Policy Targeting under Network Interference

Davide Viviano
Stanford GSB/Harvard Econ

March 29, 2023

Example

Ex Informing farmers exposed to environmental disasters to increase insurance take-up in rural China

Example

Ex Informing farmers exposed to environmental disasters to increase insurance take-up in rural China

How should we design information campaigns?

$\Rightarrow$ Choosing whom to treat:

(i) Researchers sampled ~ 185 villages in rural China and collected network information in these villages (Cai et al., 2015)

(ii) Documented spillovers in an experiment: information diffuse to friends

(iii) Treatment can be costly: treating each individual is sub-optimal

Policy Targeting

Suppose we collect data from experiment or quasi-experiment participants sampled from a large population

How do we *design* an allocation rule (policy function) that determines the treatment for the entire population?

Key features:

- Heterogeneity in treatment effects
- Network interference
- Treatment may be costly

Constraints may apply:

- Network information may not be observable in other villages
- Ethical, legal or computational constraints may apply

Other Examples

⇒ 40% of experimental paper published in 2020 top-5 mention spillovers

General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya

Dennis Egger, Johannes Haushofer, Edward Miguel, Paul Niehaus & Michael W. Walker

WORKING PAPER 26600 **DOI** 10.3386/w26600 **ISSUE DATE** December 2019

How large economic stimuli generate individual and aggregate responses is a central question in economics, but has not been studied experimentally. We provided one-time cash transfers of about USD 1000 to over 10,500 poor households across 653 randomized villages in rural Kenya. The implied fiscal shock was over 15 percent of local GDP. We find large impacts on consumption and assets for recipients. Importantly, we document large positive spillovers on non-recipient households and firms, and minimal price inflation. We estimate a local fiscal multiplier of 2.7. We interpret welfare implications through the lens of a simple household optimization framework.

Other Examples

⇒ 40% of experimental paper published in 2020 top-5 mention spillovers

General Eq
Exper

'S:

Den

Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial†

By Jessica Cohen, Pascaline Dupas, and Simone Schaner*

Both under- and over-treatment of communicable diseases are public bads. But efforts to decrease one run the risk of increasing the other. Using rich experimental data on household treatment-seeking behavior in Kenya, we study the implications of this trade-off for subsidizing life-saving antimalarials sold over-the-counter at retail drug outlets. We show that a very high subsidy (such as the one under consideration by the international community) dramatically increases access, but nearly one-half of subsidized pills go to patients without malaria. We study two ways to better target subsidized drugs: reducing the subsidy level, and introducing rapid malaria tests over-the-counter. (*JEL* D12, D82, I12, O12, O15)

How large economic stimuli generate individual and aggregate responses is a central question in economics, but has not been studied experimentally. We provided one-time cash transfers of about USD 1000 to over 10,500 poor households across 653 randomized villages in rural Kenya. The implied fiscal shock was over 15 percent of local GDP. We find large impacts on consumption and assets for recipients. Importantly, we document large positive spillovers on non-recipient households and firms, and minimal price inflation. We estimate a local fiscal multiplier of 2.7. We interpret welfare implications through the lens of a simple household optimization framework.

Other Examples

$\Rightarrow$ 40% of experimental paper published in 2020 top-5 mention spillovers

General Eq

Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial†

's:

Tagging and Targeting of Energy Efficiency Subsidies

Hunt Allcott
Christopher Knittel
Dmitry Taubinsky

AMERICAN ECONOMIC REVIEW
VOL. 105, NO. 5, MAY 2015
(pp. 187-91)

Download Full Text PDF

Article Information

Abstract

A corrective tax or subsidy is "well-targeted" if it primarily affects choices that are more distorted by market failures. Energy efficiency subsidies are designed to correct multiple distortions: externalities, credit constraints, "landlord-tenant" information asymmetries, imperfect information, and inattention. We show that three important energy efficiency subsidies are primarily taken up by consumers who are wealthier, own their own homes, and are more informed about and attentive to energy costs. This suggests that these subsidies are poorly targeted at the market failures they were designed to address. However, we show that "tagging" can lead to large efficiency gains.

'AS, AND SIMONE SCHANER*

mmunicable diseases are pub-
run the risk of increasing the
n household treatment-seeking
ations of this trade-off for sub-
over-the-counter at retail drug
idy (such as the one under con-
unity) dramatically increases
lized pills go to patients with-
etter target subsidized drugs:
roducing rapid malaria tests
, O12, O15)

ponses is a central question in
one-time cash transfers of about
l villages in rural Kenya. The
e impacts on consumption and
assets for recipients. Importantly, we document large positive spillovers on non-recipient households and firms, and minimal price inflation. We estimate a local fiscal multiplier of 2.7. We interpret welfare implications through the lens of a simple household optimization framework.

Other Examples

⇒ 40% of experimental paper published in 2020 top-5 mention spillovers

General Eq

Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial†

:S:

Tagging and Targeting of Energy Efficiency

AS, AND SIMONE SCHANER*

Published: 12 September 2012

A 61-million-person experiment in social influence and political mobilization

Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler

Nature **489**, 295–298(2012) | Cite this article

18k Accesses | **1056** Citations | **1853** Altmetric | Metrics

more informed about and attentive to energy costs. This suggests that these subsidies are poorly targeted at the market failures they were designed to address. However, we show that "tagging" can lead to large efficiency gains.

l villages in rural Kenya. The e impacts on consumption and assets for recipients. Importantly, we document large positive spillovers on non-recipient households and firms, and minimal price inflation. We estimate a local fiscal multiplier of 2.7. We interpret welfare implications through the lens of a simple household optimization framework.

Problem Description

- Finite pop: n units connected under adjacency matrix A (unobserved)
- RCT: collect $n_e \ll n$ individual level info
  - $R_i \in \{0, 1\}$: sampling indicator

Problem Description

- Finite pop: n units connected under adjacency matrix A (unobserved)
- RCT: collect $n_e \ll n$ individual level info
  - $R_i \in \{0, 1\}$: sampling indicator
  - $D_i \in \{0, 1\}$: randomized treatment for sampled units
  - $Y_i \in \mathbb{R}$: outcome of interest for sampled units
  - $Z_i := (X_i, \tilde{X}_i) \in \mathcal{Z}, X_i \in \mathcal{X}$: covariates for sampled units

Problem Description

- Finite pop: n units connected under adjacency matrix A (unobserved)
- RCT: collect $n_e \ll n$ individual level info
  - $R_i \in \{0,1\}$: sampling indicator
  - $D_i \in \{0,1\}$: randomized treatment for sampled units
  - $Y_i \in \mathbb{R}$: outcome of interest for sampled units
  - $Z_i := (X_i, \tilde{X}_i) \in \mathcal{Z}, X_i \in \mathcal{X}$: covariates for sampled units
- Collect sampled units' neighbors' info
  - $\mathcal{N}_i = \{j : A_{i,j} = 1\}$: set of neighbors of each individual
  - $R_i^f = 1\{\sum_{k \in \mathcal{N}_i} R_k > 0\}$: neighbors' sampling indicator
  - $(D_{\mathcal{N}_i}, Z_{\mathcal{N}_i})$: neighbors' covariates and assignments.

Problem Description

- Finite pop: n units connected under adjacency matrix A (unobserved)
- RCT: collect $n_e \ll n$ individual level info
  - $R_i \in \{0,1\}$: sampling indicator
  - $D_i \in \{0,1\}$: randomized treatment for sampled units
  - $Y_i \in \mathbb{R}$: outcome of interest for sampled units
  - $Z_i := (X_i, \tilde{X}_i) \in \mathcal{Z}, X_i \in \mathcal{X}$: covariates for sampled units
- Collect sampled units' neighbors' info
  - $\mathcal{N}_i = \{j : A_{i,j} = 1\}$: set of neighbors of each individual
  - $R_i^f = 1\{\sum_{k \in \mathcal{N}_i} R_k > 0\}$: neighbors' sampling indicator
  - $(D_{\mathcal{N}_i}, Z_{\mathcal{N}_i})$: neighbors' covariates and assignments.

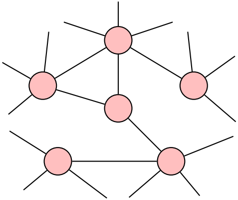
$\Rightarrow$ Construct a targeting rule (policy)

$$\pi : \mathcal{X} \mapsto \{0,1\}, \quad \pi \in \Pi$$

that determines whom should be treated.

Policy Choice: Illustration

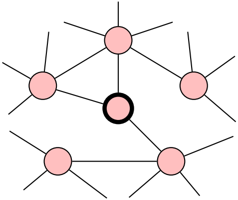
Policy Choice: Illustration



Policy Choice: Illustration

Experiment → Policy → Targeting

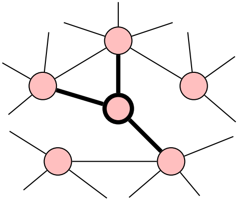
$R_i \sim_{i.i.d.} \mathrm{Bern}(n_e/n)$



Policy Choice: Illustration

Experiment → Policy → Targeting

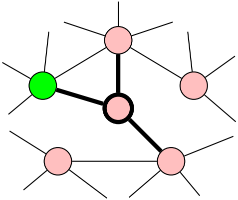
$$R_i \sim_{i.i.d.} \mathrm{Bern}(n_e/n)$$



Policy Choice: Illustration

Experiment → Policy → Targeting

$$D_i | Z_i, R_i, R_i^f \sim \mathcal{P}(Z_i, R_i, R_i^f)$$

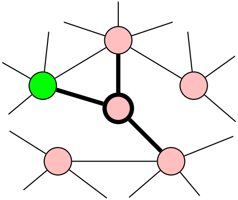


$$\Big[(Y_i, Z_i, Z_{\mathcal{N}_i}, D_i, D_{\mathcal{N}_i})R_i, R_i\Big]_{i=1}^{n}$$

Policy Choice: Illustration

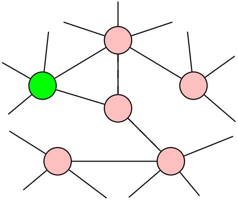
Experiment → Policy → Targeting

$$D_i | Z_i, R_i, R_i^f \sim \mathcal{P}(Z_i, R_i, R_i^f)$$



$$\Big[(Y_i, Z_i, Z_{\mathcal{N}_i}, D_i, D_{\mathcal{N}_i})R_i, R_i\Big]_{i=1}^n$$

$$\pi(X_i)$$

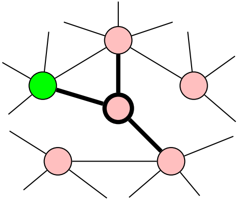


$$(X_i)_{i=1}^n \subseteq Z$$

Policy Choice: Illustration

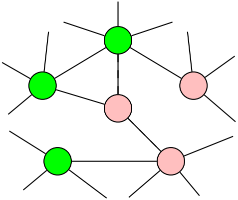
Experiment → Policy → Targeting

$$D_i | Z_i, R_i, R_i^f \sim \mathcal{P}(Z_i, R_i, R_i^f)$$



$$\Big[(Y_i, Z_i, Z_{\mathcal{N}_i}, D_i, D_{\mathcal{N}_i})R_i, R_i\Big]_{i=1}^n$$

$$\pi(X_i)$$



$$(X_i)_{i=1}^n \subseteq Z$$

Illustrative Example

Experiment:

$$Y_i = D_i \times X_i \times \gamma_1 + \sum_{k \in \mathcal{N}_i} D_k \times \gamma_2 + \varepsilon_i .$$

Illustrative Example

Experiment:

$$Y_i = D_i \times X_i \times \gamma_1 + \sum_{k \in \mathcal{N}_i} D_k \times \gamma_2 + \varepsilon_i.$$

The "oracle" method maximizes the empirical welfare ▸ Example

$$W(\pi) = \frac{1}{n} \sum_{i=1}^{n} \Big(\pi(X_i) \times X_i \times \gamma_1 + \sum_{k \in \mathcal{N}_i} \pi(X_k) \times \gamma_2 - c \times \pi(X_i) \Big).$$

Illustrative Example

Experiment:

$$Y_i = D_i \times X_i \times \gamma_1 + \sum_{k \in \mathcal{N}_i} D_k \times \gamma_2 + \varepsilon_i.$$

The "oracle" method maximizes the empirical welfare ▸ Example

$$W(\pi) = \frac{1}{n} \sum_{i=1}^{n} \Big(\pi(X_i) \times X_i \times \gamma_1 + \sum_{k \in \mathcal{N}_i} \pi(X_k) \times \gamma_2 - c \times \pi(X_i) \Big).$$

In practice (and in the paper):

- Unknown dependence with neighbors' treatments and heterogeneity
- Constraints $\pi \in \Pi$
- Policy constraints that depend on treatments in experiment D

Contributions

- Extend treatment choice literature (Manski, 2004) to spillover effects
  - Identify welfare, leveraging sampling variation as in Abadie et al., 2020
  - Semi-parametric (and cross-fitting) procedures with spillovers
  - Mixed-integer linear program formulation for optimization
- Regret analysis with interference $\sup_{\pi \in \Pi} W(\pi) - W(\hat{\pi})$
  - (Minimax) rate of convergence as function of n_e and maximum degree
  - Symmetrization for non-asymptotic regret bounds with interference + bounds on Rademacher complexity with spillovers

Main conditions

- Spillovers are local
- Policies depend on (arbitrary) individual characterics (it allows for unobserved A)

Literature

- Statistical Treatment Choice [E.g., Manski (2004); Kitagawa and Tetenov (2018, 2019); Athey and Wager (2021); Zhou et al. (2018); ...]
  - $\Rightarrow$ None study the problem with interference
- Inference under interference: [E.g., Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); ...]
  - $\Rightarrow$ None study statistical treatment choice

Literature

- Statistical Treatment Choice [E.g., Manski (2004); Kitagawa and Tetenov (2018, 2019); Athey and Wager (2021); Zhou et al. (2018); ...]
  - $\Rightarrow$ None study the problem with interference
- Inference under interference: [E.g., Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); ...]
  - $\Rightarrow$ None study statistical treatment choice
- Welfare analysis with social interactions + Seeding [E.g., E.g., Galeotti et al. (2020); Akbarpour et al. (2018); Banerjee et al. (2013); Banerjee et al. (2018); Su et al. (2019); Bhattacharya et al. (2019); Graham et al. (2010); ...]
  - $\Rightarrow$ I study choosing whom to treat as a statistical treatment choice
  - $\Rightarrow$ Here, heterogeneity and constraints on policy space (+ regret analysis)

Content

Sampling and experiment

- Pop: n units with fixed network A and characteristics $Z = (Z_i)_{i=1}^n$

Sampling and experiment

- Pop: n units with fixed network A and characteristics $Z = (Z_i)_{i=1}^n$
- Sampling indicators: $R_i | A, Z \sim_{i.i.d.} \mathrm{Bern}(n_e/n)$
- Assignments:

$$D_i = f\left(Z_i, R_i, (1 - R_i) 1\Big\{\sum_{k \in \mathcal{N}_i} R_k > 0\Big\}, \varepsilon_{D_i}\right)$$

where $\varepsilon_{D_i} \sim_{i.i.d.} \mathcal{D}$ and exogenous, and f either known or unknown.

Sampling and experiment

- Pop: n units with fixed network A and characteristics $Z = (Z_i)_{i=1}^n$
- Sampling indicators: $R_i | A, Z \sim_{i.i.d.} \mathrm{Bern}(n_e/n)$
- Assignments:

$$D_i = f\left(Z_i, R_i, (1 - R_i)1\Big\{\sum_{k \in \mathcal{N}_i} R_k > 0\Big\}, \varepsilon_{D_i}\right)$$

where $\varepsilon_{D_i} \sim_{i.i.d.} \mathcal{D}$ and exogenous, and f either known or unknown.

$\Rightarrow$ Example

- Randomly select a subset of individuals from population
- Randomize treatments to such individuals and their friends
- Treatments for the remaining units equal to baseline $D_i = 0$.

$\Rightarrow$ Extension with locally dependent R_i possible (e.g., sample small villages first)

Outcomes

$$Y_i = r\Big(D_i, T_i, Z_i, |\mathcal{N}_i|, \varepsilon_i\Big), \quad T_i = g_n\Big(\sum_{k \in \mathcal{N}_i} D_k, Z_i, |\mathcal{N}_i|\Big)$$

for $T_i \in \mathcal{T}_n$, $r(\cdot)$ unknown and $g_n(\cdot)$ known

Outcomes

$$Y_i = r\Big(D_i, T_i, Z_i, |\mathcal{N}_i|, \varepsilon_i\Big), \quad T_i = g_n\Big(\sum_{k\in\mathcal{N}_i} D_k, Z_i, |\mathcal{N}_i|\Big)$$

for $T_i \in \mathcal{T}_n$, $r(\cdot)$ unknown and $g_n(\cdot)$ known

Interference

(i) Anonymous and exogenous interference ($r(\cdot)$ is unknown)

(ii) $g_n(\cdot)$ is exposure mapping (Aronow and Samii, 2017);

(iii) $g_n(\cdot)$ control treatments overlap, most agnostic is $T_i = \sum_{k\in\mathcal{N}_i} D_k$

Outcomes

$$Y_i = r\Big(D_i, T_i, Z_i, |\mathcal{N}_i|, \varepsilon_i\Big), \quad T_i = g_n\Big(\sum_{k \in \mathcal{N}_i} D_k, Z_i, |\mathcal{N}_i|\Big)$$

for $T_i \in \mathcal{T}_n$, $r(\cdot)$ unknown and $g_n(\cdot)$ known

Interference

(i) Anonymous and exogenous interference ($r(\cdot)$ is unknown)

(ii) $g_n(\cdot)$ is exposure mapping (Aronow and Samii, 2017);

(iii) $g_n(\cdot)$ control treatments overlap, most agnostic is $T_i = \sum_{k \in \mathcal{N}_i} D_k$

Unobservables

(A) Unconfoundedness: $(\varepsilon_i)_{i=1}^n \perp (\varepsilon_{D_i}, R_i)_{i=1}^n | A, Z$

(B) Network exogeneity $\varepsilon_i | A, Z \sim \mathcal{E}_{|\mathcal{N}_i|, Z_i}$ (not necessary)

(C) $\varepsilon_i | A, Z$ dependent for one/two-degree neighbors only

E.g. $\varepsilon_i = \Big(\eta_i, \sum_{k \in \mathcal{N}_i} \eta_k\Big)$ for exogenous $i.i.d.$ η_i

Policy Targeting

- Limited (network) info: policymaker only observes a subset of covariates X_i for all $i \in \{1, \cdots, n\}$

Policy Targeting

- Limited (network) info: policymaker only observes a subset of covariates X_i for all $i \in \{1, \cdots, n\}$

- Welfare: She implements policy and generates welfare

$$W_{A,Z}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[r\Big(\pi(X_i), T_i(\pi), Z_i, |N_i|, \varepsilon_i \Big) \Big| A, Z \right],$$

where $T_i(\pi) = g_n\Big(\sum_{k \in \mathcal{N}_i} \pi(X_i), Z_i, |\mathcal{N}_i| \Big)$.

Policy Targeting

- Limited (network) info: policymaker only observes a subset of covariates X_i for all $i \in \{1, \cdots, n\}$

- Welfare: She implements policy and generates welfare

$$W_{A,Z}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[r\Big(\pi(X_i), T_i(\pi), Z_i, |N_i|, \varepsilon_i \Big) \Big| A, Z \right],$$

where $T_i(\pi) = g_n\Big(\sum_{k \in \mathcal{N}_i} \pi(X_i), Z_i, |\mathcal{N}_i| \Big)$.

- Comments
  - No assumptions on X_i (might depend on network data if available)
  - Welfare defn implies no carry-overs from previous experiment
  - Policies $\pi(D_i, X_i) = D_i + (1 - D_i)\pi(X_i)$ are possible (Appendix B)

Estimands

Defn Conditional Mean

$$m(d, t, z, l) = \mathbb{E}\Big[r(d, t, Z_i, |N_i|, \varepsilon_i)\Big| Z_i = z, |N_i| = l\Big]$$

Estimands

Defn Conditional Mean

$$m(d, t, z, l) = \mathbb{E}\Big[r(d, t, Z_i, |N_i|, \varepsilon_i)\Big| Z_i = z, |N_i| = l\Big]$$

Defn Propensity Score:

$$e(d, t, \mathbf{x}, z, \mathbf{r}, l) = P\Big(D_i, T_i = d, t \Big| (Z_i, Z_{k \in N_i}, R_{k \in N_i}, |N_i|) = (z, \mathbf{x}, \mathbf{r}, l), R_i = 1\Big)$$

$\Rightarrow$ Only function of marginal probabilities $P(D_i = 1 | R_i, R_i^f, Z_i)$

Estimands

Defn Conditional Mean

$$m(d, t, z, l) = \mathbb{E}\Big[r(d, t, Z_i, |N_i|, \varepsilon_i)\Big| Z_i = z, |N_i| = l\Big]$$

Defn Propensity Score:

$$e(d, t, \mathbf{x}, z, \mathbf{r}, l) = P\Big(D_i, T_i = d, t \Big| (Z_i, Z_{k\in N_i}, R_{k\in N_i}, |N_i|) = (z, \mathbf{x}, \mathbf{r}, l), R_i = 1\Big)$$

$\Rightarrow$ Only function of marginal probabilities $P(D_i = 1|R_i, R_i^f, Z_i)$

$\Rightarrow$ Immediate identification strategy with sampling based uncertainty

$$W_{A,Z}(\pi) = \frac{1}{n_e}\sum_{i=1}^{n} \mathbb{E}\left[R_i Y_i \frac{I_i(\pi)}{e_i(\pi)}\Big| A, Z\right],$$

where $\frac{I_i(\pi)}{e_i(\pi)} = \frac{1\{D_i=\pi(X_i), T_i(\pi)=T_i\}}{e(\pi(X_i), T_i(\pi), Z_{k\in N_i}, R_{k\in N_i}, Z_i, |N_i|)}$

Estimands

Defn Conditional Mean

$$m(d, t, z, l) = \mathbb{E}\Big[r(d, t, Z_i, |N_i|, \varepsilon_i)\Big| Z_i = z, |N_i| = l\Big]$$

Defn Propensity Score:

$$e(d, t, \mathbf{x}, z, \mathbf{r}, l) = P\Big(D_i, T_i = d, t \Big| (Z_i, Z_{k\in N_i}, R_{k\in N_i}, |N_i|) = (z, \mathbf{x}, \mathbf{r}, l), R_i = 1\Big)$$

$\Rightarrow$ Only function of marginal probabilities $P(D_i = 1 | R_i, R_i^f, Z_i)$

$\Rightarrow$ Immediate identification strategy with sampling based uncertainty

$$W_{A,Z}(\pi) = \frac{1}{n_e} \sum_{i=1}^{n} \mathbb{E}\left[R_i Y_i \frac{I_i(\pi)}{e_i(\pi)} \Big| A, Z\right],$$

where $\frac{I_i(\pi)}{e_i(\pi)} = \frac{1\{D_i = \pi(X_i), T_i(\pi) = T_i\}}{e(\pi(X_i), T_i(\pi), Z_{k\in N_i}, R_{k\in N_i}, Z_i, |N_i|)}$

$\Rightarrow$ No distributional assumptions on (A, Z) + no need of identically distributed ε_i if known propensity score

Digression: Overlap and Network Density Matter

Asm Positive overlap on the propensity score

$$e_i(\pi) \in (\delta_n, 1-\delta_n), \quad \text{for } \delta_n \in (0,1), \pi \in \Pi.$$

- $\Rightarrow$ Lack of overlap would require extrapolation
- $\Rightarrow$ Condition depends on the support of T_i (exposure mapping)
- $\Rightarrow$ Later we will study trimming strategies

Digression: Overlap and Network Density Matter

Asm Positive overlap on the propensity score

$$e_i(\pi) \in (\delta_n, 1 - \delta_n), \quad \text{for } \delta_n \in (0,1), \pi \in \Pi.$$

- $\Rightarrow$ Lack of overlap would require extrapolation
- $\Rightarrow$ Condition depends on the support of T_i (exposure mapping)
- $\Rightarrow$ Later we will study trimming strategies

Asm Network cannot be "too dense",

$$\mathcal{N}_{n,\max}^{3/2} \log(\mathcal{N}_{n,\max})/(n_e^{1/2}\delta_n) = o(1)$$

where $\mathcal{N}_{n,\max} = \max_i |\mathcal{N}_i|$.

- $\Rightarrow$ Control dependence via largest number of connections
- $\Rightarrow$ Can be relaxed using the chromatic number

Content

Estimation with known propensity score

- (Augmented) Inverse probability weighting (e.g., A&S, 2017):

$$W_n(\pi, m^c, e) = \frac{1}{n_e} \sum_{i=1}^{n} R_i \left(\frac{I_i(\pi)}{e_i(\pi)} \Big(Y_i - m_i^c(\pi) \Big) + m_i^c(\pi) \right),$$

where $m_i^c(\pi)$ is a reg adjustment for D_i, $T_i = \pi(X_i)$, $T_i(\pi)$.

Estimation with known propensity score

- (Augmented) Inverse probability weighting (e.g., A&S, 2017):

$$W_n(\pi, m^c, e) = \frac{1}{n_e} \sum_{i=1}^{n} R_i \left(\frac{I_i(\pi)}{e_i(\pi)} \Big(Y_i - m_i^c(\pi) \Big) + m_i^c(\pi) \right),$$

  where $m_i^c(\pi)$ is a reg adjustment for D_i, $T_i = \pi(X_i)$, $T_i(\pi)$.

- Optimization: estimate $\hat{\pi}_{m^c,e}$ by maximizing $W_n(\pi, m^c, e)$
  $\Rightarrow$ show that optimization admits a mixed-integer linear program for a large class of function classes Π

Regret bound: known propensity

Theorem 3.1

For Y_i having bounded third moment given (A, Z)

$$\mathbb{E}\left[\sup_{\pi \in \Pi} W(\pi) - W(\hat{\pi}_{m^c,e})\right] = \mathcal{O}\left(\frac{1}{\delta_n}\mathcal{N}_{n,\max}\sqrt{\frac{\mathcal{N}_{n,\max}\mathrm{VC}(\Pi)\log(\mathcal{N}_{n,\max})}{n_e}}\right).$$

Regret bound: known propensity

Theorem 3.1

For Y_i having bounded third moment given (A, Z)

$$\mathbb{E}\left[\sup_{\pi \in \Pi} W(\pi) - W(\hat{\pi}_{m^c,e})\right] = \mathcal{O}\left(\frac{1}{\delta_n}\mathcal{N}_{n,\max}\sqrt{\frac{\mathcal{N}_{n,\max}\mathrm{VC}(\Pi)\log(\mathcal{N}_{n,\max})}{n_e}}\right).$$

- No restrictions on n_e relative to n (valid for $n \gg n_e$)
- Maximum degree controls dependence and function class' complexity
  - $\Rightarrow$ Symmetrization for non-asymptotic regret bounds under interference
  - $\Rightarrow$ Upper bounds on the Rademacher complexity obtained from compositions of functions.
- Bound is informative only under "enough" sparsity

Thm Rate in Thm 3.1 is maximin optimal with respect to n_e

Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators ($n^{-1/2}$);
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators $(n^{-1/2})$;
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Alg: Network Cross-fitting

- Partition sampled units into groups such that in each group two sampled units are not friends or have no shared friend
- Estimation via standard cross-fitting within each group

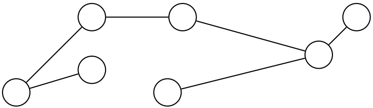
Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators $(n^{-1/2})$;
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Alg: Network Cross-fitting

- Partition sampled units into groups such that in each group two sampled units are not friends or have no shared friend
- Estimation via standard cross-fitting within each group

Thm Estimation does not affect rate in regret bound if nuisance functions converge at rate $\mathcal{N}_{n,\max}^{-1/2} n^{-1/4}$ on $i.i.d.$ data



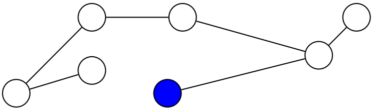
Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators $(n^{-1/2})$;
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Alg: Network Cross-fitting

- Partition sampled units into groups such that in each group two sampled units are not friends or have no shared friend
- Estimation via standard cross-fitting within each group

Thm Estimation does not affect rate in regret bound if nuisance functions converge at rate $\mathcal{N}_{n,\max}^{-1/2} n^{-1/4}$ on $i.i.d.$ data



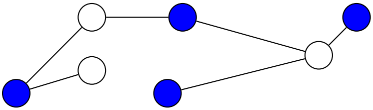
Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators $(n^{-1/2})$;
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Alg: Network Cross-fitting

- Partition sampled units into groups such that in each group two sampled units are not friends or have no shared friend
- Estimation via standard cross-fitting within each group

Thm Estimation does not affect rate in regret bound if nuisance functions converge at rate $\mathcal{N}_{n,\max}^{-1/2} n^{-1/4}$ on *i.i.d.* data



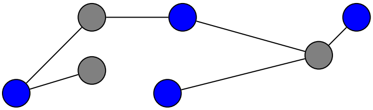
Estimation with Generic ML Algorithms

- Rate characterization also depends on convergence rate of $\hat{m}, \hat{e}$:
  - $\Rightarrow$ Minimax regret rate achieved with parametric estimators $(n^{-1/2})$;
- Can we improve the rate of convergence for semiparametric ones?
  - $\Rightarrow$ Network dependence invalidates standard cross-fitting (ML) estimators

Alg: Network Cross-fitting

- Partition sampled units into groups such that in each group two sampled units are not friends or have no shared friend
- Estimation via standard cross-fitting within each group

Thm Estimation does not affect rate in regret bound if nuisance functions converge at rate $\mathcal{N}_{n,\max}^{-1/2} n^{-1/4}$ on $i.i.d.$ data



Regret with estimation error

$$\mathcal{A}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\sup_{d,s}\Big(\hat{m}_i(d,s,Z_i,|N_i|) - m(d,s,Z_i,|N_i|)\Big)^2\Big|R_i = 1, A, Z\Big]$$

$$\mathcal{B}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\sup_{d,s}\Big(\frac{1}{\hat{e}_i(d,s,Z_{k\in N_i},Z_i,|N_i|)} - \frac{1}{e(d,s,Z_{k\in N_i},Z_i,|N_i|)}\Big)^2\Big|R_i = 1, A, Z\Big]$$

Regret with estimation error

$$\mathcal{A}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\sup_{d,s}\Big(\hat{m}_i(d,s,Z_i,|N_i|) - m(d,s,Z_i,|N_i|)\Big)^2\Big|R_i = 1, A, Z\Big]$$

$$\mathcal{B}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\sup_{d,s}\Big(\frac{1}{\hat{e}_i(d,s,Z_{k\in N_i},Z_i,|N_i|)} - \frac{1}{e(d,s,Z_{k\in N_i},Z_i,|N_i|)}\Big)^2\Big|R_i = 1, A, Z\Big]$$

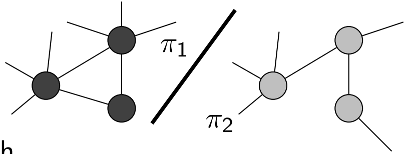
Theorem 3.3

For Y_i having bounded third moment given (A, Z)

$$\mathbb{E}\left[\sup_{\pi\in\Pi} W(\pi) - W(\hat{\pi}_{m^c,e})\right] = \mathcal{O}\left(\frac{1}{\delta_n}\sqrt{\frac{\mathcal{N}^3_{n,\max}\mathrm{VC}(\Pi)\log(\mathcal{N}_{n,\max})}{n_e}} + \underbrace{\sqrt{\mathcal{A}_n\mathcal{B}_n}}_{\mathrm{EstErr}_n}\right),$$

where $\mathrm{EstErr}_n = \mathcal{O}(\mathcal{N}^2_{n,\max} n_e^{-\zeta}/\delta_n)$, and $n_e^{-\zeta}$ is convergence rate of nuisance functs on sample of i.i.d. data.

Practical considerations

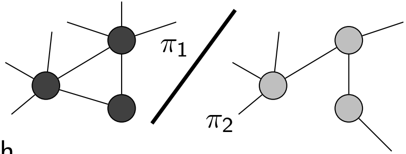


Approximate optimization

- Find (K) maximum cuts
- Estimate $\hat{\pi}_1, \hat{\pi}_2$ over each sub-graph
- Choose $\hat{\pi}_1, \hat{\pi}_2$ with largest empirical welfare on whole sample
  - $\Rightarrow$ Exponential worst-case improvement in K for comp complexity

Practical considerations

Approximate optimization



- Find (K) maximum cuts
- Estimate $\hat{\pi}_1, \hat{\pi}_2$ over each sub-graph
- Choose $\hat{\pi}_1, \hat{\pi}_2$ with largest empirical welfare on whole sample
  - $\Rightarrow$ Exponential worst-case improvement in K for comp complexity

Approximate Estimation

- Find K maximum cuts
- Estimate $\hat{m}_i, \hat{e}_i$ using information from all except cut of i
  - $\Rightarrow$ Same regret guarantees if K separated subgraphs exist

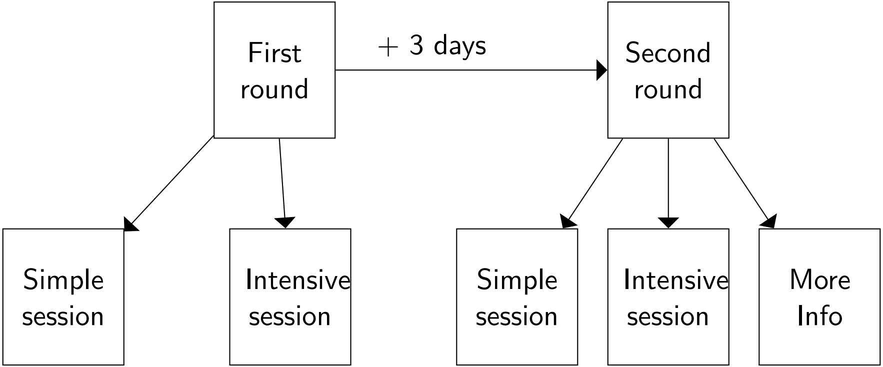
Content

Cai et al. (2015): Sampling

- Randomly select 185 villages (from ~ 50 areas) in the experiment
- Observe network information from $> 90\%$ of participants
- Focus on individuals for which we also observe neighbors' treatments and covariates (~ 4000 obs)

- Less than 50% of connections are in same village, but more than 99% of connections are in same area
- Two-stage design allows tests for endogenous spillovers

Cai et al. 2015: Design



⇒ Focus on simple vs intensive session

⇒ Follow Cai et al (2015)'s main model specification, also controlling for education, rice area and risk aversion heterogeneity

Out-of-sample comparisons

Table: *Out-of-sample* welfare improvement for a classification tree upon empirical welfare-maximization targeting rule in Athey and Wager (2020) that does not account for network effects in the design of the policy.

	Educ & Rice-ar	Educ & Risk-av	Rice-ar & Risk-av
$C = 1\%$	0.146	0.084	0.289
$C = 3\%$	0.159	0.093	0.201
$C = 5\%$	0.093	0.111	0.143

Out-of-sample comparisons

Table: *Out-of-sample* welfare improvement for a classification tree upon empirical welfare-maximization targeting rule in Athey and Wager (2020) that does not account for network effects in the design of the policy.

	Educ & Rice-ar	Educ & Risk-av	Rice-ar & Risk-av
$C = 1\%$	0.146	0.084	0.289
$C = 3\%$	0.159	0.093	0.201
$C = 5\%$	0.093	0.111	0.143

Table: Estimated coefficients of the policy $\pi(X) = 1\{X^\top\beta + \beta_0 > 0\}$

	Rice Ar	Risk Av	Educ	Wel Impr		
				$C = 1\%$	3%	5%
NEWM	-0.068	0.395	-0.397	0.074	0.085	0.093
EWM	-0.003	-0.041	-0.473			

Assumptions in the context of the empirical example

- Local spillovers ("information effects"):
  "*By varying the information available about peers' decisions and randomizing default options, we show that the network effect is driven by the diffusion of insurance knowledge rather than the purchase decisions.*" (Cai et al., 2015)

Assumptions in the context of the empirical example

- Local spillovers ("information effects"):
  "*By varying the information available about peers' decisions and randomizing default options, we show that the network effect is driven by the diffusion of insurance knowledge rather than the purchase decisions.*" (Cai et al., 2015)
- Maximum degree: here, $\mathcal{N}_{n,\max} \leq 5$
- Prop score: known propensity score (no need of correctly specified $m(\cdot)$)

Assumptions in the context of the empirical example

- Local spillovers ("information effects"):
  "*By varying the information available about peers' decisions and randomizing default options, we show that the network effect is driven by the diffusion of insurance knowledge rather than the purchase decisions.*" (Cai et al., 2015)
- Maximum degree: here, $\mathcal{N}_{n,\max} \leq 5$
- Prop score: known propensity score (no need of correctly specified $m(\cdot)$)
- Sampling: sampling indicators independently drawn across 185 villages $\Rightarrow$ unbiased welfare estimator for implementation is rural China + local dependence of R_i

Content

Controlling Overlap

Overlap can behave poorly as the maximum degree increases.

- Issue: some nodes can have many connections ($\delta_n \approx 0$).

Controlling Overlap

Overlap can behave poorly as the maximum degree increases.

- Issue: some nodes can have many connections ($\delta_n \approx 0$).

- Trimming: $W(\pi) \propto \boxed{W\left(\pi \middle| |\mathcal{N}_i| \leq \kappa_n\right)} + \mathcal{O}\Big(P_n(|\mathcal{N}_i| > \kappa_n)\Big)$

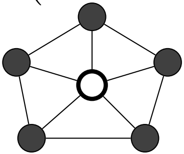
Controlling Overlap

Overlap can behave poorly as the maximum degree increases.

- Issue: some nodes can have many connections ($\delta_n \approx 0$).

- Trimming: $W(\pi) \propto \boxed{W\left(\pi \middle| |\mathcal{N}_i| \leq \kappa_n\right)} + \mathcal{O}\left(P_n(|\mathcal{N}_i| > \kappa_n)\right)$

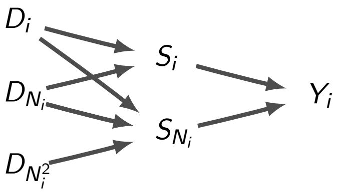
$\Rightarrow$ Choice s.t. $P(|\mathcal{N}_i| > \kappa_n) \approx 0$



$\Rightarrow$ Idea: ignore the direct effect on largely connected nodes, but incorporate the spillovers that they generate.

Other Extensions

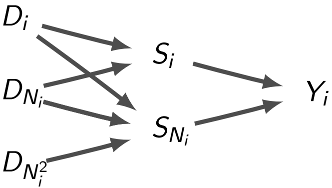
- Non-compliance: incentive $\pi(X_i)$ generates spillovers on selection into treatment.



$\Rightarrow$ Simple-to-estimate expression that depends on $P(S_i = 1|\cdot)$.

Other Extensions

- Non-compliance: incentive $\pi(X_i)$ generates spillovers on selection into treatment.



$\Rightarrow$ Simple-to-estimate expression that depends on $P(S_i = 1|\cdot)$.

- Different target population:
  (i) reweighting mechanism for empirical welfare
  (ii) If weights are unknown, control *expected* regret

- Higher order dependence of order M
  - Rates of convergence depend on higher order terms $\mathcal{N}_{n,\max}^{M/2}$

Conclusion

- I introduce a method for treatment choice with spillovers (and leverage sampling based uncertainty for policy learning)
- I derive regret guarantees for known and unknown propensity score and provide estimation and optimization procedures
- I study the method's performance in an empirical application (and simulation studies)

Conclusion

- I introduce a method for treatment choice with spillovers (and leverage sampling based uncertainty for policy learning)
- I derive regret guarantees for known and unknown propensity score and provide estimation and optimization procedures
- I study the method's performance in an empirical application (and simulation studies)

Thanks! Questions? Link: dviviano.github.io/projects

Theoretical Argument

We want to bound

$$\mathcal{R}(\hat{\pi}) \leq 2\mathbb{E}\Big[\sup_{\pi \in \Pi} \Big| W(\pi) - W_n(\pi) \Big|\Big].$$

- Network interference: let $\Gamma_i = (|N_i|, Z_{k \in N_i}, Z_i, Y_i, D_i, D_{k \in N_i})$

$$\mathbb{E}\Big[\sup_{\pi \in \Pi} \Big| \frac{1}{n} \sum_{i=1}^{n} f_\pi(\Gamma_i) - \mathbb{E}[f_\pi(\Gamma_i)] \Big|\Big] \not\leq 2\mathbb{E}\Big[\sup_{\pi \in \Pi} \Big| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_\pi(\Gamma_i) \Big|\Big]$$
$$\neq 2\mathbb{E}[\mathrm{Rad}_n(\Pi)]$$

1. New symmetrization argument:
   - Group individuals into group of conditionally independent units
   - Leverage sampling uncertainty R_i for symmetrization
2. Upper bounds on the Rademacher complexity:
   - Invoke Ledoux-Telagrand contraction inequality for spillovers;
   - Characterize the bound as a function of the square root of the maximum degree and the covering number of Π.

(back)