Investigating Causal Effects of SNAP and WIC on Food Insecurity Using FoodAPS

Helen Jensen, Brent Kreider, and Oleksandr Zhylyevskyy

Iowa State University

SEA 2017 Annual Meeting

November 19, 2017

Research funded by a grant to NBER from USDA

Background

Supplemental Nutrition Assistance Program (SNAP)

- Targets low-income persons
- Provides targeted benefits to households for food purchase
- Eligibility:
  - Income ≤ 130% poverty (before deductions)
  - Employment requirements

In 2016, **35.7%** of low-income households were **food insecure**

Special Supplemental Nutrition Program for Women, Infants and Children (WIC)

- Target population is low-income, nutritionally at-risk pregnant, breastfeeding, other post-partum women, infants, children < 5 y.o.
- Provides "vouchers" for foods in WIC package
- Eligibility:
  - Income ≤ 185% poverty
  - Or, automatic income eligibility (participation in Medicaid, TANF or selected other programs)

Research Focus

To what extent does participation in both SNAP and WIC increase household food security compared to participation in SNAP alone or in WIC alone?

- **Econometric objective**: Derive sharp bounds on average treatment effects (ATEs) of joint program participation when it is endogenous and can be misreported
  - Bounds must be logically consistent with observed data and any imposed statistical or behavioral assumptions
- **Additional objective**: exploit available administrative data on program participation to tighten inference on ATEs

Methodological Challenge

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

- **Nonrandom selection**: unobservables simultaneously affect food security and program participation
  - ➢ OLS produces **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

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

- Participation is no longer binary
- Dimensionality of measurement error increases

Our methodological approach:

- Introduce a multinomial, **partially-ordered** treatment variable to model participation
- Extend partial identification methods of Kreider & Hill (2009); Kreider, Pepper, Gundersen & Jolliffe (2012) to account for selection and measurement error in a unified framework

FoodAPS

National Household **Food A**cquisition and **P**urchase **S**urvey:

- Sample of **4,826** households who participated during one week between April 2012 and January 2013
- FoodAPS features of particular value for our research:
  - FoodAPS contains ***administratively verified*** info on SNAP participation for a subset of households
  - FoodAPS-GC provides local food environment data: we can construct monotone instrumental variables (**MIVs**) related to household food environment
- FoodAPS also collects info on food-at-home, away-from-home purchases, food security, demographics, health, diet, income, self-reported SNAP and WIC receipt

Participation in SNAP and WIC

- Our sample (*N* = 460) includes FoodAPS households with:
  - income ≤ 130% poverty, and
  - a pregnant woman, or a child aged < 5 years

- Weighted sample distribution by reported participation when SNAP participation indicator does not [does] incorporate administrative data:

SNAP \ WIC	*No*	*Yes*
No	15.3% [13.0%]	16.6% [13.6%]
Yes	31.4% [33.6%]	36.7% [39.7%]

Food Security Across Participation Subsamples

- Weighted prevalence of food security status by food program participation [modified using admin data]:

Proportion food secure:

SNAP	WIC: No	WIC: Yes
No	53.2% [55.1%]	54.5% [50.5%]
Yes	52.2% [51.6%]	58.5% [59.5%]

- Food security measure is based on USDA's 10-item, 30-day-referenced adult food security scale

Our Approach: Notation

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

$S^* = 0$: neither SNAP nor WIC

$S^* = 1$: SNAP alone

$S^* = 2$: WIC alone

$S^* = 3$: both SNAP and WIC

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

Potential outcomes framework:

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

$Y = 1$ if household is food secure, $Y = 0$ otherwise

$\boldsymbol{X}$: covariates

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 prevalence of food security would change if household were to participate in both SNAP and WIC vs. in SNAP alone

- There are no regression orthogonality conditions to satisfy

- Covariates are used to specify subpopulations

Decomposition Strategy

ATE cannot be point-identified (without assumptions) even if $S = S^*$

We decompose formulas into what is vs. isn't identified

Simplify notation: $ATE_{31} = P[Y(3)=1] - P[Y(1)=1]$

Consider decomposition:

$$P[Y(3)=1] = \underbrace{P[Y(3)=1 \mid S^* = 3]}_{\text{identified}}\underbrace{P(S^* = 3)}_{\text{identified}} + \underbrace{P[Y(3)=1 \mid S^* \neq 3]}_{\text{not identified}, \in [0,1]}\underbrace{P(S^* \neq 3)}_{\text{identified}}$$

Data cannot identify $P[Y(3)=1 \mid S^* \neq 3]$ because it refers to unobserved **counterfactual**

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} is "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

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

Tighting Bounds

Without assumptions, bounds on ATEs are wide

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

1) Misreporting process
2) Selection process: conventional **monotonicity assumptions**

Consider **restricting misreporting** process:

– Exploit logical constraints on probabilities and auxiliary data to restrict $\{\theta\}$:

FoodAPS SNAP verification

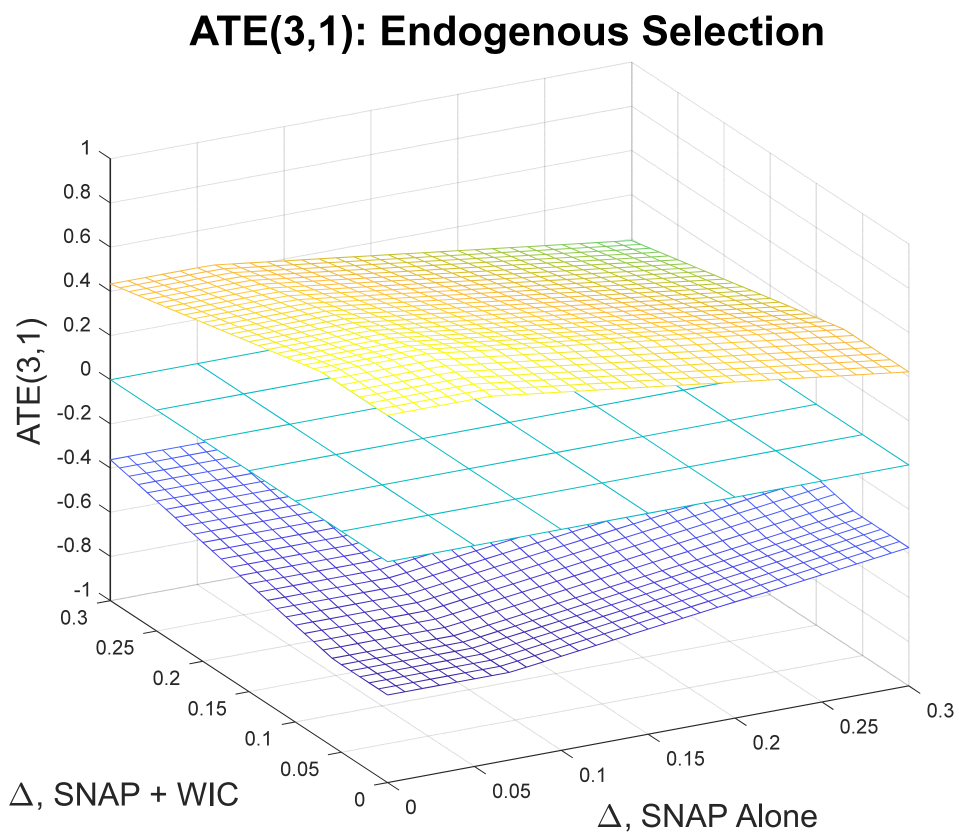
E.g., $\theta_0^{-1,1} \leq \min\{P(Y=0, S\neq 1, V_{SNAP}\neq 0), P(S^*=1)\}$

- exploits both self-reported and administrative data in FoodAPS

– "No false positives" assumption

Bounds on ATE under Endogenous Selection

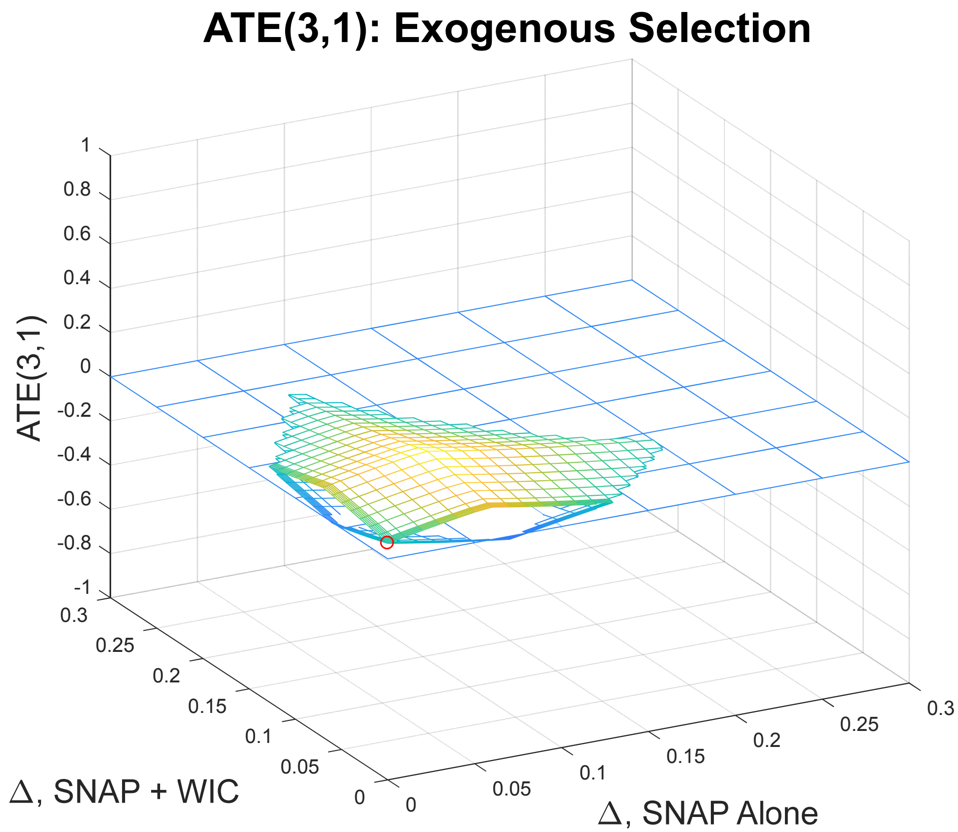
Bounds on ATE of participating in SNAP+WIC vs. SNAP alone:



Exogenous Selection

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

Bounds on ATE of participating in SNAP+WIC vs. SNAP alone:



Exogenous Selection: Identification Decay

Bounds on ATE of participating in SNAP+WIC vs. SNAP alone:

SNAP:		$\Delta_1 = 0$			$\Delta_1 = 0.01$			$\Delta_1 = 0.03$		
both:		LB	UB	width	LB	UB	width	LB	UB	width
$\Delta_3 = 0$	p.e.	[0.070,	0.070]	0.000	[0.056,	0.085]	0.029	[0.030,	0.113]	0.083
	CI	[-0.066,	0.206]		[-0.065,	0.206]		[-0.074,	0.217]	
$\Delta_3 = 0.01$	p.e.	[0.056,	0.080]	0.024	[0.042,	0.095]	0.054	[0.016,	0.123]	0.107
	CI	[-0.068,	0.204]		[-0.071,	0.208]		[-0.085,	0.224]	
								(sign of ATE not identified)		
$\Delta_3 = 0.03$	p.e.	[0.029,	0.099]	0.069	[0.015,	0.114]	0.099	[-0.011,	0.142]	0.152
	CI	[-0.081,	0.208]		[-0.089,	0.217]		[-0.109,	0.239]	

Identification **deteriorates rapidly** with more underreporting

Remark: true SNAP status is still unknown for unmatched households

MTS and MTR

Monotone treatment selection (**MTS**):

$$P[Y(j)=1 \mid S^{*}=3] \leq P[Y(j)=1 \mid S^{*}=k] \leq$$

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

- Decision to participate is monotonically related to food security. Households choose to participate in programs when they possess attributes that are bad for food security

Monotone treatment response (**MTR**):

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

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

- Participation in more programs would not harm food security (but might not help)

MIV and IV

Monotone instrumental variable (**MIV**):

We construct and use as MIVs:

$$(1)\ v = \frac{\text{actual food-at-home expenditure}}{\text{TFP-based food expenditure}}$$

$$(2)\ v = \text{income-to-poverty ratio}$$

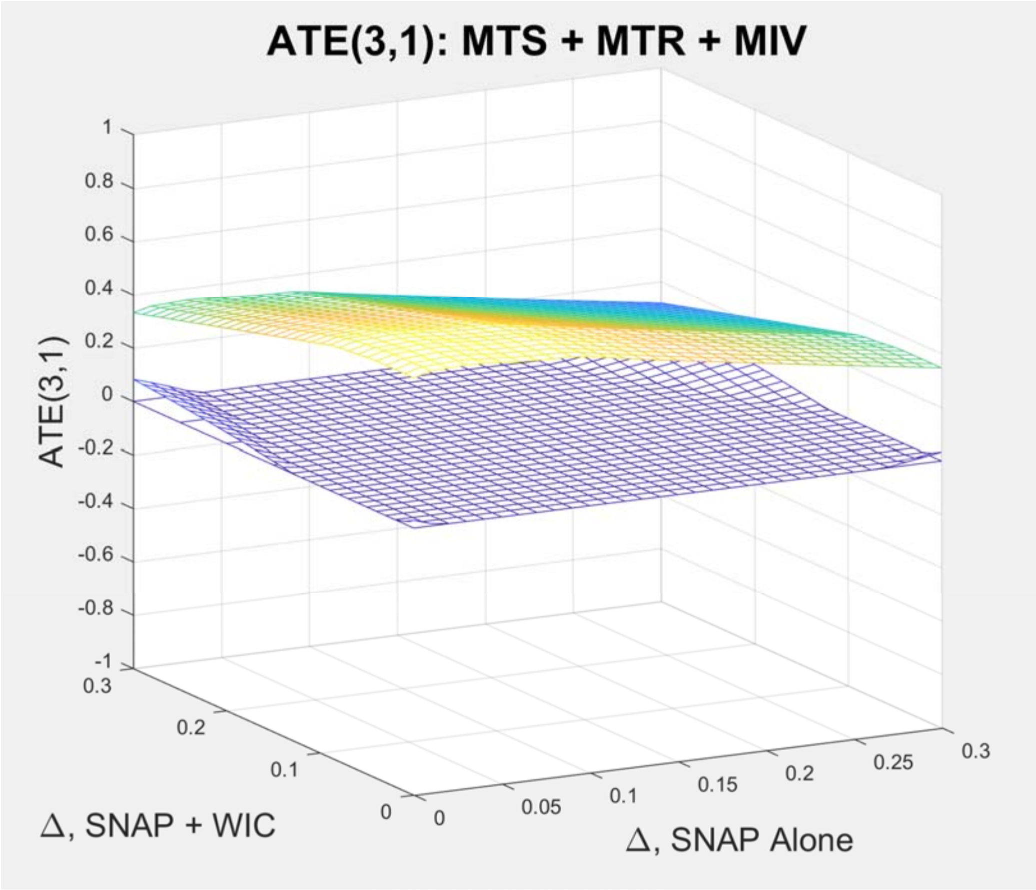
Assumption: higher *v* does not harm food security on average

Instrumental variable (**IV**): a special case of MIV

- We use SNAP Policy Database

Endogenous Selection with MTS + MTR + MIV

Bounds on ATE of participating in SNAP+WIC vs. SNAP alone:



Summary of Preliminary Results

Baseline case: Under exogenous selection and no misreporting:

- ATE of participating in both SNAP and WIC vs. in SNAP alone appears to be positive
- Results are similar when compared to WIC alone

Reporting errors: FoodAPS helps to handle misreporting of SNAP, but not WIC

Selection process: By combining assumptions, we are occasionally able to sign ATE_{31}. We are currently refining methodology to provide tighter bounds and sign ATE under a larger range of misreporting

Thank you!

SNAP Verification Status

A fraction of households in FoodAPS was matched to administrative records. In such cases, we can verify whether a household received SNAP benefits in past month

Verification Status	Sample Fraction (Weighted)
Matched households:	
Confirmed participation	57.6%
Confirmed nonparticipation	2.6%
Unmatched households:	
Not matched to administrative data	37.5%
Withheld consent to be matched	2.3%

Motivation for Our Methodology

Compare with a simple parametric approach:

$$Y_i = \gamma \cdot S_i + x_i \cdot \beta + \varepsilon_i$$

Outcome | Treatment | Covariates | 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, allows for heterogeneous response to treatment across i

Restrictions on Selection Process

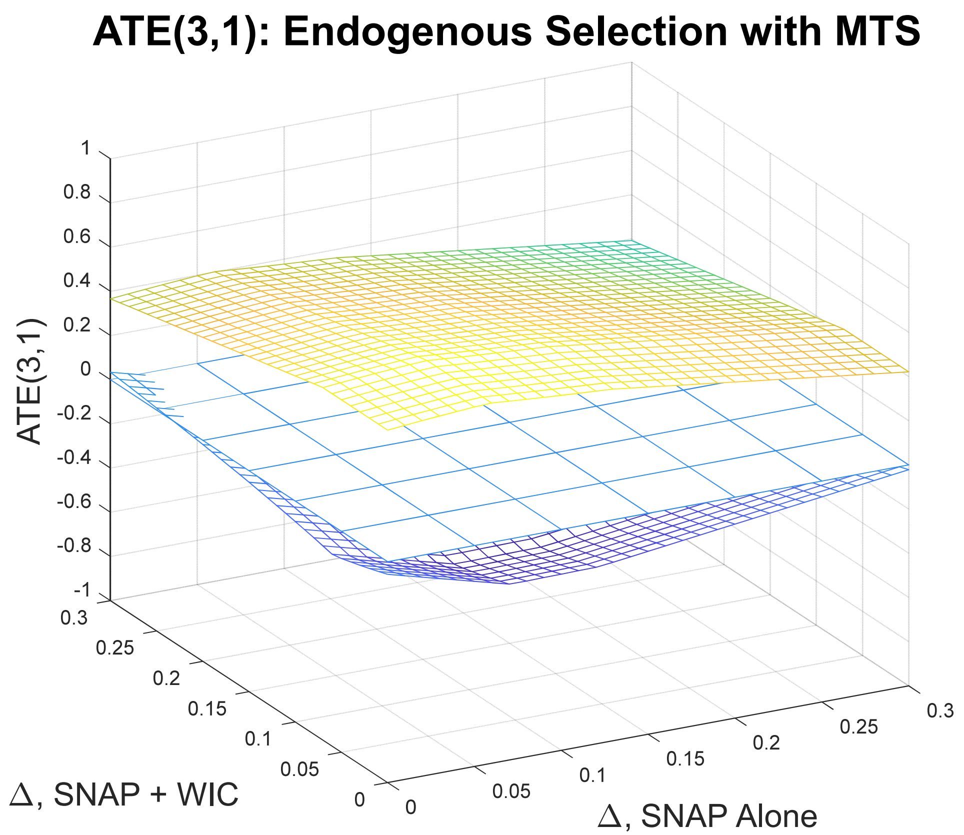
To **restrict selection process**, we can employ:

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

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

Endogenous Selection with MTS

Bounds on ATE of participating in SNAP+WIC vs. SNAP alone:



Appendix: Supplementary Data Sources

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

- Coverage: every state, every month, 1996–*present*
- 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