Identifying Causal Effects in Experiments with Social Interactions and Non-compliance

Francis J. DiTraglia[1] Camilo García-Jimeno[2] Rossa O'Keeffe-O'Donovan[1]
Alejandro Sanchez[3]

[1]University of Oxford

[2]Federal Reserve Bank of Chicago

[3]University of Pennsylvania

March 16th, 2020

Empirical Example with Potential for Indirect Treatment Effects

Crepon et al. (2013; QJE)

- Large-scale job-seeker assistance program in France.
- Randomized offers of intensive job placement services.

Displacement Effects of Labor Market Policies

"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."

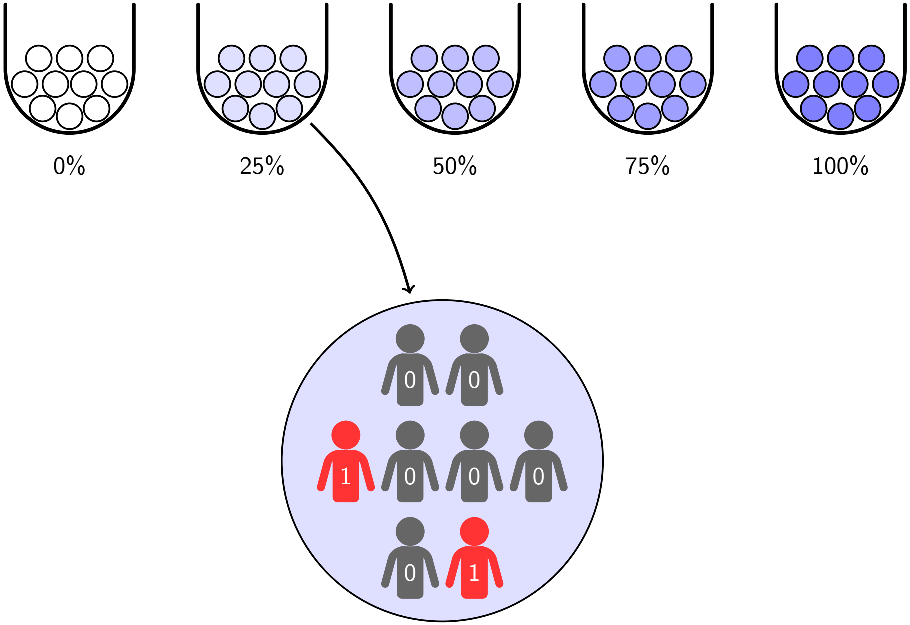
Ingredients for Studying Social Interactions Without Network Data

Partial Interference Assumption

- Each individual belongs to a single, known group g.
- Potential spillovers within but not between groups.

Randomized Saturation Design

- Saturation $\equiv$ fraction of individuals offered treatment.
- Groups randomly assigned saturation $S_g \in \mathcal{S}$.
- Individuals randomly assigned treatment offer $Z_{ig} \in \{0, 1\}$.


0%
25%
50%
75%
100%
0
0
1
0
0
0
0
1

Non-compliance in Randomized Saturation Experiments

Non-compliance

Those offered treatment don't always take it up; those not offered treatment sometimes obtain it anyway.

Existing Literature

Assumes perfect compliance or estimates intent-to-treat effects.

This Paper

1. Randomized saturation experiments with one-sided non-compliance.
2. Correlated random coefficients model to allow for unobserved heterogeneity.
3. Identify avg. direct and indirect effects for the treated; indirect for the untreated.
4. Consistent estimation under large-groups/many-groups asymptotics.
5. Apply to the Crepon et al (2013) experiment.

Notation

Sample Size and Indexing

- Groups: $g = 1, \ldots, G$
- Individuals in g: $i = 1, \ldots, N_g$
- Average group size: $\bar{N}$

Observables

- Z_{ig} = binary treatment offer to (i, g)
- D_{ig} = binary treatment take-up of (i, g)
- Y_{ig} = outcome of (i, g)
- S_g = saturation of group g
- $\bar{D}_{ig}$ = take-up fraction in g excluding (i, g)

Overview of Assumptions

(i) Experimental Design: Randomized Saturation ✓

(ii) Potential Outcomes: Correlated Random Coefficients Model

(iii) Treatment Take-up Behavior: "IOR" Assumption

(iv) Exclusion Restriction for (Z_{ig}, S_g)

Assumptions on Potential Outcomes: Correlated Random Coefficients

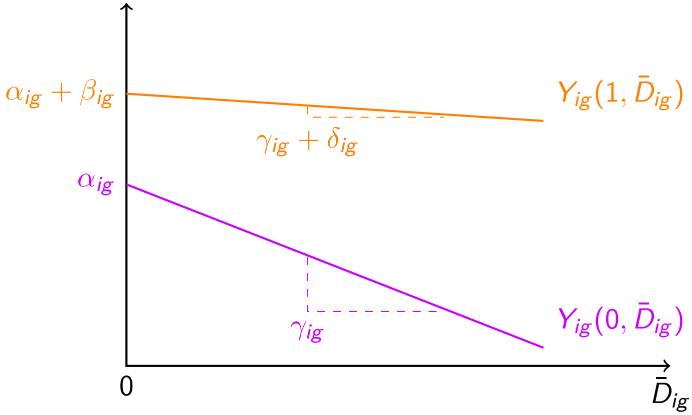
$$Y_{ig}(\boldsymbol{D}) = Y_{ig}(\boldsymbol{D}_g) = Y_{ig}(D_{ig}, \bar{D}_{ig}) = \mathbf{f}(\bar{D}_{ig})' \left[(1 - D_{ig})\boldsymbol{\theta}_{ig} + D_{ig}\boldsymbol{\psi}_{ig}\right]$$

- $\mathbf{f}$ is a vector of known functions, bounded on $[0, 1]$
- $\boldsymbol{\theta}_{ig}$ and $\boldsymbol{\psi}_{ig}$ are RVs that may be dependent on $(D_{ig}, \bar{D}_{ig})$.

This Talk

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

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



Indirect Effects

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

Untreated: γ_{ig}

Direct Effects

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

Assumptions on Treatment Take-up

One-sided Non-compliance

Only those offered treatment can take it up.

Individualistic Offer Response (IOR)

$$D_{ig}(\boldsymbol{Z}) = D_{ig}(\boldsymbol{Z}_g) = D_{ig}(Z_{ig}, \bar{Z}_{ig}) = D_{ig}(Z_{ig})$$

Notation

$C_{ig} = 1$ iff (i, g) is a complier; $\bar{C}_{ig} \equiv$ share of compliers among (i, g)'s neighbors.

$$\boxed{\text{(IOR)} + \text{(1-Sided)} \Rightarrow D_{ig} = C_{ig} \times Z_{ig}}$$

Exclusion Restriction

Notation

- $\mathbf{B}_g$ = random coefficients for everyone in group g.
- $\boldsymbol{C}_g$ = complier indicators for everyone in group g
- $\boldsymbol{Z}_g$ = treatment offers for everyone in group g

Exclusion Restriction

(i) $S_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_g, N_g)$

(ii) $\boldsymbol{Z}_g \perp\!\!\!\perp (\boldsymbol{C}_g, \mathbf{B}_g) | (S_g, N_g)$

IV Does Not Identify Interpretable Causal Effects

For simplicity: Bernoulli Offers

Unoffered Individuals

$$
\begin{aligned}
Y_{ig} &= \alpha_{ig} + \cancel{\beta_{ig} D_{ig}} + \gamma_{ig} \bar{D}_{ig} + \cancel{\delta_{ig} D_{ig} \bar{D}_{ig}} \\
&= \mathbb{E}[\alpha_{ig}] + \mathbb{E}[\gamma_{ig}] \bar{D}_{ig} + (\alpha_{ig} - \mathbb{E}[\alpha_{ig}]) + (\gamma_{ig} - \mathbb{E}[\gamma_{ig}]) \bar{D}_{ig} \\
&= \alpha + \gamma \bar{D}_{ig} + \varepsilon_{ig}
\end{aligned}
$$

IV Estimand

$$
\gamma_{IV} = \frac{\text{Cov}(Y_{ig}, S_g)}{\text{Cov}(\bar{D}_{ig}, S_g)} = \gamma + \frac{\text{Cov}(\varepsilon_{ig}, S_g)}{\text{Cov}(\bar{D}_{ig}, S_g)} = \ldots = \gamma + \frac{\text{Cov}(\gamma_{ig}, \bar{C}_{ig})}{\mathbb{E}(\bar{C}_{ig})}
$$

Identification: $Y_{ig} = \alpha_{ig} + \beta_{ig} D_{ig} + \gamma_{ig} \bar{D}_{ig} + \delta_{ig} D_{ig} \bar{D}_{ig}$

Why not OLS?

Random coefficients $\mathbf{B}_{ig}$ are correlated with $\bar{D}_{ig}$ and D_{ig}.

Why not S_g as IV for $\bar{D}_{ig}$?

Random coefficients $\mathbf{B}_{ig}$ may be correlated with $\bar{C}_{ig}$ (and N_g).

Theorem 1

$(Z_{ig}, \bar{D}_{ig}, S_g) \perp\!\!\!\perp (\mathbf{B}_{ig}, C_{ig}) | (\bar{C}_{ig}, N_g) \implies \bar{D}_{ig}$ is conditionally exogenous given $(\bar{C}_{ig}, N_g)$.

Identification – Unoffered Individuals: $Y_{ig}(0, \bar{D}_{ig}) = \alpha_{ig} + \gamma_{ig}\bar{D}_{ig}$

"Localized" Indirect Effect

$$\mathbb{E}\left\{ \begin{bmatrix} \alpha_{ig} \\ \gamma_{ig} \end{bmatrix} \middle| (\bar{C}_{ig}, N_g) \right\} = \mathbb{E}\left\{ \begin{bmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{bmatrix} \middle| (Z_{ig} = 0, \bar{C}_{ig}, N_g) \right\}^{-1} \mathbb{E}\left\{ \begin{bmatrix} 1 \\ \bar{D}_{ig} \end{bmatrix} Y_{ig} \middle| (Z_{ig} = 0, \bar{C}_{ig}, N_g) \right\}$$

Why?

Use $Y_{ig}(0, \bar{D}_{ig}) = \alpha_{ig} + \gamma_{ig}\bar{D}_{ig}$ and apply Theorem 1 twice:

$$\begin{aligned} \mathbb{E}\left\{ \begin{bmatrix} 1 \\ \bar{D}_{ig} \end{bmatrix} Y_{ig} \middle| (Z_{ig} = 0, \bar{C}_{ig}, N_g) \right\} &= \mathbb{E}\left\{ \begin{bmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{bmatrix} \begin{bmatrix} \alpha_{ig} \\ \gamma_{ig} \end{bmatrix} \middle| (Z_{ig} = 0, \bar{C}_{ig}, N_g) \right\} \\ &= \mathbb{E}\left\{ \begin{bmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{bmatrix} \middle| (Z_{ig} = 0, \bar{C}_{ig}, N_g) \right\} \mathbb{E}\left\{ \begin{bmatrix} \alpha_{ig} \\ \gamma_{ig} \end{bmatrix} \middle| (\bar{C}_{ig}, N_g) \right\} \end{aligned}$$

Identification – Unoffered Individuals: $Y_{ig}(0, \bar{D}_{ig}) = \alpha_{ig} + \gamma_{ig}\bar{D}_{ig}$

Notation

$$\mathbf{Q}_0(\bar{c}, n) \equiv \mathbb{E}\left\{ \begin{bmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{bmatrix} \middle| (Z_{ig} = 0, \bar{C}_{ig} = \bar{c}, N_g = n) \right\}$$

Rank Condition

Identifying local indirect effects requires $\mathbf{Q}_0(\bar{c}, n)$ to be invertible.

Bernoulli Offers

$$\mathbf{Q}_0(\bar{c}, n) = \frac{1}{\mathbb{E}(1 - S_g)} \begin{bmatrix} \mathbb{E}\{1 - S_g\} & \bar{c}\,\mathbb{E}\{S_g(1 - S_g)\} \\ \bar{c}\,\mathbb{E}\{S_g(1 - S_g)\} & \bar{c}^2\,\mathbb{E}\left\{S_g^2(1 - S_g)\right\} + \frac{\bar{c}}{n-1}\mathbb{E}\left\{S_g(1 - S_g)^2\right\} \end{bmatrix}$$

Identification – Unoffered Individuals: $Y_{ig}(0, \bar{D}_{ig}) = \alpha_{ig} + \gamma_{ig}\bar{D}_{ig}$

Localize

$$\begin{bmatrix} \alpha(\bar{c}, n) \\ \gamma(\bar{c}, n) \end{bmatrix} \equiv \mathbb{E}\left\{ \begin{bmatrix} 1 & \bar{D}_{ig} \\ \bar{D}_{ig} & \bar{D}_{ig}^2 \end{bmatrix} \middle| (Z_{ig} = 0, \bar{C}_{ig} = \bar{c}, N_g = n) \right\}^{-1} \mathbb{E}\left\{ \begin{bmatrix} 1 \\ \bar{D}_{ig} \end{bmatrix} Y_{ig} \middle| (Z_{ig} = 0, \bar{C}_{ig} = \bar{c}, N_g = n) \right\}$$

Then Average

$$\begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \equiv \mathbb{E}\begin{bmatrix} \alpha_{ig} \\ \gamma_{ig} \end{bmatrix} = \int \begin{bmatrix} \alpha(\bar{c}, n) \\ \gamma(\bar{c}, n) \end{bmatrix} dF(\bar{c}, n)$$

Identification Results

- Point identify average direct and indirect causal effects for compliers.
- Point identify average indirect effects for never-takers and population.
- Design should feature as many interior saturations as basis functions in potential outcome equations.

Estimation – Unoffered Individuals: Let $\mathbf{X}'_{ig} = (1, \bar{D}_{ig})$

Estimate Share of Compliers: $\widehat{C}_g = \bar{D}_g / \bar{Z}_g$

Localize

$$\begin{bmatrix} \widehat{\alpha}(\bar{c}) \\ \widehat{\gamma}(\bar{c}) \end{bmatrix} = \left[\sum_{g=1}^{G} \sum_{i=1}^{N_g} K_h(\widehat{C}_g - \bar{c})(1 - Z_{ig}) \mathbf{X}_{ig} \mathbf{X}'_{ig} \right]^{-1} \left[\sum_{g=1}^{G} \sum_{i=1}^{N_g} K_h(\widehat{C}_g - \bar{c})(1 - Z_{ig}) \mathbf{X}_{ig} Y_{ig} \right]$$

Then Average

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\gamma} \end{bmatrix} = \frac{1}{G} \sum_{j=1}^{G} \left(\frac{N_j}{\bar{N}} \right) \begin{bmatrix} \widehat{\alpha}(\widehat{C}_j) \\ \widehat{\gamma}(\widehat{C}_j) \end{bmatrix}$$

Application – Crepon et al. (2013; QJE)

Randomized Job Placement Assistance

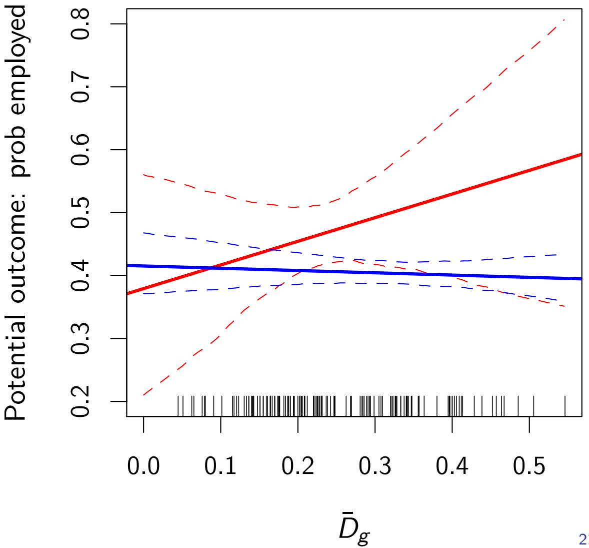
- $Y_{ig} = 1$ if employed after 8 months
- $D_{ig} =$ take-up
- $Z_{ig} =$ offer
- $G = 235$ cities
- $N = 11,806$ unemployed
- $\mathcal{S} = \{0\%, 25\%, 50\%, 75\%, 100\%\}$
- 1-sided non-compliance

Selected Results

1. No evidence that IOR is violated.
2. Indirect Effects for Compliers:
   - None if treated
   - Negative if untreated
3. "Naïve" IV gives very strange results.
4. Other Results
   - No indirect effects for never-takers
   - Similar results with $Y_{ig} =$ income

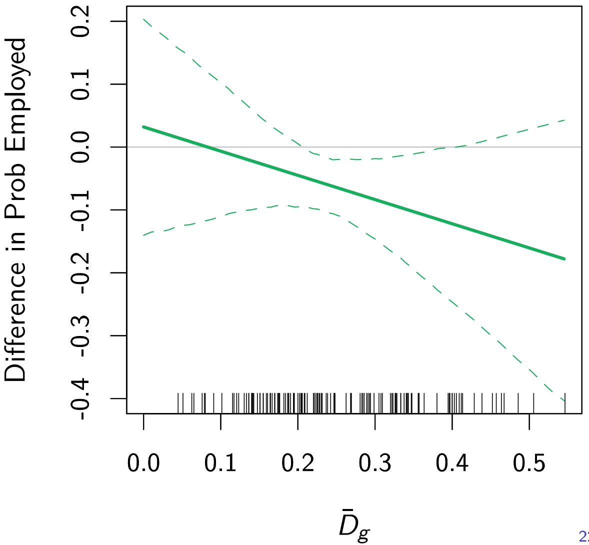
Long-Term Employment – Compliers when Treated: $(\alpha + \beta) + (\gamma + \delta)\bar{D}_g$

- IV Estimator
- Our Estimator
- 90% Bootstrap CIs



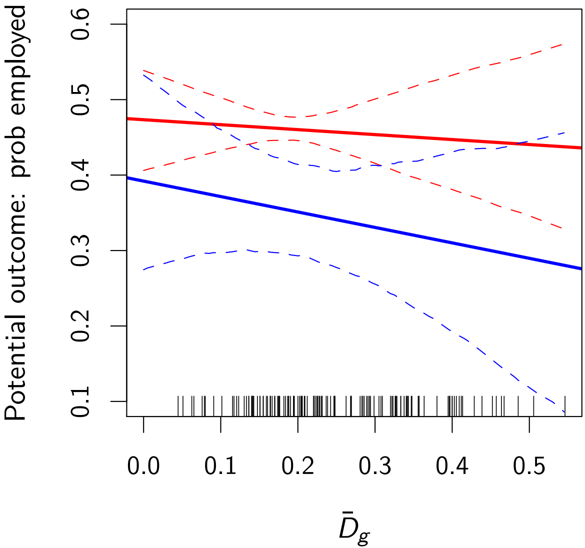
Long-Term Employment – Compliers when Treated

- Our Estimator - IV Estimator
- 90% Bootstrap CIs



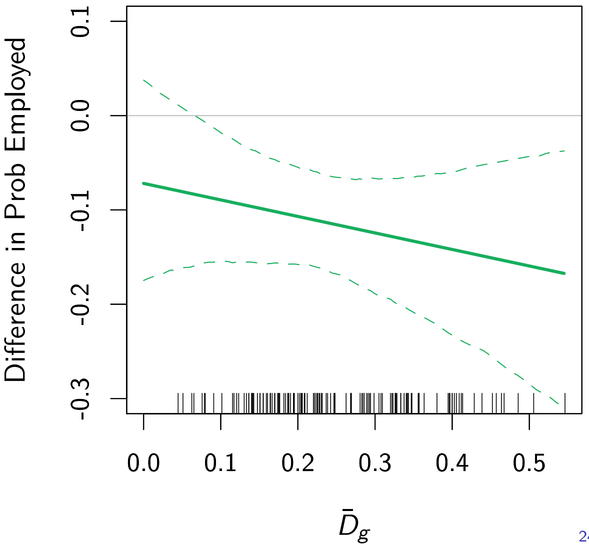
Long-Term Employment – Compliers when Untreated: $\alpha + \gamma \bar{D}_g$

- IV Estimator
- Our Estimator
- 90% Bootstrap CIs



Long-Term Employment: Compliers when Untreated

- Our Estimator - IV Estimator
- 90% Bootstrap CIs



Conclusion

1. Randomized saturation experiments with one-sided non-compliance.
2. Correlated random coefficients model to allow for unobserved heterogeneity.
3. Identify avg. direct and indirect effects for the treated; indirect for the untreated.
4. Consistent estimation under large-groups/many-groups asymptotics.
5. Apply to the Crepon et al (2013) experiment.

A Simple Testable Implication of IOR

Recall: IOR Assumption

Person (i, g)'s take-up D_{ig} depends only on her *own* treatment offer Z_{ig}.

A "Regression-based" Test

$\mathbb{E}[D_{ig}|Z_{ig} = 1, S_g = s]$ should not depend on s.

Implementation

- Regress D_{ig} on an intercept and saturation dummies for those offered treatment.
- Test the joint null that all saturation dummies are zero.
- Equivalently test that share of compliers is the same across saturation "bins"
- Crepon Experiment: p-value of 0.62

A Simple Testable Implication of IOR