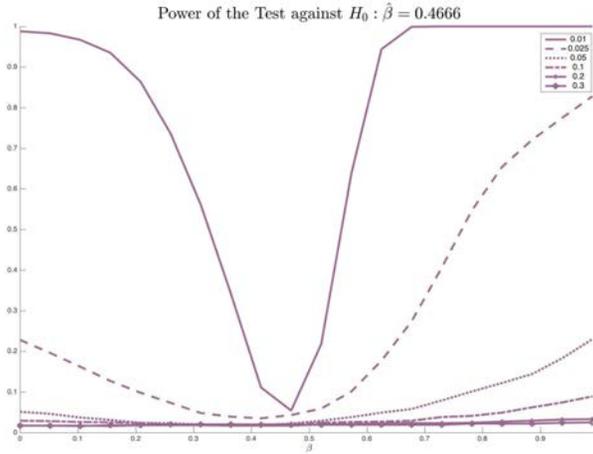
Figure 3: Power

effectively equals size, meaning there is no information in the test statistic. To investigate the cause of this low power, we look at the first stage regression:

$$RGY = \pi_1 RX + \pi_2 RGX + \pi_3 RGGX + V \qquad (19)$$

Figure 4 plots the median F-statistic for the coefficient on the instrument, π_3. This shows that the instrument is very weak for the 5% density network and closely mirrors Figure 3.
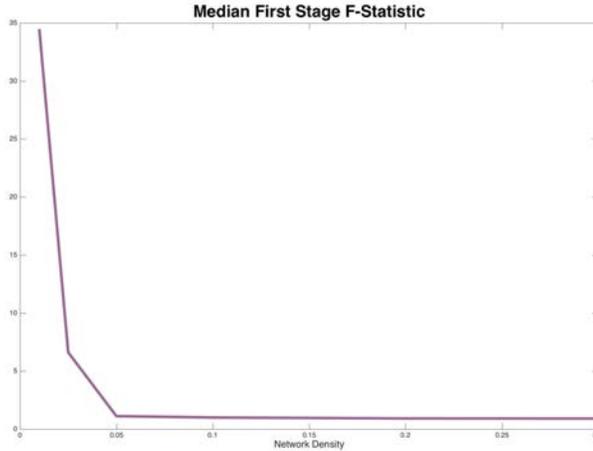

Figure 4: First Stage F-Statistic

The normal cause of a weak instrument problem is that the chosen instrument is only slightly correlated with the endogenous variable (Nelson & Startz, 1990a). Here, the reason the instrument is so weak is because GGX is highly collinear with $[\iota, X, GX]$. Figure 5 plots

7

the log of the median condition number for $W'PW$. The condition number is a measure of collinearity, and numbers above 3.4 (log(30)) are evidence of high multicollinearity (Belsley *et al.*, 1980). Here we can clearly see the relationship between collinearity, the weak instrument, and the lack of power of the estimates. The reason we are getting weak identification for higher density networks is because GGX is not providing much new information.
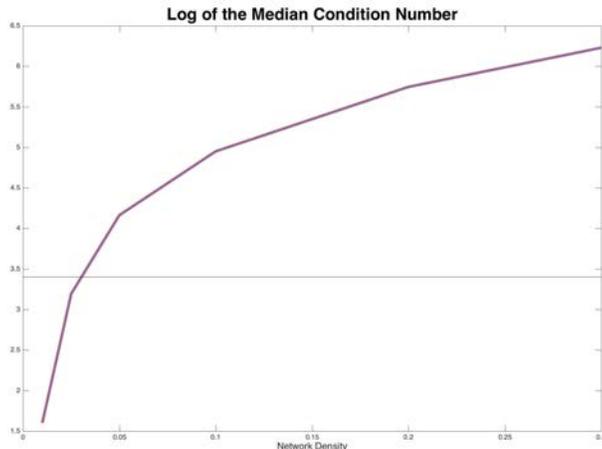

Figure 5: Log Condition Number

3.1 Anderson-Rubin Test

To further investigate the weak instrument hypothesis, we can run an Anderson-Rubin (1949) test, which should have exact finite sample size up to Monte Carlo error. Figure 6a plots the size of the Anderson-Rubin test. As expected the test has the proper 5% size across all densities. Figure 6b plots the power of the Anderson-Rubin test by varying the true β uniformly between $[-1, 1]$ and setting $\beta_0 = 0$. This shows the well known fact that the Anderson-Rubin test has limited power. However, a comparison to Figure 3 shows that the power of the Anderson-Rubin test is not much worse than the power of 2SLS. This, together with the multicollinearity results, suggests that the difficulties with instrumental variable inference in denser networks reflect a low information content rather than a failure of the usual asymptotic approximations.
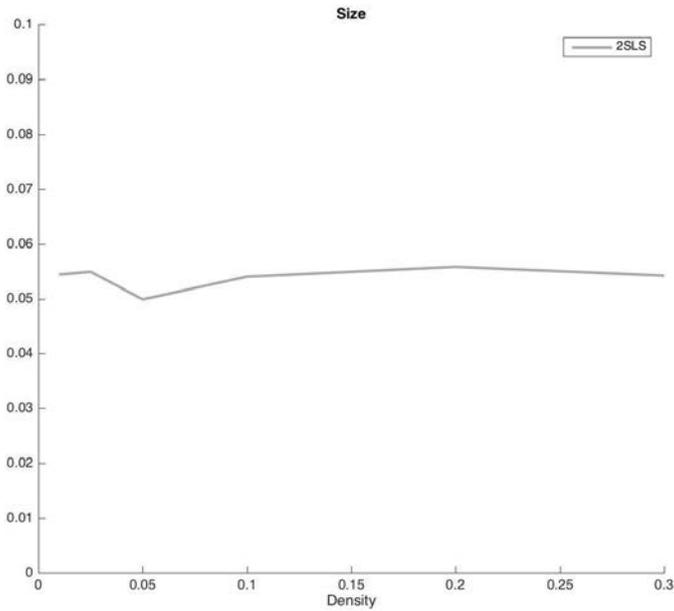
(a) Size
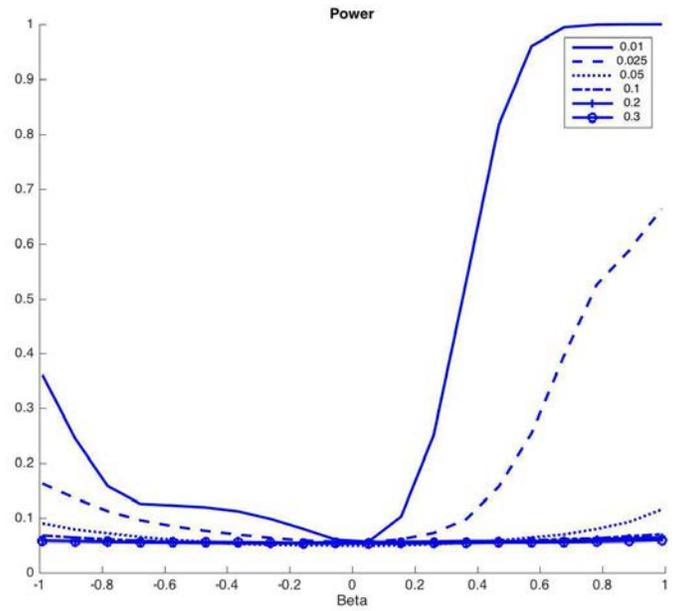
(b) Power

Figure 6: Anderson-Rubin Test

3.2   Size with True Error Variance

To investigate the causes of the test being undersized, we recompute the standard errors using the true error variance, σ^2, in place of the estimated error variance. Figure 7 plots the size of the test for both standard 2SLS and using the true σ^2. When we use the true error variance, the test has proper size, suggesting that the reason 2SLS is undersized is that it is overestimating the error variance. Thus the asymptotic distribution gives a good approximation to the empirical distribution with $N = 240$.
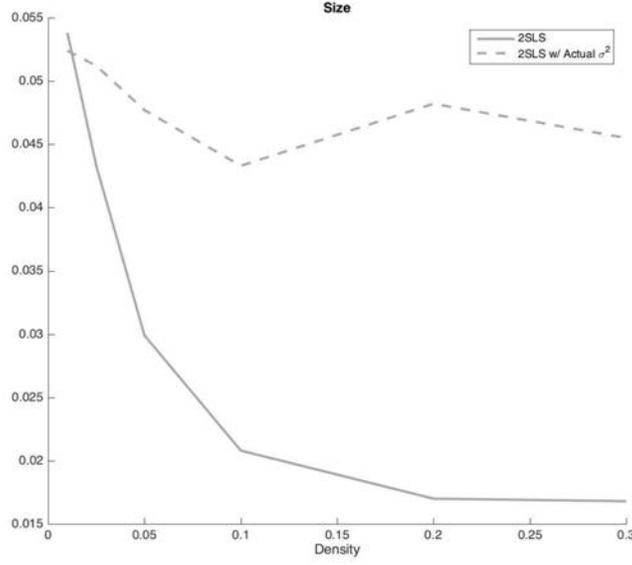
Figure 7: Size with True Error Variance

4 Alternative Estimation Strategy (Residual Series Expansion Estimation)

As seen in the previous section, any estimation strategy that relies on variation in GGX will have difficulty with dense networks. To avoid this, we exploit another part of Equation 10 to identify β:

$$
\begin{aligned}
Y = {} & \gamma X + (\beta\gamma + \delta)GX + \beta(\beta\gamma + \delta)GGX \\
& + \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)G^k X + \frac{\alpha}{1-\beta}\iota + \sum_{k=0}^{\infty} \beta^k G^k \mathcal{E} \\
= {} & \gamma X + (\beta\gamma + \delta)GX + \beta(\beta\gamma + \delta)GGX \\
& + \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)G^k X + \frac{\alpha}{1-\beta}\iota \\
& + \mathcal{E} + \beta G\mathcal{E} + \beta^2 G^2 \mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E}
\end{aligned}
\tag{20}
$$

Notice that β shows up in the series expansion of the error term. The goal of our strategy is to estimate β from this series expansion. To do this, we must first get a good approximation of the error term. Since $\mathcal{E}$ is uncorrelated with X, the regression of Y on $[\iota, X, GX, GGX]$

will generate the best linear predictor of Y.[3] As we have already seen, the omitted higher order terms are highly correlated with the included terms. Because the higher order terms are projected onto the included terms, and also because powers of β fade toward zero, the residuals will approximate the series expansion with $\mathcal{E}$.

$$Y = \lambda_0 \iota + \lambda_1 X + \lambda_2 GX + \lambda_3 GGX + \eta \tag{21}$$

$$e = Y - \hat{Y} = \sum_{k=3}^{\infty} \beta^{k-1}(\beta\gamma + \delta)(G^k X - \mu_{0k} - \sum_{l=1}^{3} \mu_{lk} G^{l-1} X) \tag{22}$$

$$+ \mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E}$$

$$\approx \mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E} \tag{23}$$

where the μ coefficients come from the projection of the higher order terms onto the included terms,

$$\forall k \in \{3, \ldots, \infty\} \quad G^k X = \mu_{0k}\iota + \mu_{1k}X + \mu_{2k}GX + \mu_{3k}GGX + \nu \tag{24}$$

The key for recovering a good approximation of Equation 23 is that the omitted higher order terms are well approximated by the included terms, in other words, that the R^2 of each Equation 24 is high.

Multiplying Equation 23 by G gives,

$$Ge \approx G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1}\mathcal{E} \tag{25}$$

and by combining Equations 23 and 25 we can write,

$$e = \beta Ge + \mathcal{E} \tag{26}$$

The least squares estimator of β in Equation 26 is biased since $Cov(Ge, \mathcal{E}) \neq 0$, but we can explicitly write out the probability limit of the OLS estimator.

$$plim[\hat{\beta}_{OLS}] = \frac{Cov(e, Ge)}{Var(Ge)} \tag{27}$$

We evaluate Equation 27 by writing out the numerator and denominator and then taking

---

[3]Any subset of $G^p X$ will also work, see Subsection 4.1

advantage of the fact that $\mathcal{E}$ is the only source of uncertainty and that σ^2 cancels between the numerator and the denominator. This leaves β as the only unknown. After dropping high-order terms in the expansions we solve numerically for β.

$$Cov(e, Ge) = Cov\left(\left(\mathcal{E} + \beta G\mathcal{E} + \beta^2 GG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^k \mathcal{E}\right), \left(G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1} \mathcal{E}\right)\right)$$

(28)

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \beta^{i+j-1} Cov(G^i \mathcal{E}, G^j \mathcal{E})$$

(29)

$$= \sum_{k=0}^{\infty} \beta^k \sum_{l=0}^{k} Cov(G^l \mathcal{E}, G^{k-l+1} \mathcal{E})$$

(30)

and

$$Var(Ge) = Var(G\mathcal{E} + \beta GG\mathcal{E} + \beta^2 GGG\mathcal{E} + \sum_{k=3}^{\infty} \beta^k G^{k+1} \mathcal{E})$$

(31)

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \beta^{i+j-2} Cov(G^i \mathcal{E}, G^j \mathcal{E})$$

(32)

$$= \sum_{k=0}^{\infty} \beta^k \sum_{l=1}^{k+1} Cov(G^l \mathcal{E}, G^{k-l+2} \mathcal{E})$$

(33)

Rearranging Equation 27 and combining on β, gives us:

$$Cov(e, Ge) - plim[\hat{\beta}_{OLS}] Var(Ge) = 0$$

(34)

$$\Rightarrow \left(\sum_{k=0}^{\infty} \beta^k \sum_{l=0}^{k} Cov(G^l \mathcal{E}, G^{k-l+1} \mathcal{E})\right) - \left(plim[\hat{\beta}_{OLS}] \sum_{k=0}^{\infty} \beta^k \sum_{l=1}^{k+1} Cov(G^l \mathcal{E}, G^{k-l+2} \mathcal{E})\right) = 0$$

(35)

$$\Rightarrow \sum_{k=0}^{\infty} \beta^k \left(\sum_{l=0}^{k} Cov(G^l \mathcal{E}, G^{k-l+1} \mathcal{E}) - plim[\hat{\beta}_{OLS}] \sum_{l=1}^{k+1} Cov(G^l \mathcal{E}, G^{k-l+2} \mathcal{E})\right) = 0$$

(36)

which is a polynomial with respect to β. Since

$$Cov(G^p\mathcal{E}, G^q\mathcal{E}) \approx \frac{\sum diag(G^p * G^q)}{N}\sigma_\mathcal{E}^2 \tag{37}$$

we can approximate the polynomial and solve for its root, which will be our estimate of $\hat{\beta}_{RSEE}$. Since Equation 36 is an infinite series, we need to choose an expansion point. Our testing has shown that $k = 3$ works well in practice, so the formula becomes:

$$plim[\hat{\beta}_{OLS}] \approx \beta + \frac{Cov(\mathcal{E}, G\mathcal{E}) + \beta Cov(\mathcal{E}, GG\mathcal{E}) + \beta^2 Cov(\mathcal{E}, GGG\mathcal{E}) + \beta^3 Cov(\mathcal{E}, GGGG\mathcal{E})}{Var(G\mathcal{E}) + \beta^2 Var(GG\mathcal{E}) + 2\beta^2 Cov(G\mathcal{E}, GGG\mathcal{E}) + 2\beta^4 Cov(GG\mathcal{E}, GGGG\mathcal{E})} \tag{38}$$

4.1 Number of Terms in First-Stage Regression

As previously shown, the goal of the first-stage regression is for the residuals to approximate the series expansion of $\mathcal{E}$. Because we include a small number of terms in Equation 21, the residuals might not be a good approximation. However, due to the collinearity between the various powers of the adjacency matrix, there is not much new information in the higher order terms. Figure 8 plots the squared correlation coefficient between the residuals and the true series expansion of $\mathcal{E}$ for three different first-stage equations. While only including two terms does worse, especially for low density networks, there is hardly any difference once we move beyond four terms. The first-stage equation is doing a very good job of approximating the error term.
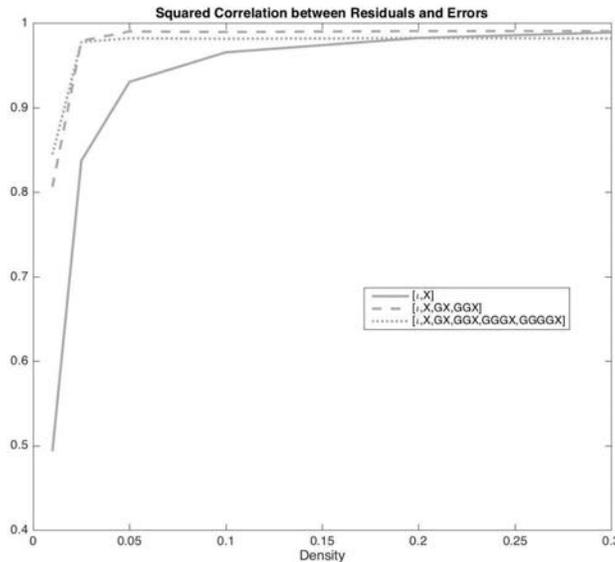

Figure 8: Comparison of First-Stage Equations

13

4.2 Standard Error Estimation

Standard errors can be estimated either via the delta method or by the bootstrap. Notice from Equation 38 that we solve $\hat{\beta}_{OLS} = f(\beta)$, for some implicit function f. Therefore $\hat{\beta}_{RSEE} = f^{-1}(\hat{\beta}_{OLS})$ and we can apply the delta method:

$$\sqrt{n}\left(f^{-1}(\hat{\beta}_{OLS}) - f^{-1}(\beta)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\hat{\beta}_{OLS}}[f^{-1}]'(\beta)^2\right)$$

$$\Rightarrow \sqrt{n}\left(\hat{\beta}_{RSEE} - \beta\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2_{\hat{\beta}_{OLS}}}{f'(\beta)^2}\right)$$

Alternatively, we generate bootstrapped standard errors. For an estimated $\hat{\beta}_{RSEE}$, we construct $Y - \hat{\beta}_{RSEE}GY$ and regress this on $[\iota, X, GX]$. Since this is no longer a simultaneous system, we can get estimates of α, γ, and δ via OLS.

$$Y - \hat{\beta}_{RSEE}GY = \alpha\iota + (\beta - \hat{\beta}_{RSEE})GY + \gamma X + \delta GX + \mathcal{E} \tag{39}$$

From these estimates, we generate residuals.

$$J = \left(Y - \hat{\beta}_{RSEE}GY\right) - \left(\hat{\alpha}\iota + \hat{\gamma}X + \hat{\delta}GX\right) \tag{40}$$

We then multiply the residuals by either -1 or 1 and reassign them to a new observation (with replacement). With the new residuals, we compute a new Y_b and then estimate a new $\hat{\beta}$.

$$p_i \in \{-1, 1\} \tag{41}$$

$$\mathcal{E}^b_j = J_i p_i \tag{42}$$

$$Y^b = (I_n - \hat{\beta}_{RSEE}G)^{-1}(\hat{\alpha}\iota + \hat{\gamma}X + \hat{\delta}GX + \mathcal{E}^b) \tag{43}$$

$$\Rightarrow \hat{\beta}^b_{RSEE} \tag{44}$$

Doing this a number of times gives a distribution of $\hat{\beta}$, and to construct the hypothesis test, we take the 2.5th and 97.5th percentiles of the distribution and see if the true β is contained in that region.

5   Simulation Results

We use the same simulation conditions as in Section 3 and estimate $\hat{\beta}$ using both 2SLS and RSEE. Figure 9 plots the distribution of the estimated $\hat{\beta}$s.
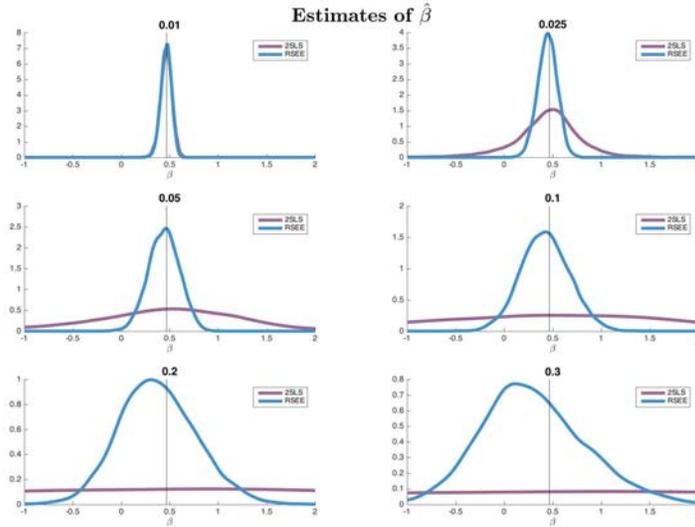

Figure 9: Distribution of $\hat{\beta}$

The RSEE estimates are very similar to 2SLS for very sparse networks and are significantly more precise for more dense networks. Figures 10 and 11 show the size and power, respectively. While 2SLS tests are undersized, our alternative tests are oversized. This is less so for the bootstrap tests, which have close to nominal size for moderately dense networks. For power, RSEE does much better, especially for more dense networks. For example, for networks with a 10 percent density the standard 2SLS tests are essentially uninformative against an alternative of 0, with size equal 0.021 and power equal 0.029. The bootstrap version of our new estimator has empirical size and power of 0.081 and 0.67, respectively.

It is often useful to consider size-adjusted power when evaluating a proposed estimator. To do this, we calculate the 2.5th and 97.5th percentiles of the empirical t-distribution. These correspond to the empirical critical values for that test. We then use those critical values when computing the power. Figure 12 plots the size-adjusted power. On a size-adjusted power, the bootstrap inference is much preferred to 2SLS results.
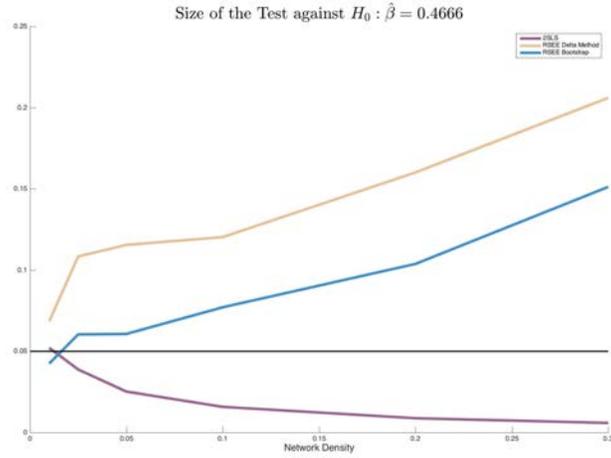
Figure 10: Size
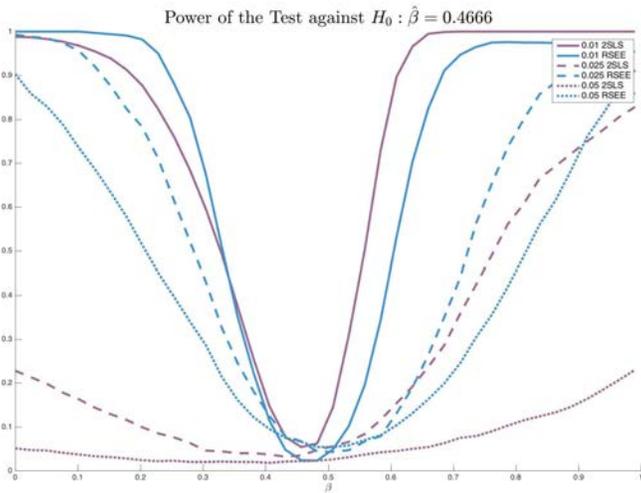
(a) Low Density
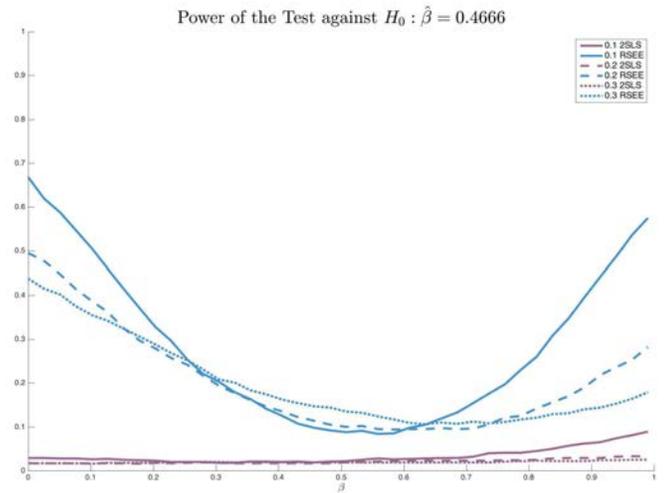
(b) High Density

Figure 11: Power
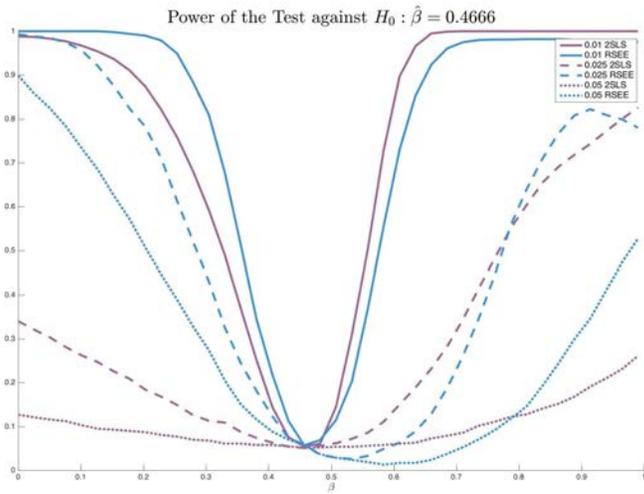
(a) Low Density
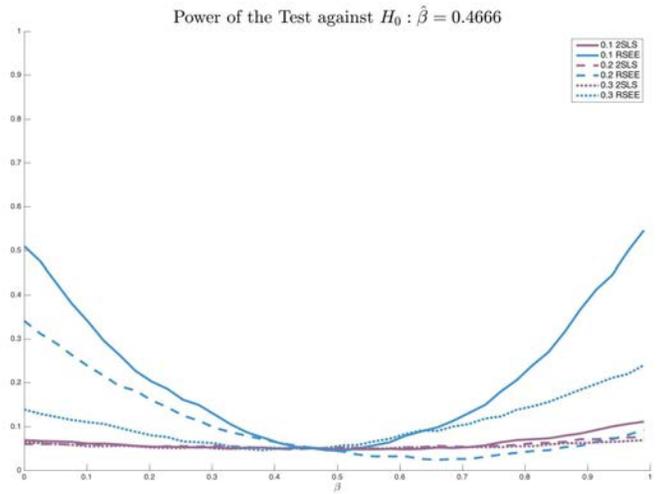
(b) High Density

Figure 12: Size-Adjusted Power

While it is clear that the RSEE has much more power, we might be concerned that we gain this power at the expense of more bias. Figure 13 shows the comparison of the median bias.
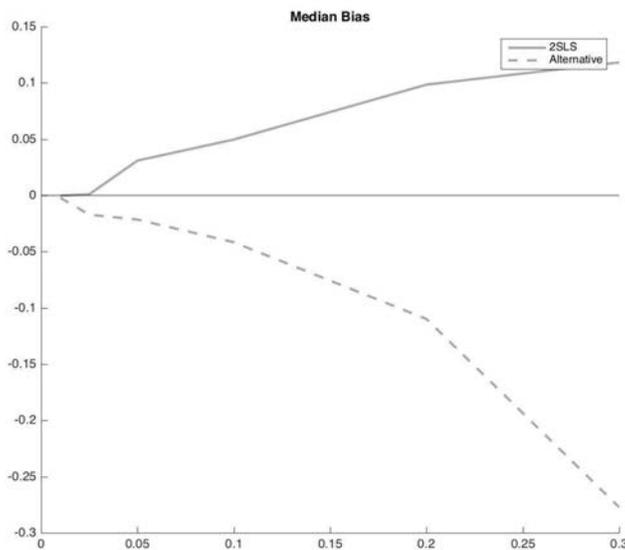
Figure 13: Median Bias

Both estimators become quite biased for very dense networks, however, the bias of RSEE is comparable to 2SLS for reasonably sparse networks.

5.1 Standard Error Estimation

The estimated standard errors from the alternative method are too small when we use the delta method. This is likely because this is for a second stage with generated regressors, since Equation 27 is a function of the residuals, e, which are a product of Equation 21. To test this, we plug in the true (series expansion of) error terms into Equation 27 and re-run the delta method. Figure 14 plots the size of both the original delta method and that without the generated regressors. This is evidence that the delta method is not providing accurate standard errors because of a generated regressors problem.
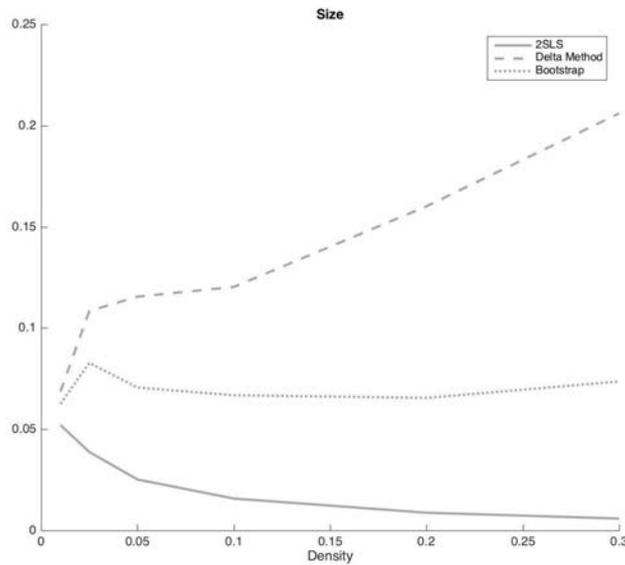

Figure 14: Comparison of Delta Methods

On the other hand, the bootstrap seems to be providing much more reasonable estimates of the standard error. Figure 15 shows the distribution of the bootstrapped standard errors relative to the empirical standard deviation. Here we can see the bootstrap is doing a reasonable job of approximating the standard error.
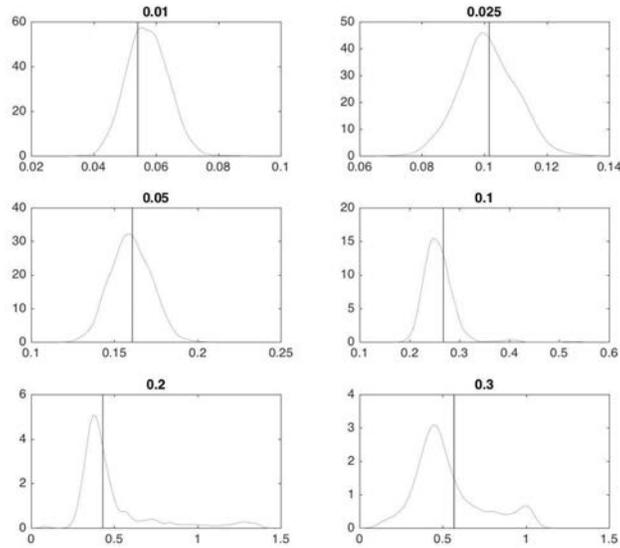
Figure 15: Distribution of Standard Errors

6 Empirical Application

To illustrate our method, we replicate the empirical results from BDF and then estimate peer effects for each school in the BDF sample. Following BDF, we use the In-School sample from Wave I of the National Longitudinal Study of Adolescent to Adult Health (Add Health). This is a nationally representative sample of 80 high schools and 52 middle schools. Each student in the sample was asked to fill out a questionnaire with questions pertaining to social and demographic characteristics, school activities and behavior, and parent education and occupation. The questionnaire also asked each student to list up to five male and five female friends. This friendship elicitation is used to construct the network.[4] The dependent variable is an index of participation in recreational activities and the covariates include characteristics of the student as well as their parents. We present a preliminary replication of the estimates from BDF, differing in that we have approximately 10,000 more observations. Table 1 shows that the summary statistics in BDF and our sample are quite similar.

---

[4]While the limit of ten total friends means that some friendships might be censored, in practice very few students list all ten.

	BDF		Our Sample	
Variable	Mean	SD	Mean	SD
Recreational Activities	2.122	1.267	1.823	1.453
Age	14.963	1.682	15.023	1.698
Female	0.535	0.499	0.507	0.500
Race is white only	0.619	0.486	0.578	0.494
Born in the US	0.928	0.259	0.909	0.287
Mother Present	0.929	0.257	0.919	0.273
Father Present	0.779	0.415	0.760	0.427
Grade 6 to 8	0.263	0.440	0.258	0.438
Grade 9 or 10	0.406	0.491	0.407	0.491
Grade 11 or 12	0.331	0.471	0.334	0.472
Parents' labor force participation	0.965	0.184	0.892	0.310
Mother No HS	0.097	0.296	0.108	0.310
Mother is HS grad	0.284	0.451	0.314	0.464
Mother more than HS but no college	0.276	0.447	0.150	0.357
Mother College grad	0.206	0.404	0.271	0.445
Mother went to school but unknown level	0.066	0.248	0.076	0.265
Father No HS	0.081	0.273	0.087	0.282
Father is HS grad	0.211	0.408	0.224	0.417
Father more than HS but no college	0.240	0.427	0.114	0.317
Father College grad	0.178	0.383	0.259	0.438
Father went to school but unknown level	0.069	0.253	0.076	0.264
Number of Observations	55208		65913	

Table 1: Summary Statistics

We follow BDF and construct one large block diagonal friendship network where each individual school network is placed on the diagonal. By combining all of the schools we have constructed a very sparse network: in our sample the density is $5.67 * 10^{-5}$. We then estimate the model using 2SLS and RSEE. Table 2 compares the results between BDF and our sample. Once again, there are some differences between the two samples, with the most important being that the estimated peer effect in our sample is smaller by a fraction of a standard error. Most of the other estimated coefficients are also fairly similar. The third column of results estimates the model using RSEE. There is not much different between the 2SLS and RSEE coefficient estimates, which is not surprising since it is a very sparse network. The estimated standard error on the peer effect is smaller by an order of magnitude when using RSEE.

	Variable	BDF 2SLS Coef	SE	Our Sample 2SLS Coef	SE	RSEE Coef	SE
Own Characteristics	Age	−0.022	0.011	0.083	0.003	−0.053	0.007
	Female	0.213	0.015	0.212	0.013	0.183	0.011
	Race is white only	−0.106	0.020	−0.092	0.016	−0.085	0.013
	Born in the US	−0.052	0.033	0.057	0.024	0.012	0.020
	Mother Present	−0.013	0.036	−0.035	0.031	−0.083	0.025
	Father Present	−0.018	0.029	−0.025	0.026	−0.044	0.022
	Grade 9 or 10	0.011	0.096	−0.213	0.032	−0.034	0.031
	Grade 11 or 12	0.021	0.102	−0.433	0.036	−0.049	0.038
	Mother is HS grad	−0.005	0.027	0.049	0.023	0.074	0.019
	Father is HS grad	0.047	0.029	0.081	0.025	0.099	0.022
	Mother more than HS but no college	0.146	0.029	0.202	0.027	0.230	0.022
	Father more than HS but no college	0.167	0.030	0.185	0.029	0.243	0.025
	Mother College grad	0.137	0.033	0.298	0.027	0.348	0.021
	Father College grad	0.127	0.031	0.278	0.027	0.333	0.023
	Mother went to school but unknown level	−0.010	0.038	−0.020	0.032	−0.024	0.027
	Father went to school but unknown level	−0.067	0.038	−0.019	0.033	−0.041	0.030
	Parents' labor force participation	0.083	0.040	0.124	0.025	0.088	0.021
Friends' Characteristics	Age	−0.061	0.020	−0.062	0.017	−0.060	0.004
	Female	0.008	0.048	−0.031	0.044	−0.062	0.022
	Race is white only	−0.019	0.045	−0.067	0.040	0.069	0.021
	Born in the US	0.042	0.066	0.041	0.060	0.115	0.036
	Mother Present	0.107	0.064	0.158	0.082	0.136	0.056
	Father Present	−0.109	0.053	−0.125	0.066	−0.151	0.047
	Grade 9 or 10	−0.034	0.186	0.021	0.090	0.050	0.039
	Grade 11 or 12	0.100	0.194	0.114	0.122	0.267	0.043
	Mother is HS grad	−0.047	0.050	−0.000	0.060	0.128	0.041
	Father is HS grad	0.172	0.055	0.130	0.065	0.216	0.044
	Mother more than HS but no college	−0.038	0.068	0.155	0.072	0.236	0.047
	Father more than HS but no college	0.091	0.067	0.070	0.084	0.283	0.051
	Mother College grad	−0.031	0.081	0.057	0.090	0.242	0.044
	Father College grad	0.124	0.061	0.153	0.083	0.337	0.047
	Mother went to school but unknown level	−0.094	0.070	−0.072	0.084	−0.013	0.060
	Father went to school but unknown level	0.150	0.075	0.261	0.087	0.202	0.062
	Parents' labor force participation	0.151	0.073	0.077	0.067	0.012	0.046
Peer Effect	Endogenous Effect	0.467	0.256	0.362	0.154	0.384	0.006
	Number of Observations	55208		65913		65913	

Table 2: Full Sample Results

To highlight the benefits of our method for less dense networks, we estimate the peer effects for each school. Figure 16 shows a histogram of the individual school network densities. Most of the schools are around the 1% density level, but all of them are significantly more dense than the combined network by a factor of nearly 1,000. For each school we estimate peer effects using both 2SLS and RSEE.[5]

---

[5]Some of the schools are completely homogeneous in one of the included covariates (Race, Born in US, etc.) so for those schools that covariate is dropped.
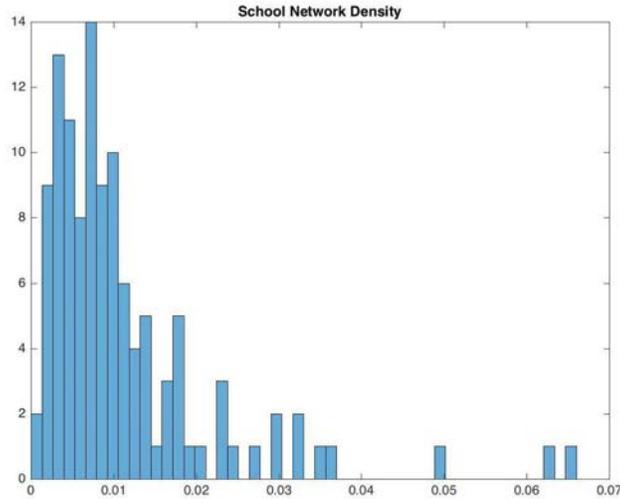
Figure 16: Histogram of School Network Density

Figure 17 plots the kernel density of the peer effect coefficients. The RSEE estimates are nicely distributed around 0.3, with an interquartile range of 0.19, while the 2SLS estimates are quite disperse, with an interquartile range of 0.96. Figure 18 plots the kernel density of the peer effect standard errors. The RSEE estimates are much more precise, with a median standard error of 0.087, while the corresponding median 2SLS standard error is 0.60. This means that for any reasonable estimate of the peer effect (between -1 and 1), 2SLS is unlikely to be able to reject the null hypothesis of no peer effects. In other words, 2SLS will be uninformative about peer effects in individual schools.
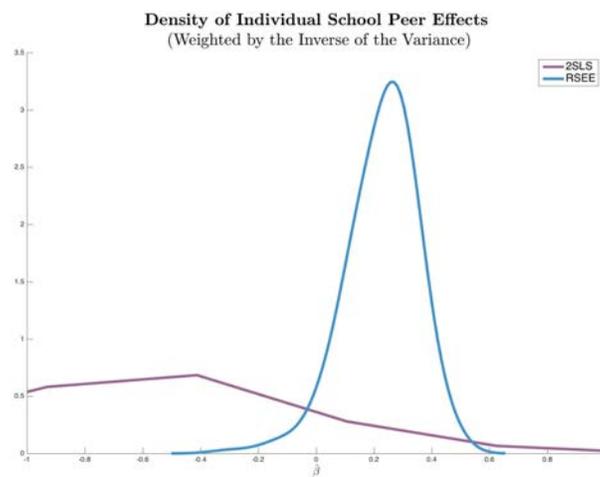

Figure 17: Kernel Density of Peer Effect Coefficients
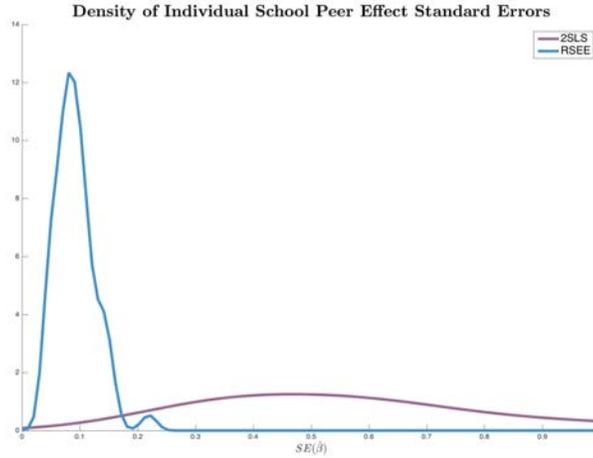
Figure 18: Kernel Density of Peer Effect Standard Errors

Our method will allow for the estimation of individual school peer effects. This is important when the unit of interest is an individual school and will allow for testing for heterogeneity between schools. A Wald test of joint equality between the individual peer effects and the full network peer effect (0.384) is strongly rejected ($\chi^2 = 793$), which suggests there is substantial heterogeneity in peer effects between the schools.

7   Conclusion

In this paper we replicate the findings of Bramoullé *et al.* (2009) and highlight the weak instrument problem that results in limited power for denser networks. We propose an alternative estimation technique that does not rely on a weak instrument and show it performs much better than 2SLS for dense networks.

References

Anderson, T. W., & Rubin, Herman. 1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, **20**(1), 46–63.

Arduini, Tiziano, Patacchini, Eleonora, & Rainone, Edoardo. 2016. Identification and Estimation of Network Models with Heterogeneous Externalities.

Belsley, David A., Kuh, Edwin, & Welsch, Roy E. 1980. Detecting and Assessing Collinearity. *Pages 85–191 of: Regression Diagnostics*. John Wiley & Sons, Inc.

Blume, Lawrence E., Brock, William A., Durlauf, Steven N., & Ioannides, Yannis M. 2011. Chapter 18 - Identification of Social Interactions. *Pages 853–964 of:* Jess Benhabib, Alberto Bisin and Matthew O. Jackson (ed), *Handbook of Social Economics*, vol. 1. North-Holland.

Bramoullé, Yann, Djebbari, Habiba, & Fortin, Bernard. 2009. Identification of peer effects through social networks. *Journal of Econometrics*, **150**(1), 41–55.

Calvo-Armengol, Antoni, Patacchini, Eleonora, & Zenou, Yves. 2009. Peer Effects and Social Networks in Education. *Review of Economic Studies*, **76**(4), 1239–1267.

De Giorgi, Giacomo, Pellizzari, Michele, & Redaelli, Silvia. 2010. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 241–275.

Fortin, B., & Boucher, V. 2015. Some challenges in the empirics of the effects of networks. *Oxford Handbook on the Economics of Networks, Oxford: Oxford University Press, forthcoming*.

Fortin, Bernard, & Yazbeck, Myra. 2015. Peer effects, fast food consumption and adolescent weight gain. *Journal of Health Economics*, **42**(July), 125–138.

Gibbons, Stephen, & Overman, Henry G. 2012. Mostly Pointless Spatial Econometrics?*. *Journal of Regional Science*, **52**(2), 172–191.

Halleck Vega, Solmaria, & Elhorst, J. Paul. 2015. The Slx Model. *Journal of Regional Science*, **55**(3), 339–363.

Halliday, Timothy J., & Kwak, Sally. 2012. What is a peer? The role of network definitions in estimation of endogenous peer effects. *Applied Economics*, **44**(3), 289–302.

Hsieh, Chih-Sheng, Kippersluis, Van, & Hans. 2015. Smoking Initiation: Peers and Personality. Sept.

Lee, Lung-fei. 2003. Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews*, **22**(4), 307–335.

Lee, Lung-fei, & Liu, Xiaodong. 2010. Efficient Gmm Estimation Of High Order Spatial Autoregressive Models With Autoregressive Disturbances. *Econometric Theory*, **26**(01), 187–230.

Lee, Lung-fei, Liu, Xiaodong, & Lin, Xu. 2010. Specification and estimation of social interaction models with network structures: Estimation of social network models. *Econometrics Journal*, **13**(2), 145–176.

Lin, Xu. 2010. Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, **28**(4), 825–860.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, **60**(3), 531–542.

Nelson, Charles R., & Startz, Richard. 1990a. The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One. *The Journal of Business*, **63**(1), S125–S140.

Nelson, Charles R., & Startz, Richard. 1990b. Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, **58**(4), 967–976.

Patacchini, Eleonora, Rainone, Edoardo, & Zenou, Yves. 2016 (Jan.). *Heterogeneous Peer Effects in Education.* SSRN Scholarly Paper ID 2757612. Social Science Research Network, Rochester, NY.

Staiger, Douglas, & Stock, James H. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**(3), 557–586.

Yeung, Ryan, & Nguyen-Hoang, Phuong. 2016. Endogenous peer effects: Fact or fiction? *The Journal of Educational Research*, **109**(1), 37–49.