

Identifying the Effects of Food Stamps on Child Health Outcomes When Participation is Endogenous and Misreported*

Brent Kreider
Department of Economics
Iowa State University
bkreider@iastate.edu

John Pepper
Department of Economics
University of Virginia
jvpepper@virginia.edu

Craig Gundersen
Department of Agricultural
and Consumer Economics
University of Illinois
cggunder@illinois.edu

Dean Jolliffe
World Bank
djolliffe@worldbank.org

July 15, 2011

Abstract: The literature assessing the efficacy of the Supplemental Nutrition Assistance Program (SNAP), formerly known as the Food Stamp Program, has long puzzled over positive associations between food stamp receipt and various undesirable health outcomes such as food insecurity. Assessing the causal impacts of food stamps, however, is hampered by two key identification problems: endogenous selection into food stamp reciprocity and extensive systematic underreporting of participation status. Using data from the National Health and Nutrition Examination Survey (NHANES), we extend partial identification bounding methods to account for these two identification problems in a single unifying framework. Specifically, we derive informative bounds on the average treatment effect of SNAP on child food insecurity, general poor health, obesity, and anemia across a range of different assumptions used to address the selection and classification error problems. In particular, to address the selection problem we apply relatively weak nonparametric assumptions on the latent outcomes, selected treatments, and observed covariates. To address the classification error problem, we formalize a new approach that uses auxiliary administrative data on the size of the SNAP caseload to restrict the magnitudes and patterns of SNAP reporting errors. Layering successively stronger assumptions, an objective of our analysis is to make transparent how the strength of the conclusions varies with the strength of the identifying assumptions. Under the weakest restrictions, there is substantial ambiguity: we cannot rule out the possibility that SNAP increases or decreases poor health. Under stronger but plausible assumptions used to address the selection and classification error problems, we find that commonly cited relationships between food stamps and poor health outcomes provide a misleading picture about the true impacts of the program. Our tightest bounds identify favorable impacts of food stamps on child health.

Keywords: Food Stamp Program, Supplemental Nutrition Assistance Program, food insecurity, health outcomes, partial identification, treatment effect, nonparametric bounds, nonclassical measurement error

JEL classification numbers: C14, C21, I38

* The authors received valuable comments from Arie Beresteanu, Ed Olsen, Mark Prell, David Smallwood, Jeff Smith, and seminar participants at Claremont McKenna, Duke, University of California-Riverside, the University of Virginia, the Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA), and the Food and Nutrition Service of the USDA and meetings of APPAM, the Econometric Society, the Economic Demography Workshop at PAA, and the Conference on Inference in Partially Identified Models with Applications. Laura Tiehen generously provided us with state-level data on SNAP administrative rules. Gundersen and Kreider acknowledge financial support from the Food Assistance and Nutrition Research Program of ERS, USDA, contract number 59-5000-7-0109. Pepper acknowledges financial support from the Bankard Fund for Political Economy. The views expressed in this paper are solely those of the authors.

I. Introduction

The Supplemental Nutrition Assistance Program (SNAP), formerly known as the Food Stamp Program, is by far the largest food assistance program in the United States and, as such, constitutes a crucial component of the social safety net in the United States. In any given month during 2009, SNAP provided assistance to more than 15 million children (Leftin, Gothro, and Eslami, 2010), and it is estimated that nearly one in two American children will receive assistance during their childhood (Rank and Hirschl, 2009). As a consequence, policymakers expect food stamps to have major beneficial impacts on numerous health and nutrition challenges facing the nation, particularly for low-income children who constitute half of the recipients. Paradoxically, however, the vast empirical literature examining the impact of food stamps on health reveals little supporting evidence regarding the efficacy of the program in promoting food security and alleviating health problems. Children residing in households receiving food stamps are substantially more likely to suffer from an array of health-related problems than observationally similar nonparticipating children (see, e.g., Currie, 2003, and Nord *et al.*, 2010).

While food stamp receipt is associated with adverse health- and nutrition-related outcomes, drawing inferences on the efficacy of the program is complicated by two fundamental identification problems. A *selection problem* arises because the decision to participate in SNAP is unlikely to be exogenous. To the contrary, unobserved factors such as expected future health status, parents' human capital characteristics, financial stability, and attitudes towards work and family are all thought to be jointly related to participation in the program and health outcomes (Currie, 2003). Families may decide to participate precisely because they expect to be food insecure or in poor health.

A nonrandom *measurement error problem* arises because large fractions of food stamp recipients fail to report their program participation in household surveys. Using administrative data matched with data from the Survey of Income and Program Participation (SIPP), for example, Bollinger and David (1997) find that errors in self-reported food stamp reciprocity exceed 12 percent and are related to respondents' characteristics including their true participation status, health outcomes, and demographic attributes. Meyer, Mok, and Sullivan (2009) provide evidence of extensive underreporting in the SIPP, the Current Population Survey (CPS), and the Panel Study of Income Dynamics (PSID).

While these identification problems have long been known to confound inferences on the impact of SNAP, credible solutions remain elusive. Most studies treat selection as exogenous and receipt as accurate, leading Currie (2003, p. 240) to assert that “many studies have [...] simply ‘punted’ on the issue of identification.” A few recent exceptions address the selection problem using instrumental variables within a linear response model (see, for example, Gundersen and Oliveira, 2001; DePolt, Ribar, and Moffitt, 2009; and Hoynes and Schanzenbach, 2009). Gundersen and Kreider (2008) formally allow for the possibility of misclassified program participation, but they focus on identifying descriptive statistics rather than causal parameters.

In this paper, we consider what can be inferred about impacts of SNAP when formally accounting for the ambiguity created by the selection and measurement error problems. This study is the first to simultaneously address both of these identification problems within a single methodological framework. To do so, we extend recently developed partial identification methods that allow one to consider weaker assumptions than required under conventional parametric approaches (see e.g., Manski, 1995; Pepper, 2000; Kreider and Pepper, 2007 and 2008; Gundersen and Kreider, 2008; Kreider and Hill, 2009; Molinari, 2008 and 2010). Gundersen, Kreider, and Pepper (forthcoming) apply some of the methods developed in this paper to study effects of the National School Lunch Program.

This partial identification approach is especially well-suited for studying the impact of SNAP where classical methodological prescriptions are often untenable. The literature evaluating the impact of means-tested assistance programs typically relies on linear response models coupled with an assumption that some observed instrumental variable (IV), often based on cross-state and time variation in program rules and regulations, affects program participation but otherwise has no effect on the potential outcomes. Yet SNAP is mostly defined at the federal level and has not substantively changed since the early 1980s, so many of the key program rules and regulations are not as useful as instrumental variables. Moreover, as is now widely recognized, the classical linear response model assumption is difficult to justify when considering programs that are thought to have heterogeneous effects (Moffitt, 2005). Finally, the implicit assumption of accurate classification of participation status is known to be violated, yet addressing the problem of classification errors in a binary regressor is difficult. The assumption of non-mean-reverting

errors cannot apply with binary variables, and the systematic underreporting of SNAP participation violates the classical assumption that measurement error arises independently of the true value of the underlying variable (see, e.g., Bollinger, 1996).

The methods applied in this paper do not require the linear response model, the classical measurement error model, or an instrumental variable assumption. Instead, we focus on weaker models that are straightforward to motivate in practice and result in informative bounds on the health consequences of SNAP. Given the largely paradoxical findings in the existing literature coupled with the methodological challenges in addressing these identification problems, deriving informative bounds under assumptions that may share some consensus seems like an important step.

Using data from the National Health and Nutrition Examination Survey (NHANES), we assess the impact of food stamps on the health of children, an important subpopulation that comprises half of all recipients and whose well-being is followed closely by policymakers and program administrators. A primary strength of the NHANES is the wealth of health-related information provided in the survey. We exploit these data by assessing the impact of SNAP on food insecurity, obesity, poor general health, and anemia.

After describing the data in Section II, we formally define the empirical questions and the identification problems in Section III. Our analysis is complicated by two distinct identification problems: (a) the selection problem that arises because the data cannot reveal unknown counterfactuals (e.g., the outcomes of a nonparticipant in an alternate state of the world in which food stamps are received) and (b) the measurement error problem that arises because the data cannot reveal respondents with misclassified food stamp participation.

In Section IV, we focus on the selection problem, abstracting away from classification errors. Following Manski (1995) and Pepper (2000), we begin by examining what can be learned without imposing any assumptions on the selection process, and then consider the identifying power of several alternative assumptions. We first consider the *monotone treatment selection* (MTS) restriction (Manski and Pepper, 2000) that formalizes the commonplace assumption that the decision to participate in SNAP is monotonically related to poor latent health outcomes. We then consider the *Monotone Instrumental*

Variable (MIV) assumption that the latent probability of a poor health outcome is nonincreasing in household income (adjusted for family composition). Requiring no *a priori* exclusion restriction, the MIV assumption can be plausible in many applications where the standard independence assumption is a matter of considerable controversy. Finally, in parts of the analysis, we consider a *Monotone Treatment Response* (MTR) assumption that participation in SNAP does not worsen health status. While recipients appear to be worse off on average than eligible nonrecipients, many have argued that participating in SNAP would not cause health or food security to deteriorate (e.g., Currie, 2003). Section IV concludes with a brief analysis of data from the 2003 CPS to assess whether the results are consistent across the surveys and to consider standard instrumental variables made available in the rich CPS covariate data.

In Section V, we introduce classification errors in the model. In doing so, we make two notable contributions to the literature. First, departing from the usual treatment effects literature that formally acknowledges ambiguity associated with counterfactuals but not ambiguity associated with misreporting, our methods simultaneously account for both problems. Second, we develop new methods that utilize administrative information on the size of the NSLP caseload to derive informative constraints on the classification error problem.

We draw conclusions in Section VI and emphasize three findings. First, the ambiguity associated with the selection and classification error problems can be substantially mitigated by applying some basic restrictions including MTS and no-false positive errors (as discussed below, the empirical literature on SNAP suggests that errors of commission are negligible). Second, under the joint MIV-MTS assumption, we find that SNAP reduces food insecurity rates. This result holds even for modest degrees of misclassification error. Finally, under the joint MIV-MTS-MTR assumption, we find that SNAP leads to a decline in food insecurity rates and other poor health outcomes even when allowing for high rates of classification error.

II. Data

To study the impact of SNAP on child nutritional health, we use data from the 2001-2006 NHANES. The NHANES, conducted by the National Center for Health Statistics, Centers for Disease Control

(NCHS/CDC), is a program of surveys designed to collect information about the health and nutritional status of adults and children in the United States through interviews and direct physical examinations. The survey currently includes a national sample of about 5,000 persons each year, about half of whom are children. Vulnerable groups, including Hispanics and African-Americans, are oversampled. Given the wealth of health-related information, NHANES has been widely used in previous research on health- and nutrition-related child outcomes (recent work includes, e.g., Gundersen *et al.*, 2008).

We focus our analysis on households with children eligible to receive SNAP. To be eligible for assistance, a household's gross income before taxes in the previous month cannot exceed 130 percent of the poverty line, net monthly income (gross income minus a standard deduction and expenses for care for disabled dependents, medical expenses, and excessive shelter costs) cannot exceed the poverty line, and assets must be less than \$2,000. Since the NHANES does not provide sufficient information to measure net income and assets, we focus on gross income eligibility. Given our focus on children, however, this data limitation should not lead to substantial errors in defining eligibility (Gundersen and Offut, 2005). In contrast, the asset test could be important for a sample that includes a high proportion of households headed by an elderly person (Haider, Jackowitz, and Schoeni, 2003). Virtually all gross income eligible households are also net income eligible.

Our preliminary sample is comprised of 4690 children between the ages of two and 17 who reside in households with income less than 130% of the federal poverty line. Children under the age of two are not included in the sample because there is no commonly accepted way to establish BMI percentiles for children this young. After dropping additional observations for which information is missing about height and weight, we obtain our final sample of 4,418 income-eligible children.

For each observation, we observe a number of socioeconomic and demographic characteristics including ratio of income to the poverty line. Our sample has an average household income level equal to 75 percent of the poverty line. To assess the characteristics of our sample relative to other national estimates, we examined data from the 2003 Current Population Survey (CPS), December Supplement (see Section IV.D for further details). These data also indicate that income eligible children lived in families with average income equal to 75 percent of the poverty line.

A. Self-Reported Food Stamp Receipt Indicator

Beyond demographic information, we also observe a self-reported measure of food stamp receipt over the past year. SNAP participants receive benefits for the purchase of food in authorized retail food outlets where the benefit amount depends on net income. Households with a net income of zero receive the maximum benefit, and benefits decline with income: for every additional dollar, the amount of SNAP benefits is reduced by 30 cents (except earned income, in which case the reduction is 24 cents). In 2010, the average monthly benefit was \$288/month for a family of four, with the maximum benefit of \$668. These benefits can represent a substantial component of low-income households' total income.

In this survey, about 46% of the eligible households claim to be receiving benefits. Thus, less than half of eligible households report receipt. This nonparticipation is ascribed to four main factors. First, there may be stigma associated with receiving food stamps. Stigma encompasses a wide variety of sources, from a person's own distaste for receiving food stamps to the fear of disapproval from others when redeeming food stamps to the possible negative reaction of caseworkers (Moffitt, 1983). Second, transaction costs can diminish the attractiveness of participation. To receive SNAP, households must personally verify their income and expenses and must visit a caseworker on a periodic basis to recertify their eligibility. The initial visit and subsequent recertifications can be time consuming. Third, the benefit level can be quite small for relatively higher income families – sometimes as low as \$10 a month.

Finally, food stamp receipt is thought to be underreported. Evidence of pervasive underreporting has surfaced in two types of studies, both of which compare self-reported information with official records. The first type has compared aggregate statistics obtained from self-reported survey data with those obtained from administrative data. These studies suggest the presence of substantial underreporting in many different surveys including the CPS, the SIPP, the PSID, and the Consumer Expenditure Survey (CES) (Trippe, Doyle, and Asher, 1992; Bitler, Currie, and Scholz, 2003; and Meyer, Mok, and Sullivan, 2009). Meyer, Mok, and Sullivan (2009, Table 13), for example, find that self-reports in the CPS reflect just over half the number of food stamp recipients identified in administrative data. Other studies have compared individual reports of food stamp participation status in surveys with matched reports from

administrative data. Using this method, researchers can identify both errors of commission (reporting benefits not actually received) and errors of omission (not reporting benefits actually received). As discussed above, Bollinger and David (1997, Table 2) find that 12.0 percent of responses in the SIPP involve errors of omission while only 0.3 percent involve errors of commission (see also Marquis and Moore, 1990).

B. Outcomes

A primary strength of the NHANES is the detailed information provided on dietary and health-related outcomes, with distinct components of the survey providing information from self-reports, medical examinations, physiological measurements, and laboratory tests. Since no single measure is thought to completely capture health and nutritional well-being, the detailed and varied health measures available in the NHANES make it a unique and important survey for studying the impact of nutritional programs on well-being.

Because alleviating food insecurity is a central goal of SNAP (U.S. Department of Agriculture, 1999), much of our attention focuses on this measure of nutritional health. The extent of food insecurity in the United States has become a well-publicized issue of concern to policymakers and program administrators. In 2009, 14.7% of the U.S. population reported that they suffered from food insecurity at some time during the previous year (Nord *et al.*, 2010). These households were uncertain of having, or unable to acquire, enough food for all their members because they had insufficient money or other resources.

To calculate these official food insecurity rates in the U.S., defined over a 12 month period, a series of 18 questions are posed in the Core Food Security Module (CFSM) for families with children. Each question is designed to capture some aspect of food insecurity and, for some questions, the frequency with which it manifests itself. Examples include “I worried whether our food would run out before we got money to buy more” (the least severe outcome); “Did you or the other adults in your household ever cut the size of your meals or skip meals because there wasn't enough money for food?” and “Did a child in the household ever not eat for a full day because you couldn't afford enough food?”

(the most severe outcome). A complete listing of the food insecurity questions is presented in Appendix Table 1. Following official definitions, we use these 18 questions to construct a comparison of children in food secure households (two or fewer affirmative responses) with children in food insecure households (three or more affirmative responses).

In addition to studying the impact of SNAP on food insecurity rates, we also examine three other outcome variables: obesity, anemia, and an indicator of fair or poor general health. Based on guidelines provided by the Center for Disease Control and Prevention, a child is classified as obese if his or her body mass index (BMI) (kg/m^2) is at or above the 95th percentile for the child's age and gender. In the NHANES, heights and weights used to calculate BMI are obtained by trained personnel (i.e., not self-reported). A child is classified as having anemia if, based on a blood test, the child is both iron deficient and has an abnormally low hemoglobin level. The indicator of fair or poor general health is reported by the child's parent. In this paper, we treat these health outcomes as accurately measured. While errors in measuring obesity and anemia are likely to be minimal (data on height and weight were collected by trained personnel and anemia is measured using a blood test), this assumption may be violated for the poor health and food insecurity outcomes. In general, measurement error in the outcome variables would widen the bounds established in this paper.

Together, these four measures reflect a wide range of health related outcomes that might be impacted by SNAP. All four outcomes are also known to be associated with a range of negative physical, psychological, and social consequences that have current and future implications for health, including reduced life expectancy. With a maximum pair-wise correlation of only 0.12 (between food insecurity and poor general health), these four outcomes are related but clearly measure different aspects of well-being. The outcomes have also attracted different levels of attention in the existing food stamp literature. Food insecurity and obesity are of central concern to policymakers and researchers studying the impact of SNAP on health (see, for example, Kaushal, 2007, and Meyerhoefer and Pylypchuk, 2008). To the best of our knowledge, this paper is the first to investigate (using any methods) the impacts of food stamps on self-reported general health and anemia.

Table 1 displays means and standard errors for the variables used in this study. The estimates in this table (and elsewhere in the paper) are weighted to account for the complex survey design used in the NHANES. Consistent with previous work on this topic, SNAP recipients tend to have worse health outcomes than eligible nonparticipants. For example, 45% of children reported as SNAP recipients are food insecure, nine percentage points higher than the 36% food insecurity rate among eligible nonparticipants (a statistically significant difference at better than the 5% level). Compared with eligible nonrecipients, SNAP recipients are also slightly more likely to be obese, be in fair or poor general health, and have anemia.

III. Identifying the Average Treatment Effect

Our interest is in learning about the average treatment effect (ATE) of SNAP participation on each of our health-related outcomes among food stamp eligible households. Focusing on binary outcomes, the ATE is given by

$$\text{ATE}(1,0 | X \in \Omega) = P[H(1)=1 | X \in \Omega] - P[H(0)=1 | X \in \Omega] \quad (1)$$

where H is the realized health outcome, $H(1)$ denotes the health of a child if he or she were to receive food stamps, $H(0)$ denotes the analogous outcome if the child were not to receive food stamps, and $X \in \Omega$ denotes conditioning on observed covariates whose values lie in the set Ω . Thus, the average treatment effect reveals how the mean outcome would differ if all eligible children received food stamps versus the mean outcome if all eligible children did not receive food stamps.

In what follows, we simplify the notation by suppressing the conditioning on subpopulations of interest captured in X . For this analysis, we focus on children who are eligible for SNAP. In much of the literature examining the impact of SNAP, other observed covariates are motivated as a means of controlling for factors influencing a family’s decision to take up food stamps. In the usual regression framework, researchers attempt to “correctly” choose a set of control variables for which the exogenous selection assumption applies. Inevitably, however, there is much debate about whether the researcher

omitted “important” explanatory variables. In contrast, conditioning on covariates in our approach serves only to define subpopulations of interest as there are no regression orthogonality conditions to be satisfied. The problem is well-defined regardless of how the subpopulations are specified (Pepper, 2000).

As discussed above, two identification problems arise when assessing the impact of SNAP on children’s health outcomes. First, even if SNAP participation were observed for all eligible households, the potential outcome $H(1)$ is counterfactual for all children who did not receive food stamps, while $H(0)$ is counterfactual for all children who did receive food stamps. This is referred to as the selection problem. Using the Law of Total Probability, this identification problem can be highlighted by writing the first term of Equation (1) as

$$P[H(1) = 1] = P[H(1) = 1 | FS^* = 1]P(FS^* = 1) + P[H(1) = 1 | FS^* = 0]P(FS^* = 0) \quad (2)$$

where $FS^* = 1$ denotes that a child is in a household that truly receives food stamps and $FS^* = 0$ otherwise. If food stamp receipt is observed, the sampling process identifies the selection probability, $P(FS^* = 1)$, the probability an eligible child does not receive food stamps, $P(FS^* = 0)$, and the expectation of outcomes conditional on the outcome being observed, $P[H(j) = 1 | FS^* = j]$, $j = 1, 0$. Still, the sampling process cannot reveal the mean outcome conditional on the outcome being counterfactual, $P[H(1) = 1 | FS^* = 0]$. Thus, $P[H(1) = 1]$ is not point-identified by the sampling process alone.

Second, true participation status may not be observed for respondents. This is referred to as the measurement or classification error problem. Instead of observing FS^* , we observe a self-reported indicator, FS , where $FS = 1$ if a child is in a household that reports receiving food stamps and 0 otherwise. Without assumptions restricting the nature or degree of classification errors, the sampling process does not reveal useful information on food stamp receipt, FS^* , and thus all of the probabilities on the right hand side of Equation (2) are unknown.

To highlight this measurement problem, let the latent variable Z^* indicate whether a report is accurate, where $Z^* = 1$ if $FS^* = FS$ and $Z^* = 0$ otherwise. Using this variable, we can further decompose the first term of Equation (1) as

$$P[H(1) = 1] = P[H(1) = 1, FS^* = 1] - \theta_1^+ + \theta_1^- \quad (3)$$

$$+ P[H(1) = 1 | FS^* = 0][P(FS = 0) + (\theta_1^+ + \theta_0^+) - (\theta_1^- + \theta_0^-)],$$

where $\theta_j^+ = P(H = j, FS = 1, Z^* = 0)$ and $\theta_j^- = P(H = j, FS = 0, Z^* = 0)$ denote the fraction of false positive and false negative classifications of food stamp recipients, respectively, for children realizing health outcome $j = 1, 0$. The first term, $P[H(1) = 1 | FS^* = 1]P(FS^* = 1)$, is not identified because of the classification error problem. The second term is not identified because of both the selection and classification error problems. As above, the data cannot reveal the counterfactual outcome distribution, $P[H(1) = 1 | FS^* = 0]$, regardless of whether participation is measured accurately, and, in the presence of classification errors, the sampling process does not reveal the proportion of respondents that received assistance, $P(FS^*)$.

IV. The Selection Problem

The literature evaluating the effect of SNAP on health has implicitly assumed that respondents accurately self-report of program participation. To provide a direct comparison to the existing literature, we begin by focusing on this special case and study what can be learned about the average treatment effect using existing methods. In Section V below, we develop new methods for simultaneously addressing the selection and classification error problems.

A natural starting point is to ask what can be learned in the absence of any assumptions invoked to address the selection problem (see Manski, 1995 and Pepper, 2000). Since the latent probability $P[H(1) = 1 | FS^* = 0]$ must lie within $[0, 1]$, it follows that

$$P(H = 1, FS^* = 1) \leq P[H(1) = 1] \leq P(H = 1, FS^* = 1) + P(FS^* = 0) \quad (4)$$

where, in the absence of classification errors, the lower and upper bounds are identified by the sampling process. An analogous result applies for $P[H(0) = 1]$. In this worst-case scenario where there is no additional identifying information, the data alone cannot reveal whether SNAP leads to better or worse mean health outcomes (see Manski, 1995, for further details).

A. Monotone Instrumental Variable Models

To derive more informative inferences about the impact of SNAP on health, prior information to address the selection problem must be brought to bear. While the exogenous selection assumption $P[H(1) = 1] = P[H(1) = 1 | FS^*]$ maintained in much of the literature seems untenable, there are a number of middle ground assumptions that narrow the bounds by restricting the relationship between SNAP participation, health outcomes, and observed covariates. In this section we apply two *monotone instrumental variable* (MIV) assumptions that certain observed covariates are known to be monotonically related to the latent response variable.

First, we consider the *Monotone Treatment Selection* (MTS) assumption (Manski and Pepper, 2000) that children receiving food stamps are likely to have worse latent health outcomes on average than nonparticipants. This selection model formalizes the most common explanation for the positive association between participation and poor health: unobserved factors associated with poor health are thought to be positively associated with the decision to take-up the program (Gundersen and Oliveira, 2001 and Currie, 2003). For example, families may participate precisely because they expect to be food insecure. Formally, the MTS assumption is given by

$$P[H(j) = 1 | FS^* = 0] \leq P[H(j) = 1 | FS^* = 1] \quad \text{for } j = 0, 1. \quad (5)$$

That is, for latent potential outcomes $H(0)$ and $H(1)$, eligible households that receive food stamps, $FS^*=1$, have no better latent health outcomes on average than eligible households that have not taken up food stamps, $FS^* = 0$. While the MTS assumption serves to reduce the upper bound on the ATE, the

assumption alone does not identify the sign of the average treatment effect (see Manski and Pepper, 2000).

Second, we consider the relatively innocuous assumption that the latent probability of negative health outcomes weakly decreases with income adjusted for family composition. A large body of empirical research supports the idea of a negative gradient between reported income and the health outcomes studied in this paper (e.g., Nord *et al.* 2010 for food insecurity; Case, Lubotsky, and Paxson, 2002, and Deaton, 2002, for general health; Newacheck, 1994, for anemia; Shrewsbury and Wardle, 2008, and Jolliffe, forthcoming, for obesity). To formalize this idea, let v be the monotone instrumental variable such that

$$u_1 < u < u_2 \text{ implies } P[H(t) = 1 | v = u_2] \leq P[H(t) = 1 | v = u] \leq P[H(t) = 1 | v = u_1] \text{ for } t = 1, 0. \quad (6)$$

These conditional probabilities can be bounded using the various nonparametric models described throughout this paper. Let $LB(u)$ and $UB(u)$ be the known lower and upper bounds evaluated at $v = u$, respectively, given the available information. Then the MIV assumption formalized in Manski and Pepper (2000, Proposition 1) implies:

$$\sup_{u \leq u_2} LB(u) \leq P[H(t) = 1 | v = u] \leq \inf_{u \geq u_1} UB(u).$$

Bounds on the unconditional latent probability, $P[H(t) = 1]$, can then be obtained using the law of total probability.

Following the approach developed in Kreider and Pepper (2007), we estimate these MIV bounds by first dividing the sample into equally sized groups (more than 200 observations per cell) delineated by an increasing ratio of income to the poverty line. Then, to find the MIV bounds on the rates of poor health outcomes, one takes the average of the plug-in estimators (weighted to account for the survey design) of lower and upper bounds across the different income groups observed in the data. We use 20 groups, although the qualitative results in this paper are unchanged when we use 15 or 25 income groups.

Since this MIV estimator is consistent but biased in finite samples (see Manski and Pepper, 2000 and 2009), we employ Kreider and Pepper's (2007) modified MIV estimator that accounts for the finite sample bias using a nonparametric bootstrap correction method. Chernozhukov, Lee, and Rosen (2009) formalize an entirely different approach for estimation, inference, and bias correction that involves applying a precision correction to their estimated boundary functions before taking the intersections. Recent work by Hirano and Porter (2009) raises some concerns about applying bias corrections when estimating intersection bounds, showing that it may be impossible to completely eliminate bias and that reducing bias too much leads to large increases in variance. With more than 200 observations per cell, however, the bias correction in our application plays a relatively modest role in our estimates (see tables below).

B. Results for the no errors case

For each of the four outcomes, Table 2 presents bias-corrected bounds, confidence intervals, and estimated finite-sample biases under a variety of different models for the no errors case. In the first row, we make no assumptions about how eligible households select themselves into food stamp reciprocity. The width of the ATE bounds always equals 1, and the bounds on the ATE always include zero (see Manski, 1995). These wide bounds highlight a researcher's inability to make strong inferences about the efficacy of the food stamps without making assumptions that address the problem of unknown counterfactuals. In the absence of restrictions that address the selection problem, we cannot rule out the possibility that SNAP has a large positive or negative impact on the likelihood on poor health outcomes. These bounds can be narrowed substantially, however, under common monotonicity assumptions on treatment selection (MTS) and relationships between the latent outcome and observed instrumental variables (MIV).

To narrow the bounds, we apply the MTS and joint MTS-MIV assumptions. These results are presented in the middle two rows of Table 2, respectively. The MTS assumption alone is not strong enough to identify the sign of the impact of SNAP on the health outcomes, but it does dramatically reduce the upper bounds on the average treatment effects. For example, the upper bound for food insecurity falls

from 0.555 to 0.093 such that $ATE \in [-0.445, 0.093]$. The upper bounds on the other three outcomes drop even further, falling to 0.015 for poor health, 0.012 for obesity, and 0.003 for anemia. Thus, under the MTS assumption alone, the estimated bounds rule out the possibility that SNAP leads to large increases in poor health, obesity and anemia. Instead, SNAP may lead to substantial reductions in these adverse health outcomes, at worst having slightly deleterious effects.

Perhaps the most important results are found when we combine the MTS assumption with the MIV assumption that the probability of poor health weakly decreases with family resources, as measured by the ratio of income to the poverty line. The upper bound is negative for all four outcomes in this joint MTS-MIV model, though the confidence interval includes zero for obesity and anemia. For example, the estimates suggest that SNAP reduces the prevalence of food insecurity by at least 12.8 points and poor health by 6.1 points. These numbers suggest that food stamps have substantial beneficial effects. In the absence of SNAP, our estimates of $P[H(0)=1]$ indicate that at least 45.9 percent of eligible children would be food insecure, with 11.2 percent in poor health. Thus, the estimates indicate that the program has reduced the prevalence of food insecurity by at least 28% ($= 12.8/45.9$) and poor health by at least 54% ($=6.1/11.2$).

C. Monotone Treatment Response Model

Despite the observed positive correlations in the data between SNAP participation and unfavorable outcomes, there is a general consensus among policymakers and researchers that SNAP does not increase the rate of food insecurity (Currie, 2003). Given this general consensus, we consider the identifying power of the *Monotone Treatment Response* (MTR) assumption (Manski, 1995 and 1997; Pepper 2000) that formalizes the common idea that SNAP does not lead to a reduction in health status: $H(1) \leq H(0)$. In this case, the average treatment effect of receiving food stamps must be nonpositive.

Since MTR rules out the possibility of deleterious effects of food stamps on health by assumption, it is not helpful in resolving the health outcomes paradox. Nevertheless, for outcomes in which MTR is thought to be credible, it can help shed light on the magnitudes of any identified beneficial effects of the program. In particular, MTR may interact with the MTS and MIV assumptions to bound

the ATE further away from zero than can be attained under the MTS-MIV assumptions alone.

For the food insecurity, general health, and anemia outcomes, the MTR assumption seems relatively innocuous in that it is difficult to imagine how receiving food stamps would lead to worse health outcomes. Previous work has demonstrated that each additional dollar in benefits leads to marginal increases in food expenditures (e.g., Breunig and Dasgupta, 2002; Levedahl, 1995). If food is a normal good, SNAP should weakly increase the consumption of food and, in turn, decrease the incidence of food insecurity and poor health. For obesity, however, the assumption is more tenuous. Better access to nutritious foods through SNAP may lead to healthier eating and less obesity, but potential increases in caloric intake could result in weight gains.

The bottom rows of Table 2 present results under the joint MTS-MIV-MTR assumption. While the MTR assumption is less credible for obesity than for the other outcomes, we present MTS-MIV-MTR estimates for each outcome to make transparent the identifying power of the MTR assumption in each case. Without measurement error, the MTR assumption does not notably reduce the estimated upper bounds on the ATE relative to the estimates derived under the MTS-MIV assumptions alone. With classification errors (see Section V), however, the MTR assumption turns out to have substantial identifying power in this application.

D. Sensitivity Analysis Using Data from the CPS

To assess the sensitivity of our findings to another data source, we estimate these models using analogous data from the December Supplement of the 2003 CPS. The CPS has been widely applied to evaluate the association between food stamps and food insecurity (see, for example, Jensen, 2002; Wilde and Nord, 2005; and Gundersen and Kreider, 2009), and is used by the USDA to establish the official food insecurity rates for the United States (e.g., Nord *et al.*, 2010).

Using the sampling design in Gundersen and Kreider (2008), our data for this sensitivity analyses includes 2,707 households with children reporting incomes less than 130 percent of the poverty line. As with the NHANES, we observe a self-reported measure of food stamp receipt over the past year, food insecurity over the past year, and the income to poverty ratio. The CPS does not include information on

the other three outcomes (poor general health, obesity, and anemia) revealed in the NHANES data. The summary statistics from the CPS data are similar to what we find in the NHAMES (see Table 2 in Gundersen and Kreider, 2008): just over 40% of the households report receiving food stamps, and the food insecurity rate among these self-report recipients is 17.9 percentage points higher than among eligible nonrecipients (52.3% versus 34.4%).

The last column of Table 2 presents estimates using the CPS data. The estimates are similar to what we find using data from the NHANES, although the estimated upper bound under the MIV-MTS model is positive. The estimated upper bound remains negative, however, when we impose the MIV-MTS-MTR assumption. In part, differences in estimates reflect the smaller sample size in the CPS data which leads to less precise estimates and larger bias corrections in the MIV models. Differences may also reflect the fact that classification errors in the CPS have been found to be more extensive than in the NHANES (Meyer *et al.*, 2009).

Although these data have very limited information on health outcomes compared with the NHANES, they are rich enough to allow us to construct some standard instrumental variables for SNAP participation used in the existing literature. In particular, state identifiers in the CPS allow us to apply a more traditional instrumental variable (IV) assumption based on cross-state variation in program eligibility rules. To do so, we merge the Urban Institute's database of state program rules (see Finegold *et al.*, 2006) with the CPS data to create two instrumental variables: an indicator for whether the state uses a simplified semi-annual reporting requirement for earnings (47%) and an indicator for whether cars are exempt from the asset test (30%). Suppose these two variables are mean independent of the latent health outcomes but related to participation. When combined with the traditional linear response model, the ATE is point identified and the Wald estimator of the average treatment effect ranges from -0.23 when the indicator for whether cars count in the asset test is used as an IV to -0.62 when the indicator for whether the state uses a simplified reporting requirement is used as an IV.

Notice that the -0.62 estimate lies outside of the worst-case bounds reported in Table 2, suggesting that either the IV assumption or the linear response model assumption is invalid. If instead of applying the linear response model we estimate Haikha and Vytlacil's (2011) nonparametric threshold

crossing model, SNAP is found to reduce food insecurity by at least 3 percentage points when using the asset test instrument and at least 5 points when using the reporting requirement instrument. The estimate based on the reporting requirement instrument is significantly different than zero at the five-percent level, but the asset test instrument is not statistically significantly different than zero.

Combing these traditional instruments with sufficiently strong assumptions reveals consistent evidence that SNAP reduces the rate of food insecurity. Yet while these estimated negative average treatment effects are qualitatively similar to our primary results found using the MIV and MTS assumptions, we caution against drawing strong conclusions on the efficacy of SNAP based on these findings alone. In particular, there may be good reasons to doubt the excludability assumption that the instruments are mean independent of the latent food insecurity outcome. State-specific food insecurity rates are well documented (e.g. Nord *et al.*, 2010), and states may endogenously design SNAP within USDA guidelines partially in response to food insecurity levels. Thus, these state program rules may not be independent of the state food insecurity rate. Finally, while the Shaikh and Vytlačil threshold cross-model provides a constructive middle-ground model that allows one to impose some additional structure, the linear response model imposes the strong homogeneity restriction on the response function that seems unlikely to hold in practice. In fact, one of the linear IV estimates lies outside of the worst case bounds suggesting that this assumption may be violated.

V. The Selection and Classification Error Model: A Unified Approach

While our findings thus far imply that SNAP plays an important role in improving children’s health, we have not yet accounted for classification errors. In this section, we introduce new methods that explicitly acknowledge the presence of SNAP reporting errors and incorporate auxiliary administrative data on the size of the SNAP caseload to restrict the magnitudes and patterns of such errors.

With classification errors, FS^* is not observed and the Manski (1995) worst-case selection bounds are not identified. In particular, defining $\Theta \equiv (\theta_1^- + \theta_0^+) - (\theta_0^- + \theta_1^+)$, we augment the Manski bounds as follows:

$$\begin{aligned}
& [-P(H = 1, FS = 0) - P(H = 0, FS = 1)] + \Theta \\
& \leq ATE(1,0) \leq \\
& [P(H = 1, FS = 1) + P(H = 0, FS = 0)] + \Theta.
\end{aligned} \tag{7}$$

Thus, without restrictions on the measurement error process, the false reporting rates θ are not identified, and the data are uninformative about the ATE. We use two sources of information to restrict Θ . First, auxiliary data on size of the SNAP caseload provides informative restrictions on the classification error components, θ . Second, the relevant validation literature provides informative restrictions on the magnitude and patterns of the classification error problem.

A. The Classification Error Model

To draw inferences in light of the classification error problem, we exploit two sources of additional information. First, we combine readily available auxiliary data on the size of the caseload from the administrative data collected by the USDA with survey data from the NHANES to estimate the true participation rate, $P(FS^* = 1)$. In Proposition 1 below, we show how knowledge of the true and the self-reported rates implies meaningful restrictions on the classification error probabilities, θ . In particular, knowledge of $P^* \equiv P(FS^* = 1)$ and $P \equiv P(FS = 1)$ implies the following three restrictions:

$$(\theta_1^- + \theta_0^-) - (\theta_1^+ + \theta_0^+) = \Delta \tag{8a}$$

$$\theta_i^- \leq \min\{P(H = i, FS = 0), P(FS^* = 1)\} \equiv \theta_i^{UB-}, \quad i = 1, 0 \tag{8b}$$

$$\theta_i^+ \leq \min\{P(H = i, FS = 1), P(FS^* = 0)\} \equiv \theta_i^{UB+}, \quad i = 1, 0 \tag{8c}$$

where $\Delta \equiv P^* - P$. Equation (8a) restricts the net fraction of false negative reports to equal the difference in the true and self-reported participation rates. Equations (8b) and (8c) place meaningful upper bounds on the fraction of false negative and positive reports.

Second, the range of studies examining the validity of self-reports provide additional information on the degree of misreporting. Evidence from validation studies find errors of commission to be negligible, with the overall rate of misreporting estimated to be no greater than about 25 percent. To

incorporate information on the overall rate of misreporting, we consider the identifying power of a restriction on the maximum amount of data corruption in the spirit of Horowitz and Manski (1995). That is, let

$$P(Z^* = 0) \leq Q_u \quad (9)$$

where Q_u is a known upper bound on the degree of SNAP misclassification. Given knowledge of P^* , this value must logically lie within the range $[|P^* - P|, 1]$. In the polar case that Q_u is set equal to 1, the researcher is setting no restriction on the proportion of false reports in the data beyond that implied by restrictions (8a)-(8c). We refer to this as the “arbitrary errors model.” In the other polar case that Q_u is set equal to $|P^* - P|$, the researcher is imposing a “no excess errors” restriction that there are no data errors beyond the proportion necessary to generate the discrepancy (distance) between the true participation rate, P^* , and reported rate, P . For the case of systematic underreporting in our application, $\Delta \geq 0$, setting Q_u equal to its minimum allowed value $|P^* - P|$ is equivalent to imposing a “no false positives” assumption that respondents do not falsely claim to participate in the program (and similarly a “no false negatives” assumption in applications involving $\Delta \leq 0$). The no false positives assumption serves as a useful benchmark for the receipt of SNAP in our application since validation data suggest very few instances of households falsely claiming to receive food stamps (e.g., Bollinger and David, 1997, and Marquis and Moore, 1990). Middle-ground positions are obtained by setting Q_u between $|P^* - P|$ and 1.

The restrictions in Equations (8) and (9) imply informative bounds on the unknown parameter, Θ , where the upper bound is found by maximizing $(\theta_1^- + \theta_0^+)$ and minimizing $(\theta_0^- + \theta_1^+)$, and vice versa for the lower bound. In particular, we derive the following bounds on Θ :

Proposition 1: *Given restrictions (8a)-(8c) and (9),*

$$\Theta \in [\max\{-Q_u, -2\theta_1^{UB+} - \Delta, -2\theta_0^{UB-} + \Delta\}, \min\{Q_u, 2\theta_1^{UB-} - \Delta, 2\theta_0^{UB+} + \Delta\}].$$

See Appendix B for a proof of this result.

Using this proposition, we can directly bound the average treatment effect when there are classification errors. In particular, bounds on the ATE follow directly by combining the Proposition 1

bound on Θ with Equation (8). Notice that allowing for ambiguity created by the reporting error problem (weakly) widens the treatment effect bounds. At the same time, Θ might be bounded to lie in a strictly positive or negative range. Thus, the upper bound on the ATE can decrease, or the lower bound increase, even as the overall width of the ATE bound expands. Applying Proposition 1 above to evaluate the impacts of the National School Lunch Program, Gundersen, Kreider, and Pepper (forthcoming) provide an example in which it is easier to identify the sign of their ATE of interest in the presence of classification errors (when the error patterns are constrained as described above) than under the standard implicit assumption of perfectly measured data.

Additional analysis is required to address the classification error problem under the MTS assumption. Under this assumption, the ATE can be written as the difference of conditional means:

$$ATE(1,0) \leq P(H = 1|FS^* = 1) - P(H = 1|FS^* = 0).$$

In the absence of classification errors, this upper bound is simply the difference in the observed poor health rate among recipients and nonrecipients. With classification errors, we can write

$$ATE(1,0) \leq \frac{P(H=1,FS=1)+\theta_1^- - \theta_1^+}{P(FS^*=1)} - \frac{P(H=1,FS=0)+\theta_1^+ - \theta_1^-}{P(FS^*=0)}, \quad (10)$$

where information on the true participation rate, $P(FS^* = 1)$, implies bounds on these conditional probabilities. In particular, we can narrow the Proposition 1 bounds as follows:

Proposition 2: Given the MTS assumption in Equation (5) and the classification error model restrictions in Equations (8) and (9), it follows that

$$ATE(1,0) \leq \begin{cases} \frac{P(H = 0)}{P(FS^* = 0)} & \text{if } 0 < P(FS^* = 1) < P(H = 1, FS = 1) \\ \frac{P(H = 1, FS = 1) + \theta_1^{UB-*}}{P(FS^* = 1)} - \frac{P(H = 1, FS = 0) - \theta_1^{UB-*}}{P(FS^* = 0)} & \text{if } P(H = 1, FS = 1) \leq P(FS^* = 1) < 1 \end{cases}$$

where $\theta_1^* \equiv \min\{Q_u, P(H = 1, FS = 0), P(FS^* = 1) - P(H = 1, FS = 1)\}$.

See Appendix B for a proof this result.

Except for the true participation rate, $P(FS^* = 1)$, all of the probabilities in Propositions 1 and 2 can be consistently estimated using data from the NHANES. To infer $P(FS^* = 1)$, we combine auxiliary data on the size of the caseload with data from the NHANES on the size of the eligible population. Administrative data from the USDA reveals that from 2001-2006 there was an average of nearly 11 million children receiving food stamps per year (calculated using annual reports from Rosso; 2002; Genser, 2003; Cunnyngham and Brown, 2004; Poikolainen, 2005; Barrett, 2006; and Wolkwitz, 2007). From the NHANES, we estimate that about 22 million children were eligible to receive assistance. Thus, the implied participation rate is about 0.50, 4 points higher than the reported rate of 0.456. We note that a net false negative reporting rate of 4% is consistent with the results Meyer *et al.* (2010) when evaluating misreporting the SIPP, but much smaller than found in the CPS and PSID. Using the CPS, for example, the participation rate is estimated to be around 70% (Cunnyngham, 2005). Given this variability in the estimated participation rates and the possibility that errors in classifying eligible children may bias the estimated participation rates, we assess the sensitivity of the bounds to variation in the true participation rate.

B. Results

Our analytical approach allows us to trace out sharp bounds on the ATE under different assumptions about selection and measurement error. To do so, we evaluate the bounds as a function of

the unknown SNAP participation rate, P^* , under various assumptions about the selection process. By layering successively stronger assumptions, our analysis reveals how the strength of the conclusions varies with the strength of the identifying assumptions. We begin in Section B.1 by focusing on bounding the impact of SNAP participation on food insecurity. We then extend the discussion to the three other health outcomes in Section B.2.

B.1. Food Insecurity

Figure 1 traces out the estimated Proposition 1 and 2 bounds – that is, the worst-case and MTS bounds – for the ATE on the food insecurity rate across all values of P^* between 0 and 1. The accompanying table highlights these results for P^* equal to (a) the NHANES self-reported participation rate of $P = 0.456$, (b) our preferred estimated true participation rate of 0.50 based on administrative data from the USDA, and (c) a higher rate of 0.70 chosen to be consistent with the participation rate found using the CPS (Cunningham, 2005). The solid lines in the figures trace out the estimated arbitrary error bounds (i.e., $Q_u = 1$) in which there are no restrictions imposed on the nature or degree of errors except those implied by the knowledge of P and P^* as captured by Equations (8a)-(8c). The dashed lines display the estimated bounds under the further restriction of no excess errors: $Q_u = |P^* - P|$. Recall that for $P^* \geq P$ in the right-hand-side of the figure, which represents underreporting (the most relevant cases in our application), this no excess error model is equivalent to no false positive reports. The table also provides Imbens-Manski (2004) confidence intervals that cover the true value of the ATE with 90% probability.

As noted above, the worst-case bounds on the ATE if SNAP receipt is accurately reported ($P^* = P$ and no excess errors) have a width of one and always include zero. For food insecurity, these worst-case no errors bounds are $[-0.445, 0.555]$ as depicted in the figure by the solid vertical line at $P^* = P$ (see also Table 2). Allowing for classification errors notably increases the width of these bounds. For example, suppose the true participation rate remains equal to the self-reported rate of 0.456, but now one only imposes the assumption of no *net* reporting errors such that the rate of false positives equals the rate of false negatives. Then, as shown in Figure 1 and the accompanying table, the ATE bounds on the food insecurity rate expand from $[-0.445, 0.555]$ to $[-0.855, 0.944]$, with a width of 1.799. If the true

participation rate is 0.50 (the rate consistent with the USDA administrative data) instead of 0.456, the bounds change to $[-0.900, 0.900]$ with a width of 1.800. These findings reveal the important negative result that the ambiguity created by classification errors can be substantial even if the true and self-reported rates are similar.

The upper bounds when P^* is near P , however, are markedly reduced by introducing the MTS assumption, especially when combined with the no excess errors assumption (no false-positives when $P^* \geq P$). With $P^* = 0.50$, for example, the no false positives bounds are $[-0.489, 0.599]$, with a width of 1.088. Adding the MTS assumption further reduces the upper bound to 0.198. Thus, the no false positives assumption decreases the ambiguity associated with measurement error from 1.800 to 1.088, a 40% reduction, and the MTS assumption further reduces the width of the bound to 0.687, more than 60% narrower than the width of the worst-case bounds when $P^* = 0.50$. While these two assumptions have substantial identifying power in this application, these wide bounds presented in Figure 1 highlight the difficulty of making strong inferences in light of the selection and measurement error problems. In the absence of additional restrictions that address the selection problem, we cannot rule out the possibility that SNAP has a large positive or negative impact on the likelihood of poor health.

To narrow the bounds, we assess the identifying power of the joint MTS-MIV and joint MTS-MIV-MTR assumptions. These results are traced out in Figure 2A and the corresponding table for the most relevant cases in our application in which P^* lies between 0.456 and 0.700. Focusing on the no-false positive classification error model, we begin by combining the MTS assumption with our MIV assumption that the probability of good health weakly increases with family resources, as measured by the ratio of income to the poverty line. In this joint MTS-MIV model, we can often sign the ATE as strictly negative without imposing the MTR assumption. In particular, we find that the bounds are strictly negative under the MTS-MIV model as long as the degree of food stamp misreporting is not too severe. Specifically, Figure 2A reveals that we can identify ATE to be negative as long as food stamp misreporting is confined to no more than about 6% of households, ranging from a 12.8 percent reduction at $P^* = 0.456$ to no effect at $P^* = 0.52$. When $P^* = 0.50$, the estimates imply that SNAP reduces food insecurity by at least 2.7 percentage points, although this upper bound is not statistically different than

zero at the ten percent significance level.

Under the joint MTR-MTS-MIV assumption, the average treatment effect is strictly negative even for large degrees of arbitrary food stamp misreporting. Under this joint assumption, our estimated bounds on the ATE vary from $[-0.808, -0.081]$ when $P^* = 0.456$ to $[-0.754, -0.081]$ when $P^* = 0.70$. In all cases, the estimates are statistically different than zero at the ten percent significance level. Thus, under this model, we find that SNAP reduces the food insecurity rate by at least 8 percentage points and perhaps much more. These results suggest that SNAP dramatically improves the likelihood of becoming food secure.

B.2. Other Health Outcomes

We also consider what can be learned about the effects of food stamps on the three other negative health outcomes: self-reported fair/poor general health, childhood obesity, and anemia. For brevity, we concentrate on results for cases when we impose the MTS-MIV and MTS-MIV-MTR models. These results are summarized in Figures 2B-2D.

As above for the case of food insecurity, we can identify strictly negative average treatment effects for each health outcome under the joint MTS-MIV assumption for sufficiently small degrees of food stamp reporting error. For example, without any errors – i.e., when $P^* = 0.456$ with no excess errors – the average treatment effects for fair/poor health, obesity, and anemia are identified to be no greater than -0.061 , -0.033 , and -0.032 , respectively. Identification of the ATE decays rapidly with P^* for each of the health outcomes, however, and each upper bound becomes positive if fewer than 4 percent of the households might misreport food stamp participation status (under either arbitrary errors or no false positives). Still, if the true participation rate is 0.50 (the rate consistent with the USDA administrative data) under no false positives, the estimated upper bounds rule out the possibility the SNAP substantially increases the incidence of these poor health outcomes. For example, the estimated bounds on the rate of poor general health is $[-0.500, 0.005]$. Thus, these results imply that SNAP may dramatically improve childhood health, as measured by poor health, obesity, and anemia, with little downside risk that the program instead has a deleterious average effect.

Finally, to shed additional light on the magnitudes of any identified beneficial effects of the program, we apply the MTR assumption that SNAP cannot lead to worse health outcomes. Under the joint MTS-MIV-MTR, we estimate strictly negative (beneficial) and substantial impacts for each health outcome across all values of $P^* \in [0.456, 0.700]$, even for the case of arbitrary reporting errors. In particular, at $P^* = 0.50$ we find that SNAP reduces the rate of poor general health by at least 0.031 (from 0.089 to 0.058), obesity by at least 0.053 (from 0.218 to 0.165), and anemia by at least 0.016 (from 0.020 to 0.004).

VI. Conclusion

The literature assessing the efficacy of SNAP has long puzzled over positive associations between food stamp receipt and various undesirable health-related outcomes such as food insecurity. These associations are often ascribed to the self-selection of less healthy households into SNAP. Identification of the causal impacts of participation on health status is also confounded by systematic underreporting of food stamp reciprocity. In this paper, we reconsidered the impact of SNAP on child food insecurity and other health outcomes by developing methods that account for both of these identification problems in a single unifying framework. Within this framework, we combine information from household self-reports with administrative data on the size of the SNAP caseload to derive formal restrictions on the magnitudes and patterns of household reporting errors. While introducing measurement error widens Manski's (1995) classic worst-case selection bounds, we show how restricting the magnitudes and patterns of errors can, in some cases, make it easier to sign the average treatment effect than under the standard implicit assumption of perfectly measured data. Acknowledging the presence of nonrandom measurement error need not necessarily hinder inference on the sign of the average treatment effect.

Our partial identification approach is well-suited for this application in which conventional assumptions strong enough to point-identify the causal impacts are not necessarily credible and there remains much uncertainty about even the qualitative impacts of SNAP. Using data from the National Health and Nutrition Examination Survey (NHANES), we make transparent how assumptions on the selection and reporting error processes shape inferences about the causal impacts of food stamp reciprocity on health outcomes. The worst-case selection bounds always include zero, and classification

errors can generate substantial additional uncertainty about the efficacy of SNAP in alleviating food insecurity and other poor health outcomes. This ambiguity, however, is substantially mitigated by applying relatively weak assumptions on the selection and classification error processes.

Our middle-ground MTS-MIV model allows us to identify strictly beneficial impacts of SNAP on food insecurity and other health outcomes as long as the degree of misreported participation status is not too large. In the absence of measurement error, the joint MTS-MIV model reveals that SNAP reduces the prevalence of food insecurity by at least 12.8 percentage points (28%), from 0.459 to 0.331. The strength of this finding naturally weakens when we allow for misclassified participation status. Nevertheless, we can still identify that SNAP reduces the prevalence of food insecurity by at least 2.7 percentage points (6.5%), from 0.417 to 0.390, when allowing for errors of omission in participation status to be consistent with the estimated true participation rate of $P^* = 0.50$. Identification decays rapidly with the degree of misreporting, however, and confidence intervals for the average treatment effect include zero unless it can be known that only a small fraction of households misreport.

Under the stronger joint MTS-MIV-MTR model, which may be less credible for obesity than the other health outcomes, the basic conclusion that SNAP improves health outcomes holds even for large degrees of measurement error. Given errors of omission consistent with $P^* = 0.50$, SNAP is estimated to reduce the prevalence of child food insecurity by at least 8.9 percentage points, poor general health by 3.1 percentage points, obesity by 5.3 percentage points, and anemia by 1.6 percentage points. These impacts are significantly different from zero for food insecurity and poor health. These findings suggest that at least some of the potentially troubling correlations between food stamp reciprocity and poor health outcomes provide a misleading picture of the impact of SNAP. The program appears to lead to at least modest reductions in food insecurity and other poor health outcomes, with little downside risk that the program has significant deleterious effects. Our estimated bounds are also consistent with the possibility that SNAP dramatically improves the well-being of children in the United States.

References

- Barrett, A. 2006. Characteristics of Food Stamp Households: Fiscal Year 2005. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/oane/).
- Bitler, M., J. Currie, and J. Scholz. 2003. "WIC Eligibility and Participation." *Journal of Human Resources*, 38: 1139-1179.
- Bollinger, C. 1996. "Bounding Mean Regressions When A Binary Variable is Mismeasured." *Journal of Econometrics*, 73(2), 387-99.
- Bollinger, C. and M. David. 1997. "Modeling Discrete Choice with Response Error: Food Stamp Participation." *Journal of the American Statistical Association*, 92 (439): 827-835.
- Breunig, R., and I. Dasgupta (2002) "Do Intra-Household Effects Generate the Food Stamp Cash-Out Puzzle?" *American Journal of Agricultural Economics*, 84(4): 552-568.
- Case, A., D. Lubotsky, and C. Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review*, 92(5): 1308-1334.
- Chernozhukov, Victor, Sokbae Lee and Adam M. Rosen, (2009). "Intersection Bounds: Estimation and Inference," Working Paper. CEMMAP Working Paper, 19/09.
- Cunningham, K. 2005. *Food Stamp Program Participation Rates: 2003*. Washington, D.C.: U.S. Department of Agriculture, Food, and Nutrition Service.
- Cunningham, K. and Brown, B. 2004. Characteristics of Food Stamp Households: Fiscal Year 2003. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/oane/).
- Currie, J. 2003. "U.S. Food and Nutrition Programs," in Means Tested Transfer Programs in the US, ed. Robert Moffitt, University of Chicago Press, Chapter 4.
- Deaton, A. 2002. "Policy Implications of the Gradient of Health and Wealth." *Health Affairs*, 21(2):13-30.
- DePolt, R.A., R.A. Moffitt and D.C. Ribar, 2009. "Food stamps, Temporary Assistance for Needy Families and food hardships in three American cities." *Pacific Economic Review*, 14, 445-473.
- Finegold, Kenneth, William Margrabe and Caroline Ratcliffe, (2006). "Food Stamp Program Rules Database Documentation. Report prepared by the Urban Institute for Economic Research Service of the USDA.
- Genser, J. 2003. Characteristics of Food Stamp Households: Fiscal Year 2002. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/oane/).
- Gundersen, C. and B. Kreider. 2008. "Food Stamps and Food Insecurity: What Can Be Learned in the Presence of Nonclassical Measurement Error?" *Journal of Human Resources*, 43(2): 352-382.

- Gundersen, C., B. Kreider, and J. Pepper. forthcoming. “The Impact of the National School Lunch Program on Child Health: A Nonparametric Bounds Analysis.” *Journal of Econometrics*.
- Gundersen, C., B. Lohman, S. Garasky, S. Stewart, and J. Eisenmann. 2008. “Food Security, Maternal Stressors, and Overweight among Low-Income U.S. Children: Results from NHANES 1999-2002.” *Pediatrics*, v122, e529–e540.
- Gundersen, C. and S. Offutt. 2005. “Farm Poverty and Safety Nets.” *American Journal of Agricultural Economics*, 87(4): 885–899.
- Gundersen, C. and V. Oliveira. 2001. “The Food Stamp Program and Food Insufficiency,” *American Journal of Agricultural Economics*, 84(3): 875-887.
- Haider, S., A. Jackowitz, and R. Schoeni. 2003. “Food Stamps and the Elderly: Why is Participation So Low?” *Journal of Human Resources*, 38(S):1080-1111.
- Hirano, Keisuke and Jack R. Porter, (2009). “Impossibility Results for Nondifferentiable Functionals,” Working Paper.
- Horowitz, J.L., and C.F. Manski, 1995. “Identification and Robustness With Contaminated and Corrupted Data.” *Econometrica*, 63(2), 281-02.
- Hoynes, H. and D. Schanzenbach. 2009. “Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program,” *American Economic Journal: Applied Economics*, 1(4): 109-139.
- Jensen, H. 2002. “Food Insecurity and the Food Stamp Program.” *American Journal of Agricultural Economics* 84(5): 1215-28.
- Jolliffe, D. Forthcoming. “Overweight and Poor? On the Relationship between Income and the Body Mass Index.” *Economics and Human Biology*.
- Imbens, G. and C. Manski. 2004. “Confidence Intervals for Partially Identified Parameters.” *Econometrica*, 72(6): 1845-1857.
- Kaushal, N. 2007. “Do Food Stamps Cause Obesity? Evidence from the Immigrant Experience.” *Journal of Health Economics*, 26: 968-991.
- Kreider, B. and S. Hill. 2009. “Partially Identifying Treatment Effects with an Application to Covering the Uninsured,” *Journal of Human Resources*, 44(2): 409-449.
- Kreider, B. and J. Pepper. 2007. “Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors.” *Journal of the American Statistical Association*, 102(478): 432-441.
- Kreider, B. and J. Pepper. 2008. “Inferring Disability Status from Corrupt Data.” *Journal of Applied Econometrics*, 23(3): 329-49.
- Leftin, J., A. Gothro, and E. Eslami. 2010. *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2009*. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/fns)

- Levedahl, J. "A Theoretical and Empirical Evaluation of the Functional Forms Used to Estimate the Food Expenditure Equation of Food Stamp Recipients." *American Journal of Agriculture Economics*, v77, 960-968. 1995.
- Manski, C. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.
- Manski, C. 1997. "Monotone Treatment Response." *Econometrica*, 65(6): 1311-1334.
- Manski, C. and J. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 997-1010.
- Manski, C. and J. Pepper. 2009. "More on Monotone Instrumental Variables," *The Econometrics Journal*, 12, July 2009, S200-216.
- Marquis, K. and J. Moore. 1990. "Measurement errors in SIPP program reports" in Proceedings of the Bureau of the Census Annual Research Conference, Wash., DC: Bureau of the Census, 721-745.
- Meyer, B., D.W. Mok, and J.X. Sullivan, 2009. The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences. Working Paper.
- Meyerhoefer, C. and V. Pylypchuk. 2008. "Does Participation in the Food Stamp Program Increase the Prevalence of Obesity and Health Care Spending?" *American Journal of Agricultural Economics*, 90(2): 287-305.
- Moffitt, R. 1983. "An Economic Model of Welfare Stigma." *American Economic Review*, 73:1023-35.
- Moffitt, R. 2005. "Remarks on the Analysis of Causal Relationships in Population Research," *Demography*, 91-108.
- Molinari, F. 2008. "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144(1), 81-117.
- Molinari, F. 2010. "Missing Treatments." *Journal of Business and Economic Statistics*, 28(1), 82-95
- Newacheck, P. 1994. "Poverty and Childhood Chronic Illness." *Archives of Pediatrics and Adolescent Medicine*, 148:1143-49.
- Nord, M., A. Coleman-Jensen, M. Andrews and S. Carlson. 2010. *Household Food Security in the United States, 2009*. (U.S. Department of Agriculture, Economic Research Report 108). Washington, DC: U.S. Government Printing Office.
- Pepper, J. 2000. "The Intergenerational Transmission of Welfare Receipt: A Nonparametric Bounds Analysis." *The Review of Economics and Statistics*, 82(3): 472-488.
- Poikolainen, A. 2005. Characteristics of Food Stamp Households: Fiscal Year 2004. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/oane/).

- Rank, M. R. and T. A. Hirschl. 2009. "Estimating the Risk of Food Stamp Use and Impoverishment During Childhood." *Archives of Pediatrics and Adolescent Medicine*, 163(11): 994-999.
- Rosso, R. 2002. Characteristics of Food Stamp Households: Fiscal Year 2001. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C. (available online at www.fns.usda.gov/oane/).
- Shaikh, A. and E. Vytlacil, 2011. "Partial Identification in Triangular Systems of Equations with Binary Dependent Variables," *Econometrica*, 79(3), 949-956.
- Shrewsbury, V and J. Wardle, 2008. Socioeconomic Status and Adiposity in Childhood: A Systematic Review of Cross-Sectional Studies 1990-2005. *Obesity* 16, 275-284.
- Trippe C., P. Doyle and A. Asher. 1992. *Trends in Food Stamp Program Participation Rates, 1976 to 1990*. US Dept. of Agriculture, Food and Nutrition Service.
- U.S. Department of Agriculture. 1999. *Food and Nutrition Service. Annual Historical Review: Fiscal Year 1997*.
- Wilde, P. and M. Nord. 2005. "The Effect of Food Stamps on Food Security: A Panel Data Approach." *Review of Agricultural Economics*, 27(3): 425-432.
- Wolkwitz, K. 2007. Characteristics of Food Stamp Households: Fiscal Year 2006. Prepared by Mathematica Policy Research, Inc. for the Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C.

Table 1: Means by Reported Food Stamp Program Participation

Variable	Income-eligible children	Recipients ($FS=1$)	Nonrecipients ($FS=0$)
Age in years	9.108 (0.099)	8.607 ^{***} (0.127)	9.527 (0.132)
Ratio of income to the poverty line	0.754 (0.011)	0.631 ^{***} (0.016)	0.857 (0.014)
Food stamp recipient	0.456 (0.022)		
Food insecure	0.400 (0.015)	0.450 ^{**} (0.023)	0.357 (0.024)
Poor or fair health	0.080 (0.005)	0.088 (0.008)	0.073 (0.007)
Obese	0.185 (0.008)	0.191 (0.014)	0.179 (0.012)
Anemia ^a	0.012 (0.002)	0.013 (0.003)	0.010 (0.003)

Notes: Sample estimates weighted using the medical exam weight. Standard errors in parentheses account for the sample design using the synthetic strata and PSU variables. The estimated means for the Food Stamp recipient population are superscripted with *, **, or *** to indicate that they are statistically significantly different from the means for the nonrecipient population, with p-values less than 0.1, 0.05, 0.01, respectively, based on Wald statistics corrected for the sample design.

^a The sample size for anemia is 3,871 (with 1,888 food stamp recipients) due to missing observations. The sample size for the other three outcomes is 4,418.

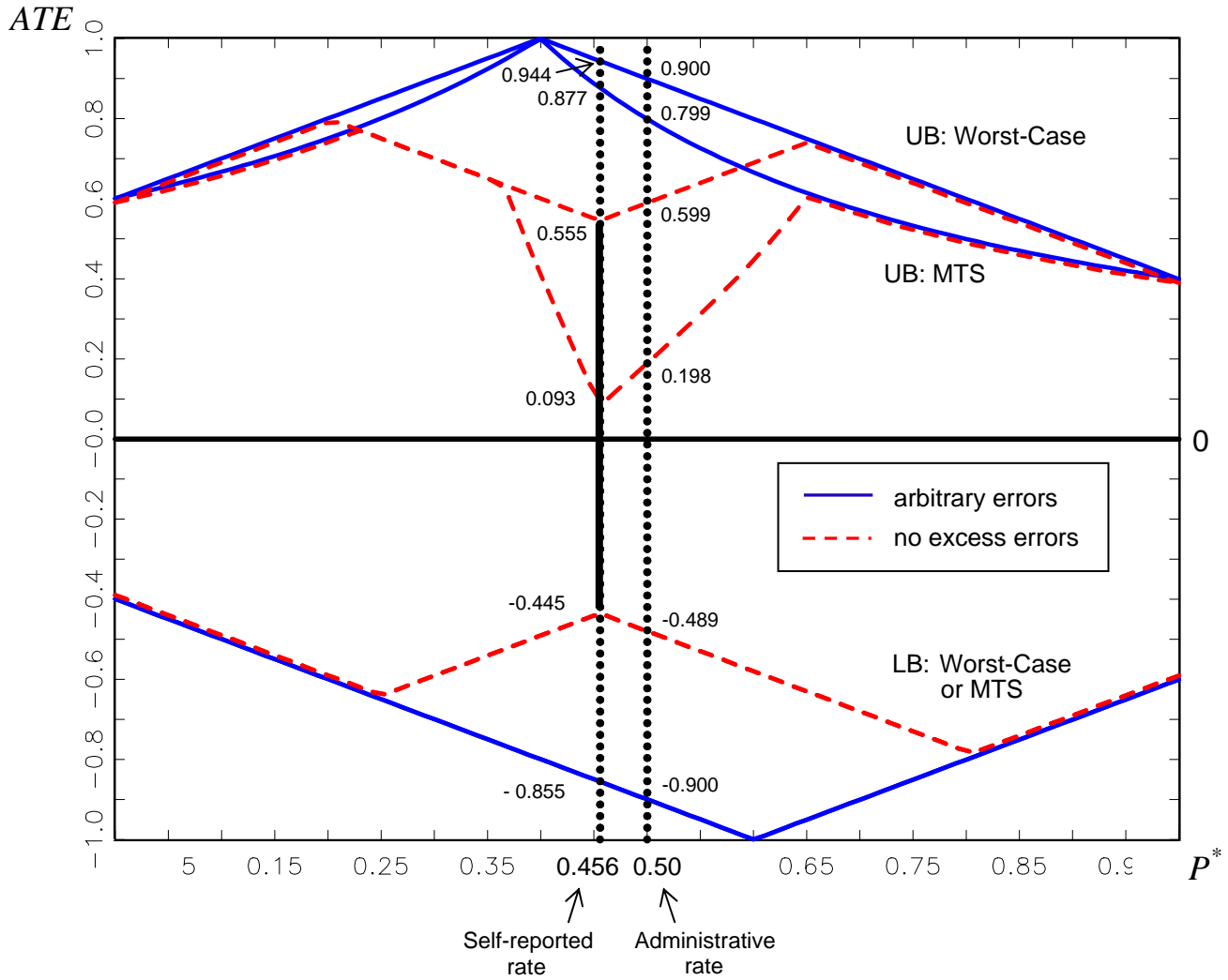
Table 2. Sharp Bounds on the ATE of SNAP Participation Under No Measurement Error

		NHANES Outcomes:							
		Food Insecurity		Poor Health		Obesity		Anemia	
Worst-Case:	p.e. [†]	[-0.445, 0.555]		[-0.455, 0.545]		[-0.466, 0.534]		[-0.460, 0.540]	
	CI [‡]	[-0.460 0.570]		[-0.470 0.559]		[-0.482 0.549]		[-0.474 0.555]	
MTS:	p.e.	[-0.445, 0.093]		[-0.455, 0.015]		[-0.466, 0.012]		[-0.460, 0.003]	
	CI	[-0.460 0.139]		[-0.470 0.048]		[-0.482 0.050]		[-0.474 0.008]	
MTS-MIV:	p.e.	[-0.366, -0.128]		[-0.398, -0.061]		[-0.411, -0.033]		[-0.391, -0.032]	
	CI	[-0.433 -0.034]		[-0.453 -0.005]		[-0.474 0.033]		[-0.450 0.006]	
	bias [*]	+0.027	-0.039	+0.019	-0.021	+0.022	-0.072	+0.008	-0.006
MTS-MIV-MTR:	p.e.	[-0.366, -0.149]		[-0.398, -0.061]		[-0.411, -0.041]		[-0.391, -0.034]	
	CI	[-0.433 -0.062]		[-0.453 -0.009]		[-0.474 0.000]		[-0.450 0.000]	
	bias	+0.027	-0.043	+0.019	-0.027	+0.022	-0.073	+0.008	-0.005
		Food Insecurity in the CPS							
Worst-Case:	p.e.	[-0.399, 0.601]							
	CI	[-0.412 0.614]							
MTS:	p.e.	[-0.399, 0.178]							
	CI	[-0.501 0.406]							
MTS-MIV:	p.e.	[-0.376, 0.138]							
	CI	[-0.412 0.205]							
	bias	+0.025	-0.083						
MTS-MIV-MTR:	p.e.	[-0.376, -0.049]							
	CI	[-0.412 0.000]							
	bias	+0.025	-0.039						

[†] Point estimates (p.e.) and [‡] 90% Imbens-Manski confidence intervals (CI) using 1,000 pseudosamples

^{*} Corrected finite sample bias

Figure 1. Sharp Bounds on the ATE for **Food Insecurity** as a Function of P^* , the Unobserved True SNAP Participation Rate: MTS-MIV and MTS-MIV-MTR Bounds



$P^* = P = 0.456$ $P^* = 0.50$ $P^* = 0.70$

Worst-Case:

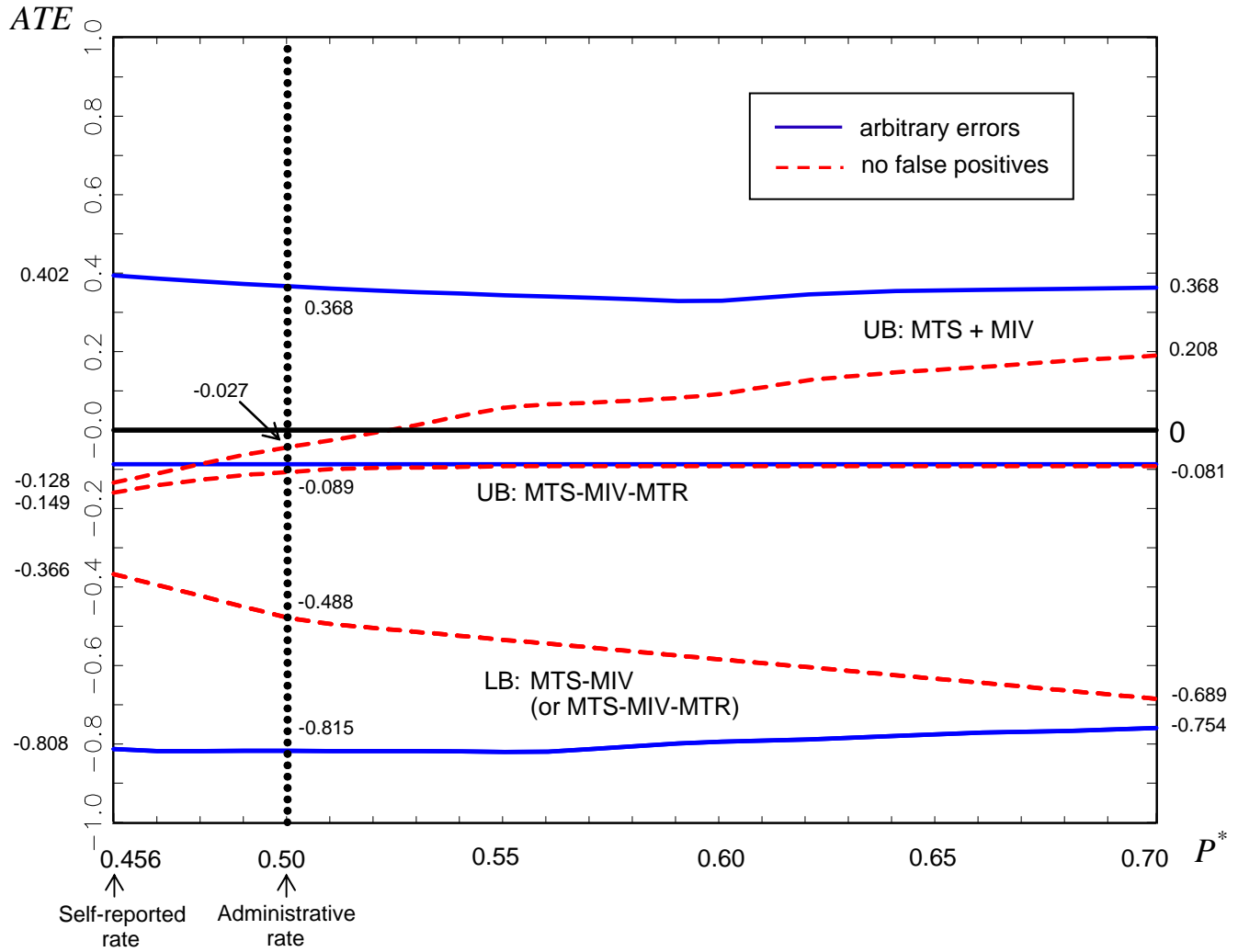
Arbitrary errors	p.e. [†]	[-0.855, 0.944]	[-0.900, 0.900]	[-0.900, 0.700]
	CI [‡]	[-0.868 0.956]	[-0.912 0.912]	[-0.913 0.712]
No excess errors	p.e.	[-0.445, 0.555]	[-0.489, 0.599]	[-0.689, 0.700]
	CI	[-0.460 0.570]	[-0.506 0.619]	[-0.706 0.712]

MTS:

Arbitrary errors	p.e.	[-0.855, 0.877]	[-0.900, 0.799]	[-0.900, 0.571]
	CI	[-0.868 0.904]	[-0.912 0.824]	[-0.913 0.589]
No excess errors	p.e.	[-0.445, 0.093]	[-0.489, 0.198]	[-0.689, 0.571]
	CI	[-0.460 0.139]	[-0.506 0.238]	[-0.706 0.589]

[†] Point estimates (p.e.) and [‡] 90% Imbens-Manski confidence intervals (CI) using 1,000 pseudosamples

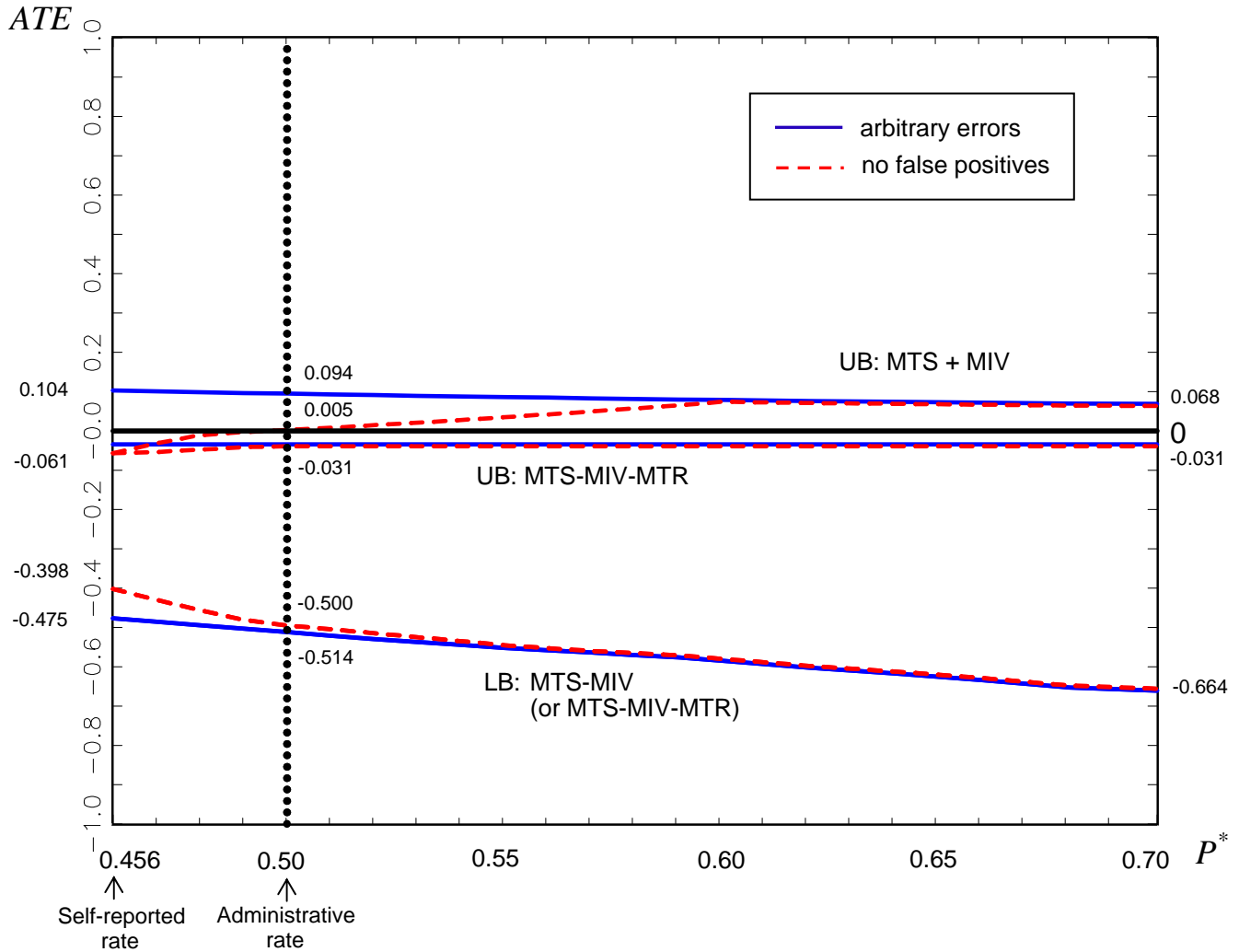
Figure 2A. Sharp Bounds on the ATE for **Food Insecurity** as a Function of P^* , the Unobserved True SNAP Participation Rate: MTS-MIV and MTS-MIV-MTR Bounds



		$P^* = P = 0.456$	$P^* = 0.50$	$P^* = 0.70$
(a) MTS-MIV, arbitrary errors	p.e.	[-0.808, 0.402]	[-0.815, 0.368]	[-0.754, 0.368]
	CI	[-0.874 0.543]	[-0.879 0.509]	[-0.816 0.511]
	bias [†]	+0.032 -0.051	+0.032 -0.046	+0.025 -0.054
(b) MTS-MIV, no false positives	p.e.	[-0.366, -0.128]	[-0.488, -0.027]	[-0.689, 0.208]
	CI	[-0.433 -0.034]	[-0.553 0.074]	[-0.723 0.353]
	bias	+0.027 -0.039	+0.025 -0.039	+0.028 -0.063
(c) MTS-MIV-MTR, arbitrary errors	p.e.	[-0.808, -0.081]	[-0.815, -0.081]	[-0.754, -0.081]
	CI	[-0.874 -0.015]	[-0.879 -0.015]	[-0.816 -0.014]
	bias	+0.032 -0.032	+0.032 -0.032	+0.025 -0.032
(d) MTS-MIV-MTR, no false positives	p.e.	[-0.366, -0.149]	[-0.488, -0.089]	[-0.689, -0.081]
	CI	[-0.433 -0.062]	[-0.553 -0.013]	[-0.723 -0.013]
	bias	+0.027 -0.043	+0.025 -0.042	+0.028 -0.032

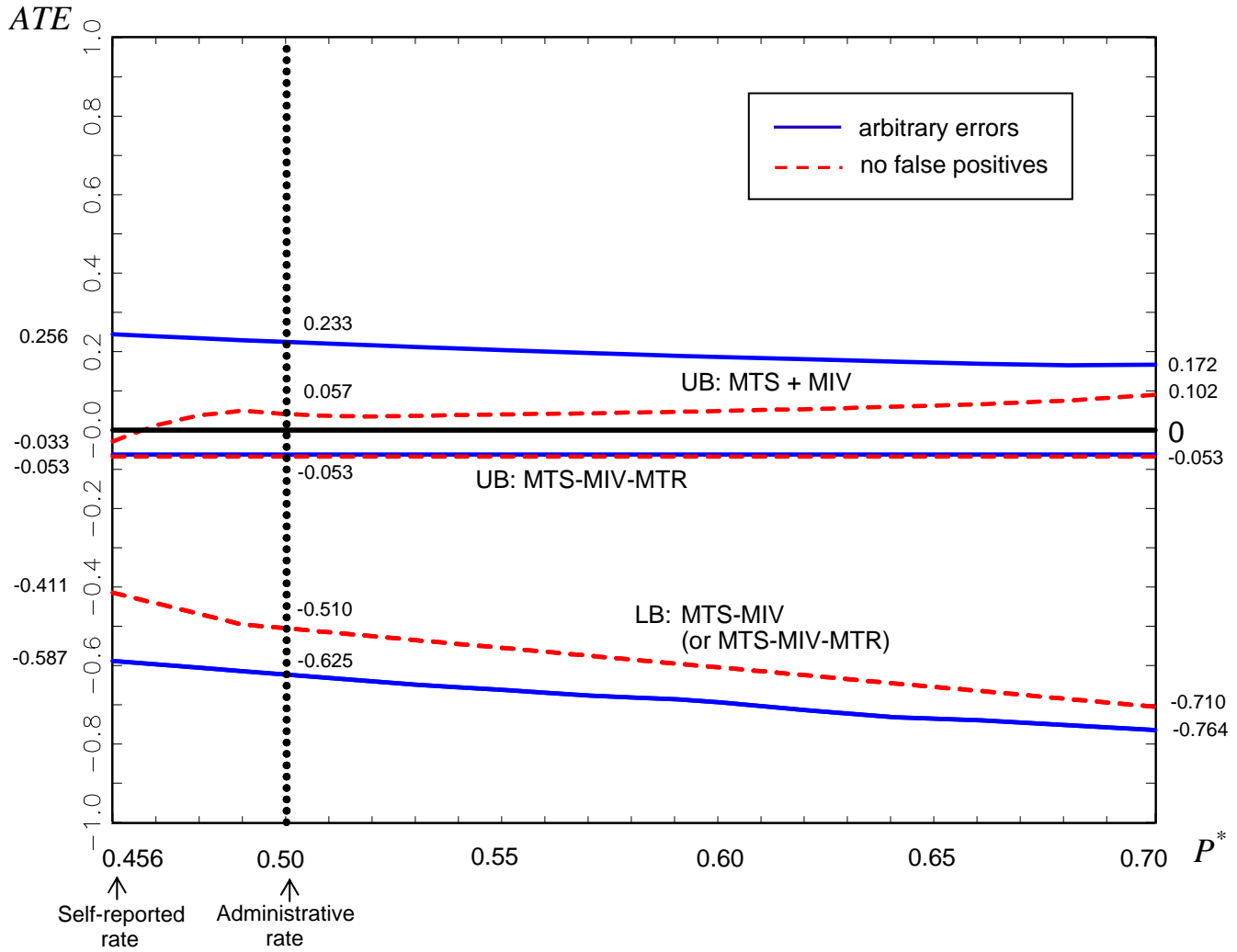
[†] corrected finite sample bias

Figure 2B. Sharp Bounds on the ATE for **Poor Health** as a Function of P^* , the Unobserved True SNAP Participation Rate: MTS-MIV and MTS-MIV-MTR Bounds



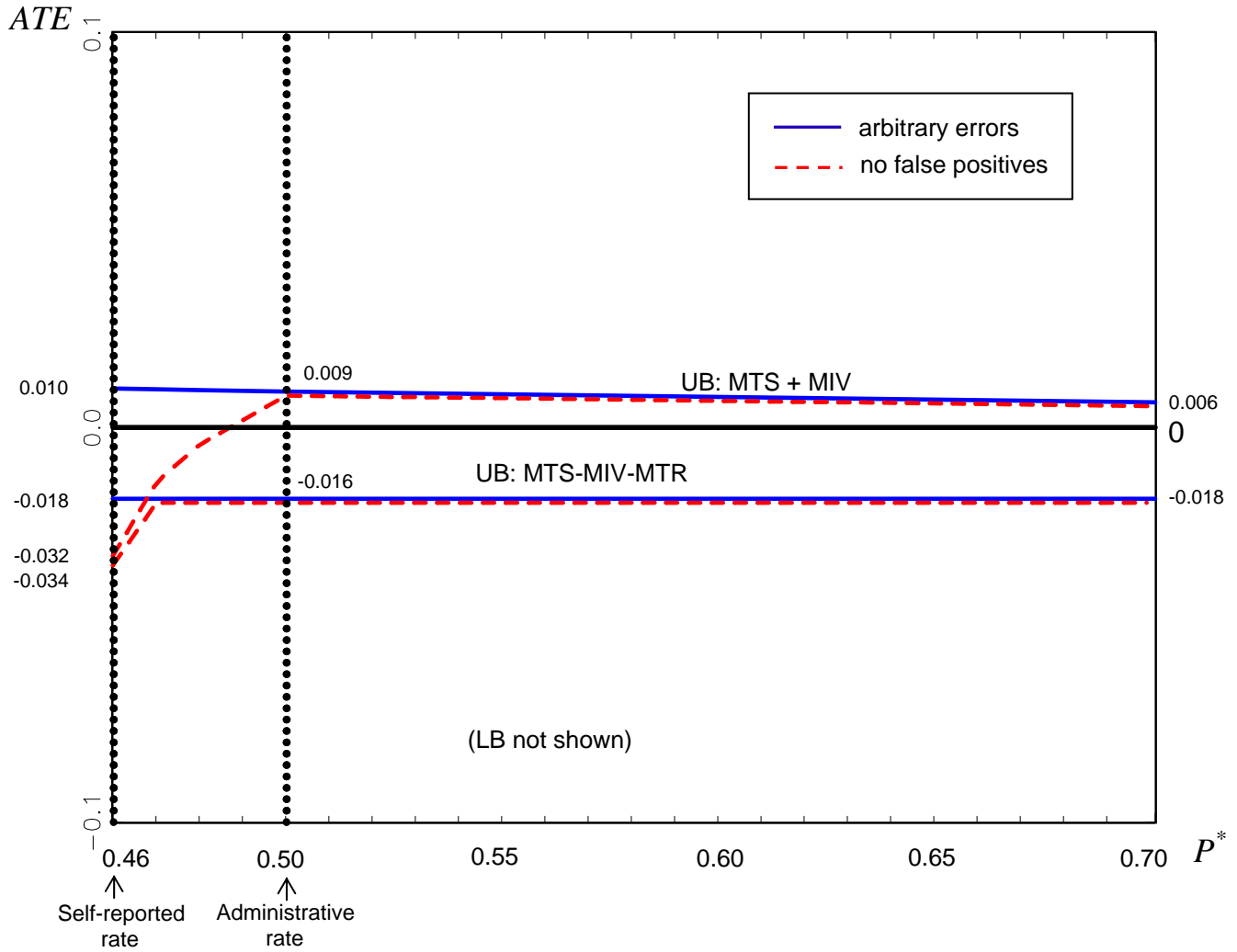
		$P^* = P = 0.456$	$P^* = 0.50$	$P^* = 0.70$
(a) MTS-MIV, arbitrary errors	p.e.	[-0.475, 0.104]	[-0.514, 0.094]	[-0.664, 0.068]
	CI	[-0.529 0.133]	[-0.572 0.122]	[-0.745 0.088]
	bias	+0.008 -0.010	+0.010 -0.009	+0.017 -0.006
(b) MTS-MIV, no false positives	p.e.	[-0.398, -0.061]	[-0.500, 0.005]	[-0.664, 0.068]
	CI	[-0.453 -0.005]	[-0.552 0.075]	[-0.739 0.100]
	bias	+0.019 -0.021	+0.022 -0.014	+0.017 -0.010
(c) MTS-MIV-MTR, arbitrary errors	p.e.	[-0.475, -0.031]	[-0.514, -0.031]	[-0.664, -0.031]
	CI	[-0.529 0.000]	[-0.572 0.000]	[-0.745 0.000]
	bias	+0.008 -0.025	+0.010 -0.025	+0.017 -0.025
(d) MTS-MIV-MTR, no false positives	p.e.	[-0.398, -0.061]	[-0.500, -0.031]	[-0.664, -0.031]
	CI	[-0.453 -0.009]	[-0.552 0.000]	[-0.739 0.000]
	bias	+0.019 -0.027	+0.022 -0.026	+0.017 -0.025

Figure 2C. Sharp Bounds on the ATE for **Obesity** as a Function of P^* , the Unobserved True SNAP Participation Rate: MTS-MIV and MTS-MIV-MTR Bounds



		$P^* = P = 0.456$	$P^* = 0.50$	$P^* = 0.70$
(a) MTS-MIV, arbitrary errors	p.e.	[-0.587, 0.256]	[-0.625, 0.233]	[-0.764, 0.172]
	CI	[-0.652 0.361]	[-0.695 0.329]	[-0.852 0.222]
	bias	+0.011 -0.074	+0.013 -0.056	+0.034 -0.020
(b) MTS-MIV, no false positives	p.e.	[-0.411, -0.033]	[-0.510, 0.057]	[-0.710, 0.102]
	CI	[-0.474 0.033]	[-0.564 0.153]	[-0.772 0.218]
	bias	+0.022 -0.072	+0.023 -0.066	+0.034 -0.047
(c) MTS-MIV-MTR, arbitrary errors	p.e.	[-0.587, -0.053]	[-0.625, -0.053]	[-0.764, -0.053]
	CI	[-0.652 0.000]	[-0.695 0.000]	[-0.852 0.000]
	bias	+0.011 -0.045	+0.013 -0.045	+0.034 -0.045
(d) MTS-MIV-MTR, no false positives	p.e.	[-0.411, -0.053]	[-0.510, -0.053]	[-0.710, -0.053]
	CI	[-0.474 0.000]	[-0.564 0.000]	[-0.772 0.000]
	bias	+0.022 -0.073	+0.023 -0.049	+0.034 -0.045

Figure 2D. Sharp Bounds on the ATE for **Anemia** as a Function of P^* , the Unobserved True SNAP Participation Rate: MTS-MIV and MTS-MIV-MTR Bounds



		$P^* = P = 0.46$	$P^* = 0.50$	$P^* = 0.70$
(a) MTS-MIV, arbitrary errors	p.e.	[-0.402, 0.010]	[-0.437, 0.009]	[-0.601, 0.006]
	CI	[-0.460 0.016]	[-0.499 0.014]	[-0.689 0.010]
	bias	+0.005 -0.003	+0.005 -0.002	+0.007 -0.002
(b) MTS-MIV, no false positives	p.e.	[-0.391, -0.032]	[-0.437, 0.009]	[-0.601, 0.006]
	CI	[-0.450 0.006]	[-0.498 0.028]	[-0.688 0.010]
	bias	+0.008 -0.006	+0.005 -0.008	+0.007 -0.002
(c) MTS-MIV-MTR, arbitrary errors	p.e.	[-0.402, -0.018]	[-0.445, -0.016]	[-0.601, -0.018]
	CI	[-0.460 0.000]	[-0.499 0.000]	[-0.689 0.000]
	bias	+0.005 -0.004	+0.005 -0.004	+0.007 -0.004
(d) MTS-MIV-MTR, no false positives	p.e.	[-0.391, -0.034]	[-0.437, -0.018]	[-0.601, -0.018]
	CI	[-0.450 0.000]	[-0.498 0.000]	[-0.688 0.000]
	bias	+0.008 -0.005	+0.005 -0.004	+0.007 -0.004

Appendix A:

Table A1

Food Insecurity Questions in the Core Food Security Module

1. “We worried whether our food would run out before we got money to buy more.” Was that **often, sometimes**, or never true for you in the last 12 months?
2. “The food that we bought just didn’t last and we didn’t have money to get more.” Was that **often, sometimes**, or never true for you in the last 12 months?
3. “We couldn’t afford to eat balanced meals.” Was that **often, sometimes**, or never true for you in the last 12 months?
4. “We relied on only a few kinds of low-cost food to feed our children because we were running out of money to buy food.” Was that **often, sometimes**, or never true for you in the last 12 months?
5. In the last 12 months, did you or other adults in the household ever cut the size of your meals or skip meals because there wasn’t enough money for food? (**Yes/No**)
6. “We couldn’t feed our children a balanced meal, because we couldn’t afford that.” Was that **often, sometimes**, or never true for you in the last 12 months?
7. In the last 12 months, did you ever eat less than you felt you should because there wasn’t enough money for food? (**Yes/No**)
8. (If yes to Question 5) How often did this happen—**almost every month, some months but not every month**, or in only 1 or 2 months?
9. “The children were not eating enough because we just couldn’t afford enough food.” Was that **often, sometimes**, or never true for you in the last 12 months?
10. In the last 12 months, were you ever hungry, but didn’t eat, because you couldn’t afford enough food? (**Yes/No**)
11. In the last 12 months, did you lose weight because you didn’t have enough money for food? (**Yes/No**)
12. In the last 12 months, did you ever cut the size of any of the children’s meals because there wasn’t enough money for food? (**Yes/No**)
13. In the last 12 months did you or other adults in your household ever not eat for a whole day because there wasn’t enough money for food? (**Yes/No**)
14. In the last 12 months, were the children ever hungry but you just couldn’t afford more food? (**Yes/No**)
15. (If yes to Question 13) How often did this happen—**almost every month, some months but not every month**, or in only 1 or 2 months?
16. In the last 12 months, did any of the children ever skip a meal because there wasn’t enough money for food? (**Yes/No**)
17. (If yes to Question 16) How often did this happen—**almost every month, some months but not every month**, or in only 1 or 2 months?
18. In the last 12 months did any of the children ever not eat for a whole day because there wasn’t enough money for food? (**Yes/No**)

Note: Responses in bold indicate an affirmative response.

Appendix B:

Proof of Proposition 1: Subject to the restrictions in Equations (8) and (9), the upper bound is found by maximizing $(\theta_1^- + \theta_0^+)$ and minimizing $(\theta_0^- + \theta_1^+)$, and vice versa for the lower bound.

For $\Delta \geq 0$. For the upper bound, first consider the case that $\theta_1^{UB-} \geq \Delta$. Then $(\theta_0^- + \theta_1^+)$ is minimized at zero, and Equation (8a) simplifies to $\theta_1^- - \theta_0^+ = \Delta$. Given Equations (8a)-(8c), we know that θ_0^+ cannot exceed $\min\{\theta_0^{UB+}, \theta_1^{UB-} - \Delta\}$ and θ_1^- cannot exceed $\min\{\Delta + \theta_0^{UB+}, \theta_1^{UB-}\}$. From Equation (9), we know that $\theta_1^- + \theta_0^+$ cannot exceed Q_u . The upper bound follows directly. Second, consider the case that $\theta_1^{UB-} < \Delta$. From Equation (8b), we know that θ_1^- cannot exceed θ_1^{UB-} and, to satisfy the restriction in Equation (8a), θ_0^- must be no less than $\Delta - \theta_1^{UB-}$. As before, θ_1^+ is minimized at zero. From Equation (8c), we know that θ_0^+ can exceed zero but Equation (8a) implies that any conjectured increase in the false positive error rate must be offset by an equivalent increase in the false negative error rate. So, in this case, the upper bound would be unchanged by increasing θ_0^+ above zero. Thus, we have the upper bound of $2\theta_1^{UB-} - \Delta$ which can be shown to be no greater than $2\theta_0^{UB+} + \Delta$.

For the lower bound, first consider the case that $\theta_0^{UB-} \geq \Delta$. Then, $(\theta_1^- + \theta_0^+)$ is minimized at zero, and Equation (8a) simplifies to $\theta_0^- - \theta_1^+ = \Delta$. Given (8a)-(8c), we know that θ_1^+ cannot exceed $\min\{\theta_1^{UB+}, \theta_0^{UB-} - \Delta\}$ and θ_0^- cannot exceed $\min\{\theta_1^{UB+} + \Delta, \theta_0^{UB-}\}$. From Equation (9), we know that $\theta_0^- + \theta_1^+$ cannot exceed Q_u . It follows that $\max\{-Q_u, -2\theta_1^{UB+} - \Delta, -2\theta_0^{UB-} + \Delta\}$ provides a lower bound on Θ . Second, consider the case that $\theta_0^{UB-} < \Delta$. From Equation (8b), we know that θ_0^- cannot exceed θ_0^{UB-} and, to satisfy the restriction in Equation (8a), θ_1^- must be no less than $\Delta - \theta_0^{UB-}$. As before, θ_0^+ is minimized at zero. From Equation (8c), we know that θ_1^+ can exceed zero but Equation (8a) implies that any conjectured increase in the false positive error rate must be offset by an equivalent increase in the false negative error. So, in this case, the lower bound would be unchanged by increasing θ_1^+ above zero. Thus, we have the lower bound of $-2\theta_0^{UB-} + \Delta$ which can be shown to be no smaller than $-2\theta_1^{UB+} - \Delta$.

For $\Delta < 0$. For the upper bound, first consider the case that $\theta_0^{UB+} \geq -\Delta$. Then $(\theta_0^- + \theta_1^+)$ is minimized at zero, and Equation (8a) simplifies to $\theta_1^- - \theta_0^+ = \Delta$. Given Equations (8a)-(8c), we know that θ_1^- cannot exceed $\min\{\theta_1^{UB-}, \theta_0^{UB+} + \Delta\}$ and θ_0^+ cannot exceed $\min\{\theta_1^{UB-} - \Delta, \theta_0^{UB+}\}$. From Equation (9), we know that $\theta_1^- + \theta_0^+$ cannot exceed Q_u . The upper bound follows directly. Second, consider the case where $\theta_0^{UB+} < -\Delta$. From Equation (8b), we know that θ_0^+ cannot exceed θ_0^{UB+} and, to satisfy the restriction in Equation (8a), θ_1^+ must be no less than $-\Delta - \theta_0^{UB+}$. As before, θ_0^- is minimized at zero. From Equation

(8c), we know that θ_1^- can exceed zero but the restriction in Equation (8a) implies that any conjectured increase in the false negative error rate must be offset by an equivalent increase in the false positive error rate. So in this case, the upper bound would be unchanged by increasing θ_1^- above zero. Thus, we have the upper bound of $2\theta_0^{UB+} + \Delta$ which can be shown to be no greater than $2\theta_1^{UB-} - \Delta$.

For the lower bound, first consider the case that $\theta_1^{UB+} \geq -\Delta$. Then, $(\theta_1^- + \theta_0^+)$ is minimized at zero, and Equation (8a) simplifies to $\theta_0^- - \theta_1^+ = \Delta$. Given (8a)-(8c), we know that θ_0^- cannot exceed $\min\{\theta_0^{UB-}, \theta_1^{UB+} + \Delta\}$ and θ_1^+ cannot exceed $\min\{\theta_0^{UB-} - \Delta, \theta_1^{UB+}\}$. From Equation (9), we know that $(\theta_0^-) + (\theta_1^+)$ cannot exceed Q_u . It follows that $\max\{-Q_u, -2\theta_1^{UB+} - \Delta, -2\theta_0^{UB-} + \Delta\}$ provides a lower bound on Θ . Second, consider the case that $\theta_1^{UB+} < -\Delta$. From Equation (8b), we know that θ_1^+ cannot exceed θ_1^{UB+} and, to satisfy the restriction in Equation (8a), θ_0^+ must be no less than $-\Delta - \theta_1^{UB+}$. As before, θ_1^- is minimized at zero. From Equation (8c), we know that θ_0^- can exceed zero but Equation (8a) implies that any conjectured increase in the false negative error rate must be offset by an equivalent increase in the false positive error. So, in this case, the lower bound would be unchanged by increasing θ_0^- above zero. Thus, we have the lower bound of $-2\theta_1^{UB+} - \Delta$ which can be shown to be no smaller than $-2\theta_0^{UB-} + \Delta$. \square

Proof of Proposition 2: The objective is to maximize θ_1^- and minimize θ_1^+ subject to each conditional probability lying between 0 and 1 and the constraints in Equations (8) and (9).

Case (i): When $P^* < P(H = 1, FS = 1)$, the first ratio in Equation (10) exceeds 1 unless θ_1^+ is at least as large as $\theta_1^{+*} \equiv P(H = 1, FS = 1) - P^*$. At this value, the first ratio still exceeds 1 unless $\theta_1^{+*} = 0$. The upper bound for this case is maximized at θ_1^{+*} and θ_1^{+*} , with the difference in Equation (10) reducing to $\frac{P(H=0)}{P(FS^*=0)}$. All constraints in Equations (8) and (9) are satisfied. Case (ii): When $P^* \geq P(H = 1, FS = 1)$, set $\theta_1^{+*} = 0$ and recall that $\theta_1^- \leq \min\{Q_u, P(H = 1, FS = 0), P(FS^* = 1)\}$ by Equations (8b) and (9). Also, the first ratio in Equation (10) exceeds 1 unless $\theta_1^- \leq P(FS^* = 1) - P(H = 1, FS = 1)$, which is nonnegative by assumption in Case (ii). Thus, set $\theta_1^{+*} = \min\{Q_u, P(H = 1, FS = 0), P(FS^* = 1), PFS^*=1-PH=1,FS=1\} = \min\{PH=1,FS=0, PFS^*=1-PH=1,FS=1\}$. Again, all of the constraints in Equations (8) and (9) are satisfied. \square