Causal Effects of Multiple Food Assistance Program Participation on Child Food Insecurity

Helen Jensen, Brent Kreider, and Alex Zhylyevskyy

Human Resources Workshop

Iowa State University

December 8, 2016

Outline

- Motivation and research goals
- Data and key variables
- Basics of our methodology
- Examples of analytical results
- Selected empirical results
- Summary and implications

Food Insecurity in U.S.

Conceptually, **food insecurity** means limited access to food needed for active and healthy life

Growing literature (see Gundersen et al., 2011) indicates food insecurity harms long-run development and health of children

National surveys (e.g., CPS) show substantial prevalence of food insecurity in low-income U.S. population

Coleman-Jensen et al. (2016): In 2015, among households with children and income below 130% poverty:

- **29%** had low food security
- **12%** had very low food security

3.2M food-insecure households

- Also, **20%** (**1.5M** households) had food-insecure children

Food Assistance Programs in U.S.

USDA operates 15 food programs (Oliveira, 2016). Five largest:

- Supplemental Nutrition Assistance Program (**SNAP**, *aka* food stamps)
- National School Lunch Program (**NSLP**)
- Special Supplemental Nutrition Program for Women, Infants, & Children (WIC)
- School Breakfast Program
- Child and Adult Care Food Program

To date, most papers focus on **only one** program. Literature on effects of **multiple** programs is small (e.g., Schmidt et al., 2016)

Many assistance recipients participate in multiple programs. E.g., **35%** of households in our sample are both on SNAP and NSLP

Question: How do food assistance programs interact in creating a food safety net? Are there synergies? Is there a redundancy?

Methodological Challenge

Identifying **causal** effect is difficult even for a single program:

- **Nonrandom selection**: unobservables simultaneously affect food security and program participation
    - ➢ Simple regression methods produce **inconsistent** estimates of causal effects
- **Nonclassical measurement error**: households systematically underreport benefits; misreporting varies across households with different attributes
    - ➢ Standard IV methods produce **inconsistent** estimates as well

Allowing for **multiple** programs adds another layer of complexity:

- Participation cannot be modeled using a binary variable
- Dimensionality of measurement error problem increases

Our approach and methodological contribution:

- We introduce a **partially-ordered** multiple treatment variable to model participation
- We extend partial identification methods of Kreider & Hill (2009), Kreider et al. (2012), which account for selection and measurement error in a single framework

Research Focus

We develop partial identification methodology to study **two** programs

In application, we focus on effects of **SNAP** and **NSLP**

Both are large food programs. In FY 2015 (Oliveira, 2016):

- 46M people participated in SNAP on average per month
- 22M children received free/reduced-price school lunches on average per day
- Annual federal expenditures on SNAP: $74B, on NSLP: $13B

Receipt of benefits is **underreported** in surveys (Meyer et al., 2015):

- 40% of SNAP benefits are not reported in CPS
- 45% underreporting rate for NSLP

Our goal is to account for selection and misreporting and quantify:

- To what extent participation in SNAP & NSLP improves food security compared to no program participation
- To what extent participation in both augments effect of either program alone

Main Data Source

Food Security Supplement (**FSS**) of Current Population Survey (**CPS**)

FSS is administered every December

FSS/CPS provides information on:

- Food security
- Food program participation
- Food expenditures
- Demographic and socioeconomic characteristics

We pool FSS/CPS data for years 2002–2010

Analytical sample: households with **school-age** children and income below **130%** of federal poverty line, ***N* = 10,390**

Supplementary Data Sources

Quarterly Food-at-Home Price Database (**QFAHPD**) provides local food prices based on food purchase transactions in Nielsen Homescan

- Timeframe: every quarter, 1999–2010
- 50+ food groups: e.g., fresh orange vegetables, low fat cheese
- 35 areas partitioning USA: 26 metro + 9 non-metro
- Food group prices are in $ per 100 grams as purchased
- Allows us to construct food expenditure-based **MIVs**

SNAP Policy Database provides state-level policies regarding SNAP eligibility, reporting requirements, use of biometric technology, etc.

- Coverage: every state, every month, 1996–
- Allows us to construct **IVs** for SNAP participation used in the literature:
  - Continuous: e.g., SNAP outreach spending per capita
  - Binary: e.g., fingerprinting, phone certification

Reported Program Participation

Sample distribution by program participation (*N* = 10,390; weighted):

NSLP \ SNAP	*yes*	*no*
yes	34.9%	35.6%
no	5.0%	24.6%

Sample of households with:

- 1+ school-age child
- Income < 130% poverty

FSS question underlying reported NSLP participation:

*During the **past 30 days**, did any children in the household (between 5 and 18 years old) receive **free or reduced-cost lunches** at school?* Yes/no

FSS questions underlying reported SNAP participation:

*In the past 12 months, since December of last year, did (you/anyone in this household) get **SNAP or food stamp benefits**?* Yes/no

*In which months of … were SNAP or food stamp benefits received? **November**?* Yes/no

*In which months of … were SNAP or food stamp benefits received? **December**?* Yes/no

Measuring Child Food Security (I)

Conceptually, **food security** means **access to enough food** for active and healthy life. It implies:

- Ready availability of nutritionally adequate and safe foods, and
- Assured ability to acquire such foods in socially acceptable ways

In practice, **food security status** is assigned using a questionnaire on food-related behaviors under lack of resources

Questions can focus on household, adults, or children. E.g.:

- *Did any of the children ever skip a meal because there wasn't enough money for food?* Yes/No
- *Did any of the children ever not eat for a whole day because there wasn't enough money for food?* Yes/No

Measuring Child Food Security (II)

FSS has 18-item **Household Food Security Survey Module**

8 items are child specific (answered by adult proxy)

Responses are scored (0-1) and summed. Summary score is used to construct **four categories** of child food security:

- High (*score = 0*), marginal (*1*), low (*2–4*), and very low (*5+*)

We use child food security status referenced to **past month**. In our sample (drawn from households with income below 130% poverty):

Food secure (**88.3%**)
(1) High
(2) Marginal

Food insecure (**11.7%**)
(3) Low
(4) Very low (**1.5%**)

Food Security by Participation

Prevalence of child food security by food program participation in our sample (*N* = 10,390; weighted):

		SNAP yes	SNAP no
NSLP	*yes*	86.3%	86.6%
	no	87.0%	93.9%

- Reference period for child food security and program participation: **past month**
- A cell shows **percentage** of households with food secure children in subsample

Lower prevalence of child food security under program participation is likely due to **selection**

Motivation for Our Methodology

A simple parametric approach:

$$\underbrace{Y_i}_{\text{Outcome}} = \gamma \cdot \underbrace{S_i}_{\text{Treatment}} + \underbrace{x_i}_{\text{Covariates}} \cdot \beta + \underbrace{\varepsilon_i}_{\text{Error term}}$$

Treatment S_i is **binary**. Say, $S_i = 1$ if i is on SNAP, 0 if not

If same unobservables affect S_i and Y_i, then $cov(S_i, \varepsilon_i) \neq 0$ and OLS is inconsistent due to **endogeneity**

Measurement error in S_i is **nonclassical**. Thus, standard IV estimation is inconsistent as well

Our **nonparametric bounding** methodology handles endogeneity, misreporting, and multiple treatments (not just *binary* S_i). Also, it allows for heterogeneous response to treatment across i

Basics of Our Approach: Notation

$\boldsymbol{S^*}$: **true** program participation status:

- $S^* = 0$: neither SNAP nor NSLP
- $S^* = 1$: SNAP alone
- $S^* = 2$: NSLP alone
- $S^* = 3$: both SNAP and NSLP

S^* is **partially ordered**

$\boldsymbol{S}$: **reported** program participation; S need not equal S^*

Potential outcomes framework:

$\boldsymbol{Y(S^*)}$: potential outcome under treatment S^*

- $Y = 1$ if children are food secure, $Y = 0$ otherwise

$\boldsymbol{X}$: covariates (some used as instruments)

Basics of Our Approach: ATE

We focus on **average treatment effects** (**ATEs**):

$$ATE_{jk} = P[Y(S^* = j) = 1 \mid X] - P[Y(S^* = k) = 1 \mid X] \text{ for } j \neq k$$

For example, consider ATE_{31}:

$$ATE_{31} = P[Y(S^* = 3) = 1 \mid X] - P[Y(S^* = 1) = 1 \mid X]$$

ATE_{31} measures by how much likelihood of child food security would change if household were to participate in both SNAP and NSLP vs. in SNAP alone

There are no regression orthogonality conditions to satisfy

Covariates are only used to specify subpopulations

Decomposition Strategy

ATE cannot be point-identified without assumptions even if $S \equiv S^*$

We decompose formulas into what is identified and what is not

Let's simplify notation: $ATE_{31} = P[Y(3)=1] - P[Y(1)=1]$

Consider decomposition:

$$P[Y(3)=1] = \underbrace{P[Y(3)=1 \mid S^* = 3]}_{\text{identified}} \underbrace{P(S^* = 3)}_{\text{identified}} + \underbrace{P[Y(3)=1 \mid S^* \neq 3]}_{\text{not identified}} \underbrace{P(S^* \neq 3)}_{\text{identified}}$$

Data cannot identify $P[Y(3)=1 \mid S^* \neq 3]$ because it refers to unobserved **counterfactual**. We only know that $P[Y(3)=1 \mid S^* \neq 3] \in [0,1]$

However, extending methods of Manski (1995), we derive worst-case bounds for $P[Y(3)=1]$, $P[Y(1)=1]$, and ATE_{31}

Addressing Misreporting

When S may deviate from S^*, define: $\theta_i^{j,k} \equiv P(Y=i, S=j, S^*=k)$

$P[Y(3)=1]$ becomes:

$$P[Y(3)=1] = P(Y=1, S=3) + \theta_1^{-3,3} - \theta_1^{3,-3}$$
$$+P[Y(3)=1 \mid S^* \neq 3]\left\{P(S \neq 3) + \sum_{j \neq 3}(\theta_1^{-j,j} + \theta_0^{-j,j} - \theta_1^{j,-j} - \theta_0^{j,-j})\right\}$$

ATE_{31} can be bounded as:

$$-P(Y=0, S \neq 1) - P(Y=1, S \neq 3) + \Theta_{3,1}^{LB}$$
$$\leq ATE_{3,1} \leq$$
$$P(Y=0, S \neq 3) + P(Y=1, S \neq 1) + \Theta_{3,1}^{UB}$$

unobserved

$$\Theta_{3,1}^{LB} \equiv \theta_1^{-3,3} - \theta_1^{3,-3} + \theta_0^{-1,1} - \theta_0^{1,-1},\ \Theta_{3,1}^{UB} \equiv -\theta_0^{-3,3} + \theta_0^{3,-3} - \theta_1^{-1,1} + \theta_1^{1,-1}$$

Tightening Bounds (I)

Without assumptions, bounds on ATEs are wide and **contain zero**

To **tighten** bounds, we can impose restrictions on:

1) Misreporting process
2) Selection process

Consider **restricting misreporting** process. We can:

- Exploit logical constraints on probabilities and auxiliary data to restrict θ's. E.g.,

$$\theta_1^{-1,1} \leq \min\{P(Y=1, S \neq 1), P(S^* = 1)\}$$

- Restrict **prevalence** of misreporting. E.g., constrain value of $\Delta_j \equiv P_j^* - P_j$
- Restrict **pattern** of misreporting. Say, impose "**No-stigma verification**" assumption: Household with $S > 0$ is presumed to provide accurate participation information. Household with $S = 0$ can misreport. This assumption implies "**no false positives**" (on average) and zeroes out several θ's

Tightening Bounds (II)

To **restrict selection process**, we can employ:

- **Exogenous selection** assumption [often does **not** hold, though]
- Monotone treatment selection (**MTS**) assumption (Manski & Pepper, 2000)
- Monotone treatment response (**MTR**) assumption (Manski, 1995)
  - We extend MTS and MTR to partially ordered unobserved treatments
- Monotone instrumental variables (**MIVs**, Manski & Pepper, 2000)
- Instrumental variables (**IVs**). Say, use IVs for SNAP (Ratcliffe et al., 2011)

We can **combine assumptions** to further tighten bounds on ATEs

Example of Analytical Results

Proposition 2(ii)(B):

Under "**no-stigma verification**" with **endogenous selection**, bounds on $ATE_{3,1}$ are as follows:

- **Lower bound**:

$$ATE_{3,1}^{LB} = -P(Y=1, S \neq 3) - P(Y=0, S \neq 1) + \max\{0, \Delta_3 - P_{000}\} + \max\{0, \Delta_1 - P_{100}\}$$

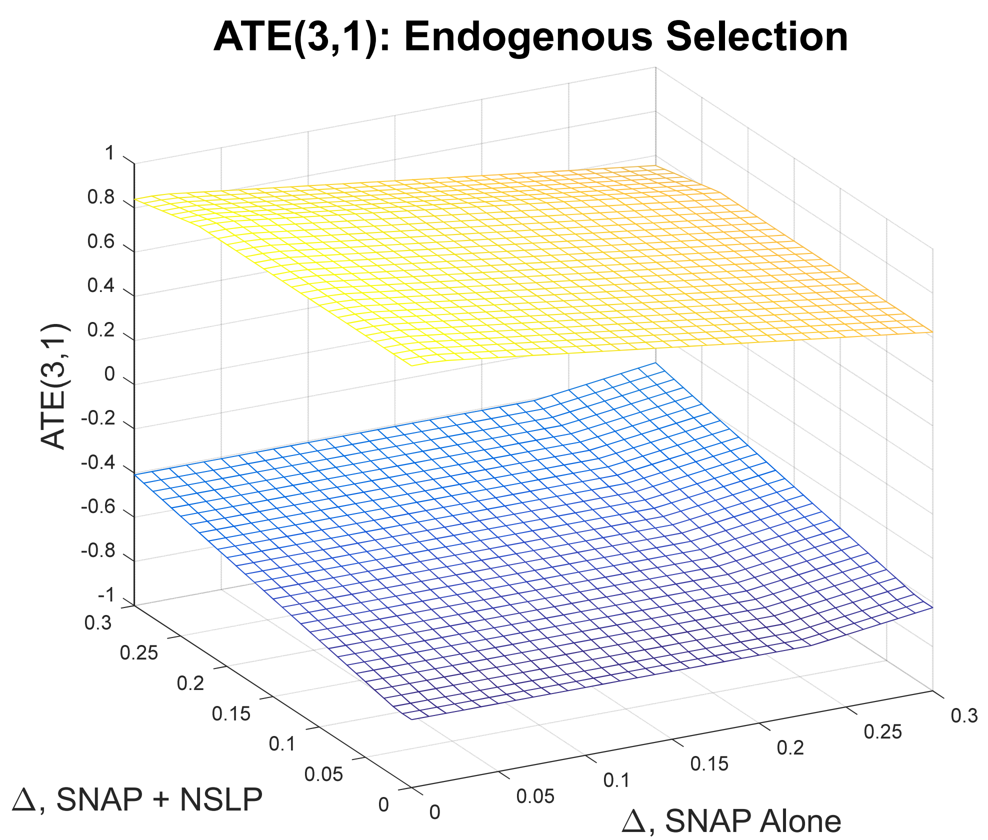
- **Upper bound**:

$$ATE_{3,1}^{UB} = P(Y=0, S \neq 3) + P(Y=1, S \neq 1) - \max\{0, \Delta_3 - P_{100}\} - \max\{0, \Delta_1 - P_{000}\}$$

$\Delta_1 \equiv P_1^* - P_1$, $\Delta_3 \equiv P_3^* - P_3$, $P_{000} \equiv P(Y=0, S=0, V=0)$,
$P_{100} \equiv P(Y=1, S=0, V=0)$

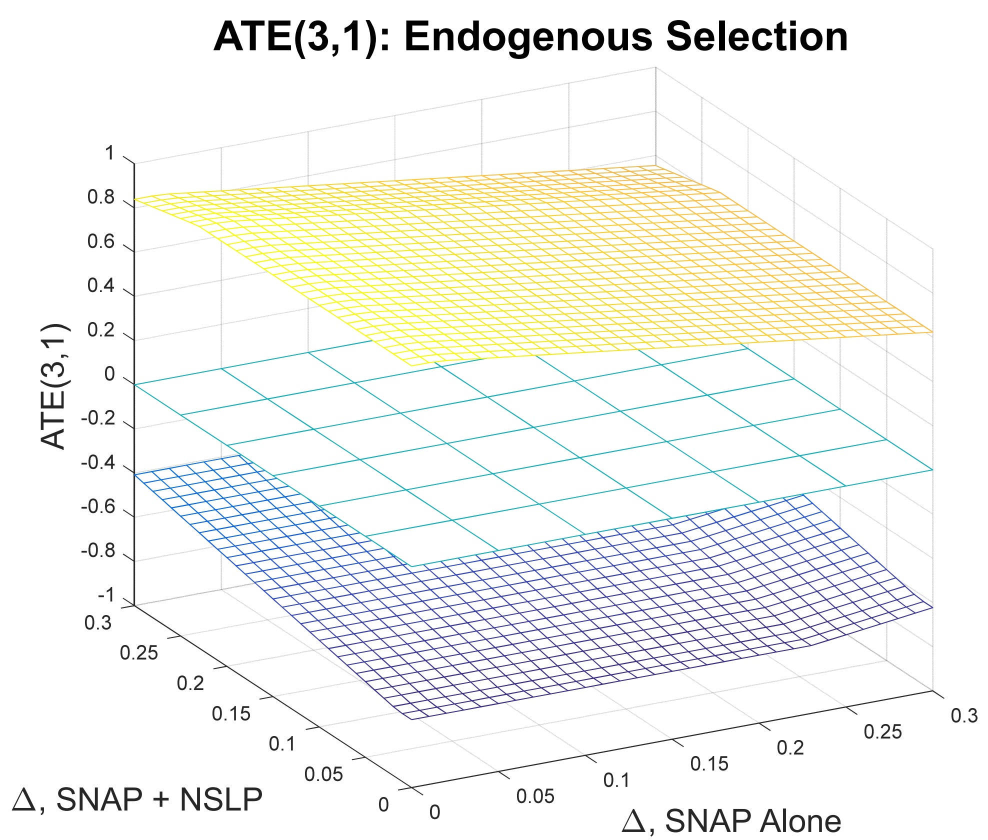
Results: Worst-Case Bounds

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Results: Worst-Case Bounds

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Exogenous Selection: Definition

Exogenous selection:

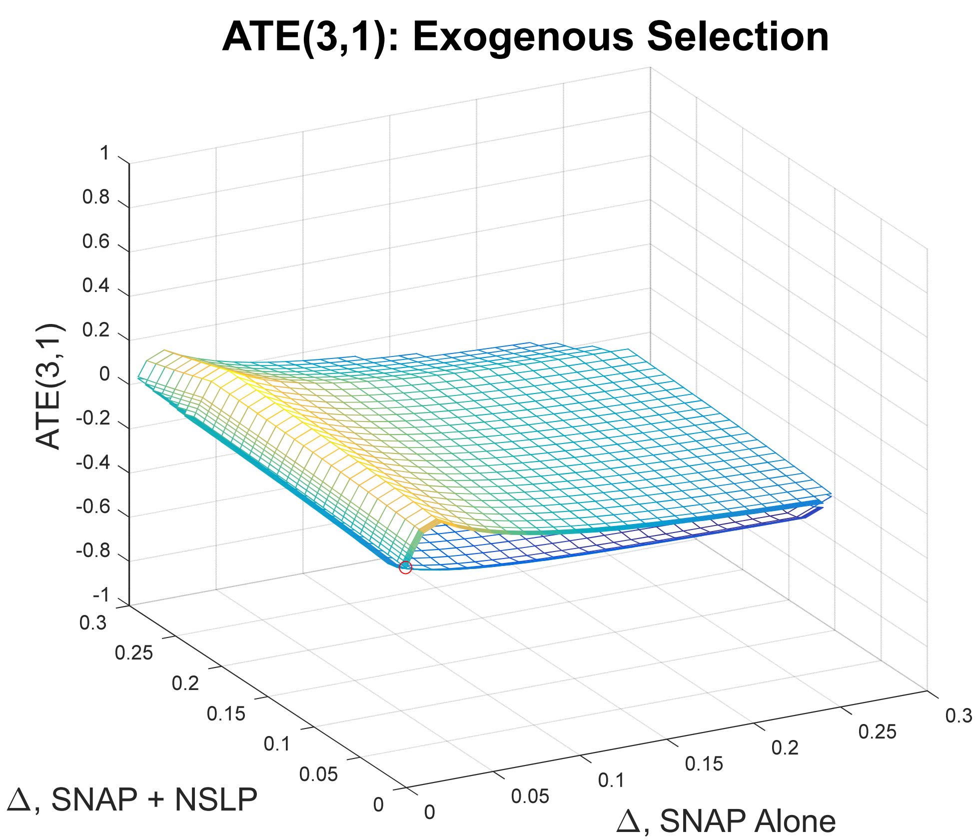
$$P[Y(j)=1]=P[Y(j)=1\,|\,S^{*}=k]\quad \forall j,k$$

Assumption means that expected potential outcomes do not depend on realized treatment

Assumption makes sense when assignment to programs is truly **random**

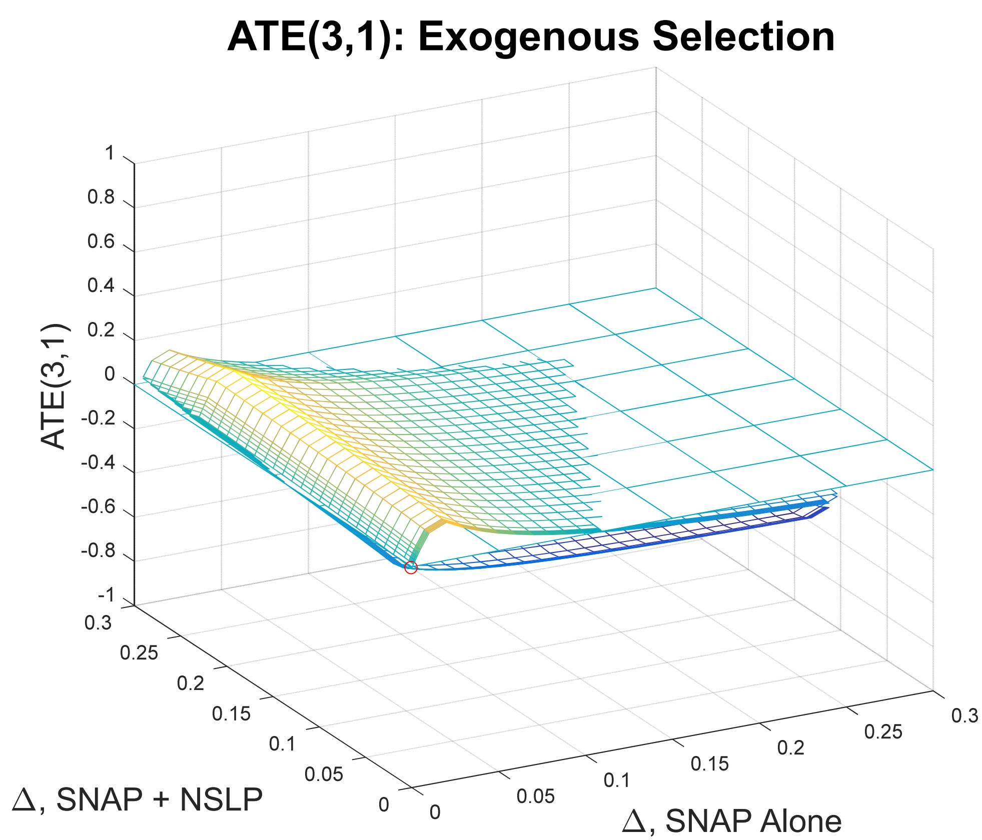
Exogenous Selection

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



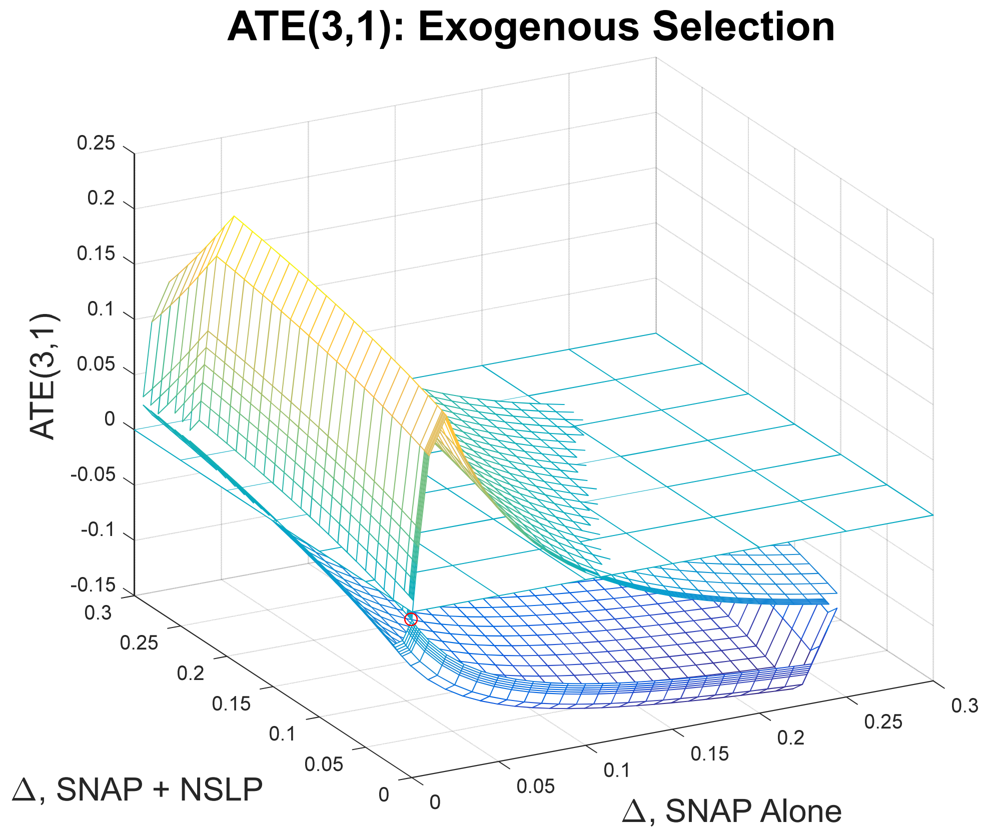
Exogenous Selection

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Exogenous Selection: Closer View

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Exog. Selection: Identification Decay

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:

		$\Delta_1 = 0$			$\Delta_1 = 0.01$			$\Delta_1 = 0.10$		
		LB	UB	width	LB	UB	width	LB	UB	width
$\Delta_3 = 0$	p.e.	[-0.007,	-0.007]	0.000	[-0.029,	0.14]	0.167	[-0.094,	0.007]	0.101
	CI	[-0.040,	0.026]		[-0.051,	0.16]		[-0.106,	0.022]	
$\Delta_3 = 0.01$	p.e.	[-0.031,	-0.004]	0.028	[-0.053,	0.14]	0.195	[-0.118,	0.010]	0.129
	CI	[-0.057,	0.022]		[-0.075,	0.17]		[-0.130,	0.025]	
$\Delta_3 = 0.10$	p.e.	[-0.010,	0.023]	0.034	[-0.032,	0.17]	0.201	[-0.097,	0.037]	0.134
	CI	[-0.036,	0.049]		[-0.054,	0.19]		[-0.108,	0.051]	

Identification deteriorates with extent of underreporting of SNAP

MTS: Definition

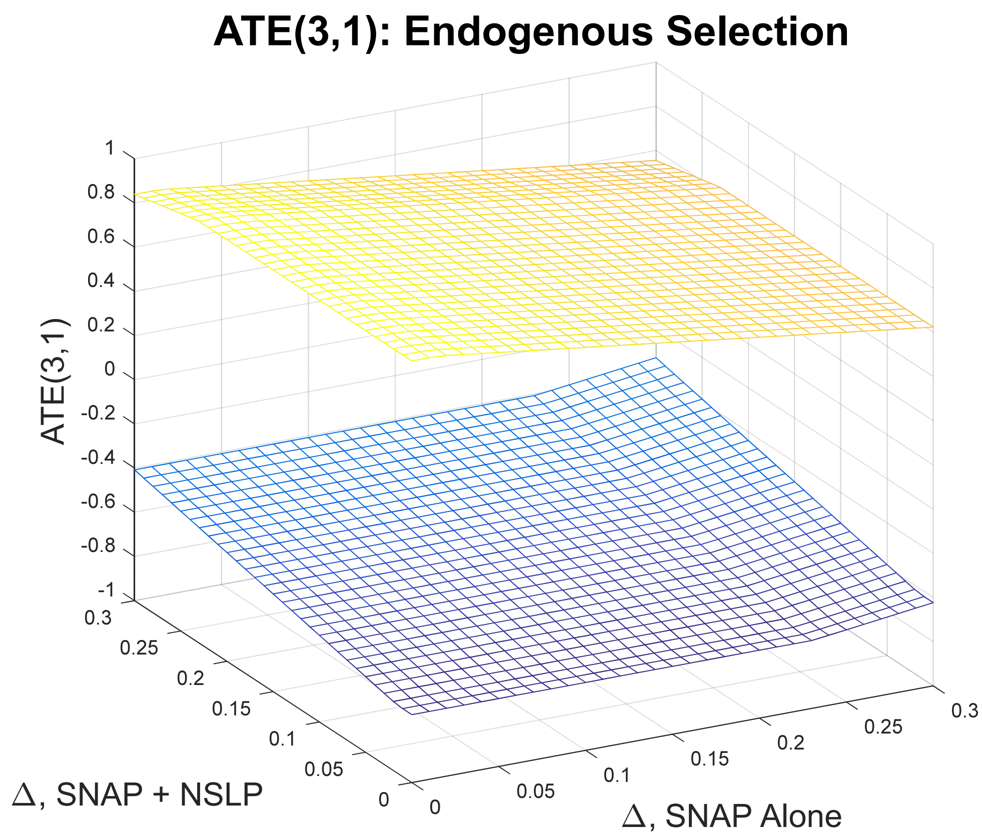
Monotone treatment selection (**MTS**):

$$P[Y(j)=1 \mid S^{*}=3] \le P[Y(j)=1 \mid S^{*}=k] \le P[Y(j)=1 \mid S^{*}=0] \quad \forall\, j;\, k=1,2$$

Under MTS assumption, decision to participate is monotonically related to food insecurity: households choose to participate in more programs in anticipation of worse food security situation

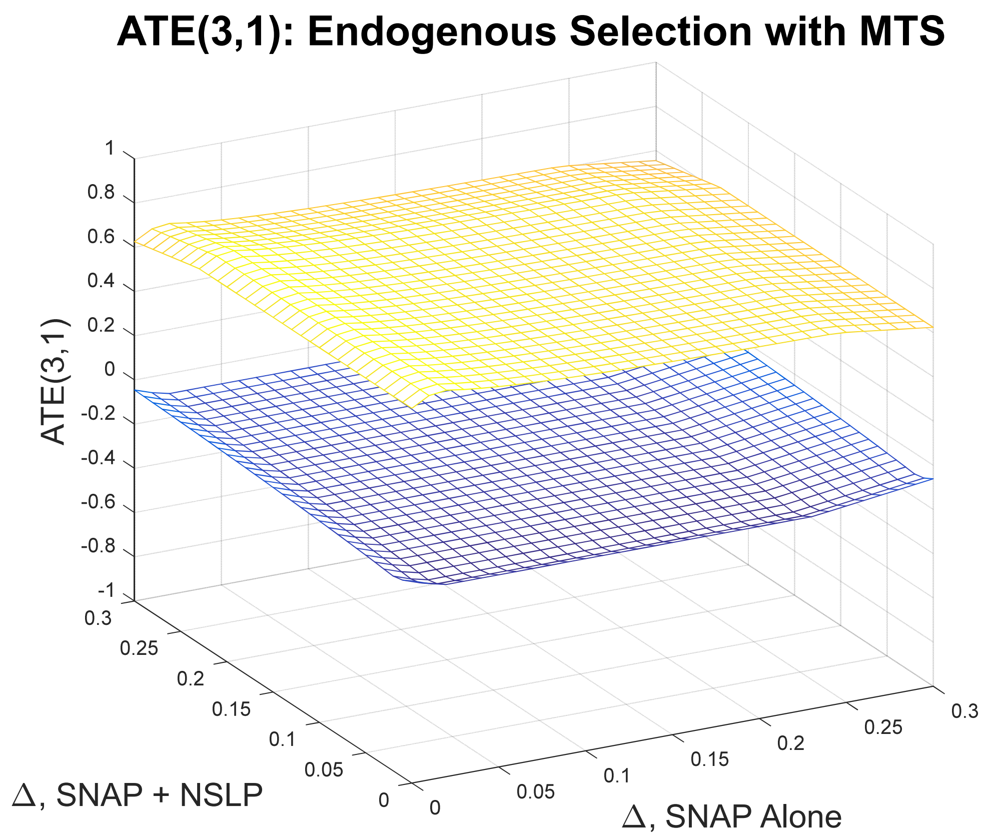
Recall: Worst-Case Bounds

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



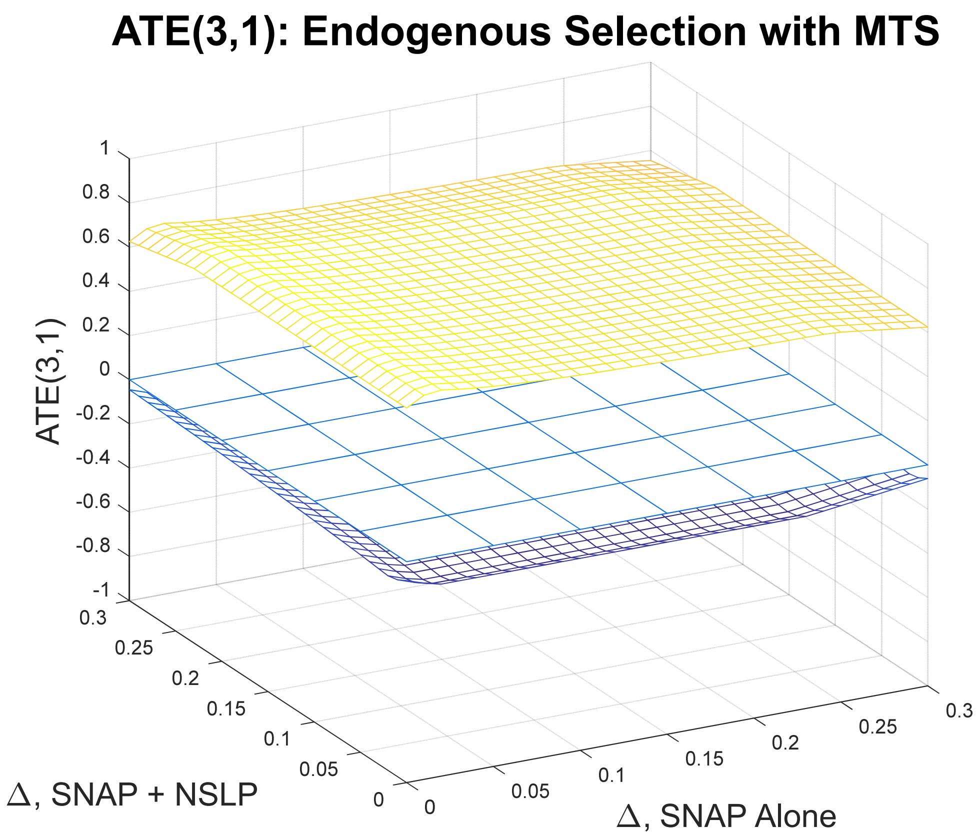
Endogenous Selection with MTS

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Endogenous Selection with MTS

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



MTR: Definition

Monotone treatment response (**MTR**):

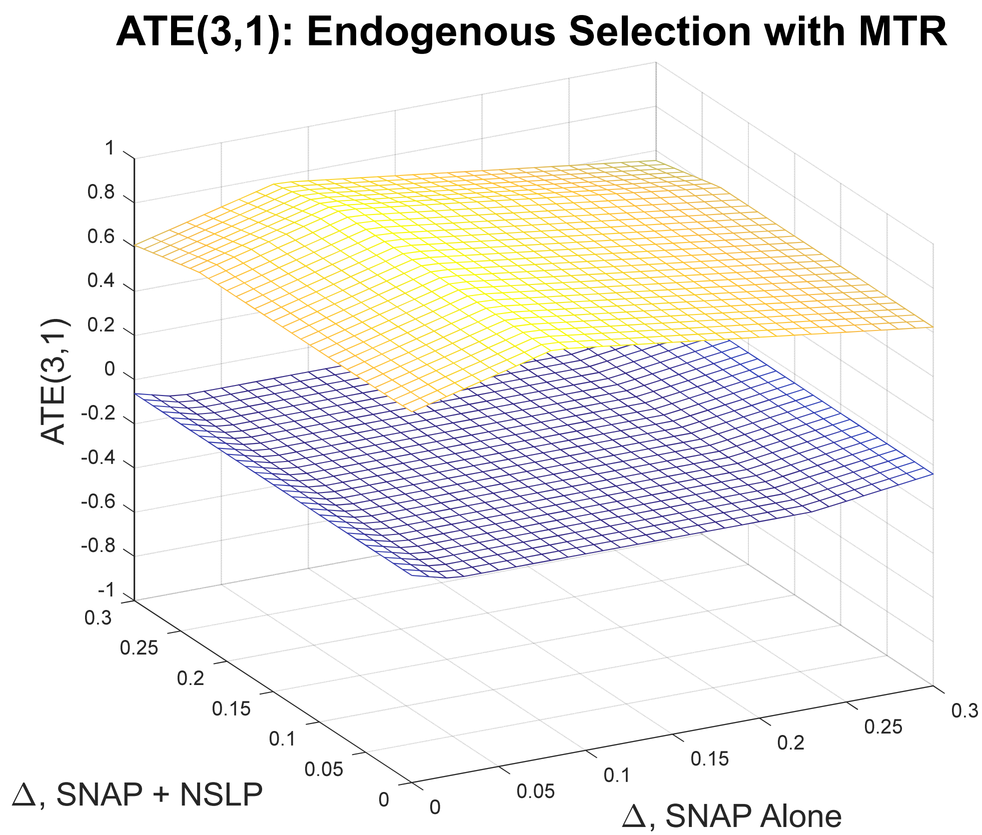
$$P[Y(3)=1 \mid S^*] \geq P[Y(1)=1 \mid S^*] \geq P[Y(0)=1 \mid S^*]$$

$$P[Y(3)=1 \mid S^*] \geq P[Y(2)=1 \mid S^*] \geq P[Y(0)=1 \mid S^*]$$

Under MTR assumption, potential participation in (more) food programs would not harm food security on average

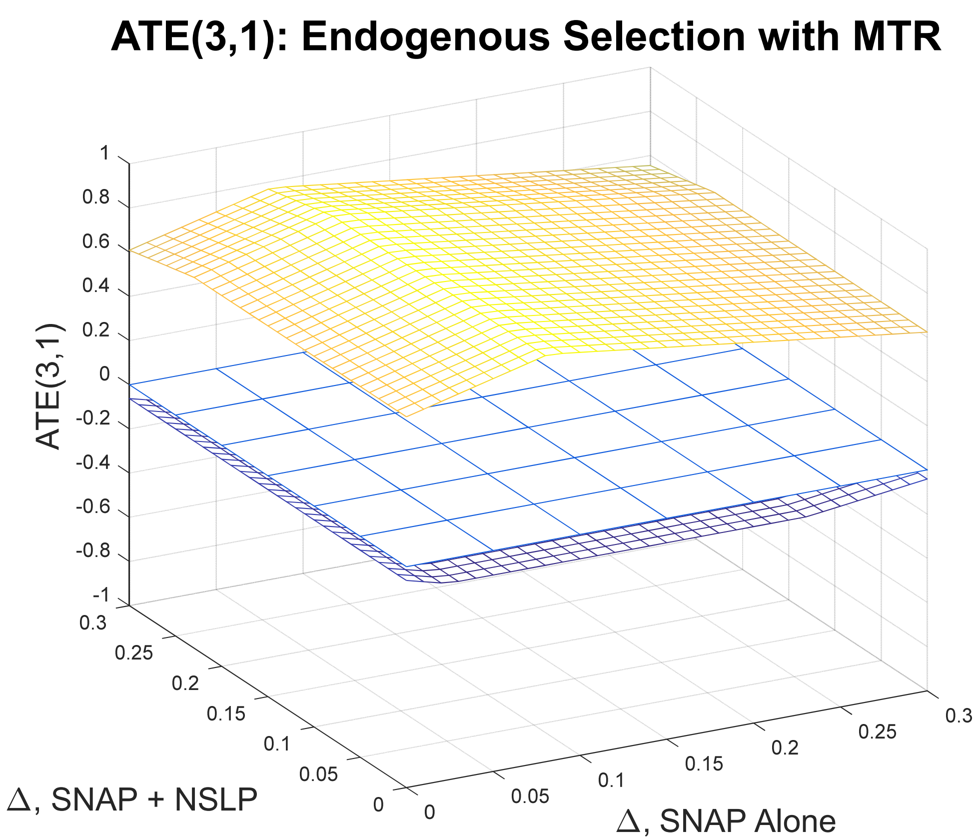
Endogenous Selection with MTR

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Endogenous Selection with MTR

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



MIV: Definition

Monotone instrumental variable (**MIV**):

$$u_1 \le u \le u_2 \Rightarrow$$

$$P[Y(j)=1 \mid v = u_1]$$

$$\le P[Y(j)=1 \mid v = u] \le$$

$$P[Y(j)=1 \mid v = u_2]$$

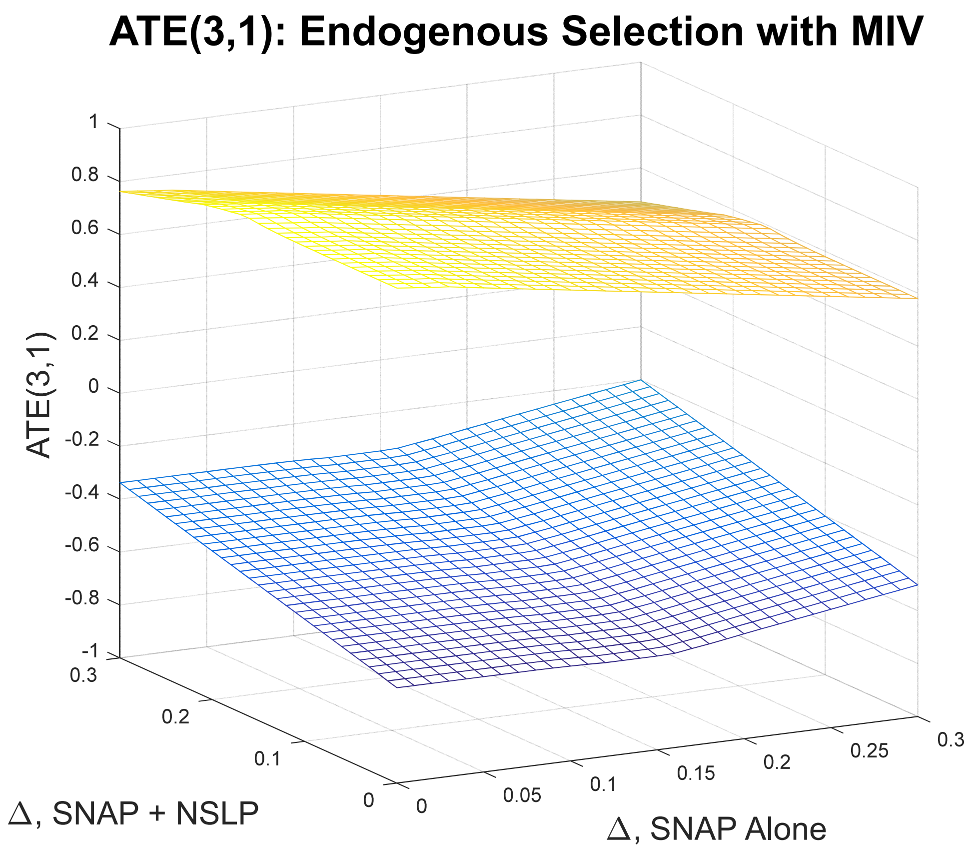
We construct and use:

$$v = \frac{\text{Usual household food expenditures}}{\text{TFP-based minimum expenditures}}$$

Assumption: higher v would not harm food security on average

Bounds under MIV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



IV: Definition

Instrumental variable (**IV**):

$$\forall u_1, u_2 :$$
$$P[Y(j)=1 \mid v=u_1] = P[Y(j)=1 \mid v=u_2]$$

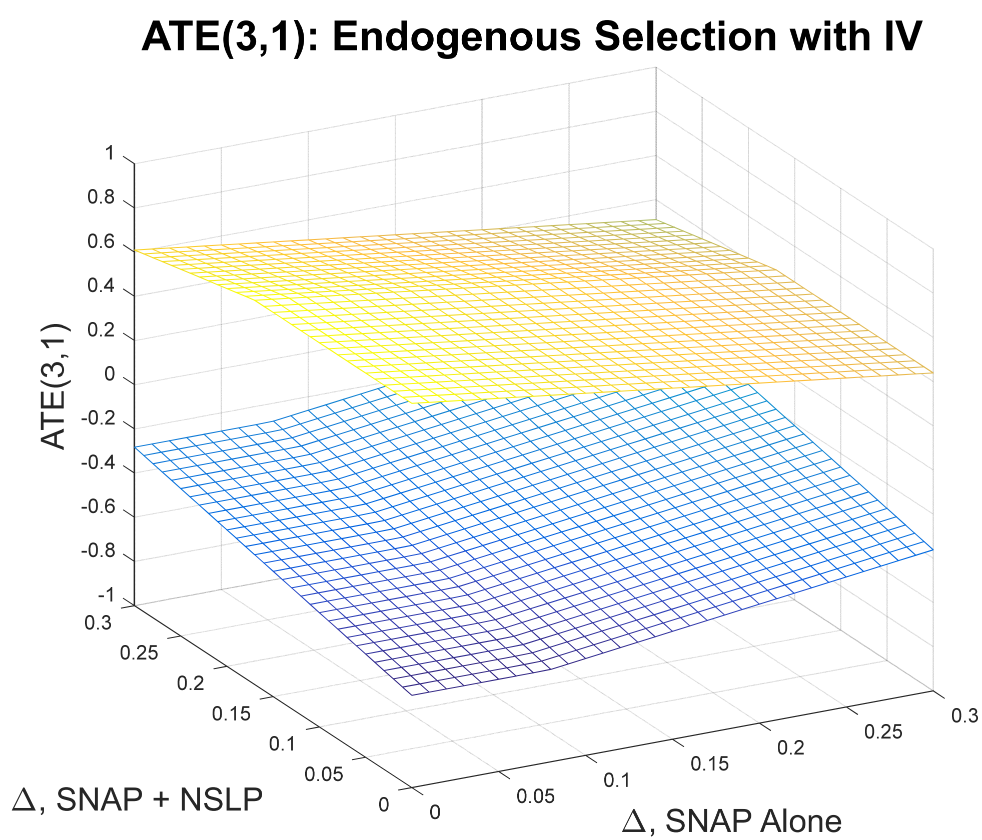
IV is a special case of MIV

We employ **SNAP Policy Database** to construct conventional IVs used in previous literature to instrument for SNAP participation. Many such IVs are binary

We create a scalar IV with many values by combining seven conventional IVs

Bounds under IV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



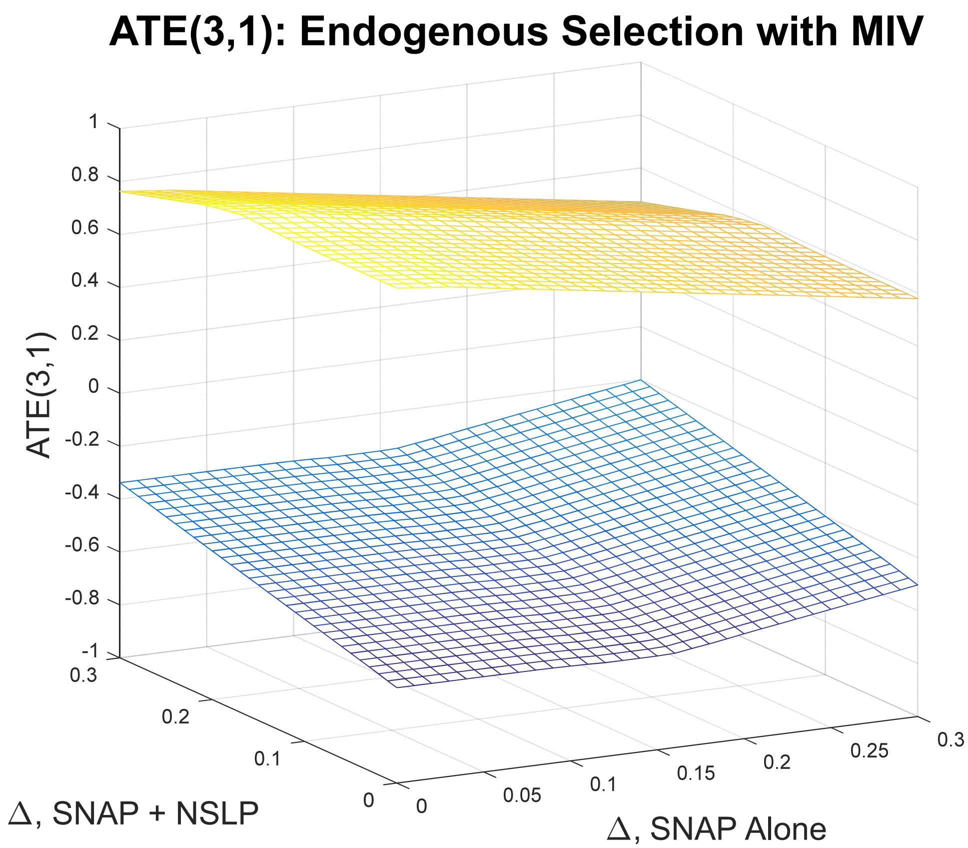
Combining Assumptions

We can combine monotonicity assumptions to **further tighten bounds**

In many cases, $ATE_{3,1}$ can be identified as **strictly positive** even in the presence of substantial classification error

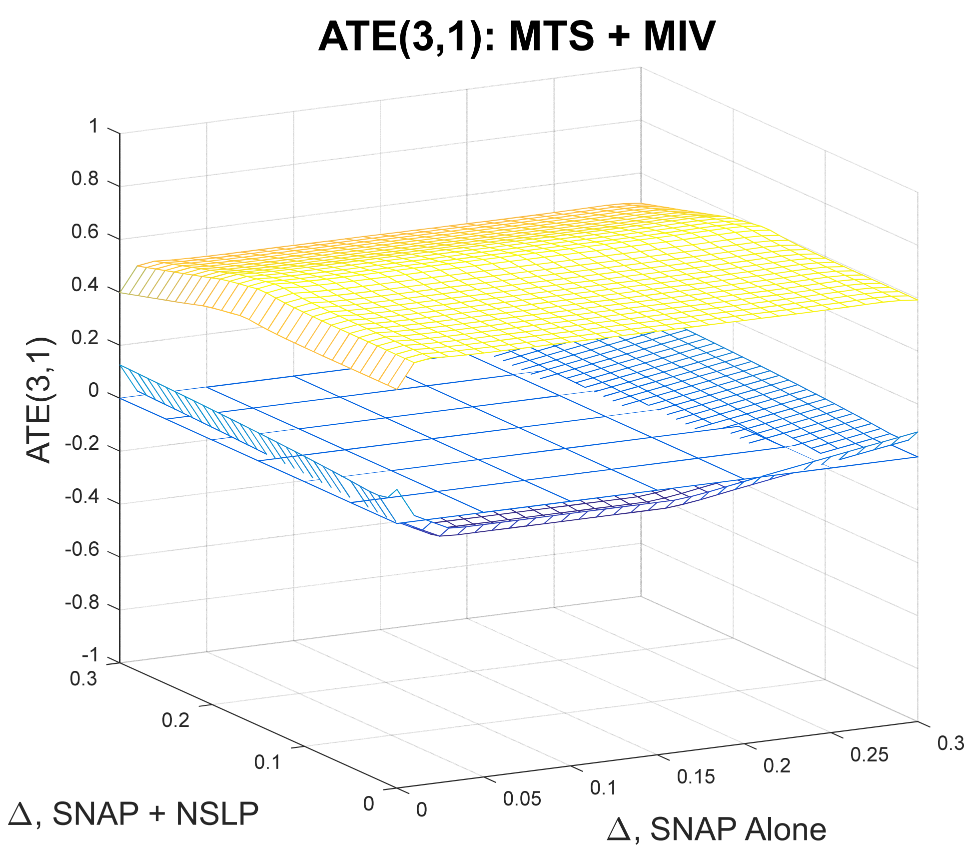
Recall: Bounds under MIV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



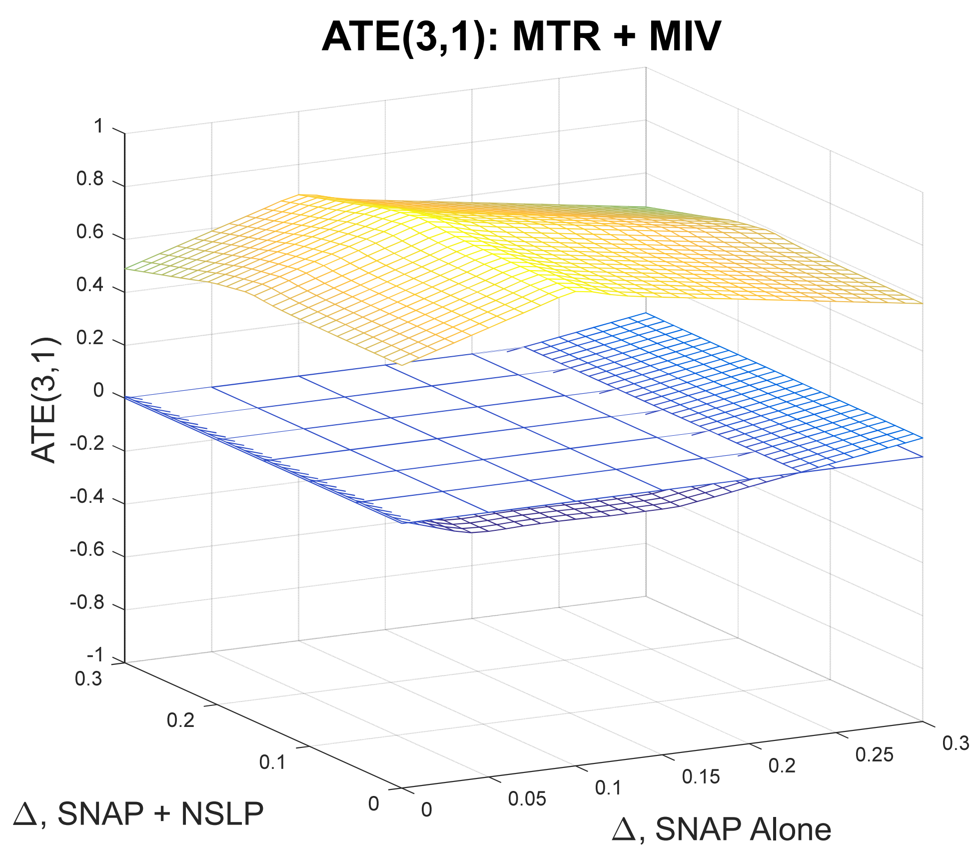
Bounds under MTS + MIV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



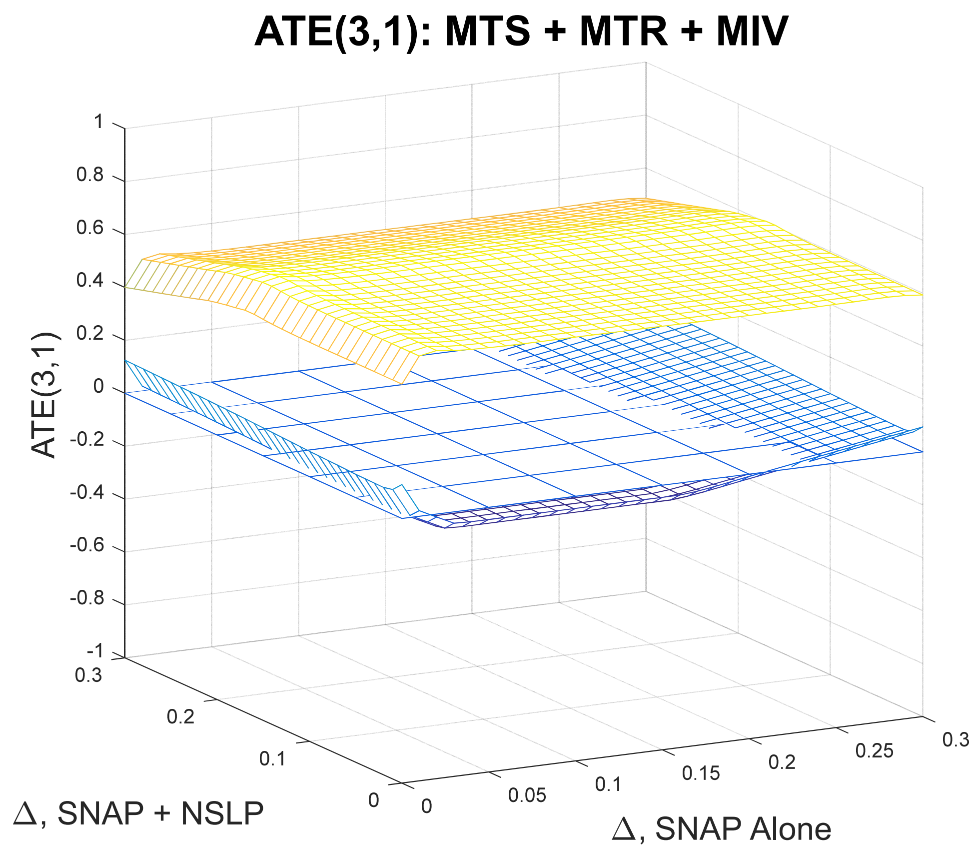
Bounds under MTR + MIV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



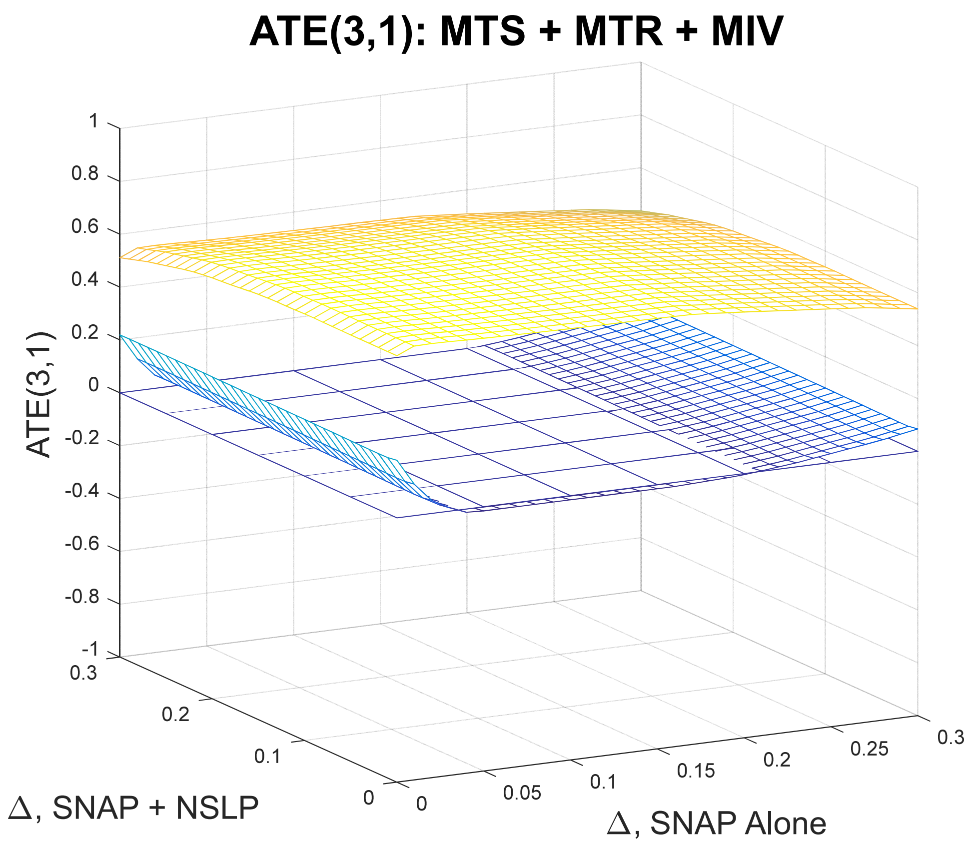
Bounds under MTS + MTR + MIV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



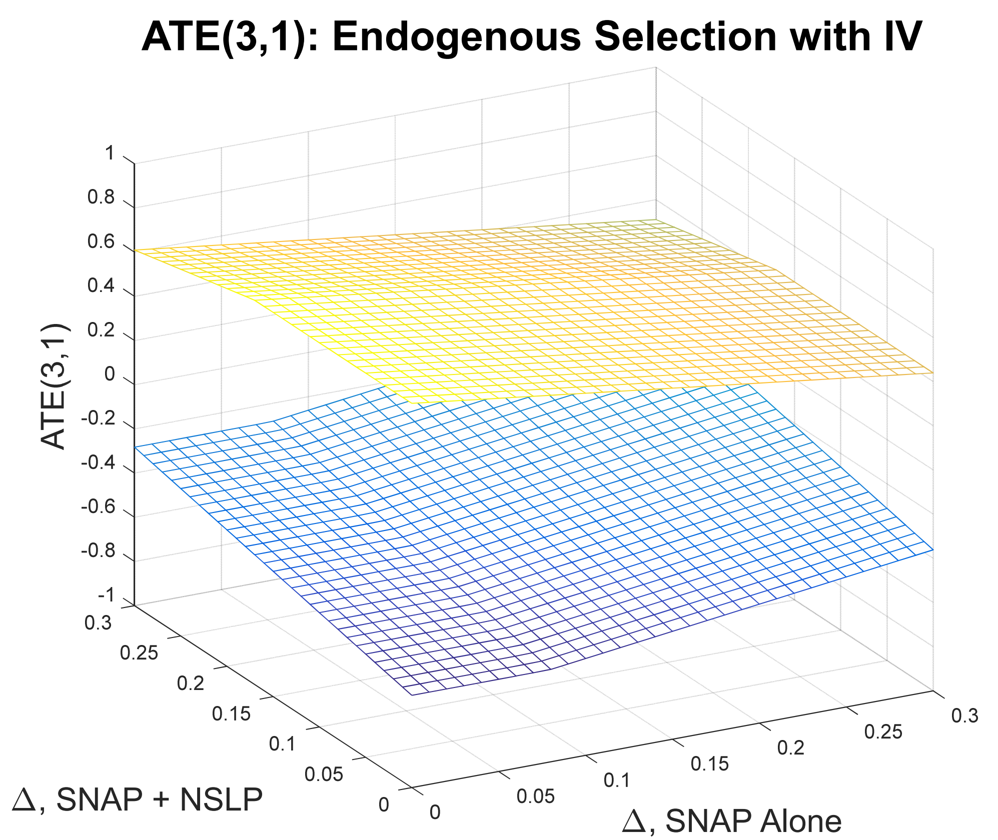
Bounds under MTS + MTR + MIV

Alternatively, consider “income” MIV:



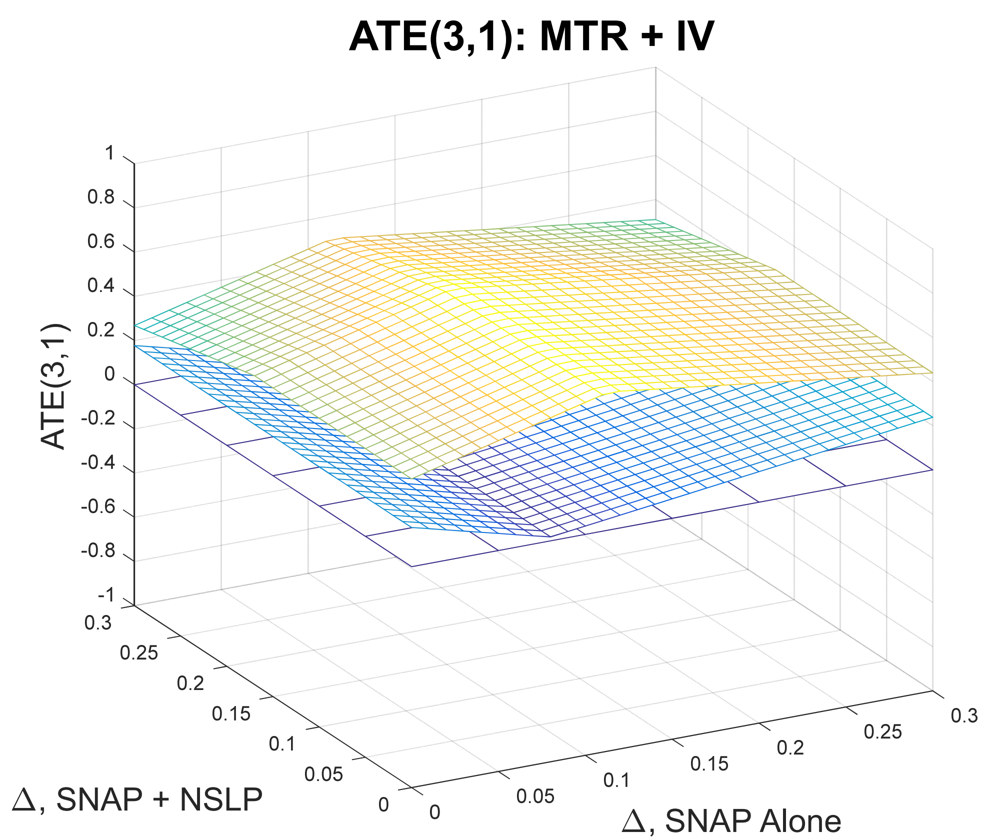
Recall: Bounds under IV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



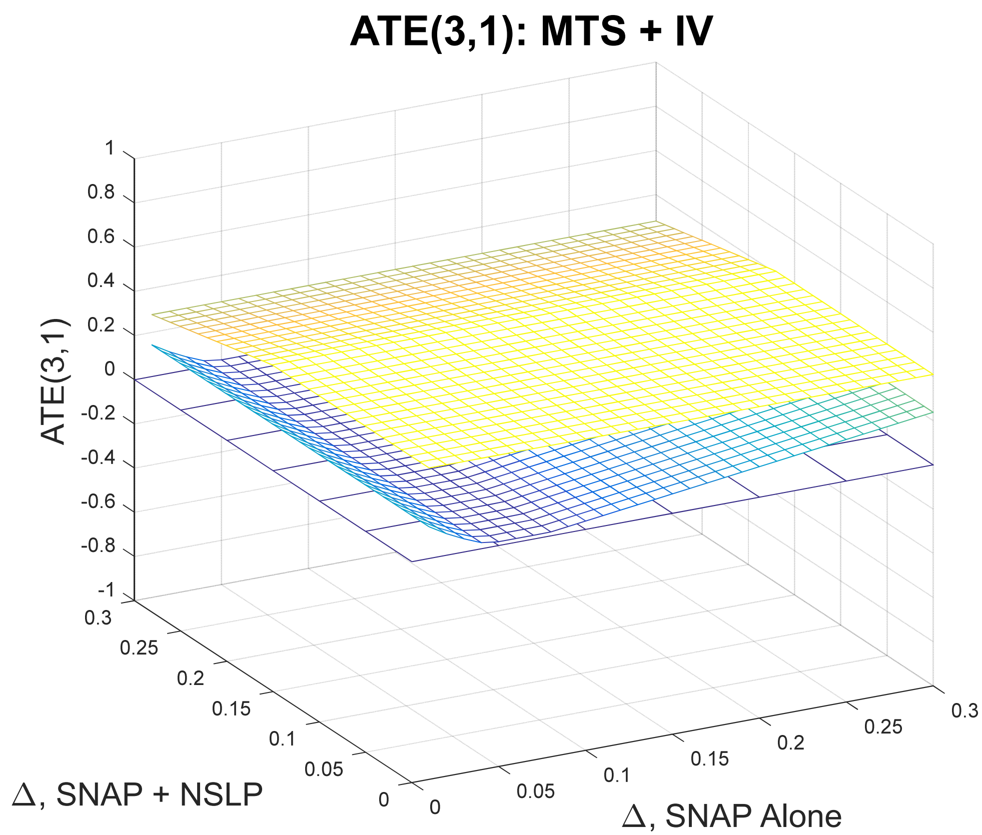
Bounds under MTR + IV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



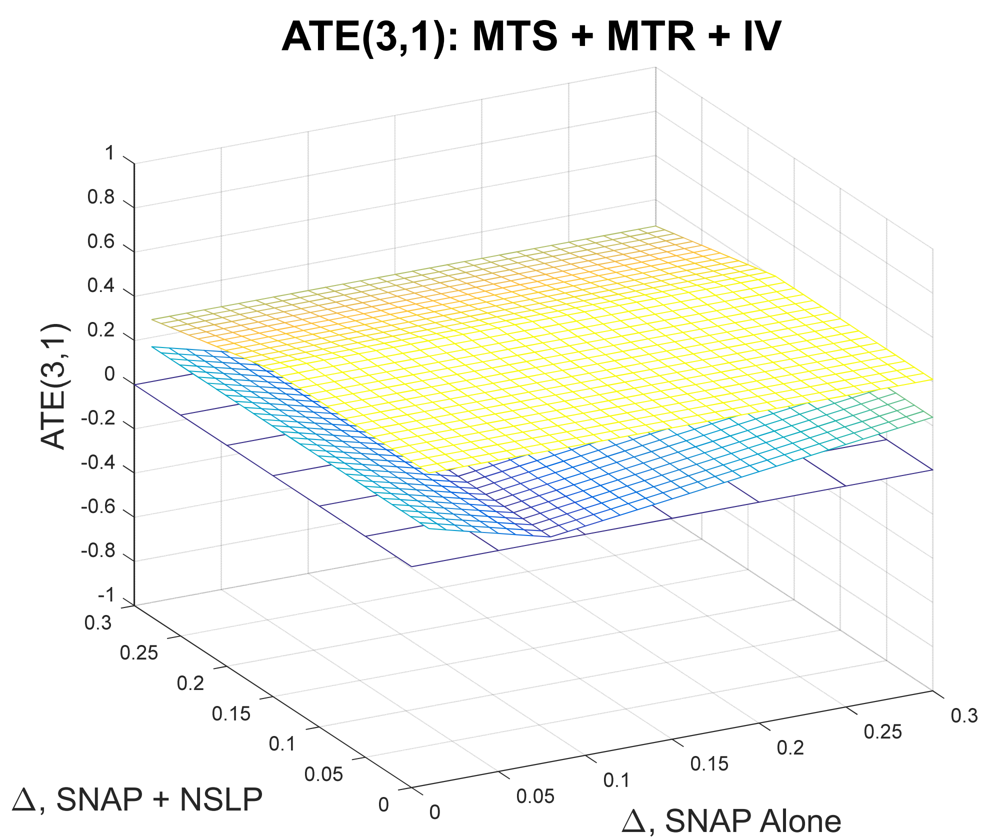
Bounds under MTS + IV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Bounds under MTS + MTR + IV

Bounds on ATE of participating in SNAP and NSLP vs. SNAP alone:



Summary

Motivating question: How do existing food programs interact in creating a food safety net?

Research objective: Quantify by how much SNAP+NSLP improves child food security relative to SNAP alone or NSLP alone and relative to nonparticipation

Policy relevance: Research informs better alignment and design of food programs

Methodology: Nonparametric bounding approach handles endogeneity, misreporting, multiple partially ordered treatments

Selected results: Bounding under MTS + IV or MTR + IV shows SNAP+NSLP improves child food security on top of effect of SNAP alone

Thank you!
Questions?