No Sh*t: Demand Estimation with Strategic Complementarities

R. Guiteras[1] J. Levinsohn[2] A. Mushfiq Mobarak[2]

[1]NCSU ARE [2]Yale University

September 2017

Outline

1. Background and Context
2. Context and Experimental Design
3. The Model
4. Estimation Strategy
5. Results
6. Policy Simulations
7. Policy Experiments

The Policy Context

- 1 billion people (15% of world population) practice open defecation
- 1.5 billion more use non-hygienic latrines
- Negative effects on health and human capital
  - Spears (India), Alzua et al (Mali)

- Latrine Subsidies are a common policy tool
  - Pursued by many of the major actors in the sanitation sector
  - Govt of India, Gates Foundation, World Bank, major NGOs
- Subsidies more effective than info or supply-side strategies
  - Guiteras, Levinsohn and Mobarak, *Science*, 2015

- Subsidies typically most effective at promoting adoption of a range of technologies, products and behaviors
  - Dupas 2014, Meredith et al 2013, JPAL "*The Price is Wrong*"

Interdependence in decision-making

- Sanitation adoption may be inter-dependent across households
  - Technological/epidemiological complementarity
  - Social norms about open defecation
  - Learning spillovers
- With inter-dependencies, optimal subsidy policies are unclear
  - Should we concentrate or disperse subsidies?
  - Whom to target with a discount?
  - Even demand elasticity is not simple to calculate
- Important, because externalities are the main reason to subsidize

What We Do, and Why

1. RCT designed ex ante to measure social spillovers
2. Structural demand model incorporating social spillover

- Reduced-form estimates insufficient to predict effect of subsidy on take-up rates
  - Fixed point problem; structural estimation necessary
- Structural model allows counter-factual policy experiments
  - What is the optimal way to distribute a subsidy budget of $100?
  - What are the effects of targeting the poor or the socially connected?
  - Target densely populated or inter-connected villages?
- Predict effects of policies *in between* the limited number of experimental cells
- Improve subsidy policy design, especially if:
  - Some households are more price sensitive, or
  - Communities have features that are more conducive to generating spillovers

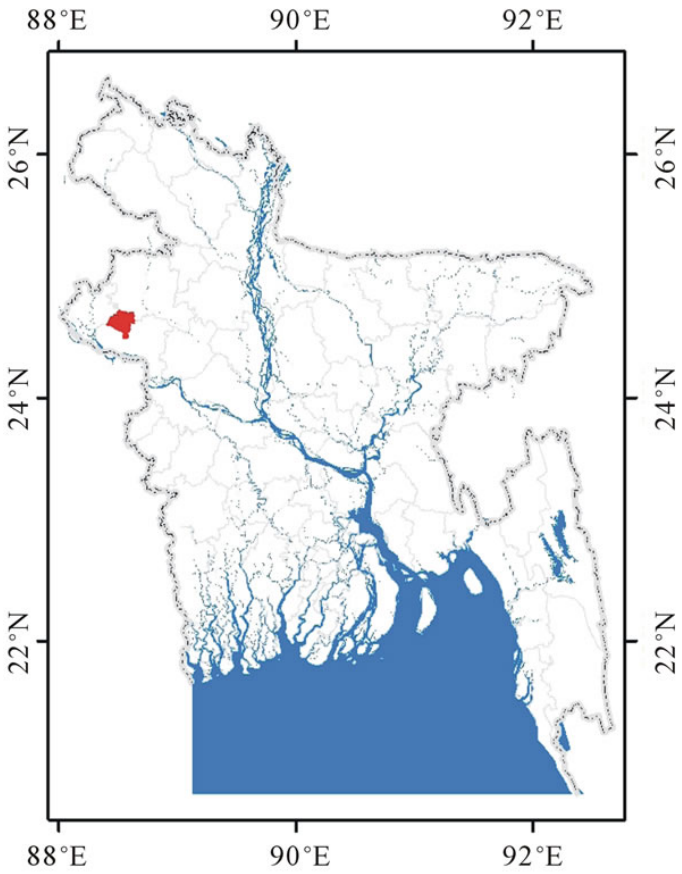
Broader Implications

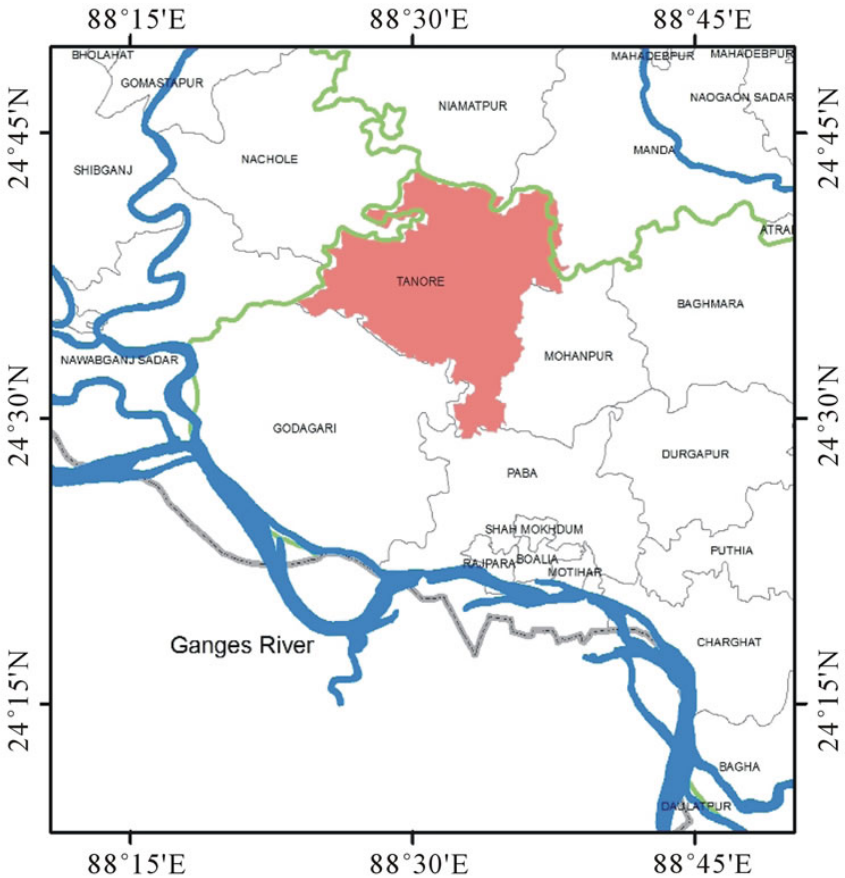
- Decision inter-dependencies for many different behaviors, products and technologies
  - Large *Peer Effects* literature
  - Agriculture (Foster and Rosenzweig 1995, Conley and Udry 2010)
  - Health Products
    - Bed nets (Dupas 2013), Deworming pills (Kremer and Miguel 2007)
    - Sanitation (Pattanayak *et al* '09), Hygiene (Oster & Thornton 2012)
  - Investment choices (Bursztyn et al 2014)
  - Work effort (Mas and Moretti 2009)
  - Migration (Akram et al 2017, Morten 2017, Meghir et al 2017)
  - Insurance (Mobarak and Rosenzweig 2012,2015; Kinnan 2017)
  - Gender norms in labor force participation (Bertrand 2011)

- Evaluating policies to promote any of these products or behaviors requires us to analyze feedback loops in adoption.

Context

- Rural Tanore, Bangladesh
  - 32% of adults regularly engaged in open defecation
  - 21% of households owned a hygienic latrine





Intervention

Information Treatment (Latrine Promotion Program, LPP - *like* CLTS)

Intervention

Latrine Subsidy (Public Lottery)

- Vouchers with 75% discount on latrine parts
- Household pays own transportation and installation
- Subsidy vouchers for ≈ half the cost of an installed hygienic latrine

- Households paid ≈ USD 6-12 depending on the exact model + USD 7-10 in transport costs
- Value of voucher: USD 16-36. (USD 36 = 2800 Taka)

- *Eligible* households: Near landless; constituted poorest 75%
- On average, 60% of eligible households won a voucher

Intervention

Latrine Subsidy (Public Lottery)

Intervention

Free Materials for Roof ("Tin")

- Lottery for corrugated iron for roof
- Worth $\approx$ USD 15
- Half of eligible households won
- Independent of subsidy lottery

Intervention

Free Materials for Roof ("Tin")

3 Distinct Features of Experimental Design

Level 3: Household-level subsidy lottery

"Average" subsidy village

		Superstructure lottery	
		Won	Lost
Subsidy lottery	Won	Won both	Won voucher, Lost tin
	Lost	Lost voucher, Won tin	Lost both

Experimental Design

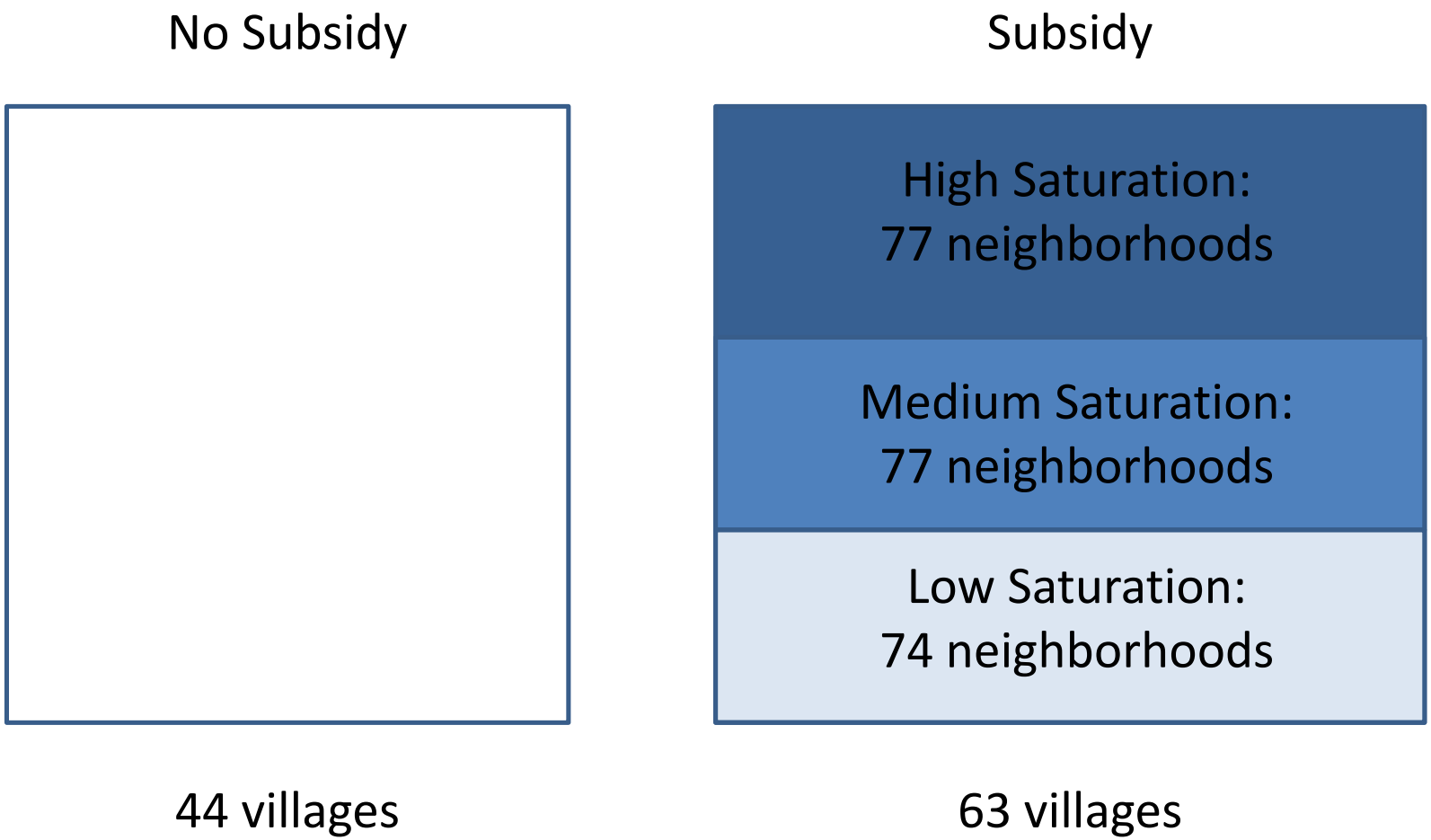
Level 1: Villages that received subsidies were randomly chosen

Randomize at village level where to offer subsidies

Experimental Design

Level 2: Vary subsidy saturation across neighborhoods

Randomize neighborhood share of households winning subsidy



Experimental Design

2 types of villages, 3 types of neighborhoods, 4 types of households

Level 3 at neighborhood level – average

		Superstructure lottery	
		Won	Lost
Subsidy lottery	Won	Won both	Won voucher, Lost tin
	Lost	Lost voucher, Won tin	Lost both

Experimental Design

2 types of villages, 3 types of neighborhoods, 4 types of households

Level 3 at neighborhood level – Low "Saturation" neighborhood

		Superstructure lottery	
		Won	Lost
Subsidy lottery	Won	Won both	Won voucher, Lost tin
	Lost	Lost voucher, Won tin	Lost both

Experimental Design

2 types of villages, 3 types of neighborhoods, 4 types of households

Level 3 at neighborhood level – Medium "Saturation" neighborhood

		Superstructure lottery	
		Won	Lost
Subsidy lottery	Won	Won both	Won voucher, Lost tin
	Lost	Lost voucher, Won tin	Lost both

Experimental Design

2 types of villages, 3 types of neighborhoods, 4 types of households

Level 3 at neighborhood level – High "Saturation" neighborhood

		Superstructure lottery	
		Won	Lost
Subsidy lottery	Won	Won both	Won voucher, Lost tin
	Lost	Lost voucher, Won tin	Lost both

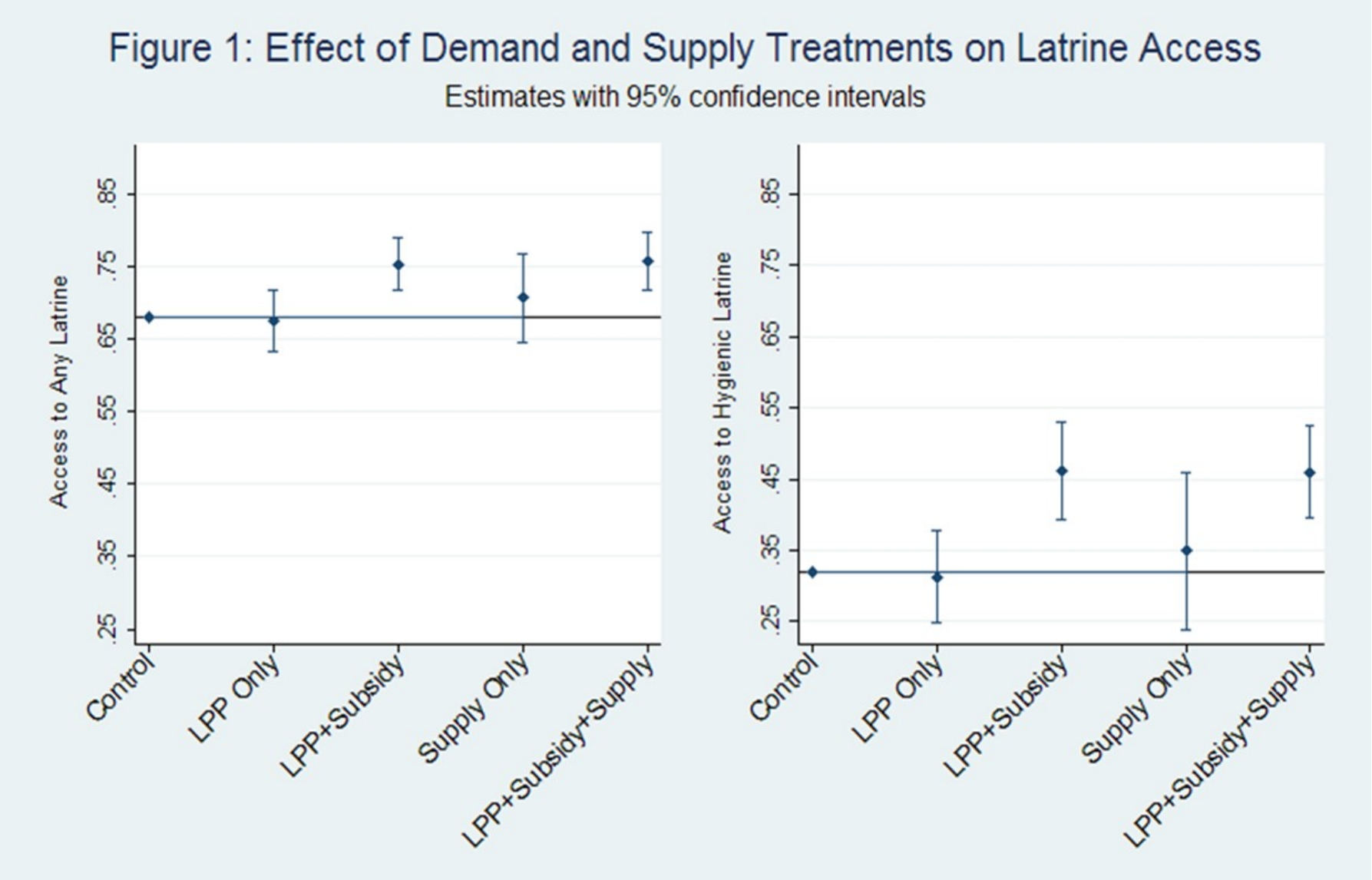
Experimental Design

Summary

- Two-stage randomization
- Randomized price allows us to estimate price elasticity
- Randomized saturation allows us to estimate social spillover

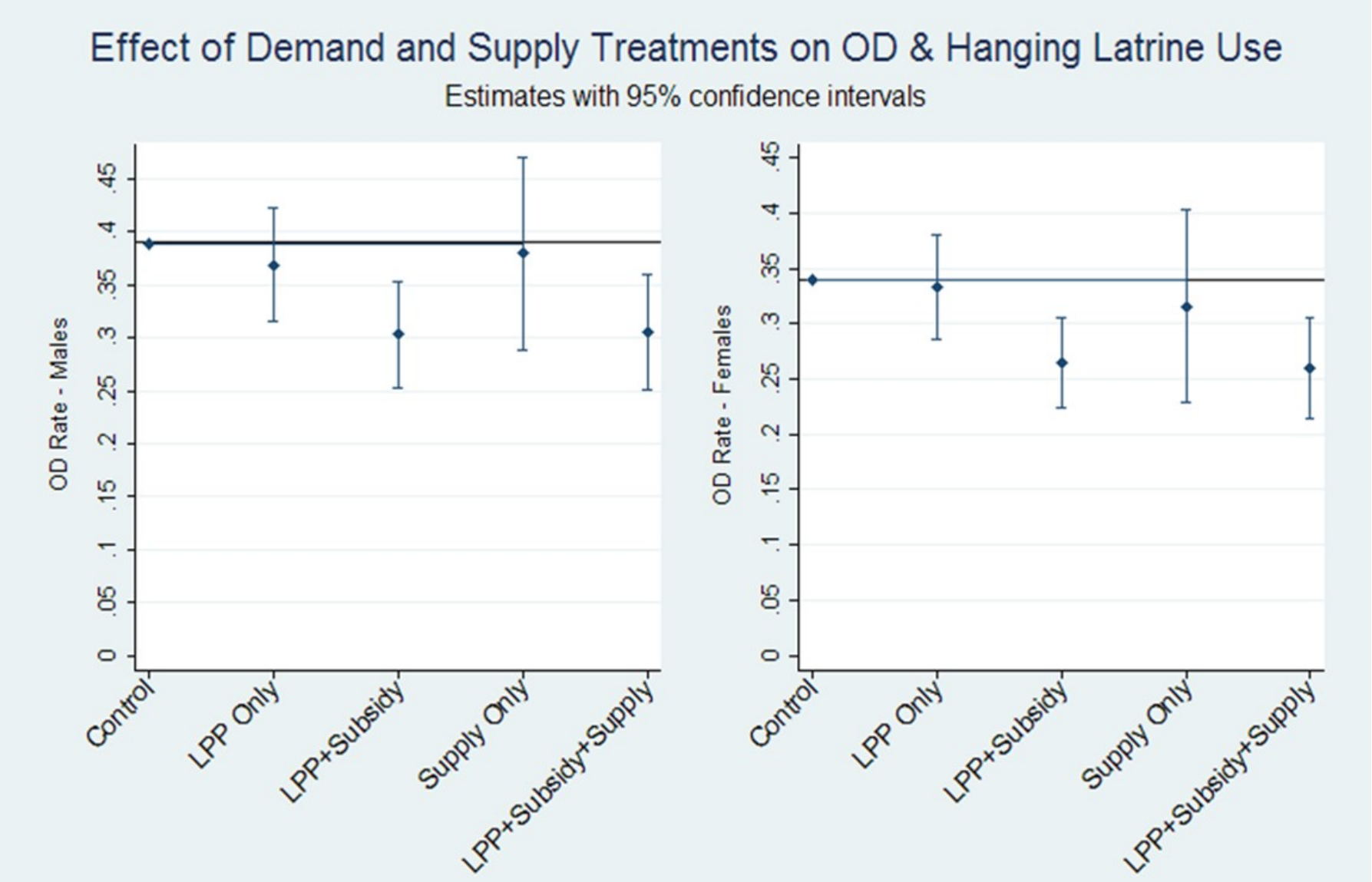
Guiteras, Levinsohn, Mobarak *Science* 2015 Results

Simple RCT program evaluation results



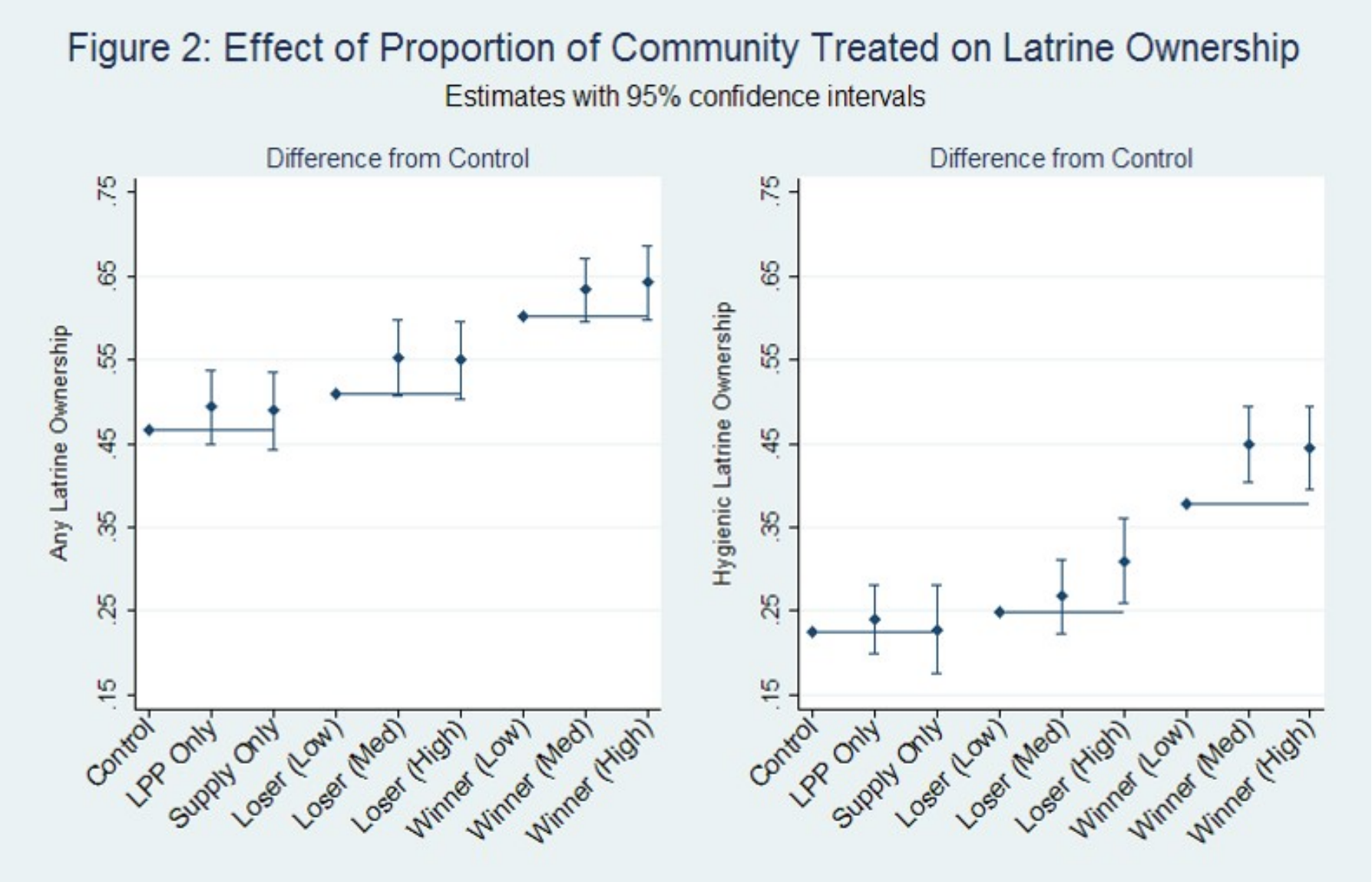
Guiteras, Levinsohn, Mobarak *Science* 2015 Results

Simple RCT program evaluation results



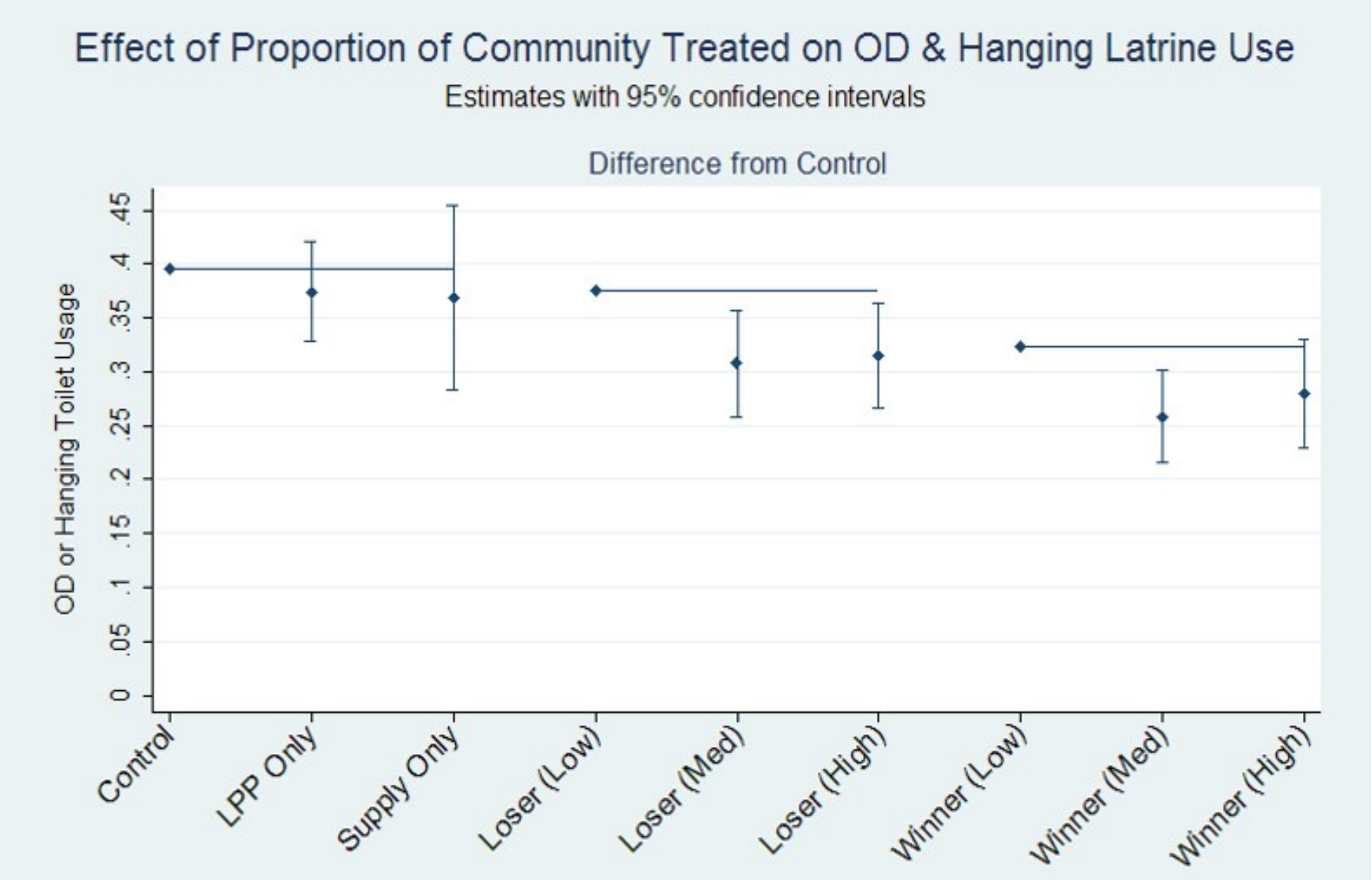
Guiteras, Levinsohn, Mobarak *Science* 2015 Results

Simple RCT program evaluation results



Guiteras, Levinsohn, Mobarak *Science* 2015 Results

Simple RCT program evaluation results



Setup

The utility household i receives depends on its own adoption decision, a_i and that of others in the neighborhood, a_j.

$$U_i = U(a_i, a_{j \neq i})$$

Strategic Complements:

$$\left.\frac{\partial U_1(\cdot)}{\partial a_1}\right|_{a_{j\neq 1}=1} > \left.\frac{\partial U_1(\cdot)}{\partial a_1}\right|_{a_{j\neq 1}=0}$$

Strategic Subsitutes:

$$\left.\frac{\partial U_1(\cdot)}{\partial a_1}\right|_{a_{j\neq 1}=1} < \left.\frac{\partial U_1(\cdot)}{\partial a_1}\right|_{a_{j\neq 1}=0}$$

Utility

We model the utility a household i residing in neighborhood (or cluster) c receives from adoption ($j = 1$) or not ($j = 0$) as:

$$U_{ijc} = f(z_{ic}, x_c, P_{ijc}, \bar{s}_c, \xi_c, \epsilon_{ijc})$$

where:

z_{ic} is a vector of observable household-level attributes,

x_c is a vector of observable neighborhood-level attributes,

P_{ijc} is the price of a latrine j faced by household i in neighborhood c,

$\bar{s}_c$ is the share of households purchasing a latrine in neighborhood c,

ξ_c is a neighborhood-level unobservable component of utility;

ϵ_{ijc} is a household-specific unobservable component of utility, distributed "logit."

The utility of not having a hygienic latrine, the outside good, is normalized to zero.

Challenges Posed by the Model for Estimation

- Typically, this model would face severe endogeneity issues:
  - P_{ijc}: Price is endogenous for all the usual reasons.
  - $\bar{s}$: The share of the neighborhood that adopts may be driven by unobservables correlated across households. (Manski's *Reflection Problem*)
- Our RCT was designed to address these
  - Prices randomized via subsidy lottery
  - Instrument for neighborhood share with randomized subsidy saturation

Overview of Estimation

$$U_{ijc} = \alpha_0 + \alpha_1 p_{ijc} + \gamma_1 \bar{s}_c + \xi_c + \epsilon_{ijc}$$

Two step estimator:

- Step 1: Binomial Logit. Regress Adopt/No Adopt on price and neighborhood-level fixed effects.

$$U_{ijc} = \alpha_0 + \alpha_1 p_{ijc} + \delta_c + \epsilon_{ijc}$$

- Step 2: Linear IV Regression of fixed effects from Step 1 on share adopting

$$\hat{\delta}_c = \gamma_0 + \gamma_1 \bar{s}_c + \xi_c$$

  And instruments are neighborhood-level randomly assigned subsidy intensities.

Overview of Estimation

$$U_{ijc} = \alpha_0 + \alpha_1 p_{ijc} + \gamma_1 \bar{s}_c + \xi_c + \epsilon_{ijc}$$

Own price semi-elasticity of demand is a function of both price and share adopting:

$$\frac{\delta s}{\frac{\delta P}{P}} = \frac{\alpha_1 s_c(1 - s_c)}{1 - \gamma_1 s_c(1 - s_c)}$$

- Varies by baseline conditions in village
- Cannot compute elasticity without estimate for coef. on $\bar{s}_c$

Estimation: Covariates and interactions

$$U_{ijc} = \alpha_0 + \alpha_1 p_{ijc} + \alpha_2 L_{ic} + \alpha_3 (p_{ijc} * L_{ic}) + \gamma_1 \bar{s}_c + \gamma_2 D_c + \gamma_3 (\bar{s}_c * D_c) + \xi_c + \epsilon_{ijc}$$

Two step estimator:

- Step 1: Binomial Logit. Regress Adopt/No Adopt on price, landless indicator ($L_i c$), interaction, and neighborhood-level fixed effects.

$$U_{ijc} = \alpha_0 + \alpha_1 p_{ijc} + \alpha_2 L_{ic} + \alpha_3 (p_{ijc} * L_{ic}) + \delta_c + \epsilon_{ijc}$$

- Step 2: Linear IV Regression of fixed effects from Step 1 on share adopting, neighborhood density (D_c), and their interaction.

$$\hat{\delta}_c = \gamma_0 + \gamma_1 \bar{s}_c + \gamma_2 D_c + \gamma_3 (\bar{s}_c * D_c) + \xi_c$$

And instruments are neighborhood-level randomly assigned subsidy saturation, plus interactions of saturation with D_c.

First-step household-level fixed effects logit
Dependent variable: Ownership of hygienic latrine.

	(1)
Price (log)	-0.745***
	(0.072)
Interact price and landless	-0.183**
	(0.089)
Landless	-0.834***
	(0.054)
Num. of neighborhoods	368
Num. of households	12,770

Price variable recentered (mean zero). True mean: 8.26.

* p<0.10, ** p<0.05, *** p<0.01.

Second step neighborhood-level IV
Dependent variable: neighborhood estimated fixed effect from first step

Share of peers adopting	1.386***
	(0.397)
Log of neighbors within 50m, para-average	0.190**
	(0.097)
First-step F stat.	18.9
Num. neighborhoods	368

Nearby neighbor measure averaged over eligible households only.

Covariates recentered (mean zero). Mean share of peers adopting: 0.333.

Mean of average neighbors: 12.483, mean of log measure: 2.453.

* p<0.10, ** p<0.05, *** p<0.01.

Recipe for simulating counter-factuals

At initial estimated values, the following fits exactly:

$$s_c^n = \frac{1}{N_c} \sum_{i \in c} \frac{e^{\alpha_0 + \alpha_1 p_{ijc} + \alpha_2 L_{ic} + \alpha_3 (p_{ijc} * L_{ic}) + \gamma_1 \bar{s}^n + \xi_c}}{1 + e^{(\cdot)}}$$

To simulate a policy, change the exogenous variables and iterate to compute a new equilibrium for each neighborhood:

$$s_c^n = \frac{1}{N_c} \sum_{i \in c} \frac{e^{\alpha_0 + \alpha_1 p_{ijc} + \alpha_2 L_{ic} + \alpha_3 (p_{ijc} * L_{ic}) + \gamma_1 \bar{s}^{n-1} + \xi_c}}{1 + e^{(\cdot)}}$$

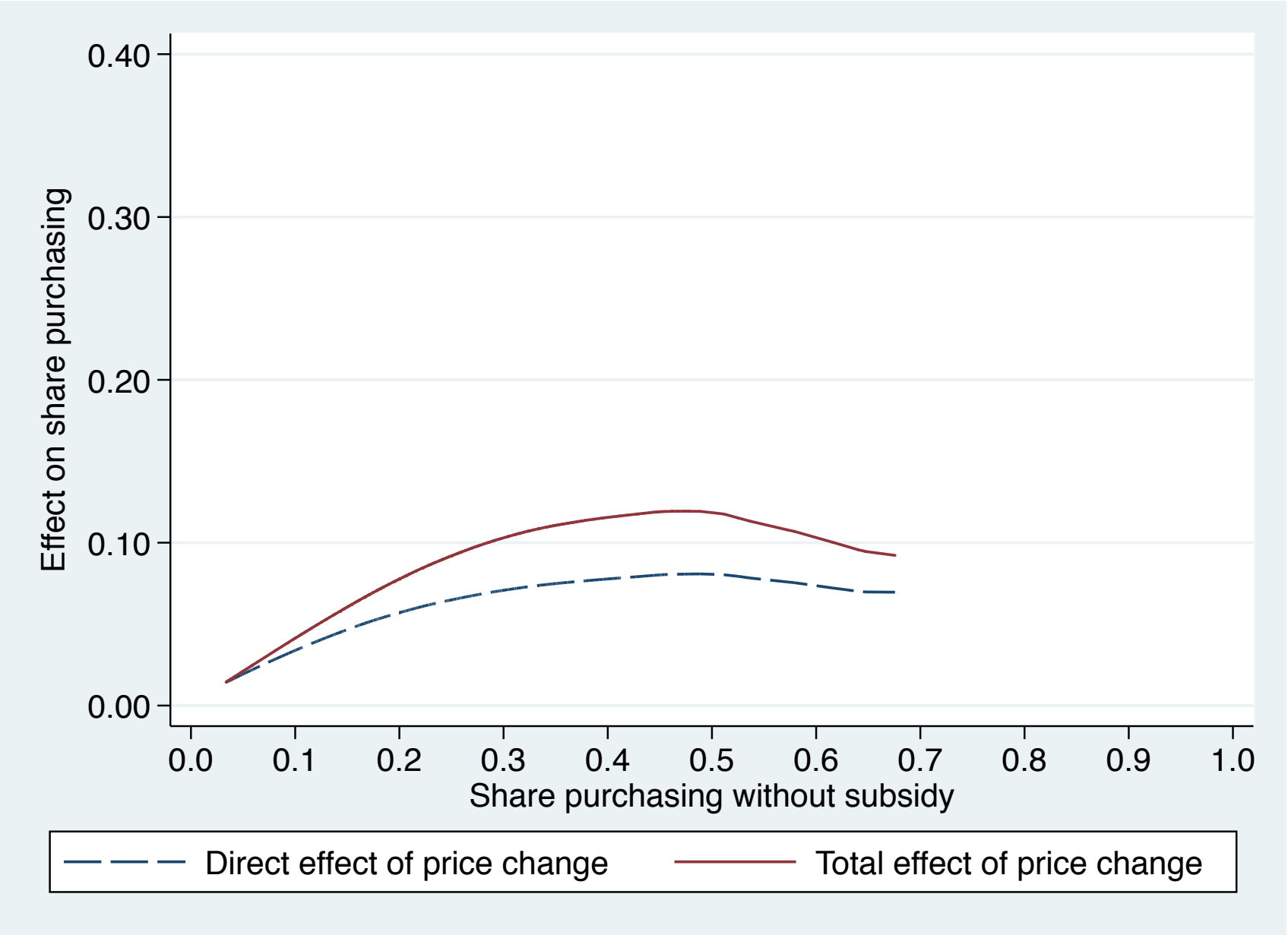
Iterating until $s_c^n - s_c^{n-1} < tol$.

Model Validation

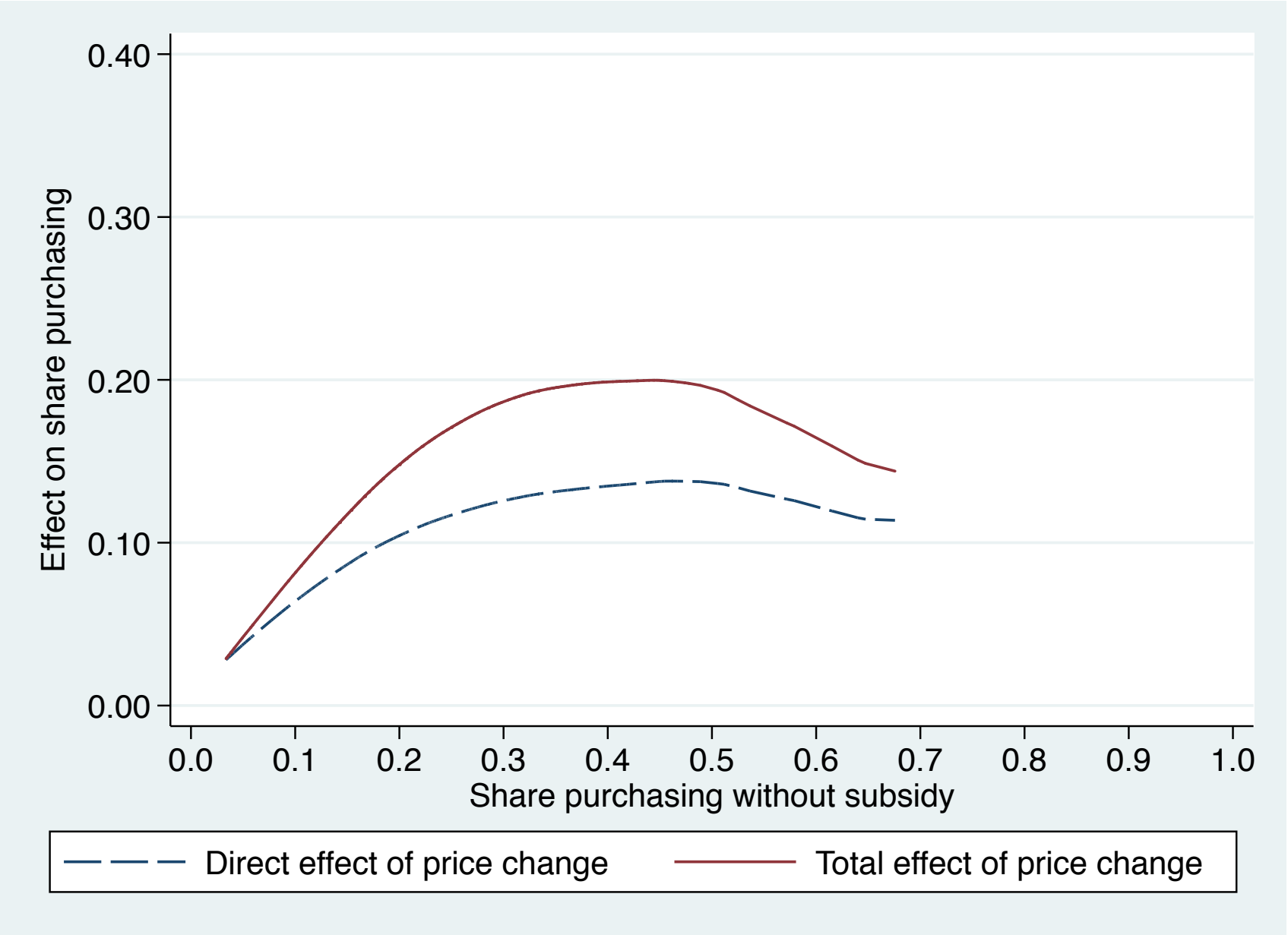
How well does the model predict the results of the RCT?

- Use the model estimates to predict the effects of the low, medium and high saturation experiments
  - i.e., perform the RCT experiments inside the model
- Compare against the reduced form results from the RCT
- Pretty good match: (8%,14%,18%) in data vs. (6%,11%,15%) in the model without any covariates
- This suggests that the model captures the essential components of latrine adoption decisions quite well
- Estimated price elasticities vary across saturation levels (0.08, 0.16, 0.25)
- Other experiments?
  - How do people respond to the latrine subsidy and the tin subsidy?

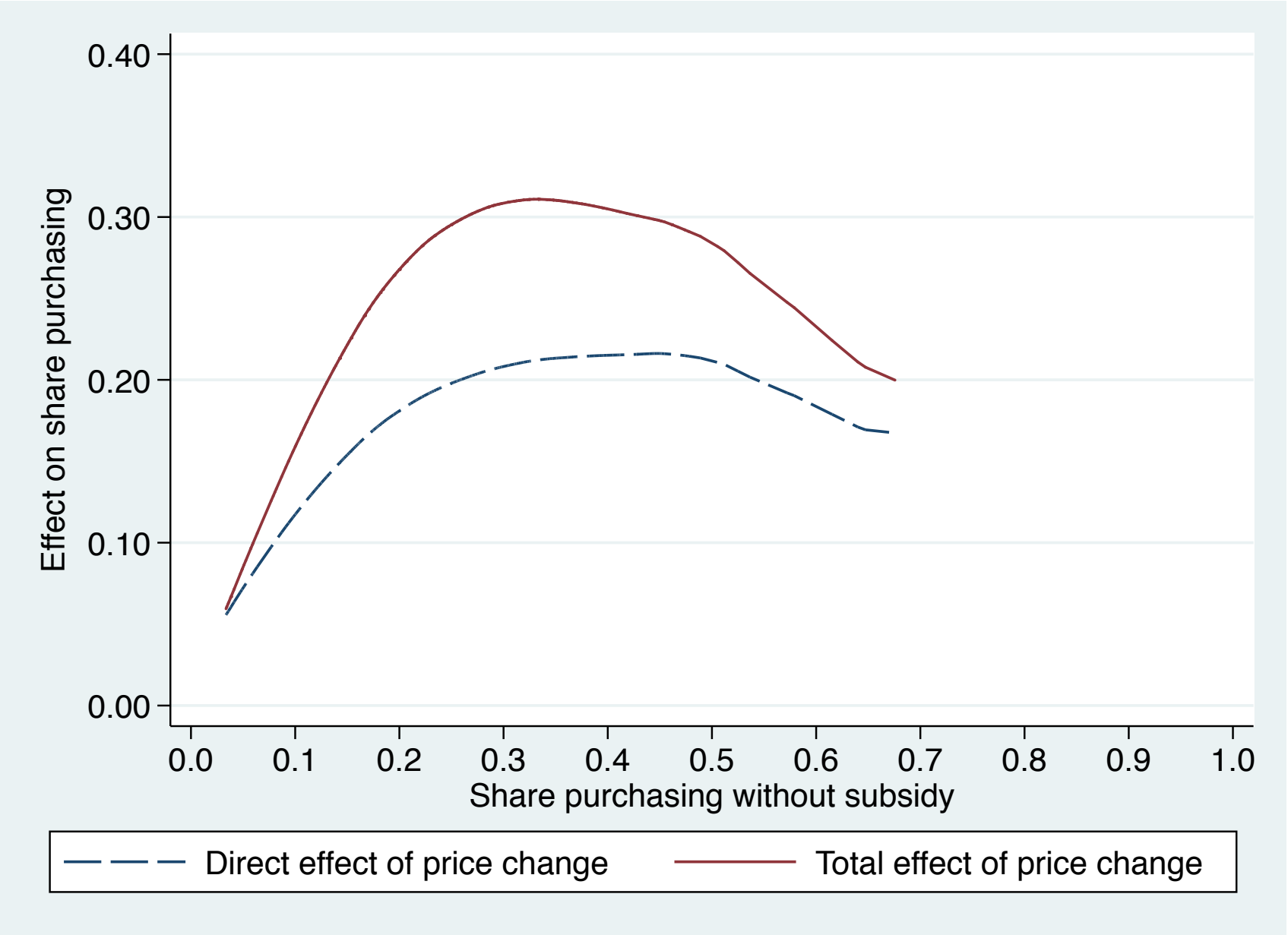
Experiment: Uniform 2000 BDT Subsidy



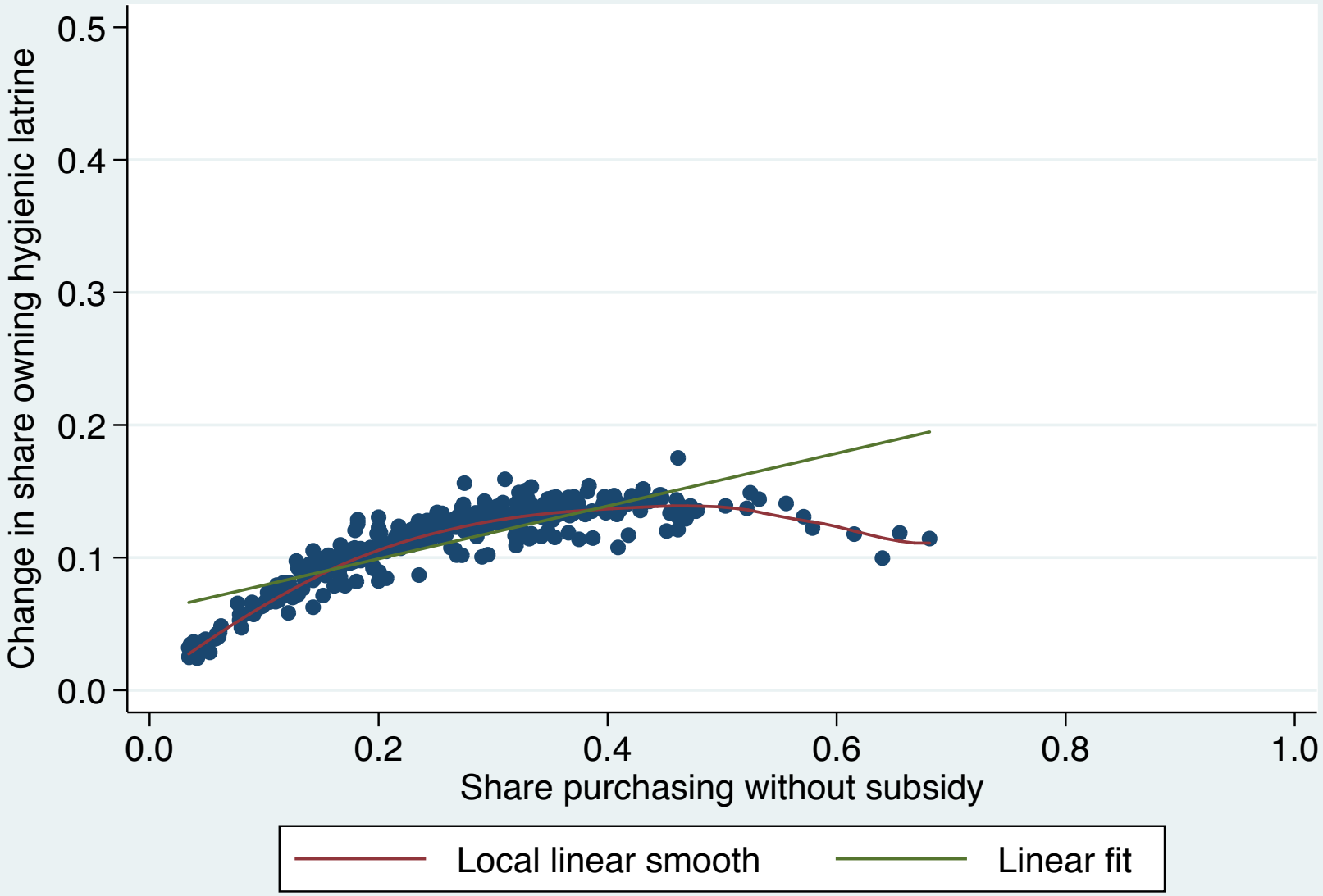
Experiment: Uniform 3000 BDT Subsidy



Experiment: Uniform 4000 BDT Subsidy

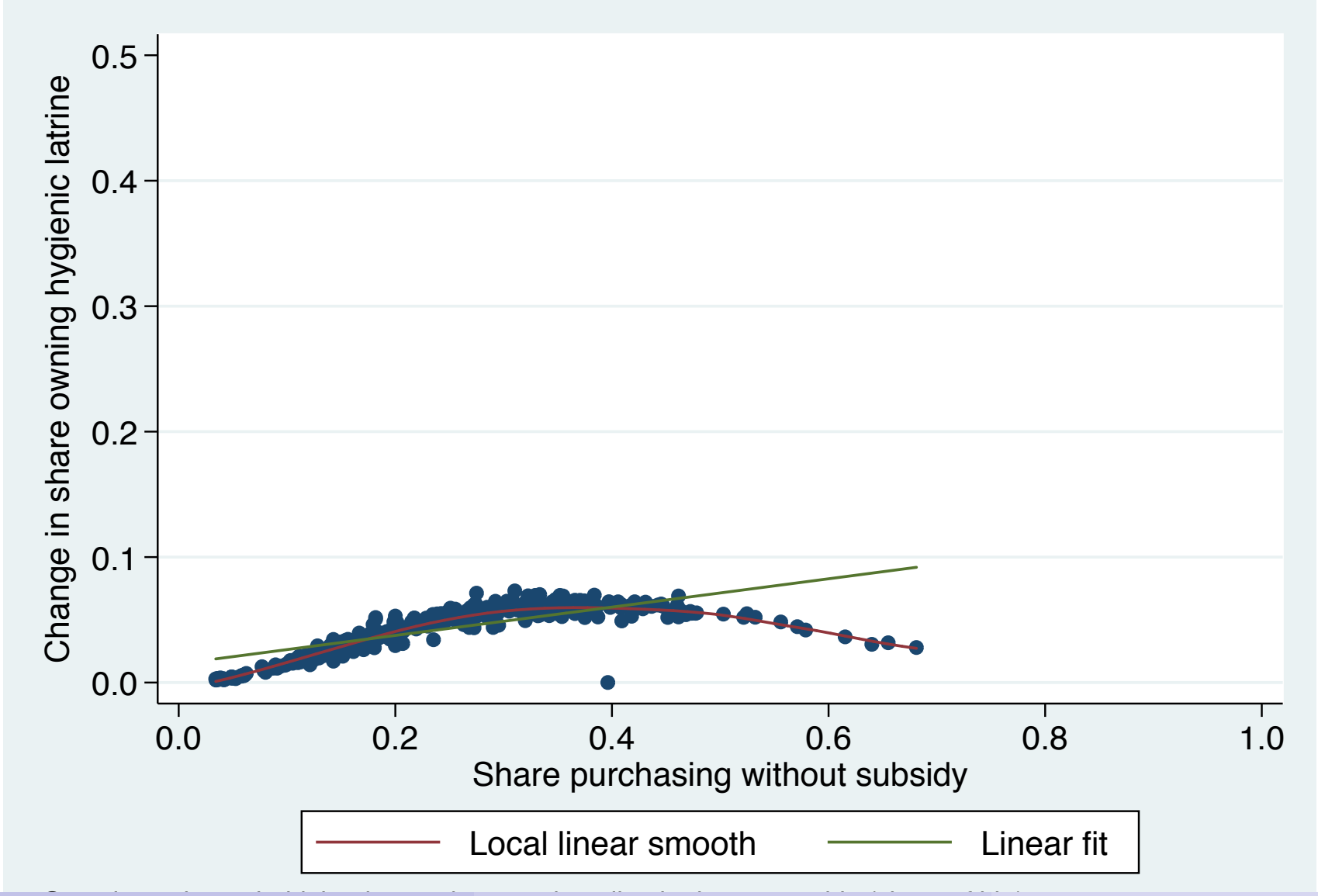


Experiment: Uniform 3000 BDT Subsidy, Direct Effect

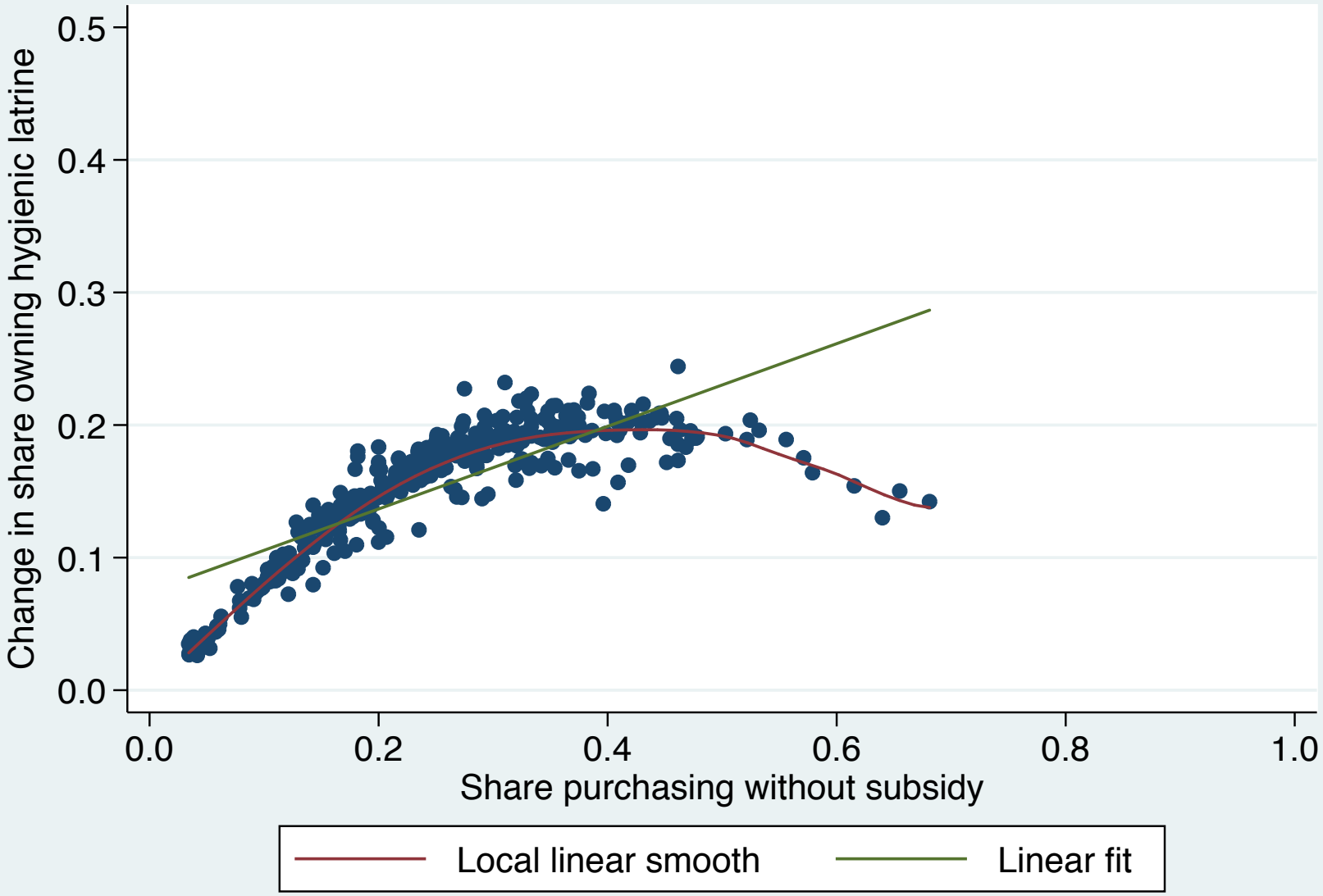


Covariates: household: land owned; group: baseline latrine ownership (share of hh.)

Experiment: Uniform 3000 BDT Subsidy, Spillover Effect

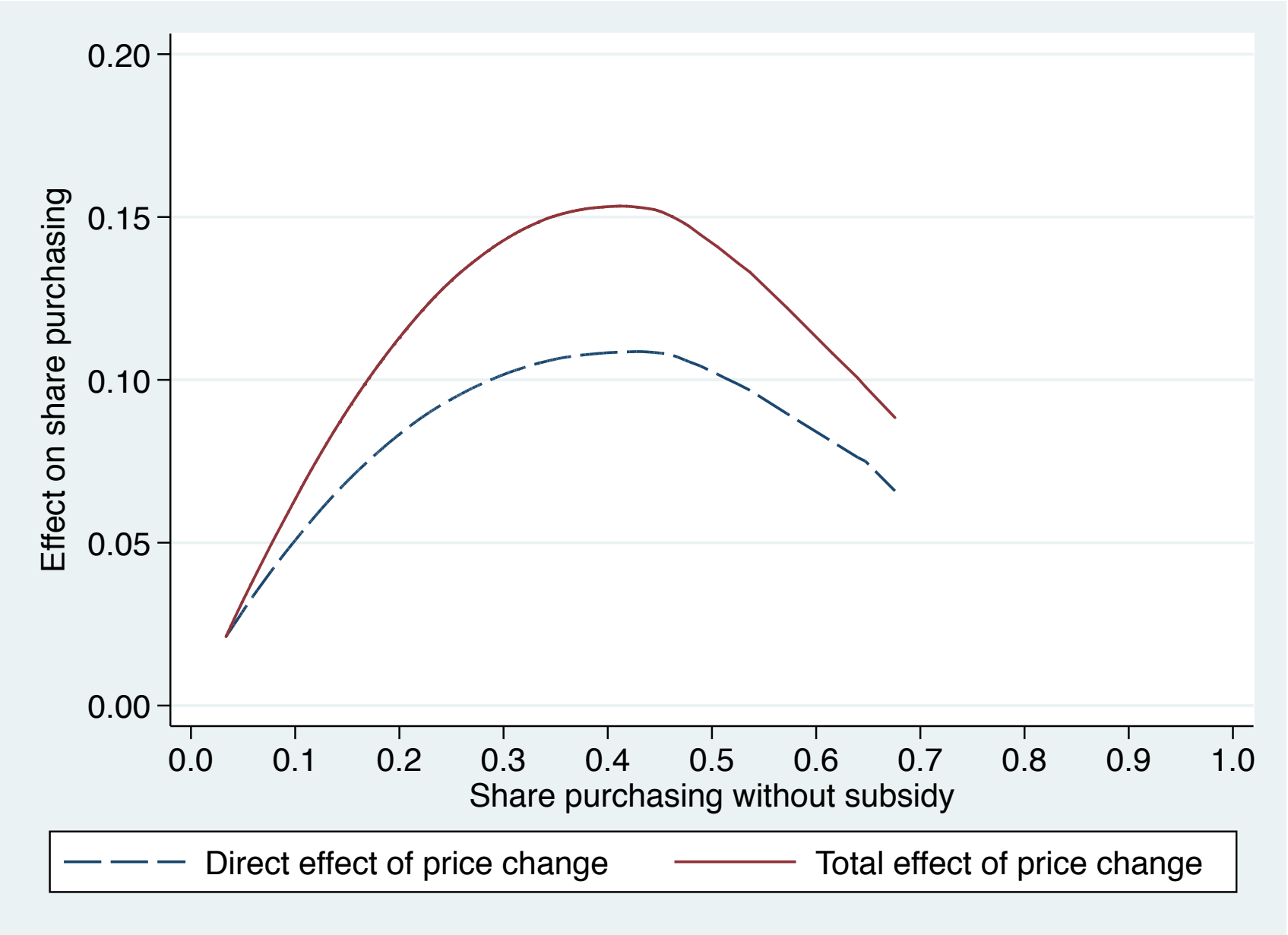


Experiment: Uniform 3000 BDT Subsidy, Total Effect

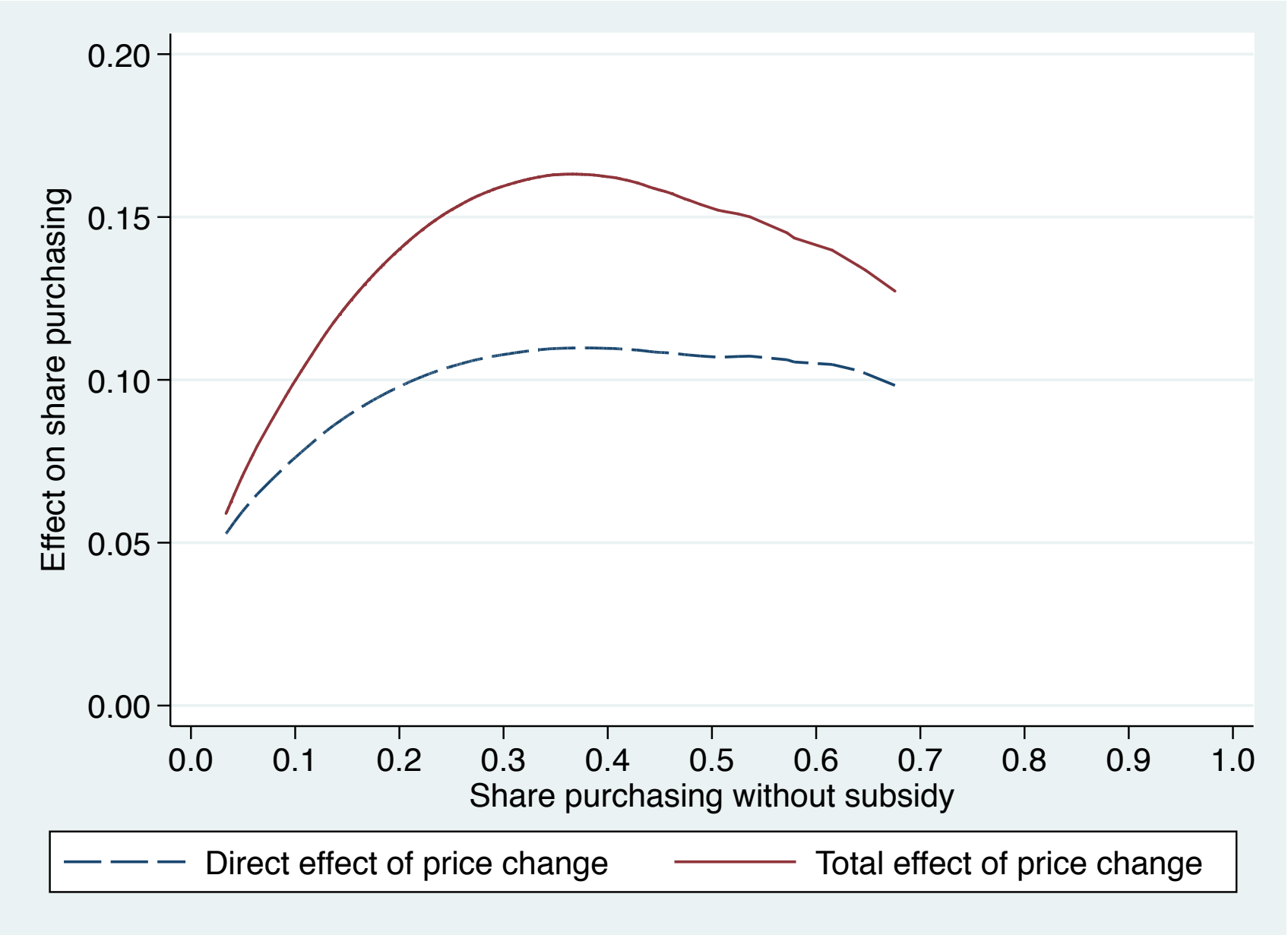


Covariates: household: land owned; group: baseline latrine ownership (share of hh.)

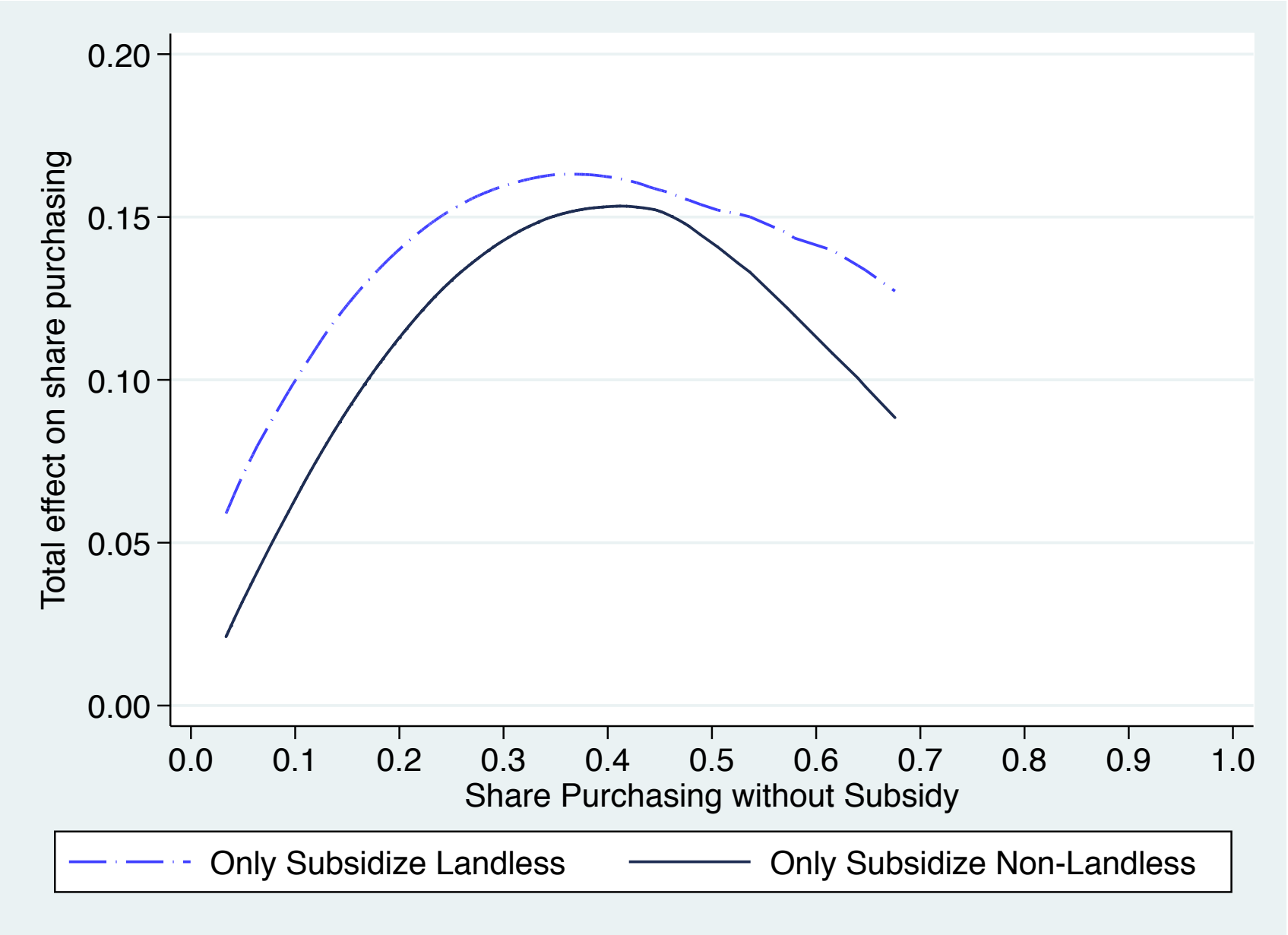
Experiment: Subsidize Landed Only



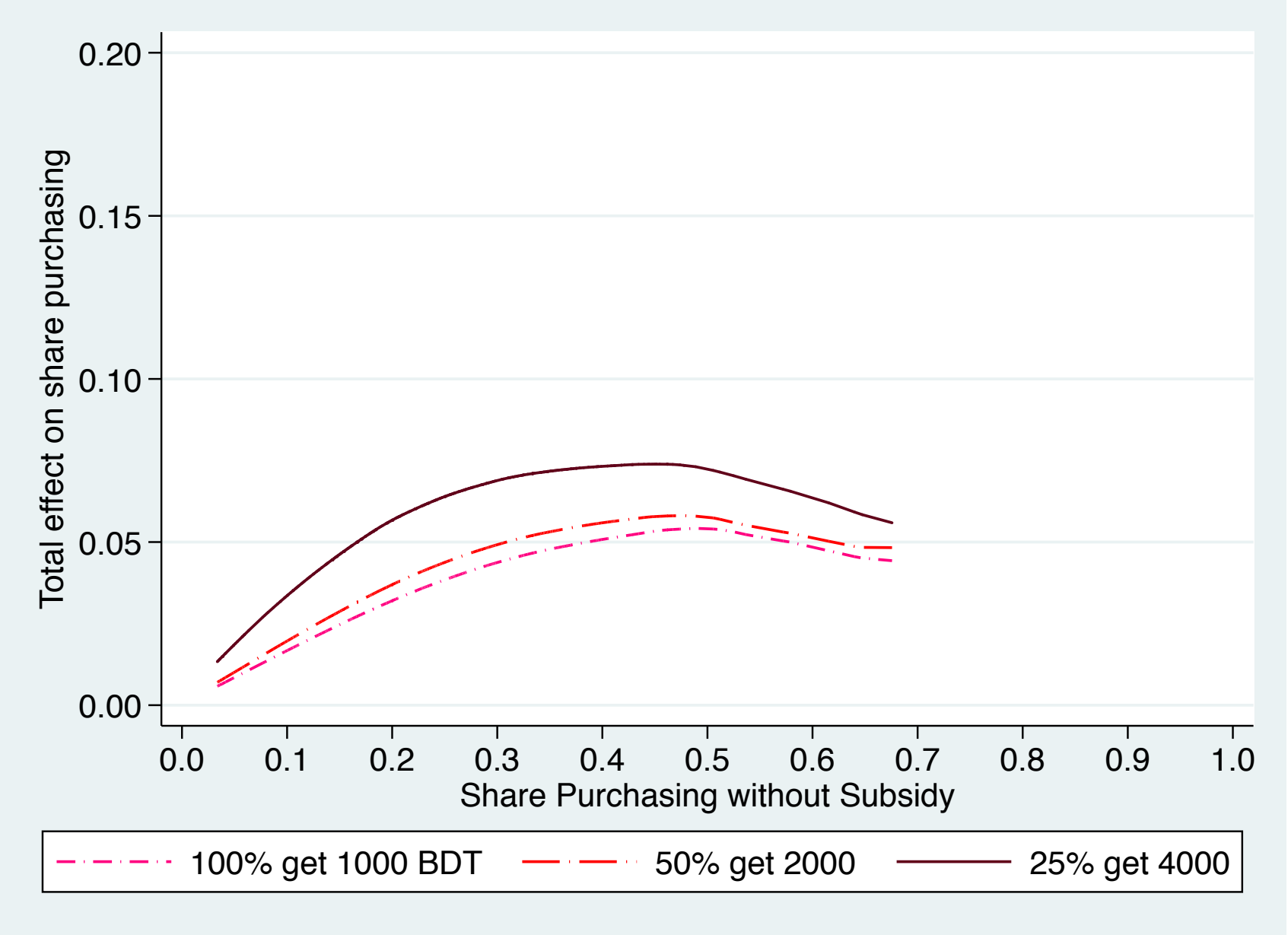
Experiment: Subsidize Landless Only



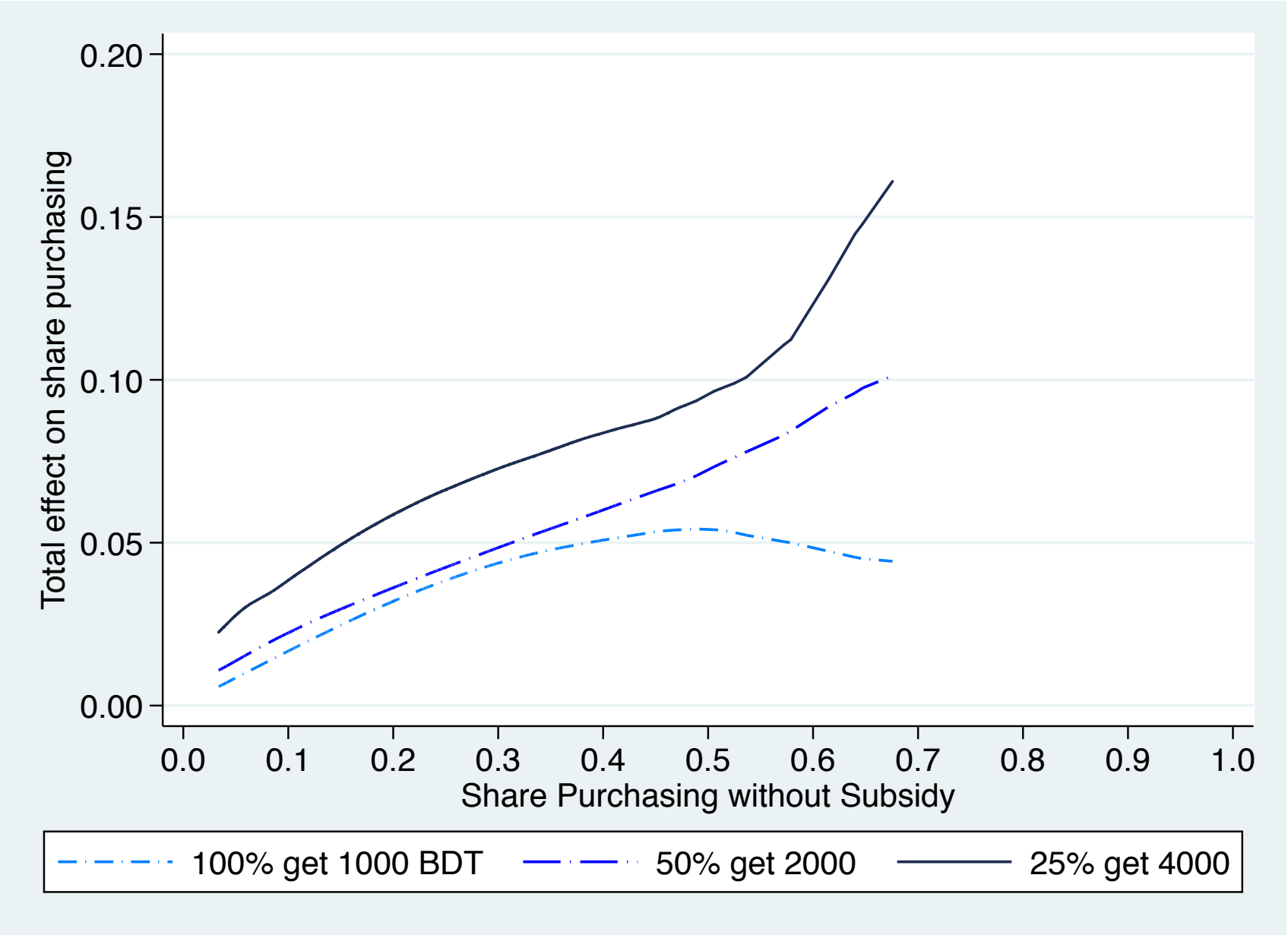
Compare Landless & Non-Landless Total Effects



Experiment: Vary Subsidies *within* neighborhoods

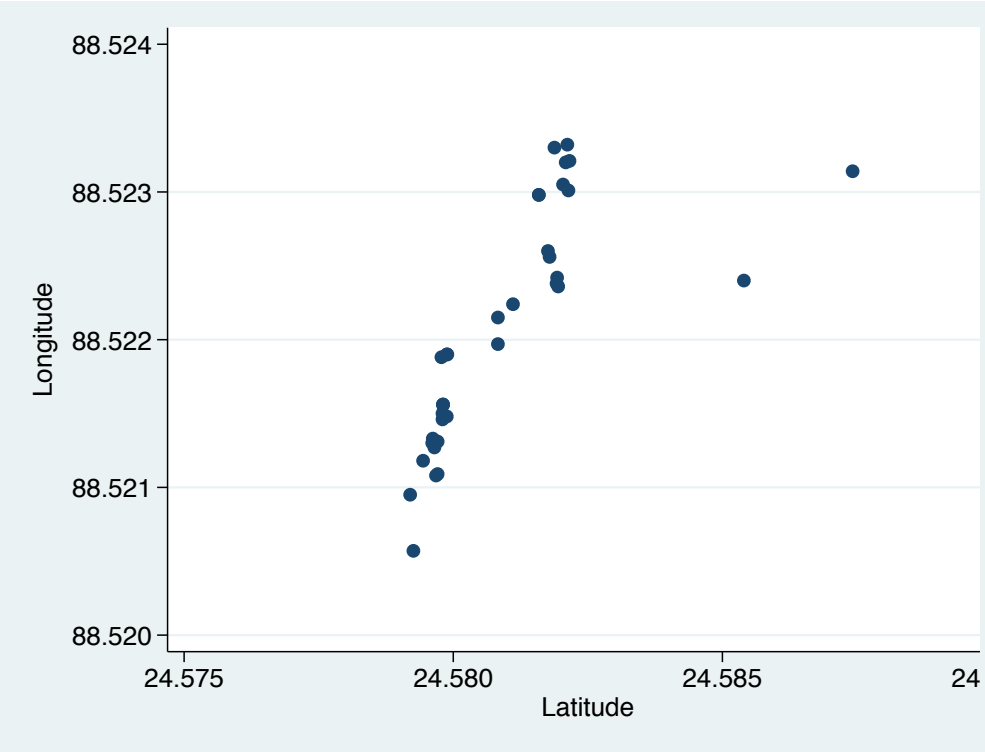


Experiment: Vary Subsidies across neighborhoods

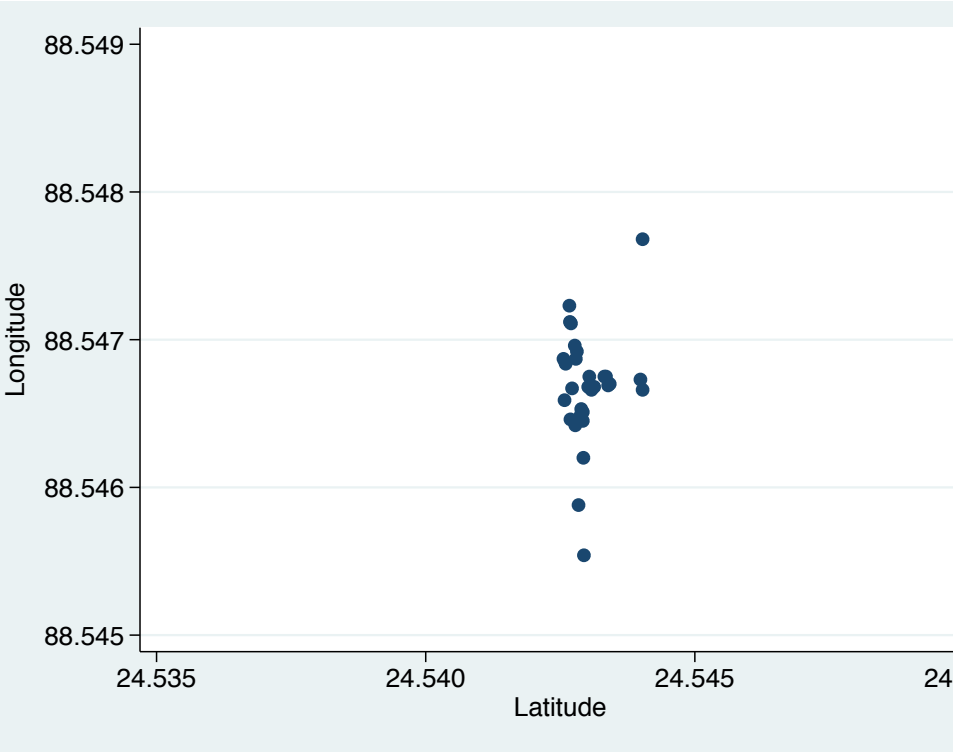


Representative High- & Low-Density neighborhoods

neighborhood at 20th Density pctile



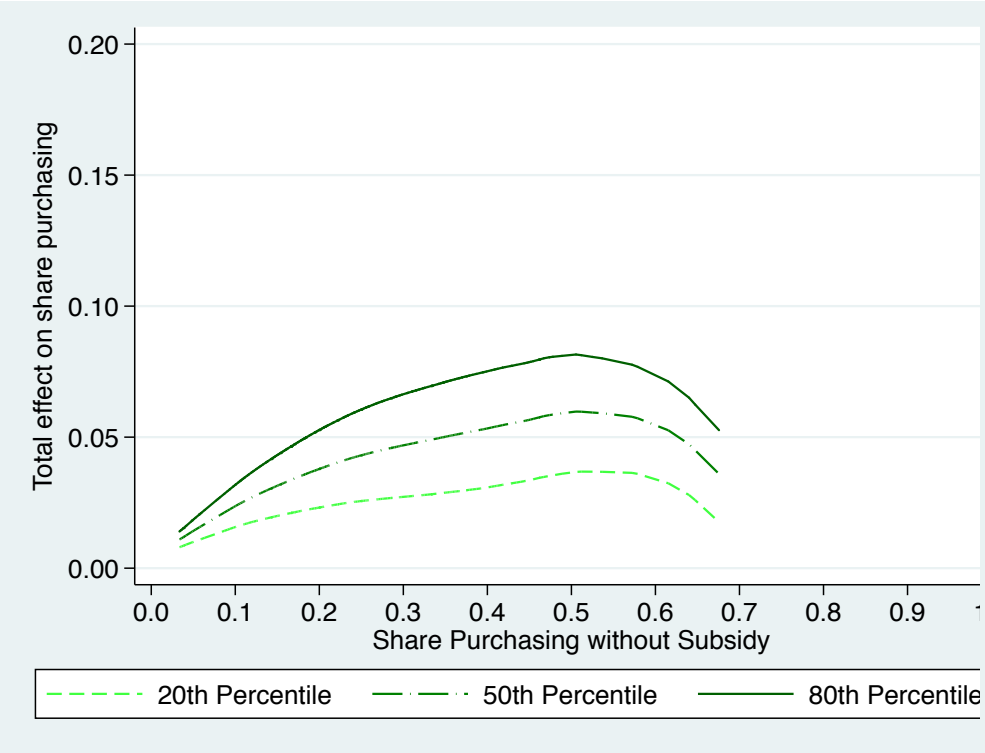
neighborhood at 80th Density pctile



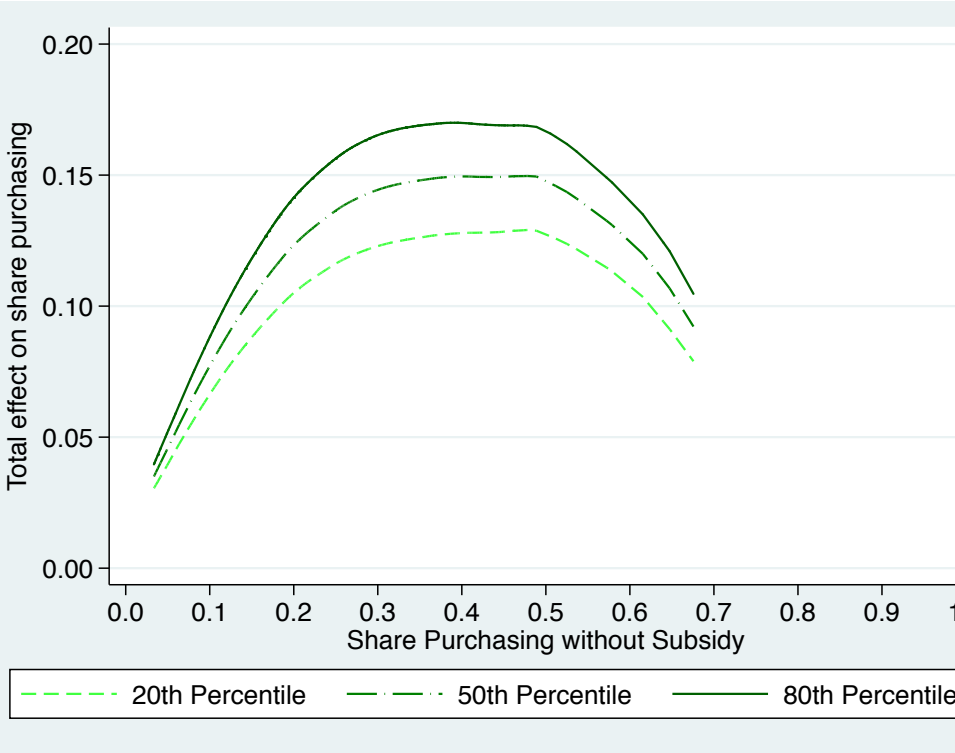
Density: mean number of neighbors within 50m of eligible households

Experiment: Give half of households large subsidy, Vary Population Density

2000 BDT Subsidy



4000 BDT Subsidy



All neighborhoods assumed to be at the 20th, 50th, or 80th percentile of density.

Targeting Subsidies to Socially Central Households

Can we take advantage of social learning?

- Prior to interventions, we conducted a social network census
- Asked every household to identify who they are connected to
- We analyzed the data and identified "Highly Connected Individuals" (HCI) in each neighborhood
- We then ran a cross-cutting experiment within subsidy villages
  - Bias the lotteries in favor of HCI in some neighborhoods
  - Leave it unbiased in others

Second step neighborhood-level IV
Dependent variable: neighborhood estimated fixed effect from first step

	(1) No Landless Int.†	(2) (1) w/ Int.	(3) (2) w/ HCI
Share of peers adopting	1.307***	1.293***	1.373**
	(0.417)	(0.418)	(0.550)
Interact peer adoption & HCI			-0.103
			(0.824)
Randomized HCI binary			-0.013
			(0.062)
First-step F stat.	33.6	33.6	17.0
Num. neighborhoods	368	368	368

† (2) and (3) have first-step interactions of landlessness and price. (1) does not.
(3) randomly assigns HCI=1 to half of non-subsidized paras on a pairwise basis.
Share of peers adopting recentered (mean zero).
Mean share of peers adopting: 0.333.
* p<0.10, ** p<0.05, *** p<0.01.

Reduced-form household-level OLS
Dependent variable: Ownership of hygienic latrine.

	(1) HCI Only	(2) Non-HCI Only
Only won latrine	0.111***	0.072***
	(0.025)	(0.024)
Only won tin	0.023	-0.021
	(0.028)	(0.028)
Won both	0.213***	0.172***
	(0.023)	(0.026)
Subsidy Med	0.048	0.058*
	(0.031)	(0.034)
Subsidy High	0.040	0.100**
	(0.032)	(0.040)
Mean ownership %, excluded group	24	32
Num. of neighborhoods	123	102
Num. of households	4,266	3,362

Excluded group: lottery-losers in low-subsidy paras.

Reduced-form household-level OLS
Dependent variable: Ownership of hygienic latrine.
by latrine-lottery winning status

	(1) HCI Won	(2) HCI Lost	(3) Non-HCI Won	(4) Non-HCI Lost
Subsidy Med	0.052	0.038	0.105***	-0.010
	(0.040)	(0.032)	(0.040)	(0.039)
Subsidy High	0.036	0.034	0.127***	0.067
	(0.039)	(0.035)	(0.043)	(0.048)
Mean excluded ownership %	42	27	36	29
Num. of neighborhoods	123	123	102	102
Num. of households	2,760	1,506	2,130	1,232

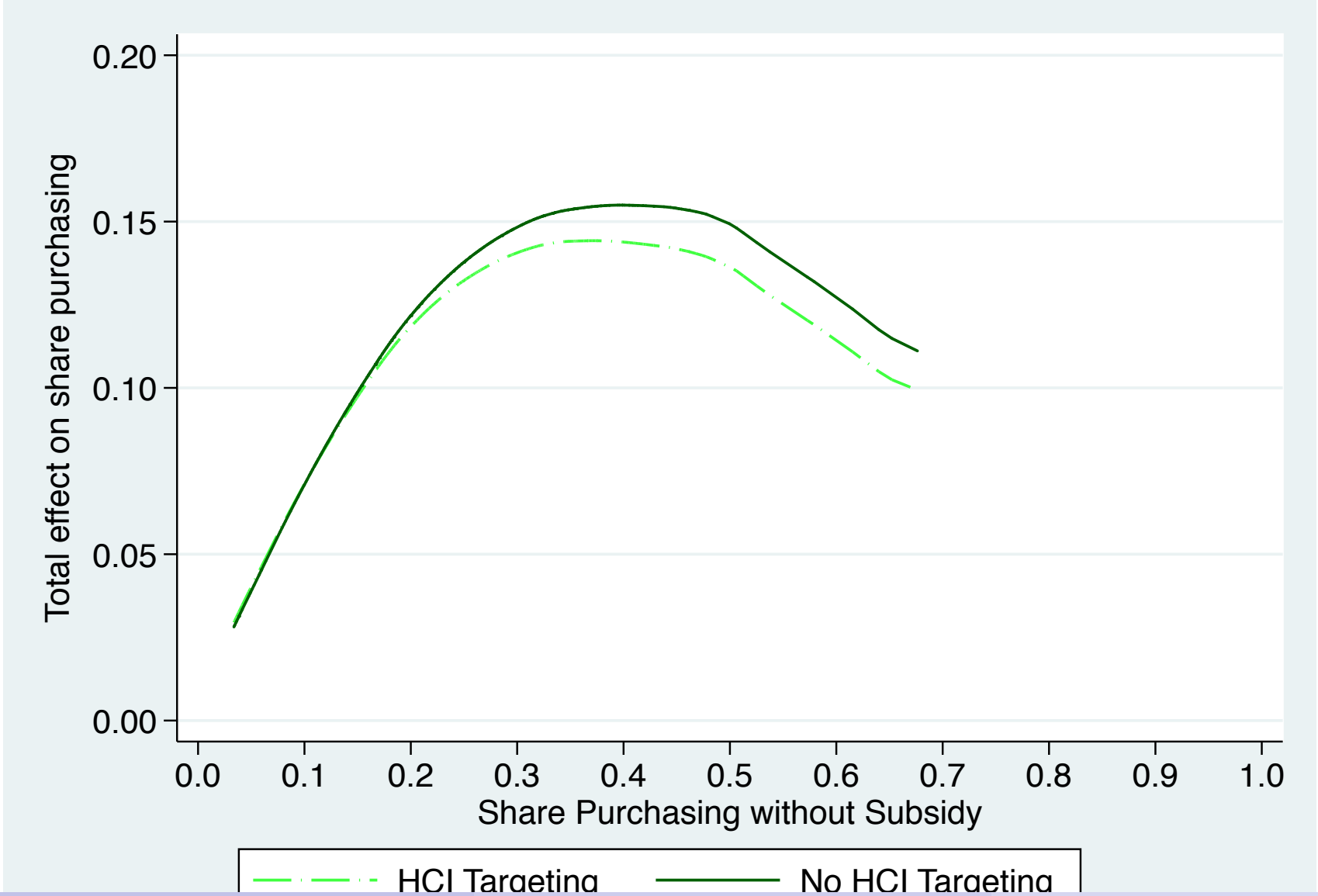
Excluded group: households in low-subsidy paras.

Why doesn't HCI targeting produce higher take-up?

Social Norms and Shame

- Behavior consistent with a model of shame
- If poor/marginal people stop defecating in the open, it's shameful for you continue doing so
  - Especially if we provided you with a subsidy
  - And you *still* don't react
- Not as shameful if richer leaders are investing

Comparison of HCI & non-HCI Experiments, Subsidize 50% at 4000 BDT each



Tapping Social Networks

Do Diffusive and Integrated Networks experience greater adoption spillovers?

1. Construct "adjacency matrix" for social connections
   - kids play together, ask for technical advice, leaders, visit
2. Calculate first eigenvalue: maximum eigenvalue of adjacency matrix.
   - Higher value implies new seeds rapidly reach other points in network.
3. Calculate Second eigenvalue
   - Lower second eigenvalue means a network is more integrated. Iterative updating procedures converge faster.

Second step neighborhood-level IV using village-level first eigenvalue, all measures

	(1)	(2)	(3)	(4)
Share of peers adopting	1.455***	1.640	1.396***	1.619
	(0.404)	(1.208)	(0.405)	(3.315)
Interact peer adoption and level		-0.008		
		(0.047)		
First eigenvalue, level	0.004	0.006		
	(0.003)	(0.015)		
Interact peer adoption and log				-0.072
				(1.045)
First eigenvalue, log			0.076	0.100
			(0.067)	(0.354)
First-step F stat.	16.1	16.1	16.5	16.5
Num. neighborhoods	368	368	368	368

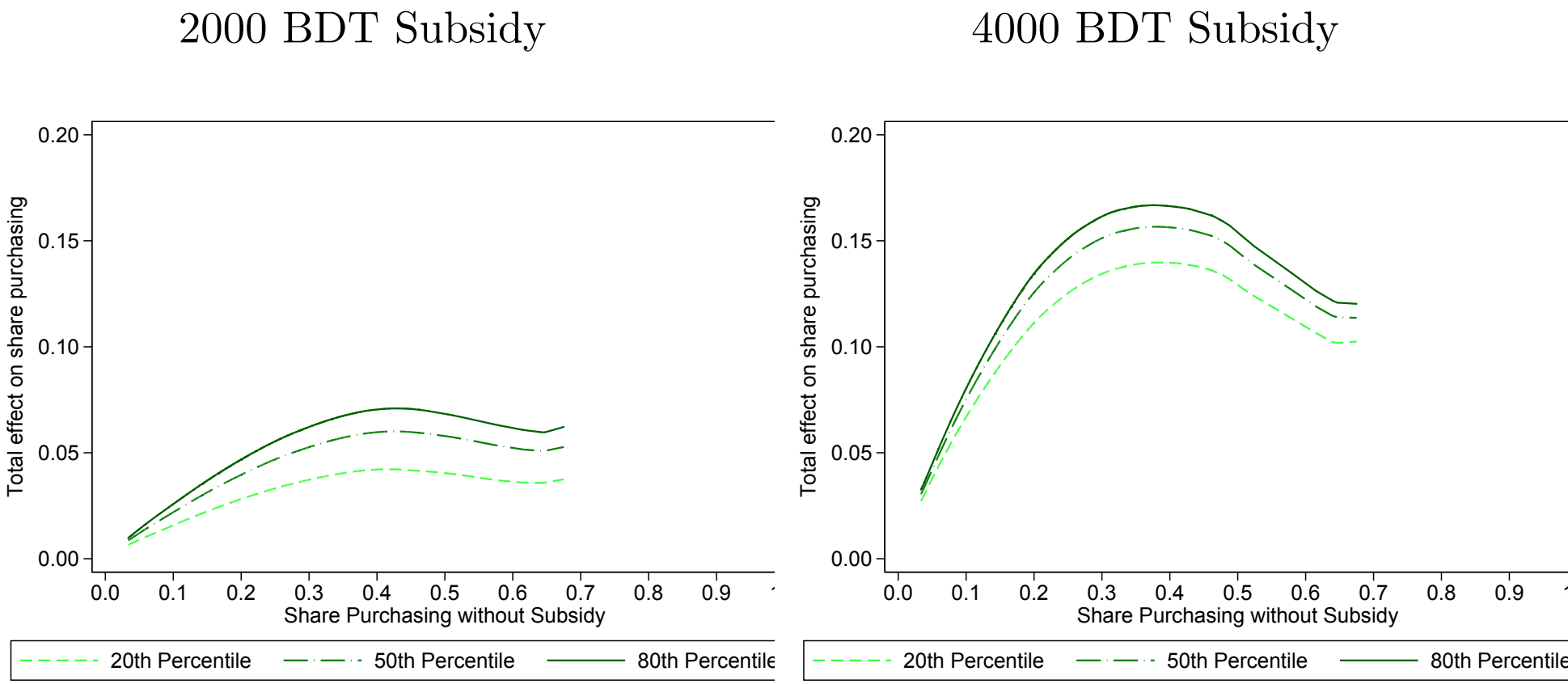
Dependent variable: neighborhood estimated fixed effect from first step

Second step neighborhood-level IV using village-level second eigenvalue, all measures

	(1)	(2)	(3)	(4)
Share of peers adopting	1.429***	9.251***	1.397***	0.083
	(0.413)	(3.218)	(0.416)	(0.788)
Interact peer adoption and level		-9.379**		
		(3.991)		
Second eigenvalue, level	0.440	3.538**		
	(0.344)	(1.393)		
Interact peer adoption and log				-7.131**
				(3.038)
Second eigenvalue, log			0.418	2.677***
			(0.283)	(1.030)
First-step F stat.	16.5	16.5	16.5	16.5
Num. neighborhoods	368	368	368	368

Dependent variable: neighborhood estimated fixed effect from first step
Second eigenvalue transformed by taking absolute value.

Experiment: Give half of households large subsidy, Vary second eigenvalue



All neighborhoods assumed to be at the 20th, 50th, or 80th percentile of second eigenvalue.

Conclusion

- Our empirical methods are applicable broadly
- Many classes of products are strategic complements or substitutes
- Two lessons for evaluating policy in settings with strategic interaction:
  - It's a fixed point problem. Structural estimation necessary.
  - Design RCTs with both market level and individual level variation
- In sanitation, these considerations provide reasons to target subsidies:
  - To the poor
  - In high density areas
  - More intensively in fewer communities
  - Larger subsidies to fewer households
- Targeting subsidies to the poor in dense areas in a concentrated way achieves other social and operational goals anyway.
  - We show it's also efficient to do so!