

Estimating Spillovers Using Panel Data

Arcidiacono, Foster, Goodpaster, and Kinsler (2012, QE)

Miguel Salema

28/10/2025

Econometrics Reading Group
University of Bern

Problems Estimating Peer Effects



Identification challenges:

- The reflection problem (Manski, 1993, REStud):
regressing y_i on $\bar{y}_{\sim i}$ $\Rightarrow$ simultaneity bias;
- Selection: high-ability students sort with high-ability peers;
- Spillovers can emanate from unobservables.

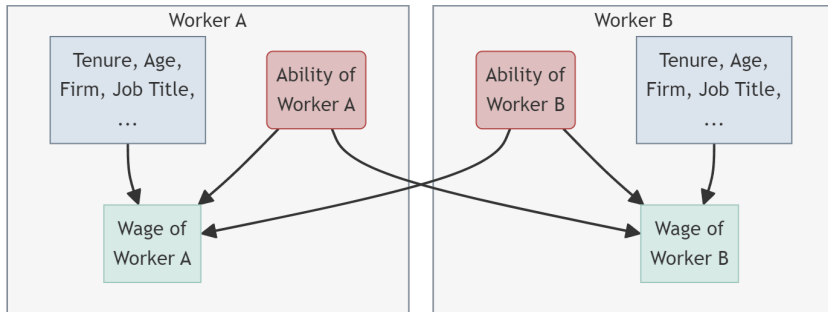
Conventional fixes (random assignment, IV, proxies) struggle when peer groups are chosen and ability is only partly observable.

Consistently Estimating Spillovers

Key contribution:

- Capture peer effects through a linear combination of individual fixed effects of peers and estimate with panel data.

A simple example: 1 team, 2 workers



In This Paper

- I The nonlinear least squares estimate of the spillover parameter is consistent and asymptotically normal.
- II Easy to implement iterative estimation that converges to the nonlinear least squares solution.
- III Application to the classroom using University of Maryland transcript data.

Model Setup I

Baseline specification:

$$y_{itn} = \mathbf{x}'_{it}\beta_1 + u_{it}\beta_2 + \frac{1}{M_{tn}} \sum_{j \in \mathbb{M}_{tn \sim i}} (\mathbf{x}'_{jt}\gamma_1 + u_{jt}\gamma_2) + e_{itn}, \quad (1)$$

where y_{itn} is the outcome for individual i at time t in peer group n , $\mathbf{x}_{it}$ and u_{it} are observed and unobserved characteristics, respectively, $M_{tn} + 1$ is the total number of individuals in peer group n at time t , $\mathbb{M}_{tn \sim i}$ is the set of peers (excludes i) and e_{itn} is the error term.

Model Setup II

Proportionality assumption: the relevance to outcomes of peer characteristics is the same to that of own characteristics.

- $\gamma_1 = \gamma\beta_1$;
- $\gamma_2 = \gamma\beta_2$.

Define $\alpha_{it} := \mathbf{x}'_{it}\beta_1 + u_{it}\beta_2$ and $\bar{\alpha}_{\sim i,tn} := \frac{1}{M_{tn}} \sum_{j \in \mathbb{M}_{tn \sim i}} \alpha_{jt}$. Rewrite equation 1:

$$y_{itn} = \alpha_{it} + \gamma \bar{\alpha}_{\sim i,tn} + e_{itn}.$$

Model Setup III

If individual ability varies over time, in general, we need several observations per individual in each time period (usually not available).

Therefore, we assume ability is either:

- Static: $\alpha_{it} = \alpha_i$;
- Cumulative: $\alpha_{it} = \alpha_i + \sum_{t' < t} \gamma \bar{\alpha}_{\sim i, t' n}$.

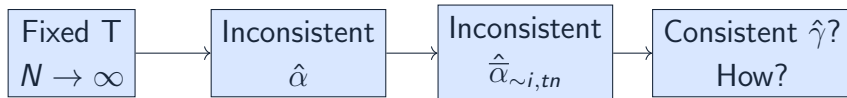
Both lead to the same number of parameters to estimate. I'll focus on static, but the algorithm accommodates both.

Nonlinear Least Squares Objective

Unknowns are $(\{\alpha_i\}_{i=1}^N, \gamma)$:

$$(\hat{\alpha}, \hat{\gamma}) = \arg \min_{\alpha, \gamma} \sum_{i=1}^N \sum_{t=1}^T \left(y_{itn} - \alpha_i - \gamma \bar{\alpha}_{\sim i, tn} \right)^2.$$

Theorem 1: T is fixed, $N \rightarrow \infty$. Fixed effects $\hat{\alpha}$ not consistently estimated (incidental parameters), but NLS for $\hat{\gamma}$ is consistent and asymptotically normal under regularity conditions.



We expected downward bias: true in two-step procedures (Mas and Moretti, 2009). However, in this method, solving $\hat{\alpha}(\gamma)$ removes measurement error.

Theorem 1 — Consistency and Asymptotic Normality

Suppose either A1 (Static) or A2 (Cumulative). Let $\mathcal{N}$ denote the number of individuals observed at least twice and assume a change in peer composition:

$\sum_{j \in \mathbb{M}_{tn \sim i}} \frac{\alpha_{jt}}{\bar{M}_{tn}} \neq \sum_{j \in \mathbb{M}_{t'n \sim i}} \frac{\alpha_{jt}}{\bar{M}_{t'n}}$ for some $t \neq t'$. Additionally, assume:

1. $\mathbb{E}(e_{itn}e_{jsk}) = 0$ for all $j \neq i, t \neq s, n \neq k$.
2. $\mathbb{E}(e_{itn}\alpha_{jt}) = 0$ for all i, j, t, n .
3. $\mathbb{E}(\alpha_{it}^4) < \infty$ for all i, t, n .
4. $\mathbb{E}(e_{itn}) = 0$ and $E(e_{itn}^4) < \infty$ for all i, t, n .
5. Either $\mathbb{E}(e_{itn}^2 | n, t) = E(e_{itn}^2 | n, t)$ for all i, j, t, n or $\text{Cov}(e_{itn}^2, N_i) = 0$, where N_i is the number of observations for individual i ; under A2 $\mathbb{E}(e_{itn}^2 | t)$ is time invariant.
6. $\gamma \in \text{int}(\Gamma)$ with $\sup \Gamma = \bar{\gamma} < M$, where M is the smallest class size.

Conclusion. If (1)–(6) hold, then $\hat{\gamma}$ is $\sqrt{\mathcal{N}}$ -consistent and asymptotically normal for fixed T . $\sqrt{\mathcal{N}}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1})$

Alternative Notation

An alternative notation that can simplify the FOC interpretation (Cardoso et al., 2022):

$$\mathbf{Y} = \mathbf{D}\alpha + \gamma \mathbf{W}\mathbf{D}\alpha + \mathbf{X}\beta + \mathbf{e}, \quad (2)$$

where $\mathbf{D}_{(M \times N)}$ is the individual design matrix, $\mathbf{W}_{(M \times M)}$ is a symmetric block diagonal weighting matrix, defined as:

$$\mathbf{W} = \text{diag}(\mathcal{W}), \text{ where the set } \mathcal{W} = \{\mathbf{w}_p \mid p \in [1; P]\}.$$

$$\mathbf{w}_p = \frac{\mathbf{1}_{n_p} \mathbf{1}_{n_p}' - I_{n_p}}{n_p - 1}.$$

Alternative Notation: Toy Panel

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{w}_1 = \frac{1}{2} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

If $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, then

$$\mathbf{D}\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_2 \\ \alpha_4 \end{bmatrix}, \quad \mathbf{W}\mathbf{D}\boldsymbol{\alpha} = \begin{bmatrix} (\alpha_2 + \alpha_3)/2 \\ (\alpha_1 + \alpha_3)/2 \\ (\alpha_1 + \alpha_2)/2 \\ \alpha_4 \\ \alpha_2 \end{bmatrix}.$$

General Identification with High Dimensionality

Computationally Infeasible

What if model 2 is high dimensional?

Consider the basic AKM model, after Abowd, Kramarz and Margolis (1999, ECMA):

$$\mathbf{Y} = \mathbf{D}\alpha + \mathbf{F}\phi + \mathbf{X}\beta + \mathbf{e},$$

where ϕ are firm FE. Because minimizing the least square function can be infeasible, we can solve it via an algorithm that for the q th iteration:

- Step 1: Conditional on α_{q-1} , estimate ϕ_q and β_q by OLS;
- Step 2: Conditional on ϕ_q and β_q , estimate α_q by OLS.

The sum of squares decreases in each step $\rightarrow$ converge to the OLS solution, regardless of the initial guess.

Peer Effects and High Dimensionality

Complication: outcome depends on functions of the individual fixed effects themselves. Let $\arg \min_{\beta, \alpha, \gamma} S(\cdot) = \mathbf{e}'\mathbf{e}$. The FOCs:

$$\frac{\partial S(\cdot)}{\partial \hat{\beta}} = \mathbf{X}'\mathbf{e} = 0,$$

$$\frac{\partial S(\cdot)}{\partial \hat{\gamma}} = \hat{\alpha}'\mathbf{D}'\mathbf{W}\mathbf{e} = 0,$$

$$\frac{\partial S(\cdot)}{\partial \hat{\alpha}} = [\mathbf{D}' + \hat{\gamma} \mathbf{D}'\mathbf{W}]\mathbf{e} = 0 \Leftrightarrow \mathbf{D}'\tilde{\mathbf{W}}\tilde{\mathbf{W}}\mathbf{D}\hat{\alpha} = \mathbf{D}'\tilde{\mathbf{W}}[\mathbf{Y} - \mathbf{X}\hat{\beta}],$$

where $\tilde{\mathbf{W}} = \mathbf{I} + \hat{\gamma}\mathbf{W}$.

FOC for α

- matrix $\mathbf{D}'\tilde{\mathbf{W}}\tilde{\mathbf{W}}\mathbf{D}$ is too big to invert;
- conjugate gradient method obtains explicit solution for $\hat{\alpha}$ given $\hat{\gamma}$ and $\hat{\beta}$.

Algorithm to Identify Peer Effects

Algorithm 1 Estimation of equation 2

Require: $n_j \neq 1, \forall j$

▷ No peer group singletons

Ensure: $\arg \min \mathbf{e}'\mathbf{e} = \{\hat{\gamma}, \hat{\alpha}, \hat{\beta}\}$

Input: $\tilde{\gamma}_0 \leftarrow 0$

1: **Initialize**

2: $\{\tilde{\alpha}_1, \tilde{\beta}_0\} \leftarrow \arg \min(\mathbf{e}'\mathbf{e} \mid \tilde{\gamma}_0)$ with OLS

3: $\{\tilde{\gamma}_1, \tilde{\beta}_1\} \leftarrow \arg \min(\mathbf{e}'\mathbf{e} \mid \tilde{\alpha}_1)$ with OLS

▷ Insert $\mathbf{D}\tilde{\alpha}_1$ as regressor; don't demean

4: **while** $\left| \frac{\tilde{\gamma}_h}{\tilde{\gamma}_{h-1}} - 1 \right| > \text{ToI}_\gamma$ **do**

5: $\tilde{\mathbf{W}}_h \leftarrow \mathbf{I} + \tilde{\gamma}_h \mathbf{W}$

6: $\tilde{\alpha}_{h+1} \leftarrow CG : \mathbf{A} = \mathbf{D}'\tilde{\mathbf{W}}_h\tilde{\mathbf{W}}_h\mathbf{D} \wedge \mathbf{b} = \mathbf{D}'\tilde{\mathbf{W}}_h[\mathbf{Y} - \mathbf{X}\tilde{\beta}_h]$

7: $\{\tilde{\gamma}_{h+1}, \tilde{\beta}_{h+1}\} \leftarrow \arg \min(\mathbf{e}'\mathbf{e} \mid \tilde{\alpha}_{h+1})$ with OLS

8: **end while**

9: **return** $\{\hat{\gamma}, \hat{\alpha}, \hat{\beta}\}$

Theorem 2: Define mapping $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ that stacks the RHS of the α -updates. A sufficient condition for f to be a contraction: $\gamma < 0.4$; algorithm converges to NLS solution.

Why Wild Bootstrap If We Have Asymptotics?

Finite-sample concerns:

1. High-dimensional nuisance FE α, δ are *generated regressors* inside a nonlinear estimator $\Rightarrow$ plug-in SEs can be noisy.
2. Nonlinearity (NLS + iteration) $\Rightarrow$ delta-method approximation may be poor in moderate $\mathcal{N}$.

Wild (cluster) bootstrap is robust to unknown heteroskedasticity and accommodates clustering naturally.

Model Extensions

Correlated effects

Model 2 can contain more FE in $\mathbf{X}\beta$.

- Add class/section FE δ_{ct} to soak up group-level shocks;
- Update δ_{ct} by section-mean residuals each iteration.

Heterogeneous Effects

- The framework extends to outcomes depending on other functions of α ;
- Heterogeneity in:
 1. Response: groups are differently affected;
 2. Sensibility: groups cause different impacts.

Monte Carlo Simulations

TABLE 1. Baseline model, $\gamma_o = 0.15$.^a

Obs. per Student	Peer Group Size		Random Assignment		Selection	
			$\sigma_\varepsilon = 1.95$	$\sigma_\varepsilon = 1.15$	$\sigma_\varepsilon = 1.95$	$\sigma_\varepsilon = 1.15$
2	2	$\hat{\gamma}$	0.151 (0.034)	0.151 (0.016)	0.140 (0.060)	0.151 (0.024)
		R^2	0.706	0.822	0.713	0.828
5	10	$\hat{\gamma}$	0.150 (0.041)	0.150 (0.021)	0.146 (0.059)	0.149 (0.033)
		R^2	0.482	0.686	0.494	0.697
10	10	$\hat{\gamma}$	0.150 (0.025)	0.148 (0.012)	0.148 (0.036)	0.152 (0.018)
		R^2	0.415	0.644	0.429	0.659

^aThe R^2 values reported in this table pertain to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student so as to manipulate the amount of variation in performance that is explained by the ability measure. Parameter values are averages over 100 simulations on a population of 10,000 students.

Application

Data

- University of Maryland transcripts, six semesters;
- Peer group: student–section (discussion group) within large classes.
- Course types: Humanities, Social Sciences, Math/Science.
- Post-restrictions sample: 246,831 student–section obs., 18,511 students.

Empirical Specification

Let Y be the grades. Estimate separately by course type (Humanities / Social Sciences / Math+Science): $Y_{itn} = \alpha_{i,n} + \delta_{c(i,t),n} + \gamma_n \bar{\alpha}_{\sim i,tn} + e_{itn}$.

Interpretation vs. Own Ability: Marginal Effect Ratio

$$\text{MER}_n = \gamma_n \frac{\text{SD}(\bar{\alpha}_{\sim i,tn})}{\text{SD}(\alpha_{i,n})}$$

Main peer-effect estimates (homogeneous γ)

TABLE 5. Peer effects results by course type: homogeneous gamma model.^a

	Humanities	Soc. Sci.	Math/Sci.
Section peer ability	0.1613 (0.016)	0.1960 (0.019)	0.0483 (0.013)
<i>N</i>	86,844	77,312	82,675
<i>R</i> ²	0.6373	0.6321	0.6861

^aThe dependent variable is the grade in the class. Class fixed effects are estimated in all specifications. Standard errors are obtained using a wild bootstrap.

TABLE 6. Standard deviations of estimated ability and marginal effects: homogeneous gamma model.^a

Course Type	Section SD	Population SD	Marginal Effect Ratio
Humanities	0.2823	0.6804	0.0669
Social science	0.3103	0.7125	0.0853
Math and science	0.5952	0.9498	0.0302

^aSection SD is the standard deviation of average peer ability across the sample of student–section observations of the given course type. Population SD is the standard deviation of ability across the sample of student–section observations in courses of the given course type. These calculations are both based on the fixed effects estimated by the homogeneous gamma model. Marginal Effect Ratio shows the ratio of the effect on grades from a 1 standard deviation increase in peer ability to the effect on grades from a 1 standard deviation increase in own ability.

Heterogeneous γ (by gender/race/SAT)

TABLE 7. Peer effects results by course type: heterogeneous gamma model.^a

Section Peer Ability	Humanities	Soc. Sci.	Math/Sci.
Overall	0.2058 (0.020)	0.2227 (0.025)	0.0940 (0.018)
Female	0.0970 (0.019)	0.0584 (0.021)	-0.0517 (0.018)
Asian	-0.0098 (0.035)	-0.0347 (0.030)	-0.0346 (0.023)
Other nonwhite	0.0375 (0.028)	0.0252 (0.031)	-0.0420 (0.025)
SAT math	0.0410 (0.012)	0.0507 (0.016)	-0.0560 (0.011)
SAT verbal	0.0222 (0.011)	0.0147 (0.014)	0.0635 (0.009)
<i>N</i>	86,844	77,312	82,675
<i>R</i> ²	0.6376	0.6323	0.6864

^aThe dependent variable is the grade in the class. Class fixed effects are estimated in all specifications. Standard errors are obtained using a wild bootstrap.

Inaccurate Interpretations in the Labour Literature?

1. Cornelissen, Dustmann and Schönberg (2017, AER):
"The coefficient of 0.148 implies that a 10 percent increase in peer quality increases wages by 1.48 percent [...]" (p. 441)
2. Hond and Lattanzio (2022):
"The figure [where $\hat{\eta} = 0.258$] shows [...] A 10 percent increase in peer quality increases the contemporaneous wage by 2.58 percent [...]" (p.16)
3. Cardoso et al. (2022):
"The key parameter of interest is estimated to be 0.2, meaning that if the quality of the peers as measured by $\bar{\alpha}_{it}$ increases by 10 percent wages will increase by 2.0 percent." (p. 12)

It's not an elasticity!

By Jensen's inequality

- The average of log ability and the log of average ability are not the same...

Peer Gap: From the Peer Pressure Literature

The interpretation must survive the following re-weighting:

$$\begin{aligned}y_{itn} &= \alpha_{it} + \eta (\bar{\alpha}_{\sim i, tn} - \alpha_{it}) + e_{itn} \\ &= (1 - \eta)\alpha_{it} + \eta \bar{\alpha}_{\sim i, tn} + e_{itn},\end{aligned}$$

that leads to $\eta < \gamma$.

What We Achieve

1. Consistent peer effects estimates;
2. Weak conditions for LLN/CLT; exogeneity of e_{itn} conditional on (α, δ) .

VS.

Limitations

1. Linear-in-means
2. Proportional relevance
3. Panel with changing peers; $T \geq 2$; sufficient within-student variation in peer composition.
4. Static/Cumulative