# The Econometrics of Social Interactions and Networks

## IOEA Academy


Camilo García-Jimeno

Federal Reserve Bank of Chicago


Cargèse, May 15th, 2023

# Norms in Agricultural Tenancy Contracts, Burke and Young (2001)

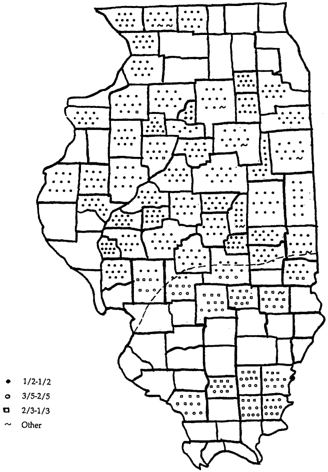



FIGURE 4. DISTRIBUTION OF SHARE CONTRACTS BY COUNTY: ILLINOIS, 1995

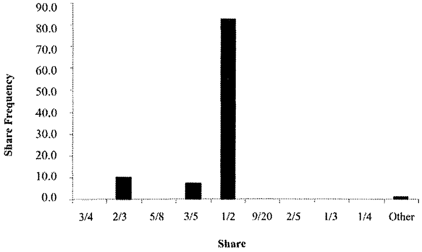



FIGURE 1. CROP SHARE FREQUENCIES IN ILLINOIS: TENANT'S SHARE OF THE CORN CROP (Frequencies in percent)

*Source:* Illinois Cooperative Agricultural Extension Service *Farm Leasing Survey*, 1995.

# Neighborhood Racial Composition Changes, Card, et al. (2008)

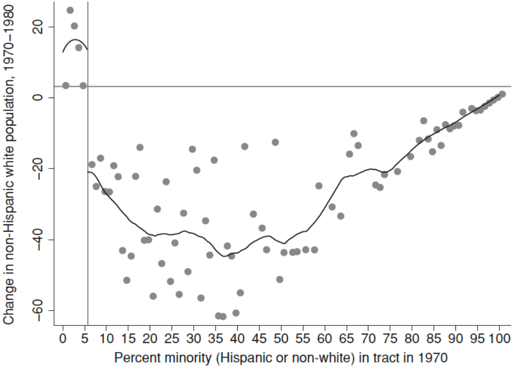
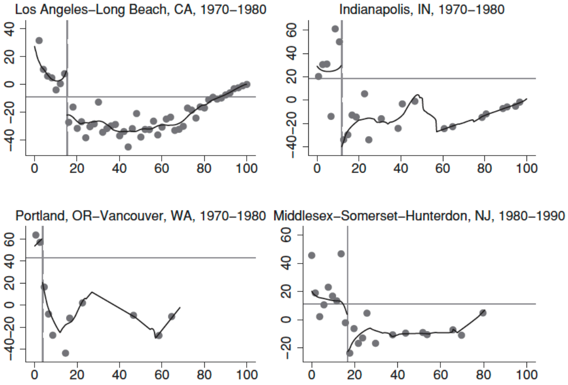



FIGURE I
Neighborhood Change in Chicago, 1970–1980

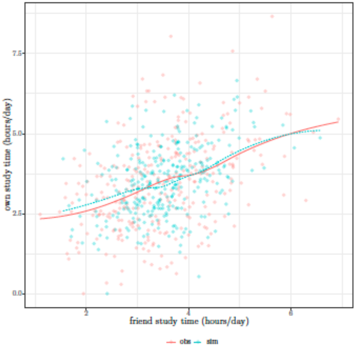


# Study Time Among College Students, Conley et al. (2023)

Table 2: Study time and GPA OLS regressions

| | *Dependent variable:* | |
|---|---|---|
| | Own study | GPA |
| | (1) | (2) |
| Male | −0.369** | −0.131* |
| | (0.171) | (0.076) |
| Black | 0.116 | −0.225** |
| | (0.214) | (0.109) |
| HS GPA | 0.413** | 0.437*** |
| | (0.188) | (0.081) |
| ACT | −0.032 | 0.040*** |
| | (0.023) | (0.013) |
| HS study | 0.043*** | 0.001 |
| | (0.008) | (0.004) |
| Expected study | −0.002 | −0.006 |
| | (0.009) | (0.003) |
| Friend study | 0.166*** | |
| | (0.039) | |
| Own study | | 0.090*** |
| | | (0.022) |
| Constant | 1.915** | 0.417 |
| | (0.759) | (0.362) |
| Observations | 574 | 571 |
| $R^2$ | 0.169 | 0.259 |

Figure A4: Fit of own study time against friend study time



# Social Interactions

- Commonality: others' choices matter for individual's outcomes and/or choices, and are *not* mediated by markets/prices.
- Economists refer to these as settings involving **social interactions**.
  - Old question in the social sciences: the interrelationship between the group and the individual.
- Wide array of phenomena: crime, teen behavior, school performance, addictions, learning, fads, public good provision, social norms, residential choice, etc.
  - Cross-group vs. within-group variance differences.
  - Individual-level variation in characteristics unable to account for variation across the population.
  - Unifying approach: social networks.

# Social Interactions: Main Ingredients

1. A social network: $\mathbf{A}$, $a_{ij}$ indicates $i$ and $j$ influence each other (study partners).
   - In some settings, $\mathbf{A}$ is known and fixed: (doesn't mean exogenous!: sorting).
   - In other settings, interest is in how $\mathbf{A}$ arises: network formation.
2. Some outcome of interest $y_i$ (academic achievement).
3. Some treatment or choice of interest $\omega_i$ (study time).
   - Others' choices $\omega_j$ matter for own choice $\omega_i$: 'endogenous (peer) effects'.
4. Observed individual characteristics $x_i$ (sex).
   - Others' characteristics $x_j$ matter for own choice $\omega_i$: 'contextual effects'.
5. Unobserved individual characteristics $\epsilon_i$ (conscientiousness).
   - Others' characteristics $\epsilon_j$ dependent with own characteristics $\epsilon_i$: 'correlated effects'.

# Micro-founded Workhorse Linear Model, Blume et al. (2015)

- Linear-quadratic utility:

$$U = \left(\gamma x_i + \epsilon_i + \delta x_{A(i)}\right)\omega_i - \frac{\phi}{2}\left(\omega_i - \omega_{A(i)}\right)^2 - \frac{1}{2}\omega_i^2$$

  defining

$$x_{A(i)} \equiv \sum_{j:a_{ij}>0} a_{ij}x_j \qquad \omega_{A(i)} \equiv \sum_{j:a_{ij}>0} a_{ij}\omega_j.$$

- $\delta$: strength of contextual effects.
- $\phi$: strength of peer effects. In this case, preference for conformity.
- $\gamma$: standard 'own' effect.
- $\epsilon_i$ unobserved by us and $i$'s social network $\Rightarrow$ game of incomplete info.

# Linear Best Reply, Blume et al. (2015)

- Leads to a linear 'best reply':

$$\omega_i = \frac{\gamma}{1+\phi}x_i + \frac{\delta}{1+\phi}x_{A(i)} + \frac{\phi}{1+\phi}\mathbb{E}[\omega_{A(i)}|x_i, \epsilon_i] + \frac{1}{1+\phi}\epsilon_i$$

  and solving for the social equilibrium,

$$\boldsymbol{\omega}^* = \frac{1}{1+\phi}\left(I - \frac{\phi}{1+\phi}\mathbf{A}\right)^{-1}[\gamma I + \delta \mathbf{A}]\,\mathbf{x} + \boldsymbol{\nu}$$

  which has the 'reduced-form' form

$$\omega_i = \pi_1 x_i + \pi_2 x_{A(i)} + \nu_i$$

# Identification I: Reduced Form

$$\omega_i = \pi_1 x_i + \pi_2 x_{A(i)} + \nu_i \tag{1}$$

- $\pi_2$ captures both direct and social network (indirect) influences on $i$'s choice:
  - Sex of study partners $x_{A(i)}$ directly affects $i$'s marginal utility of studying.
  - Sex of study partners $x_{A(i)}$ directly affects their marginal utility of studying, and thus indirectly $i$'s pressure to conform to his peers' study hours.
- From (1) we cannot separately identify $(\delta, \gamma, \phi)$.
  - $\phi \neq 0$: social feedback effects: Manski (1993) called it the 'reflection problem'.
  - But if $cov(x_{A(i)}, \nu_i) = 0$ and we can reject $H0 : \pi_2 = 0$, we have evidence of some kind of social interaction (contextual or peer).

# Identification II: Correlated Unobservables

- $cov(x_{A(i)}, \nu_i) = 0$: very strong assumption.
- $\nu_i$ is a linear combination of the unobservables of $i$ and $i$'s social network.
- Own observables likely dependent with own unobservables: students with higher high school GPA possibly more conscientious.
- Friend's unobservables likely dependent with own unobservables: $i$ sorts into a study group of conscientious students.

# Social Interactions III: Correlated Unobservables

- When the problem generating $cov(x_{A(i)}, \nu_i) \neq 0$ is sorting of similar people into groups, a possibility is to do an intervention that **randomizes group composition**.
- This breaks the dependence between others' unobservables and individual characteristics.
- See e.g., Sacerdote (2001): random assignment of roommates in a college.
- But beware of re-sorting post intervention!: see Carrell, Sacerdote, West (2013): Air Force Academy experiment.

# Social Interactions IV: Correlated Unobservables

- When the problem generating $cov(x_{A(i)}, \nu_i) \neq 0$ is the presence of common environmental factors (crime, quality of schools, quality of governance, etc), the problem is perhaps more serious.
  - E.g., Men likely to study with men, and men study near the soccer field, which is noisy and distracting.
  - Perhaps the reason for the recent work estimating 'neighborhood effects' (e.g. Chetty and coauthors).
  - Or quasi-natural experiments (instruments) that generate variation in $x_{A(i)}$:
    - Note: can look for variation in the $x_j'$'s (see e.g., Acemoglu, Garcia-Jimeno, Robinson (2015)), or in the $A(i)$ (see e.g., Garcia-Jimeno, Iglesias, Yildirim (2022)).

# Social Multipliers I: Glaeser et al (1996), (2003)

- Alternative way to look for evidence of social interactions: social multipliers.
  - In the presence of social interactions (even just contextual effects), individual level causal relationship $\neq$ aggregate level 'equilibrium' relationship.
  - E.g., If one additional hour of study time causes a 0.1 pp increase in GPA, the slope of a group-level regression of average study time on average GPA will have a slope much higher than 0.1.
  - For simplicity, suppose $\phi = 0$, $x_i \perp \epsilon_i$, and that people interact in disjoint groups $g$.
  - Individual-level reduced-form regression: $\omega_{ig} = a + bx_{ig} + \eta_{ig}$.
  - Group-level regression (average within groups): $\overline{\omega}_g = \tilde{a} + \tilde{b}\overline{x}_g + \overline{\eta}_g$
  - Glaeser et al. define the social multiplier as: $S = \frac{\tilde{b}}{b}$.

# Social Multipliers II

- Since the social interactions model is

$$\omega_{ig} = \gamma x_{ig} + \delta \overline{x}_g + \epsilon_{ig} \Rightarrow plim(b) = \gamma + \delta \frac{cov(x_{ig}, \overline{x}_g)}{var(x_{ig})}$$

- Averaging the social interactions model,

$$\overline{\omega}_g = (\gamma + \delta)\overline{x}_g + \overline{\epsilon}_g \Rightarrow plim(\tilde{b}) = \gamma + \delta$$

# Social Multipliers III

- $\Rightarrow$ Social multiplier:

$$S = \frac{\gamma + \delta}{\gamma + \delta \frac{cov(x_{ig}, \overline{x}_g)}{var(x_{ig})}}$$

- If segregation across groups is perfect, then $cov(x_{ig}, \overline{x}_g) = var(x_{ig})$, and

$$S = 1$$

- In the other extreme, if membership to groups is randomly assigned, then $cov(x_{ig}, \overline{x}_g) = 0$, and

$$S = \frac{\gamma + \delta}{\gamma}$$

- The more similar people are within a group, the lower the multiplier.

# Social Multipliers IV: Dartmouth Randomly Assigned Roomates

TABLE 1. SOCIAL MULTIPLIERS IN FRATERNITY PARTICIPATION
DARTMOUTH ROOMMATE DATA: EFFECT OF BACKGROUND CHARACTERISTICS ON PARTICIPATION IN FRATERNITIES AT THE INDIVIDUAL LEVEL AND THREE LEVELS OF AGGREGATION

| | (1) Member of fraternity or sorority | (2) Room average level membership | (3) Floor average membership | (4) Dorm average membership |
|---|---|---|---|---|
| Drank beer in high school | 0.1040 | 0.0984 | 0.1454 | 0.2320 |
| | (0.0258) | (0.0399) | (0.0812) | (0.1930) |
| Constant | 0.0482 | 0.1980 | 0.7993 | 2.2277 |
| | (0.1455) | (0.2266) | (0.4594) | (1.1421) |
| R-squared | 0.04 | 0.05 | 0.03 | 0.08 |
| Observations | 1579 | 700 | 197 | 57 |
| Average group size | 1 | 2.3 | 8.0 | 28 |

Notes: Data are for Dartmouth Freshmen. Roommates and dormmates are randomly assigned as described in Sacerdote (2001). Regressions include math and verbal SAT scores, dummy for male, family income, high school GPA. SAT scores are from Dartmouth Admissions data. Family income, use of beer, and high school GPA are self reported on the UCLA Higher Education Research Institute's Survey of Incoming Freshmen. Standard errors in parentheses.
Column (1) shows the OLS regression of individual fraternity participation on own use of beer in high school. (Own SAT scores, own high school GPA and own family income are also included but not shown). Column (2) regresses the average participation at the dorm room level on dorm room averages of high school beer use, SAT scores, HS GPAs, and family income. Columns (3) and (4) increase the level of aggregation to the dorm floor and dorm building, respectively.

# Experimental Approaches: Randomization

- Settings where researchers may conjecture that social interactions are present (e.g. spillovers of a program):
  - Distinction between **direct** and **indirect** treatment effects -analogous to own and contextual effects- (see Manski (2013) for a general framework).
- Ideally, randomization should break the dependence of own treatment $x_i$ with unobservables, but also of social network's treatments $x_{A(i)}$ with unobservables.
  - Issue 1: How to do this?
  - Issue 2: Non-compliance: Even if you randomly assign a treatment, people may not take it up (for reasons related to their unobservables).

# Social Interactions Moral Hazard: Kremer and Miguel (2007)

- An insightful example is the work by Kremer and Miguel (2007) who use data from an intervention in Kenyan schools to incentivize de-worming of children (see also Beaman et al. (2021)).
- Experimental design:
  - 3 sets of schools, staggered randomized treatment $\Rightarrow$ schools comparable.
  - Treatment: de-worming drugs and lots of training.
  - Ask parents how many other parents they know with children in the different schools.
- Idea: compare child de-worming drug take-up decisions of parents ($T_i$) in later-treated schools that have different numbers of social connections to parents of children in the early treatment schools.
  - How many other parents you know is endogenous, but how many of those are in early schools is exogenous thanks to randomization.

# Social Interactions Moral Hazard: Kremer and Miguel (2007)

$$\mathbb{P}(T_i = 1) = \delta x_{A(i)} + \beta_1 N_i + \beta_2 Z_i + \epsilon_i$$

- $x_{A(i)}$: How many parents with children in treated schools I know.
- $N_i$: How many other parents I know.
- Notice that the children in early-treated schools are **not** in this regression sample.

# What if we try estimating a peer effect with nonexperimental data?

NONEXPERIMENTAL SOCIAL EFFECT ESTIMATES (GROUPS 2 AND 3)

| | Dependent variable: Child took deworming drugs in 2001 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Explanatory variables: | | | |
| Proportion deworming drug take-up in 2001, respondent's own school (not including respondent) | 0.852*** (0.107) | | |
| # parent links with children in respondent's own school whose children received deworming | | 0.016 (0.011) | |
| # parent links with children in early treatment schools whose children received deworming and had "good effects" | | | 0.004 (0.025) |
| # parent links with children in early treatment schools whose children received deworming and had "side effects" | | | −0.152* (0.080) |
| # parent links with children in early treatment schools whose children received deworming and respondent does not know effects | | | 0.003 (0.049) |
| # parent links with children in early treatment schools whose children did not receive deworming | | | −0.006 (0.055) |
| # parent links with children in early treatment schools, respondent does not know whether they received deworming | | | −0.010 |
| Total social link controls, socio-economic controls | Yes | Yes | Yes |
| Number of observations (parents) | 1,678 | 886 | 886 |
| Mean of dependent variable | 0.61 | 0.56 | 0.56 |

# Social Interactions Moral Hazard: Kremer and Miguel (2007) results

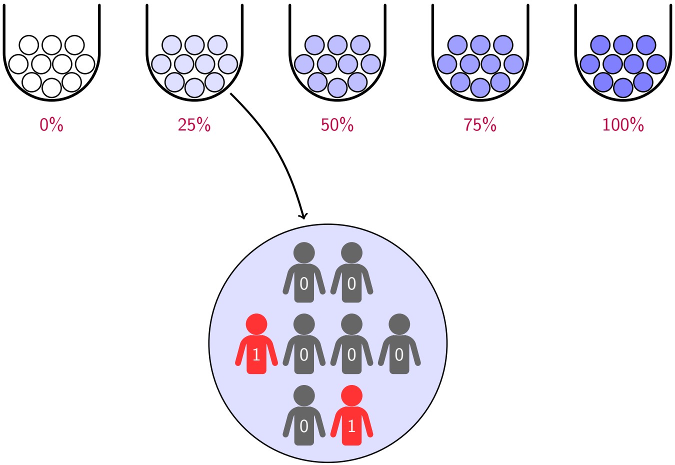
Experimental Social Effect Estimates

| | Dependent variable: Child took deworming drugs in 2001 | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Explanatory variables: | | | | | |
| # parent links with children in early treatment schools (Groups 1 and 2, not own school) | −0.031** | −0.040** | | | −0.002 |
| | (0.014) | (0.017) | | | (0.018) |
| # parent links with children in early treatment schools * Group 2 school indicator | | 0.017 | | | |
| | | (0.029) | | | |
| Proportion direct (first-order) parent links with children in early treatment schools | | | −0.098** | | |
| | | | (0.045) | | |
| # parent links with children in early treatment schools, with whom respondent speaks at least twice/week | | | | −0.030** | |
| | | | | (0.016) | |
| # parent links with children in early treatment schools, with whom respondent speaks less than twice/week | | | | −0.033 | |
| | | | | (0.033) | |
| # parent links with children in Group 1, 2, or 3 schools, not own school, with whom respondent speaks at least twice/week | | | | 0.008 | |
| | | | | (0.012) | |
| # parent links with children in Group 1, 2, or 3 schools, not own school, with whom respondent speaks less than twice/week | | | | 0.026 | |
| | | | | (0.027) | |
| # parent links with children in early treatment schools * Respondent years of education | | | | | −0.0062* |
| | | | | | (0.0032) |
| # parent links with children in Group 1, 2, or 3 schools, not own school | 0.013 | 0.012 | −0.006 | | −0.014 |
| | (0.011) | (0.017) | (0.009) | | (0.014) |
| # parent links with children not in Group 1, 2, or 3 schools | −0.007 | −0.008 | −0.005 | −0.007 | −0.008 |
| | (0.007) | (0.009) | (0.007) | (0.007) | (0.011) |
| # parent links, total | 0.019*** | 0.029*** | 0.021*** | 0.018*** | 0.013 |
| | (0.005) | (0.007) | (0.007) | (0.005) | (0.008) |

# Example: Displacement Effects of Labor Market Policies

- Crepon et al. (2013) study a large-scale job-seeker assistance program in France.
- In DiTraglia et al. (2023) we revisit it.
- They randomized offers of intensive job placement services.
- The issue:
  - *Job seekers who benefit from counseling may be more likely to get a job, but at the expense of other unemployed workers with whom they compete in the labor market. This may be particularly true in the short run, during which vacancies do not adjust: the unemployed who do not benefit from the program could be partially crowded out.*

# How can we test whether such displacement effects are present?: Randomized Saturation Designs




## Partial Interference

Spillovers within but not between groups.

## Randomized Saturation

Two-stage experimental design.

# Randomized Saturation Experiments

**Sample Size and Indexing**

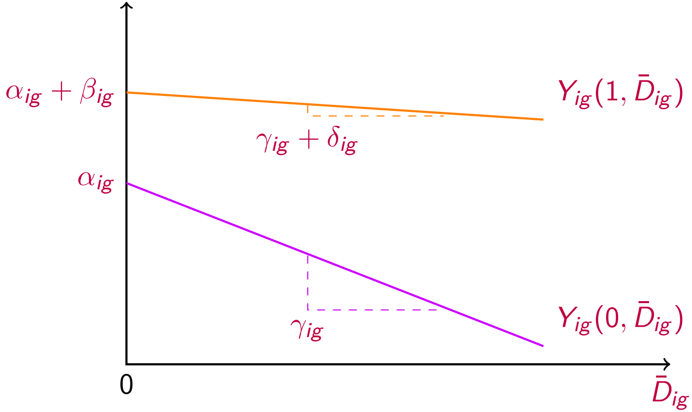
| | |
|---|---|
| Groups | $g = 1, \ldots G$ |
| Individuals in $g$ | $i = 1, \ldots, N_g$ |

**Observables**

| | |
|---|---|
| $Y_{ig}$ | outcome of $(i, g)$ |
| $Z_{ig} \in \{0, 1\}$ | treatment offer to $(i, g)$ |
| $D_{ig} \in \{0, 1\}$ | treatment take-up of $(i, g)$ |
| $\bar{D}_{ig} \in [0, 1]$ | take-up of $(i, g)$'s "neighbors" |
| $S_g \in \mathcal{S} \subseteq [0, 1]$ | saturation of group $g$ |

## Linear Model allowing for heterogeneous effects:

$$Y_{ig}(D_{ig}, \bar{D}_{ig}) = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \bar{D}_{ig} + \delta_{ig} D_{ig} \bar{D}_{ig}$$




Indirect Effects

Treated: $\gamma_{ig} + \delta_{ig}$

Untreated: $\gamma_{ig}$

Direct Effects

$\beta_{ig} + \delta_{ig} \bar{D}_{ig}$

# Randomized Saturation Experiments

- Intuitively, each layer of randomization provides exogenous variation for each effect (own and spillover).
- However: Need to be careful:
  1. IOR: Need to rule out **spillovers in take-up**:

$$D_{ig}(Z_i, Z_{-i}) = D_{ig}(Z_i).$$

  2. The "naive IV" will not work: need to condition on the share of compliers $\overline{C}_g$ in one's group.
     - Under RSD, share of neighbors offered treatment is exogenous.
     - Under IOR, share of neighbors taking up is $\overline{D}_g \approx \overline{C}_g \times S_g$.

# Testing the IOR assumption in the Crepon et al. (2013) experiment

## (IOR) + (1-Sided)

Take-up only depends on *own* offer:

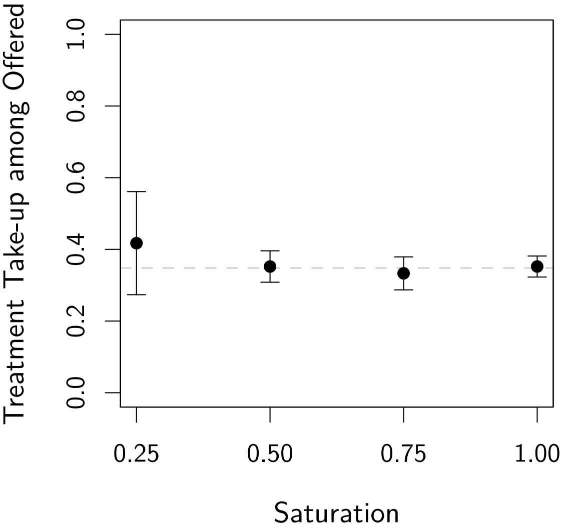
$D_{ig} = C_{ig} \times Z_{ig}$

## Testable Implication

$\mathbb{E}[D_{ig}|Z_{ig} = 1, S_g] = \mathbb{E}[D_{ig}|Z_{ig} = 1]$

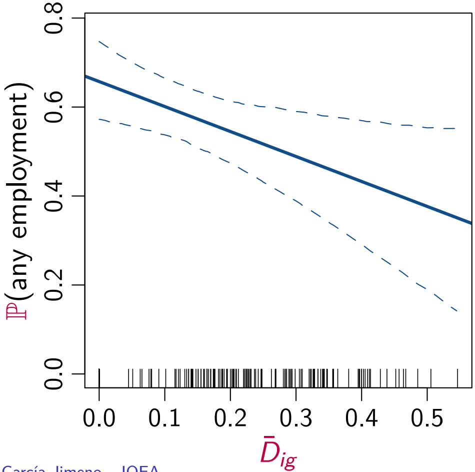
## Figure at right

Take-up among offered doesn't vary with saturation (p = 0.62)

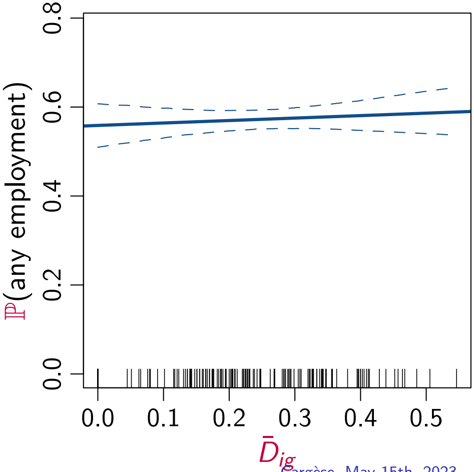


# Displacement and Shielding Effects in the Crepon et al. (2013) experiment

Untreated: $\mathbb{E}[Y_{ig}(0, \bar{d})|\text{Complier}]$




Treated: $\mathbb{E}[Y_{ig}(1, \bar{d})|\text{Complier}]$



# Non-experimental Approaches: Instrumental Variables

- Two examples using quasi-experimental variation to identify social interaction effects.
  1. Exogenous cross-sectional variation in the contextual variable $x_i$, hence in $x_{A(i)}$ (see Bramoulle, Djebbari, Fortin (2009)). State capacity and public good provision.
  2. Exogenous time series variation in network connections $A(i)$ (hence in $x_{A(i)}$): Information diffusion and collective action.

# State Capacity: Acemoglu, Garcia-Jimeno, Robinson (2015)

- Social interactions may be present in the provision of public goods and the development of state infrastructure too.
- In Acemoglu, Garcia-Jimeno, Robinson (2015) how investments in public infrastructure in a municipality may matter for investments in nearby municipalities in Colombia.
- Both model the investment choices, and use colonial state presence measures as instruments.
- Municipality chooses investment $\omega_i$ to maximize 'prosperity':

$$p_i = (\gamma_1 x_i + \epsilon_i)\omega_i + \phi\omega_i\omega_{A(i)} + \gamma_2\omega_{A(i)} + u_i - \frac{\theta}{2}\omega_i^2$$

- Alow for direct spillover effects of public goods from neighboring municipalities.

# Best Responses and Equilibrium

- Investment choice (best response):

$$\omega_i = \frac{\phi}{\theta}\omega_{A(i)} + \gamma_1 x_i + \epsilon_i \tag{2}$$

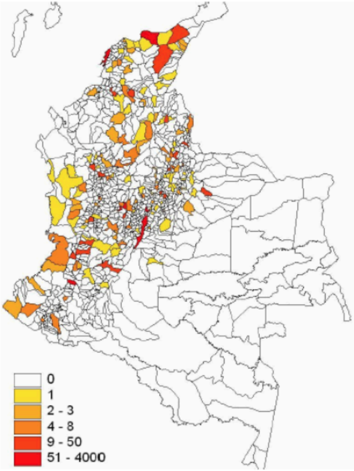
- So the *equilibrium* relationship between prosperity and investments is

$$p_i = \theta\omega_i^2 + \gamma_2\omega_{A(i)} + u_i \tag{3}$$

- Estimate **both** (2) and (3) using IVs:
  - An element of $x_i$ is colonial state presence.
  - If $x_i \Rightarrow \omega_i$, then $x_{A(i)} \Rightarrow \omega_{A(i)}$.
  - Exclusion restriction: $cov(x_{A(i)}, \epsilon_i) = 0$.

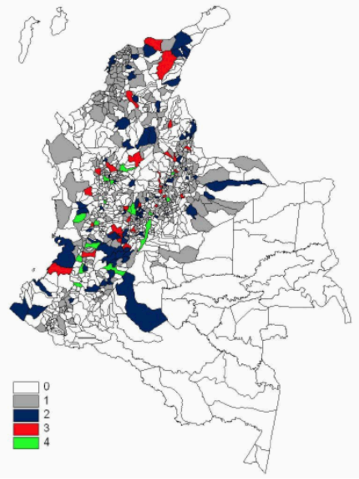
# The Spanish Colonial State in 18th Century Colombia

Crown Employees




Source: Duran y Díaz (1794)

State Agencies




Source: Duran y Díaz (1794)

# Acemoglu, Garcia-Jimeno, Robinson (2015): Estimates

TABLE 3—CONTEMPORARY STATE EQUILIBRIUM BEST RESPONSE

| State capacity measured as log of: | Equilibrium best response | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Number of state agencies | | | | Number of municipality employees | | | |
| *Panel I* | OLS (1) | IV (2) | IV (3) | Sys. GMM (4) | OLS (5) | IV (6) | IV (7) | Sys. GMM (8) |
| $ds_i/ds_j$ | 0.016 | 0.017 | 0.019 | 0.020 | 0.021 | 0.022 | 0.022 | 0.016 |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.003) |
| $ds_i/d(colonial\ state\ officials_i)$ | 0.127 | 0.128 | 0.108 | −0.040 | 0.129 | 0.130 | 0.105 | 0.087 |
| | (0.031) | (0.031) | (0.033) | (0.050) | (0.043) | (0.043) | (0.047) | (0.069) |
| $ds_i/d(colonial\ state\ agencies_i)$ | 0.003 | 0.001 | −0.016 | 0.096 | 0.017 | 0.017 | −0.002 | 0.085 |
| | (0.033) | (0.033) | (0.033) | (0.055) | (0.058) | (0.059) | (0.060) | (0.085) |
| $ds_i/d(distance\ to\ royal\ roads_i)$ | 0.008 | 0.010 | 0.007 | 0.074 | −0.035 | −0.035 | −0.038 | −0.036 |
| | (0.019) | (0.019) | (0.021) | (0.034) | (0.034) | (0.035) | (0.036) | (0.044) |

TABLE 4A—PROSPERITY AND PUBLIC GOODS STRUCTURAL EQUATION

| | State capacity measured as: log of number of municipality state agencies | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Prosperity equation | | | | | | | |
| | Life quality index | | | | Public utilities coverage | | | |
| *Panel I* | OLS (1) | IV (2) | IV (3) | Sys. GMM (4) | OLS (5) | IV (6) | IV (7) | Sys. GMM (8) |
| $dp_i/ds_i$ | 0.802 | 0.394 | 0.389 | 0.314 | 0.602 | 0.563 | 0.567 | 0.314 |
| | (0.044) | (0.135) | (0.143) | (0.041) | (0.037) | (0.127) | (0.134) | (0.041) |
| $dp_i/ds_j$ | 0.015 | 0.024 | 0.025 | 0.025 | 0.022 | 0.020 | 0.020 | 0.027 |
| | (0.004) | (0.006) | (0.006) | (0.004) | (0.004) | (0.006) | (0.006) | (0.003) |

# Collective Action and Information: Garcia-Jimeno, et al. (2022)

- Social networks play important role in information diffusion.
- Collective action (e.g., protest movements) requires lots of coordination, which in turn requires information transmission.
- We study the spread of the Temperance Crusade, an anti-liquor protest movement in the 1870s.

# Collective Action and Information: Garcia-Jimeno, et al. (2022)

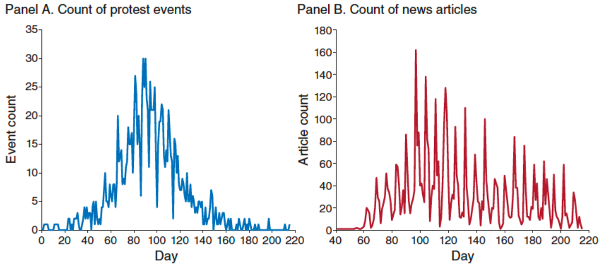



Figure 1. Temperance Crusade Activity and Its Newspaper Coverage, December 1873–July 1874

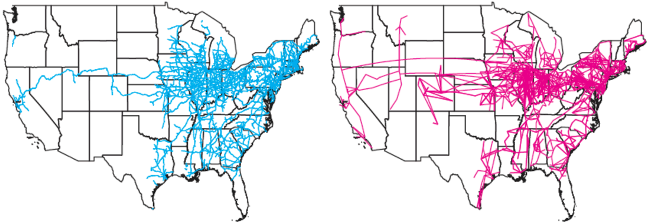


Figure 2. The Railroad and Telegraph Networks, c.1870

# Collective Action and Information: Garcia-Jimeno, et al. (2022)

- Model the protest events as pieces of information, and the railroad and telegraph networks as the links over which the information is transmitted:

$$\omega_{it} = \beta_r^\ell \omega_{R_t(i),t-\ell} + \beta_\gamma^\ell \omega_{\Gamma_t(i),t-\ell} + \epsilon_{it}$$

- Use railroad accidents that disrupt the communications network to construct **time-varying** instruments for rail and telegraph connected towns.

# Railroad Accidents and Variation in Connectivity

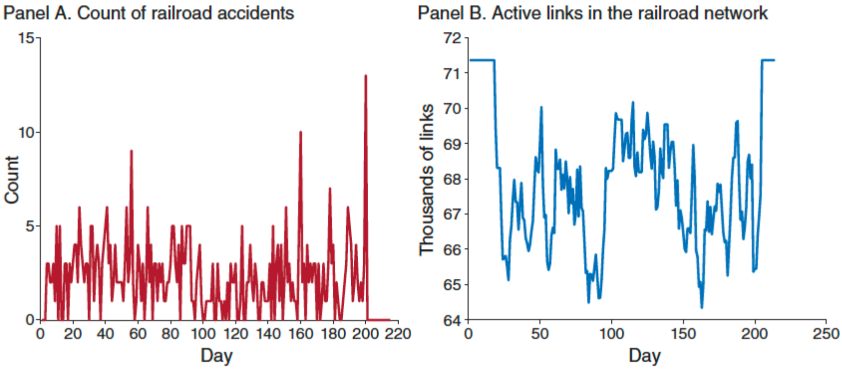



Figure 3. Variation in Railroad Accidents: Number of Accidents and Active Links in the Railroad Network, December 1873–July 1874

# Garcia-Jimeno, et al. (2022): Results

Table 2—The Effect of Information along the Rail and Telegraph Networks: Causal Estimates

| | Any crusade activity $a_{it}$—meetings, petitions, marches | | | | | |
|---|---|---|---|---|---|---|
| | 5 days | | | 3 days | | |
| | OLS (1) | IV (2) | GMM (3) | OLS (4) | IV (5) | GMM (6) |
| First lag rail | 0.004 | 0.037 | 0.038 | 0.005 | 0.032 | 0.031 |
| ($\mathbf{r}_{i,t}\mathbf{a}_{t-1}$) | (0.001) | (0.013) | (0.013) | (0.001) | (0.014) | (0.016) |
| First lag telegraph | 0.014 | 0.172 | 0.163 | 0.008 | 0.072 | 0.076 |
| ($\gamma_i\mathbf{a}_{t-1}$) | (0.006) | (0.033) | (0.054) | (0.005) | (0.015) | (0.039) |
| Second lag telegraph | 0.018 | −0.068 | −0.053 | 0.011 | 0.020 | 0.014 |
| ($\gamma_i\mathbf{a}_{t-2}$) | (0.005) | (0.031) | (0.058) | (0.004) | (0.024) | (0.034) |
| Third lag distance | 0.001 | 0.006 | 0.006 | 0.001 | −0.001 | −0.001 |
| ($\mathbf{d}_i\mathbf{a}_{t-3}$) | (0.0001) | (0.002) | (0.002) | (0.0001) | (0.001) | (0.001) |
| Number of towns | 15,934 | 15,934 | 15,934 | 15,950 | 15,950 | 15,950 |
| Maximum number of periods | 16 | 16 | 16 | 30 | 30 | 30 |
| Observations | 267,247 | 267,247 | 267,247 | 487,548 | 487,548 | 487,548 |

# Network Formation

- Economists are also interested in understanding where **A** comes from: how are social connections produced? E.g., high-school friendships, trading relations, co-authoring relationships, etc.
- A starting point is a conditionally independent dyadic model:

$$a_{ij} = f(x_i, x_j, u_i, u_j, v_{ij})$$

- $a_{ij} =$ link (friendship), $x_i =$ observed characteristics of $i$ (PSG vs. Marseille fan), $u_i =$ unobserved characteristics of $i$ (sociability), $v_{ij}$ a pair-specific unobserved shock (same church –easier–, or having a friend in common –harder–).
- In general, $x_i \not\perp u_i$.

# Network Formation: Issues

- $a_{12} \perp a_{34}$: Friendship decisions of disjoint pairs are independent: knowing that 1 and 2 are friends provides no information about whether 3 and 4 are friends.
- $a_{12} \not\perp a_{13}$: both influenced by $x_1, u_1$: E.g., 1 is a grumpy person. Observing that 1 is not friends with 2 suggests 1 is unlikely to be friends with 3.
  - $a_{12} \not\perp a_{13} | x_1$: both link decisions still influenced by $u_1$.
- $a_{12} \not\perp a_{21}$:
  - Links do not need to be symmetric (e.g., I cite you, you don't cite me).
  - $a_{12} \not\perp a_{21} | x_i, x_j, u_i, u_j$: If $v_{ij}, v_{ji}$ covary, still dependence.

$\Rightarrow$ Running a dyadic OLS/Probit regression (e.g., Santos Silva and Tenreyro (2010)) will be a problem:

1. Bias: Covariates $x_i, x_j$'s dependent with the error $u_i, u_j, v_{ij}$.
2. Inference: errors across dyad observations $(i, j)$, $(i, k)$ dependent.

# Network Formation: Example of Bias

- Model of friendships with **homophily**: teenagers more likely to be-friend other teenagers of the same race ($x_i$) (see Miyauchi (2016)).
- Also, some teenagers more sociable than others ($u_i$).

$$a_{ij} = \mathbf{1}\left\{\phi \mathbf{1}\{x_i = x_j\} + u_i + u_j + v_{ij} > 0\right\}$$

- Three students, same race, all friends with each other: $a_{12} = a_{13} = a_{23} = 1$.
- Suppose we ignore $u_i, u_j$: $\Rightarrow$ attribute all three links to a preference for homophily (large $\phi$), although maybe 1 and 3 became friends not because they are both the same race but because A is very sociable (large $u_i$).

# Solutions: Graham (2017)

- $u_i, u_j$ are **nuisance parameters**: In a relatively sparse network, each $u_i$ only matters for a relatively small number of links $\Rightarrow$ cannot just use individual fixed effects.
- When $v_{ij}$ are logit errors,

$$\ell_{ij} = \mathbb{P}(a_{ij} = 1|\phi, u_i, u_j) = \frac{exp(\phi x_{ij} + u_i + u_j)}{1 + exp(\phi x_{ij} + u_i + u_j)}$$

- Graham (2017) offers two possible solutions.

# Solutions: Graham (2017)

- 1. Conditional likelihood: Find a sufficient statistic such that conditioning on it, the likelihood of the network does *not* depend on the $u_i$ anymore.
  - The **degree sequence** $S = (d_1, d_2, ..., d_N)$ of the network happens to be such a sufficient statistic.

$$L = \mathbb{P}(\mathbf{A}|\phi, S) = \prod_{ij} \ell_{ij} = \frac{exp(\sum_{ij} a_{ij}\phi x_{ij})}{\sum_{\alpha \in D} exp(\sum_{ij} \alpha_{ij}\phi x_{ij})},$$

$$D_s = \{\text{Adjecency matrices } \mathbf{A} \text{ with degree sequence } S\}.$$

# Solutions: Graham (2017)

- 2. Fixed point problem:
  - First fix a $\phi$ and find the solution to

$$\mathbf{U}(\phi) = \text{argmax} \quad \mathbb{P}(\mathbf{A}|\phi, u_1, ...u_N).$$

  - Then find the solution to

$$\phi^* = \text{argmax} \quad \mathbb{P}(\mathbf{A}|\phi, \mathbf{U}(\phi))$$

  - Iterate.
- See Islam et al. (2022) for an application to a Bangladeshi friendship network.

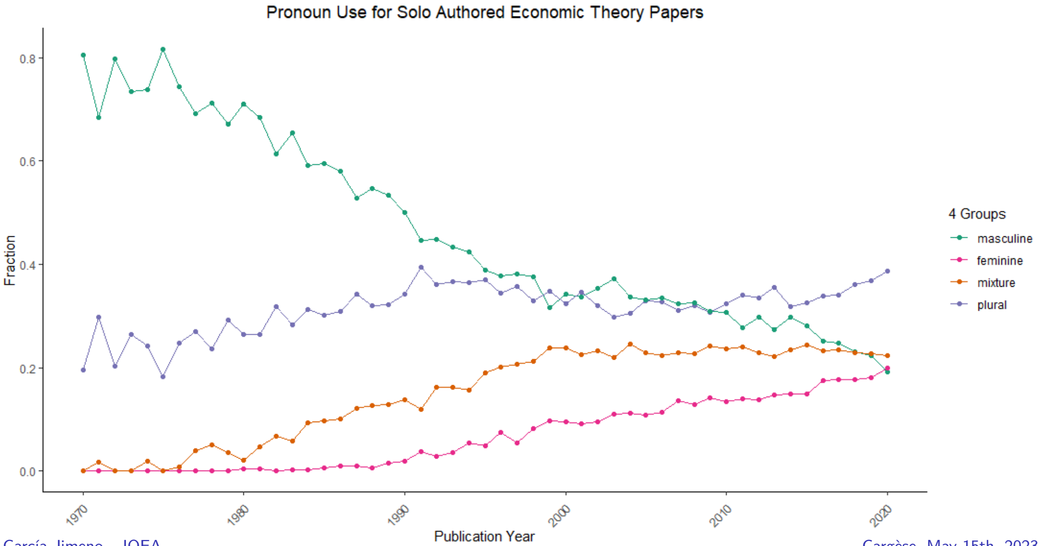
# Community Detection

- Community Detection: Methods to classify nodes in a network that are the same type –belong to the same community– (see Karrer and Newman (2010)).
- In our notation, find which nodes have the same $u_i$, when $u_i$ indexes groups.
- Methods based on statistical inference develop efficient algorithms to find the best possible partition of nodes into a fixed number of groups to maximize

$$\mathbb{P}(\mathbf{A}|\phi, u_1, ...u_N)$$

- E.g., Social norms (I'm working on): Co-authorship network of economists.
  - 2 groups: liberals and conservatives.
  - Homophily in co-authorship decisions.
  - Use the recovered classification to control for $u_i$ in a social interactions model of pronoun gender choice.

# Pronoun Gender Choice



# References


- Acemoglu, Garcia-Jimeno, Robinson (2015). "State Capacity and Economic Development: A Network Approach", *American Economic Review*, 105(8).
- Beaman, BenYishay, Magruder, Mobarak (2021). "Can Network Thory-Based Targeting Increase Technology Adoption?", working paper.
- Blume, Brock, Durlauf, Ioannides (2011). "Identification of Social Interactions", in *Handbook of Social Economics*, Benhabib et al (eds.), North-Holland.
- Blume, Brock, Durlauf, Jayaraman (2015). "Linear Social Interactions Model", *Journal of Political Economy*, 123(2).
- Bramoulle, Djebbari, Fortin (2009). "Identification of Peer Effects Through Social Networks", *Journal of Econometrics*, 150(1).
- Carrell, Sacerdote, West (2013). "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation", *Econometrica*, 81(3).
- Conley, Mehta, Stinebrickner, Stinebrikner (2023). "Social Interactions, Mechanisms, and Equilibrium: Evidence from a Model of Study time and Academic Achivement", working paper.
- Crepon, Duflo, Gurgand, Rathelot, Zamora (2013). "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment", *Quarterly Journal of Economics*, 128(2).
- DePaula (2017). "Econometrics of Network Models", in *Advances in Economics and Econometrics*, Honore, et al. (eds.). Cambridge University Press.

# References


- DiTraglia, Garcia-Jimeno, O'Keefe-O'Donovan, Sanchez (2023), "Identifying Causal Effects in Experiments with Spillovers and Non-Compliance", *Journal of Econometrics*, forthcoming.
- Garcia-Jimeno, Iglesias, Yildirim (2022), "Information Networks and Collective Action: Evidence from The Women's Temperance Crusade", *American Economic Review*, 112(1).
- Graham (2017). "An Econometric Model of Network Formation with Degree Heterogeneity", *Econometrica*, 85(4).
- Glaeser, Sacerdote, Scheinkman (1996)."Crime and Social Interactions", *Quarterly Journal of Economics*, 111(2).
- Glaeser, Sheinkman, Sacerdote (2003). "The Social Multiplier", *Journal of the European Economic Association*, 1(2-3).
- Islam, Kim, Song, Zenou (2022). "Network Formation with Unobserved Homophily: Identification and Consistent Estimation", working paper.
- Karrer, Newman (2010). "Stochastic Blockmodels and Community Structure in Networks", Physics Review E, 83(1).
- Manski (2013). "Identification of Treatment Response with Social Interactions", *Econometrics Journal*, 16(1).
- Miyauchi (2016). "Structural Estimation of Pairwise Stable Networks with Nonnegative Externality", *Journal of Econometrics*, 195(2).
- Sacerdote (2001)."Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 116(2).
- Santos Silva, Tenreyro (2010). "Currency Unions in Prospect and Retrospect", *Annual Review of Economics*, 2.