Online Appendix

Oscar Volpe

This document contains supplementary results and discussions that relate to the paper "Discrete Choice with Generalized Social Interactions". It does not include proofs of main results, which are in the paper. All notation is consistent with the main text of the paper.

Appendix A: Justification for the Remarks in the Footnotes

Remark in Footnote 10

Suppose $\mathbf{J}$ is a symmetric matrix and that its eigenvalues all have non-positive real parts. Since $\boldsymbol{\beta}(m^*)$ is a diagonal matrix with positive elements, $\mathbf{J}$ is congruent to $\boldsymbol{\beta}^{1/2}(m^*)\mathbf{J}\boldsymbol{\beta}^{1/2}(m^*)$, which is similar to $\boldsymbol{\beta}^{1/2}(m^*)[\boldsymbol{\beta}^{1/2}(m^*)\mathbf{J}\boldsymbol{\beta}^{1/2}(m^*)]\boldsymbol{\beta}^{-1/2}(m^*) = \boldsymbol{\beta}(m^*)\mathbf{J} = \mathbf{D}(m^*)$. By Sylvester's law of inertia, it follows that $\mathbf{D}(m^*)$ also has eigenvalues with non-positive real parts for any equilibrium m^*. Given this property, $0 < \prod_{k=1}^{K}\big(1-\lambda_k(m^*)\big) = \det\big(I-\mathbf{D}(m^*)\big) = \det\big(\mathbf{D}_{\mathcal{H}}(m^*)\big)$ at every equilibrium m^*. By the index theorem, the model always has a unique equilibrium.

Remark in Footnote 12

Consider the equilibrium equation $m^* = F_\varepsilon(Jm^*)$, where $J < 0$. By the previous remark, there is always a unique equilibrium. Moreover, since $h = 0$ and F_ε is symmetric about zero, the equilibrium equals $m^* = 0$. It is unstable if $\frac{\partial F_\varepsilon(Jm)}{\partial m}\big|_{m=0} = J \times f_\varepsilon(0) < -1 \Leftrightarrow J < -f_\varepsilon^{-1}(0)$.

Remark in Footnote 13

Claim. Suppose that $\operatorname{sgn}(J_{k\ell}) = \operatorname{sgn}(J_{km}J_{m\ell})$ for all $k, \ell, m \in \mathcal{K}$. Then Assumption A.1 holds.

Proof. To begin, suppose that $\operatorname{sgn}(J_{k\ell}) = \operatorname{sgn}(J_{km}J_{m\ell})$ for all $k, \ell, m \in \mathcal{K}$. I will now prove through induction that $\operatorname{sgn}(J_{j_0j_1}J_{j_1j_2}\cdots J_{j_Mj_0}) \geq 0$ for any arbitrary indices $j_0, j_1, \ldots, j_M \in \mathcal{K}$. First, note that $\operatorname{sgn}(J_{j_0j_2}) = \operatorname{sgn}(J_{j_0j_1}J_{j_1j_2})$. Next, assume $\operatorname{sgn}(J_{j_0j_m}) = \operatorname{sgn}(J_{j_0j_1}J_{j_1j_2}\cdots J_{j_{m-1}j_m})$ for some $m \in \mathbb{N}$, and write: $\operatorname{sgn}(J_{j_0j_{m+1}}) = \operatorname{sgn}(J_{j_0j_m}J_{j_mj_{m+1}}) = \operatorname{sgn}(J_{j_0j_1}J_{j_1j_2}\cdots J_{j_{m-1}j_m}J_{j_mj_{m+1}})$. By induction, it follows that $\operatorname{sgn}(J_{j_0j_M}) = \operatorname{sgn}(J_{j_0j_1}J_{j_1j_2}\cdots J_{j_M\ell})$ for any index $\ell \in \mathcal{K}$. By setting $\ell = j_0$, I obtain $\operatorname{sgn}(J_{j_0j_1}J_{j_1j_2}\cdots J_{j_Mj_0}) = \operatorname{sgn}(J_{j_0j_0})$, where $\operatorname{sgn}(J_{j_0j_0}) = \operatorname{sgn}(J_{j_0j_0}J_{j_0j_0}) \geq 0$. $\square$

Claim. Define $\mathbf{J}_{k\ell}$ to be $\mathrm{E}(J_{ij}|i \in k, j \in \ell)$, where $J_{ij} \in \{-1, 1\}$. Then $\operatorname{sgn}(\mathbf{J}_{k\ell}) = \operatorname{sgn}(\mathbf{J}_{km}\mathbf{J}_{m\ell})$ for all $k, \ell, m \in \mathcal{K}$ whenever $\mathrm{P}(J_{i_0i_1}J_{i_1i_2} = J_{i_0i_2}|i_0 \in k, i_1 \in m, i_2 \in \ell) \geq 0.5$ for all $k, \ell, m \in \mathcal{K}$.

Proof. For any indices $k, \ell, m \in \mathcal{K}$, the product $\mathbf{J}_{km}\mathbf{J}_{m\ell}$ equals:

$$\begin{aligned}
\mathbf{J}_{km}\mathbf{J}_{m\ell} &= \mathrm{E}(J_{ij}|i \in k, j \in m) \times \mathrm{E}(J_{i'j'}|i' \in m, j' \in \ell) \\
&= \mathrm{E}\big(J_{ij} \times \mathrm{E}(J_{i'j'}|i' \in m, j' \in \ell)|i \in k, j \in m\big) \\
&= \mathrm{E}(J_{ij}J_{i'j'}|i \in k; j, i' \in m; j' \in \ell) \\
&= 2\mathrm{P}(J_{ij}J_{i'j'} = 1|i \in k; j, i' \in m; j' \in \ell) - 1
\end{aligned}$$

Suppose that $\mathrm{P}(J_{i_0 i_1} J_{i_1 i_2} = J_{i_0 i_2} | i_0 \in k, i_1 \in m, i_2 \in \ell) \geq 0.5$ for all $k, \ell, m \in \mathcal{K}$. To simplify notation, I define: $\gamma = \mathrm{P}(J_{ij} J_{ji'} = J_{ii'} | i \in k; j, i' \in m)$, $\delta = \mathrm{P}(J_{ii'} J_{i'j'} = J_{ij'} | i \in k; i' \in m; j' \in \ell)$, and $\kappa = \mathrm{P}(J_{ji'} = 1 | j, i' \in m)$. Next, I decompose $\mathrm{P}(J_{ij} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell)$ so that:

$$\begin{aligned}
\mathrm{P}(J_{ij} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell) &= \kappa \times \mathrm{P}(J_{ij} J_{ji'} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell, J_{ji'} = 1) \\
&\quad + (1-\kappa) \times \mathrm{P}(J_{ij} J_{ji'} J_{i'j'} = -1 | i \in k; j, i' \in m; j' \in \ell, J_{ji'} = -1) \\
&= \kappa \times \Big[\gamma \times \mathrm{P}(J_{ii'} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell, J_{ij} J_{ji'} = J_{ii'}) \\
&\quad + (1-\gamma) \times \mathrm{P}(J_{ii'} J_{i'j'} \neq 1 | i \in k; j, i' \in m; j' \in \ell, J_{ij} J_{ji'} \neq J_{ii'}) \Big] \\
&\quad + (1-\kappa) \times \Big[\gamma \times \mathrm{P}(J_{ii'} J_{i'j'} \neq 1 | i \in k; j, i' \in m; j' \in \ell, J_{ij} J_{ji'} = J_{ii'}) \\
&\quad + (1-\gamma) \times \mathrm{P}(J_{ii'} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell, J_{ij} J_{ji'} \neq J_{ii'}) \Big] \\
&= \kappa \times \Big[\gamma \times \Big[\delta \times \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) + (1-\delta) \times \mathrm{P}(J_{ij'} \neq 1 | i \in k; j' \in \ell) \Big] \\
&\quad + (1-\gamma) \times \Big[\delta \times \mathrm{P}(J_{ij'} \neq 1 | i \in k; j' \in \ell) + (1-\delta) \times \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) \Big] \Big] \\
&\quad + (1-\kappa) \times \Big[\gamma \times \Big[\delta \times \mathrm{P}(J_{ij'} \neq 1 | i \in k; j' \in \ell) + (1-\delta) \times \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) \Big] \\
&\quad + (1-\gamma) \times \Big[\delta \times \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) + (1-\delta) \times \mathrm{P}(J_{ij'} \neq 1 | i \in k; j' \in \ell) \Big] \Big] \\
&= (1-\kappa) + (1-\gamma)(2\kappa - 1) + (1-\delta)(2\gamma - 1)(2\kappa - 1) \\
&\quad + (2\kappa - 1)(2\gamma - 1)(2\delta - 1) \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) \\
&= \frac{1}{2} \big[1 - (2\kappa - 1)(2\gamma - 1)(2\delta - 1) \big] + (2\kappa - 1)(2\gamma - 1)(2\delta - 1) \mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell)
\end{aligned}$$

Note that $\kappa, \gamma, \delta \in [0.5, 1]$ by assumption. So $(2\kappa - 1)(2\gamma - 1)(2\delta - 1)$ is bounded somewhere between 0 and 1. It follows from the last equality above that $\mathrm{P}(J_{ij'} = 1 | i \in k; j' \in \ell) \geq 0.5$ if and only if $\mathrm{P}(J_{ij} J_{i'j'} = 1 | i \in k; j, i' \in m; j' \in \ell) \geq 0.5$. This statement further implies that $\mathbf{J}_{k\ell}$ has the same sign as $\mathbf{J}_{km} \mathbf{J}_{m\ell}$. Because this result applies for any $k, m, \ell \in \mathcal{K}$, the claim is true. $\square$

Remark in Paragraph 4 of Page 13

I now justify that "both $\underline{m}^*$ and $\overline{m}^*$ are always locally stable". Consider the mapping $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1} m)$, which is defined in the proof of Property 6. Tarski's fixed point theorem implies that $\hat{\mathcal{Q}}$ has a least fixed point $\mathbf{B}\underline{m}^*$ and a greatest fixed point $\mathbf{B}\overline{m}^*$. Suppose, for sake of contradiction, that $\overline{m}^*$ is unstable. Then, as argued in the proof of Property 4, $\hat{\mathcal{Q}}$ must have another fixed point, which is strictly greater than $\mathbf{B}\overline{m}^*$. Arriving at a contradiction in this case, I conclude that $\overline{m}^*$ is a locally stable. By the same reasoning, $\underline{m}^*$ is also locally stable.

Appendix B: Adapting the Identification Results to Allow for Covariates

Suppose that the choice equation is modified to allow for exogenous covariates $X_i \in \mathbb{R}^r$. I define $\omega_i = \mathbb{1}\{X_i' c + h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell} m_n^{\ell *} + \varepsilon_i \geq 0\}$, such that $\mathrm{P}(\varepsilon_i \leq z | X_i, k, \alpha_n) = F_{\varepsilon|k}(z)$ and $\mathrm{P}(X_i \leq x | k, \alpha_n) = F_{X|k}(x)$. In equilibrium, $m_n^{k*} = \int \mathrm{E}(\omega_i | X_i, k, \alpha_n, \{m_n^{\ell *}\}_{\ell=1}^K) dF_{X|k}$, where:

$$\mathrm{E}(\omega_i | X_i, k, \alpha_n, \{m_n^{\ell *}\}_{\ell=1}^K) = F_{\varepsilon|k}\bigg(h_k + \alpha_n + X_i' c + \sum_{\ell=1}^K J_{k\ell} m_n^{\ell *} \bigg), \quad \text{for } k = 1, \ldots, K.$$

Proof of Lemma 2 (Version with Covariates)

In the presence of covariates, Lemma 2 must be adapted. I do so in the following way.

Lemma 2. *(Sufficiency Property.)* For any agent i in group k_1 and any agent j in group k_2 residing in network n: $\mathrm{E}(\omega_i|X_i, k_1, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = \mathrm{E}(\omega_i|X_i, X_j, k_1, \{m_n^{\ell*}\}_{\ell=1}^K, \mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K))$.

Proof. Since $F_{\varepsilon|k_2}$ is strictly increasing, its inverse $F_{\varepsilon|k_2}^{-1}$ exists. By this property, I can write:

$$\alpha_n = F_{\varepsilon|k_2}^{-1}\big(\mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K)\big) - h_{k_2} - X_j'c - \sum_{\ell=1}^{K} J_{k_2\ell} m_n^{\ell*}$$

By plugging this expression for α_n into the definition of $\mathrm{E}(\omega_i|X_i, k_1, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K)$, I find that:

$$\begin{aligned}\mathrm{E}(\omega_i|X_i, k_1, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = F_{\varepsilon|k_1}\Bigg(& h_{k_1} - h_{k_2} + (X_i - X_j)'c \\ & + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^{\ell*} + F_{\varepsilon|k_2}^{-1}\big(\mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K)\big)\Bigg)\end{aligned}$$

□

Proof of Theorem 1 (Version with Covariates)

For semiparametric identification (Theorem 1), I must make an additional assumption. In particular, I require that there exists some $k \in \mathcal{K}$ such that $\mathrm{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$. Note that this condition is far weaker than Assumption B.3. Nevertheless, using Proposition 5 in Manski (1988), it is enough to show that c, $\{h_{k_1} - h_{k_2}\}_{k_1,k_2}$ and $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ are identified. The proof of Theorem 1 is adapted in the following way.

Theorem 1. Suppose that Assumptions B.1 & B.2 hold, and assume that m_n^{k*} is observed for all networks n and social groups k. Also, suppose that there is some $k \in \mathcal{K}$ where $\mathrm{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$. Whenever $\{F_{\varepsilon|k}\}_{k=1}^K$ are known:

(i) Without further assumptions, $\{h_{k_1} - h_{k_2}\}_{k_1,k_2}$, $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$, and c are identified.

(ii) If $\alpha_n = W_n'd$ for some observed vector W_n, then d, $\{h_k\}_k$, $\{J_{k\ell}\}_{k,\ell}$, and c are identified.

Proof. Fixing some network n and defining the term $\zeta_n^k = h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell} m_n^{\ell*}$, I write:

$$\mathrm{E}(\omega_i|X_i, k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = F_{\varepsilon|k}(\zeta_n^k + X_i'c)$$

To recover c, I must show that $F_{\varepsilon|k}(\zeta_n^k + X_i'c) = F_{\varepsilon|k}(\hat{\zeta}_n^k + X_i'\hat{c})$ implies $c = \hat{c}$. Since $F_{\varepsilon|k}$ is known, this property holds by Proposition 5 of Manski (1988). It follows that c is identified.

Having demonstrated that c can be recovered, I now focus on the other parameters. Consider any two social groups k_1 and k_2, and then define the function $\phi : \mathbb{R} \to [0,1]$ so that:

$$\phi(\nu) = \int F_{\varepsilon|k_1}\big((X_i - X_j)'c + \nu\big) dF_{X|k_1},$$

where X_j is chosen from $\mathrm{supp}(X|k_2)$ and where the integral is evaluated over the conditional support of X_i given k_1. By construction, $\phi(\cdot)$ is nonlinear and monotonically increasing in ν.

For any network n, let $\nu_n = h_{k_1} - h_{k_2} + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^{\ell*} + F_{\varepsilon|k_2}^{-1}(\mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^{K}))$. By Lemma 2, the expected average choice $m_n^{k_1*}$ is equal to $\phi(\nu_n)$. Because $\phi(\cdot)$ is monotonic:

$$\begin{aligned} m_n^{k_1*} &= \int F_{\varepsilon|k_1}\left(h_{k_1} - h_{k_2} + (X_i - X_j)'c + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^{\ell*} + F_{\varepsilon|k_2}^{-1}(\mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^{K}))\right)dF_{X|k_1} \\ &= \int F_{\varepsilon|k_1}\left(\widehat{h_{k_1} - h_{k_2}} + (X_i - X_j)'c + \sum_{\ell=1}^{K}(\widehat{J_{k_1\ell} - J_{k_2\ell}})m_n^{\ell*} + F_{\varepsilon|k_2}^{-1}(\mathrm{E}(\omega_j|X_j, k_2, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^{K}))\right)dF_{X|k_1} \end{aligned}$$

is satisfied if and only if $\sum_{\ell=1}^{K}\left[(\widehat{J_{k_1\ell} - J_{k_2\ell}}) - (J_{k_1\ell} - J_{k_2\ell})\right]m_n^{\ell} = (h_{k_1} - h_{k_2}) - (\widehat{h_{k_1} - h_{k_2}})$ for all networks $n \in \{1, \ldots, N\}$. Since the expected average choices are nonlinear functions of one another, sufficient variation in these equilibria across different networks will ensure that:

$$h_{k_1} - h_{k_2} = \widehat{h_{k_1} - h_{k_2}} \quad \text{and} \quad J_{k_1\ell} - J_{k_2\ell} = \widehat{J_{k_1\ell} - J_{k_2\ell}},$$

for every $\ell \in \mathcal{K}$. Also, since k_1 and k_2 are chosen arbitrarily, this result holds for all $k_1, k_2 \in \mathcal{K}$. $\square$

For nonparametric identification (Theorem 2), there is nothing to modify, since these results already rely on individual-level covariates. Also, for identification in contexts with unknown expected average choices (Theorem 3), all modifications will follow directly from Theorems 1 and 2. Specifically, the IV estimands may be adapted to incorporate covariates.

Appendix C: Additional Details about the Monte Carlo Simulations

To perform simulations, I draw observations from the following data generating process:

$$\omega_i = \mathbb{1}\left\{h_k + \alpha_n + J_{k1}m_n^{1*} + J_{k2}m_n^{2*} + \varepsilon_i \geq 0\right\},$$

where $m_n^{k*} = F_{\varepsilon|k}(h_k + \alpha_n + J_{k1}m_n^{1*} + J_{k2}m_n^{2*})$ for $k \in \{1, 2\}$ and $n = 1, \ldots, N$. The idiosyncratic payoffs follow logistic distributions, i.e., $\varepsilon_i|k \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ for $k \in \{1, 2\}$. The contextual effect α_n is a continuous random variable that is evenly distributed over the interval $[-2, 2]$. The identity-specific effects are: $h_1 = 0$ and $h_2 = 1$. Finally, the interaction matrix is equal to:

$$\mathbf{J} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

For case *(ii)*, i.e., where $\alpha_n = W_n'd$ for some observed W_n, I set $W_n = \alpha_n$ and $d = 2$. Note that these parameters are chosen arbitrarily and that alternative DGP's produce similar results.

To draw agent's choices ω_i in equilibrium, I run a fixed point iteration on the equilibrium condition. This procedure leverages the fact that **J** satisfies Assumption A, which guarantees there exists at least one locally stable equilibrium (Property 4). Note that it is not necessary that Assumption A holds for this estimation procedure to be valid. However, this condition is helpful for conducting simulations as it facilitates the computation of an equilibrium. For all additional details about specifications for the simulations, I refer to the replication code.

Appendix D: Robustness Analyses

1. Verifying Random Assignment to Classrooms

Under the experimental protocols, students in each school were randomly assigned into classrooms of three different types. However, of the 79 schools in the study, 48 schools had more than three kindergarten classrooms. Since these schools had more than one classroom of each type, it is conceivable that there may have been nonrandom sorting within the same class type. As Graham (2008) argues, this scenario is unlikely. Indeed, he finds no evidence of within-class-type sorting. Nevertheless, I present a version of the IV estimates where I restrict the sample to the 31 schools that have only three classrooms. In doing so, I rule out any possibility of nonrandom assignment to classrooms of the same type. These estimates are presented below, and they appear to be consistent with the results that use the full sample.

Table A.1: IV Estimates for Schools with Fewer than 4 Classrooms

	Outcome Variable:					
	Math			Reading		
	Top 25%	Top 50%	Top 75%	Top 25%	Top 50%	Top 75%
$J_{ff} - J_{mf}$	3.494	4.215***	3.961**	3.843	5.265**	6.425
	(2.826)	(1.039)	(1.444)	(3.951)	(1.952)	(4.974)
$J_{mm} - J_{fm}$	3.251	4.088***	4.104***	3.048	5.117***	6.522
	(3.133)	(1.251)	(1.111)	(5.612)	(1.552)	(5.430)
Intercept	−0.280	−0.097	0.081	−0.008	−0.045	1.088
	(0.443)	(0.201)	(0.273)	(1.282)	(0.218)	(2.872)
Number of Classrooms	89	89	89	89	89	89
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
$F_{(df1,df2)}$ 1st-Stage ($\bar{\omega}_n^m$)	$2.23_{(2,10)}$	$3.06_{(2,19)}$	$2.78_{(2,17)}$	$2.58_{(2,12)}$	$2.82_{(2,17)}$	$8.35_{(2,13)}$
$F_{(df1,df2)}$ 1st-Stage ($\bar{\omega}_n^f$)	$2.01_{(2,10)}$	$2.26_{(2,19)}$	$2.18_{(2,17)}$	$5.71_{(2,12)}$	$2.02_{(2,17)}$	$3.61_{(2,13)}$

Notes. Estimates are obtained by computing $\hat{\beta}^{IV}_{f,m}$, which corresponds to the estimand in equation (24). For implementation, I randomly split each classroom so that half the sample is used to form endogenous variables X_n, and the remaining half is used to form instruments Z_n. *p<0.1; **p<0.05; ***p<0.01.

2. Misspecification Tests

I conduct a variety of hypothesis tests to determine whether any of the gender-specific parameters in the model depend on the observed classroom-level variables. Specifically, I test the null hypotheses that $J_{ff} - J_{mf}$, $J_{mm} - J_{fm}$, and the intercept (respectively) differ for:

1. high poverty classrooms ($\geq$ 50% FRPL) and low poverty classrooms ($<$ 50% FRPL)
2. high minority classrooms ($<$ 75% white) and low minority classrooms ($\geq$ 75% white)
3. more experienced teachers ($>$ 10 years) and less experienced teachers ($\leq$ 10 years)
4. more educated teachers (graduate deg.) and less educated teachers (no graduate deg.)
5. rural classrooms (in rural district) and urban classrooms (in urban or suburban district)

The p-values from each of these one-dimensional hypothesis tests are reported in Tables A.2 and A.3. When conducting these tests, I consider two different outcomes variables: (1) scoring in the top 50% on the math exam and (2) scoring in the top 50% on the reading exam. Recall that these percentiles are all calculated relative to the full sample of Tennessee kindergarten students who participated in the study. For both outcome variables, I find no evidence to reject the hypothesis that any of the gender-specific parameters differ across networks. These results help to justify the model specification and identification strategy.

Table A.2: Tests for Misspecification (*Outcome:* Top 50% on Math Exam)

	Large Share Poverty	Large Share Minority	Teacher Has >10yrs Experience	Teacher Has Higher Degree	Rural District
$J_{ff} - J_{mf}$	0.948	0.35	0.525	0.754	0.927
$J_{mm} - J_{fm}$	0.634	0.77	0.611	0.893	0.945
Intercept	0.568	0.198	0.767	0.606	0.92

Notes. This table reports p-values corresponding to the one-dimensional hypothesis tests for whether $J_{ff} - J_{mf}$, $J_{mm} - J_{fm}$, and the intercept (respectively) differ by observed classroom features.

Table A.3: Tests for Misspecification (*Outcome:* Top 50% on Reading Exam)

	Large Share Poverty	Large Share Minority	Teacher Has >10yrs Experience	Teacher Has Higher Degree	Rural District
$J_{ff} - J_{mf}$	0.762	0.984	0.911	0.996	0.662
$J_{mm} - J_{fm}$	0.626	0.979	0.967	0.875	0.886
Intercept	0.532	0.977	0.898	0.905	0.321

Notes. This table reports p-values corresponding to the one-dimensional hypothesis tests for whether $J_{ff} - J_{mf}$, $J_{mm} - J_{fm}$, and the intercept (respectively) differ by observed classroom features.

3. Sensitivity of IV Estimates to the Partitioning of Classrooms

To estimate the model, internal instruments are defined by randomly partitioning each classroom into two equal (or almost equal if there is an odd number of students) subsamples. In general, the estimates will be sensitive to the way in which the classrooms are partitioned. Nevertheless, as long as the model is correctly specified, this IV strategy always generates consistent estimates regardless of which partitions are realized. To justify the efficacy of this approach, I re-estimate the model $M = 1000$ times, each time randomly choosing a different partition of classrooms when constructing the instruments. Through this procedure, I can evaluate how sensitive the IV estimates are to the way that internal instruments are defined.

I report histograms of the parameter estimates for each outcome variable in Figure A.1. Observe that the point estimates appear approximately normally distributed, and the amount of dispersion is not large enough to invalidate any of the qualitative findings in the paper. Moreover, the main IV estimates reported in Table 2 do not seem to be outliers, which indicates that most alternative partitions of classrooms would give similar results. So, I interpret Figure A.1 as further evidence that the main results in the empirical application are robust.

Figure A.1: Densities of IV Estimates over Random Classroom Partitions

(a) Parameter: Intercept — (b) Parameter: $J_{mm} - J_{fm}$ — (c) Parameter: $J_{ff} - J_{mf}$