

Do Food Stamps Contribute to Obesity in Low-Income Women?

Evidence from the National Longitudinal Survey of Youth 1979

Maoyong Fan*

Department of Economics

Ball State University

Abstract:

Does the Food Stamp Program (FSP), which provides in-kind transfers to low-income Americans, cause female participants to become obese? This question is particularly important because participants are substantially more likely to be obese than are nonparticipants. This paper estimates the effects of food stamp benefits on obesity, overweight and body mass index (BMI) of low-income women. Contrary to previous results, we find little evidence that the FSP causes obesity, overweight or higher BMI. Our analysis differs from previous research in three aspects. First, we exploit a rich longitudinal data set, the National Longitudinal Survey of Youth 1979, to distinguish between full-time and part-time participation. Second, instead of making parametric assumptions on outcomes, we employ a variety of difference-in-difference matching estimators to control for selection bias. Third, we estimate both short-term (one-year participation) and long-term (three-year participation) treatment effects. Empirical results show that after controlling for selection bias and defining the treatment and comparison groups carefully, there is little evidence that food stamps are responsible for higher BMI or obesity in female participants. Our estimates are robust to different definitions of the treatment and comparison groups, and to various matching algorithms.

Key Words: Food Stamp Program, Obesity, Body Mass Index, Propensity Score Matching

* Department of Economics, Ball State University, Muncie, IN 47304 (email: mfan@bsu.edu). I would like to thank Max Auffhammer, Peter Berck, Ryan Kellogg, Hugo Salgado, Enrico Moretti, and Brian Wright for insightful comments and suggestions. I am especially indebted to my advisor Jeff Perloff for his financial support, guidance and unlimited patience. I am also grateful to seminar participants at Berkeley ERE seminar, Texas A&M University, University of Akron, and Ball State University for their helpful comments.

1. Introduction

Obesity has become one of the most serious public health problems in the United States, causing more than 400,000 premature deaths and over \$100 billion in economic losses annually.¹ In particular, low-income women² have an even higher obesity rate (43%) than the rest of the population. The high prevalence of obesity in low-income women has raised the question of whether or not the Food Stamp Program (FSP),³ which provides food to low-income households, has contributed to this serious health problem. This paper uses a new definition of FSP participation and employs a difference-in-difference (DID) propensity score matching (PSM) technique to estimate the effects of receiving food stamp benefits (conditional on all other cash income) on obesity and body mass index (BMI)⁴ for female participants.

The FSP was designed to provide a nutritional safety net for low-income households. Currently, the FSP is the largest federal food assistance program, providing over \$37.5 billion transfers to about 31.6 million people in 2008. Both the income effect and the restriction to only food purchases with food stamps could increase food consumption and levels of energy intake (Fraker et al., 1995;Whitmore, 2002). The administrative practice of distributing food stamps once a month is linked to ups and downs of calorie intake (Shapiro, 2005;Wilde and Ranney, 2000), causing human body to reserve energy and increase BMI (Kendall et al., 1996;Townsend et al., 2001). Previous studies have found that the FSP is associated with obesity among low-income women (e.g. Chen et al., 2005;Gibson, 2003). However, the positive correlation between obesity and FSP participants could reflect that female

¹ Source: National Institutes of Health: National Institute of Diabetes, Digestive and Kidney Diseases. "Statistics Related to Overweight and Obesity: The Economic Costs." December 2004.

² Family income is lower than the federal poverty level.

³ As of Oct. 1, 2008, Supplemental Nutrition Assistance Program (SNAP) is the new name for the federal Food Stamp Program.

⁴ Body Mass Index is calculated as the ratio of the weight in kilograms to height in meters squared. It is a measurement of body fat for adults. According to the World Health Organization, a BMI of less than 18.5 is considered underweight and a BMI of greater than 25 (30) is considered overweight (obese).

participants were initially more obese. In addition, previous research has failed to distinguish between different types of participants. To identify the causal effects of the FSP on obesity, we need to measure spells of participation carefully and control for selection into the program.

We use a longitudinal data set, National Longitudinal Survey of Youth 1979 (NLSY79), to define a new FSP participation variable that is based on the interviewee's recalled history of monthly program participation. Instead of using the traditional binary dummy variable that indicates whether an individual participated in the past calendar year, we concentrate on individuals who were in the program for the entire year (or full-time participants). This is because all theoretical links between the FSP and obesity indicate that full-time participants are the most intensively treated people and their BMIs are more likely to be affected by the FSP. We separately examine part-time participants who were in the program for less than the entire year.

Potentially, we face a selection bias problem in evaluating the effects of the FSP on obesity. Both obesity and FSP participation may be affected by factors that are unobservable to the econometrician, such as a love for food, food insecurity, depression and so forth. If so, OLS estimates will be biased and inconsistent. In addition, the FSP participation variable may capture the difference in pre-treatment BMI if BMI is a determinant of participation.⁵ The NLSY79 shows that the obesity rate of FSP participants is higher before they enter the program than that of eligible nonparticipants. So failing to control for the initial BMI may result in upward biased estimates. The DID PSM method takes advantage of the panel structure of the NLSY79 to control for the pre-treatment BMI by examining the changes in obesity.

⁵ Some studies show that overweight and obese women earn less than similar women who have normal BMI. For example, Averett and Korenman (1996), Pagan and Davila (1997), and Cawley (2000).

This study focuses on female FSP participants for three reasons. First, females are the only group for which multiple previous studies show a positive correlation between FSP participation and obesity. Second, using doctor-measured BMI in multiple waves from the National Health and Nutrition Examination Survey, we do not find systematic differences in obesity between male FSP participants and income eligible nonparticipants for all age groups. Third, the size of the male treatment group consisting of full-time participants is too small to do matching analysis.

Three measures of weight are examined in this paper: obesity, overweight, and BMI. The empirical results show little evidence that full-time or part-time participation causes obesity, overweight, or higher BMI. Given the fact that one of the main purposes of the FSP is to eliminate hunger and increase nutrition intake, we also examine whether the FSP is effective in pushing underweight participants to normal weight. Our results show no evidence that underweight FSP participants are more likely to become normal weight than otherwise similar nonparticipants. The results of this paper suggest that there are other reasons other than the FSP that contribute to obesity in low-income women.

This paper has several advantages over previous studies. First, we distinguish between full-time and part-time participants so that the treatment is more clearly defined than in the past. Second, we estimate long-term participation to avoid the possible mean reversion problem in the evaluation of changes in BMI. Third, we take advantage of the panel structure of the NLSY79 to eliminate all time-invariant confounding factors. Fourth, by doing matching, we avoid the possible mistake of imposing specific functional forms of obesity and BMI. Finally, we use multiple matching methods to test the sensitivity of our results and reach the same conclusion.

The remainder of the paper proceeds as follows. Section 2 discusses the hypothetical links between the FSP and obesity and reviews past studies. Section 3 briefly describes the

data set. Section 4 presents the econometric method we use for our purpose. Section 5 presents the matching results and various robustness checks. Section 6 concludes.

2. The Hypothetical Links between the Food Stamp Program and Obesity, and Previous Studies

The FSP was designed to promote food consumption and nutrition intake of low-income Americans by providing in-kind benefits that can be redeemed in authorized stores.⁶ Over 67 percent of non-elderly adult participants are women. Figure 1 shows that female FSP participants have much higher obesity rates than eligible nonparticipants among working age women. Facing an obesity epidemic among low-income women,⁷ policy-makers and the public are questioning whether the FSP is too successful in boosting food consumption in participants. We first describe how the FSP participation is linked to weight gain. Then we review studies examining the effects of FSP participation on weight gain and obesity for adult female participants.

2.1. The Hypothetical Links between the FSP and Obesity

The theoretical relationship between FSP participation and obesity is not clear yet, largely due to missing information regarding energy expenditure. However, there are two hypothetical links between FSP participation and energy intake/food expenditure, which provide partial information to bridge the gap between FSP participation and BMI, and more severely, obesity. First, both the income effect of food stamps and the restriction that food stamps can only be used on food purchases may increase food consumption and levels of

⁶ Alcoholic drinks, tobacco products, vitamin and medicine, and prepared foods cannot be purchased with food stamps.

⁷ Compared with the overall population, the low-income females have much higher obesity rate (43% versus 34%). Source: Nutrition and Health Characteristics of Low-Income Populations: Volume I, Food Stamp Program Participants and Nonparticipants. <http://www.ers.usda.gov/publications/efan04014-1/efan04014-1.pdf>

energy intake. Second, the practice that food stamp benefits are provided once a month may be linked to binge eating which leads to weight gain over time. It is possible that none, one or both links play a role in increasing obesity among FSP participants. Examining both links is helpful in understanding the hypothesized connections between FSP participation and obesity.

The income effect of food stamps for inframarginal participants is theoretically intuitive. Given more resources, the participants will purchase more food (since it is a normal good) and more of other goods. The income effect is illustrated in figure 2(a): consumption allocation moves from A to B. Marginal FSP participants that have a high demand for other goods may be constrained by the restriction of food stamp benefits on food purchases and end up with consuming at the kink point as shown in figure 2(b).⁸ The restriction distorts participants' choices and forces them to spend more money on food and thus, consume more food than they otherwise would if they receive an equivalent cash transfer. Fraker et al. (1995) study several food stamp cash-out experiments that randomly assigned some FSP participants to receive equivalent cash benefits and find that every dollar of food stamps increases food expenditures more than the same amount of cash by 18 to 28 cents. This result does not necessarily lead to excessive calorie intake and obesity because food stamp benefits also allow people to switch to more expensive foods that were previously out of reach. However, studies show that FSP participants choose high calorie foods, not healthful foods, to meet their caloric needs. Whitmore (2002) studies Alabama and San Diego food stamp cash-out experiments and finds that food stamp recipients are more likely to consume

⁸ Whitmore (2002) estimates that between 20% and 30% food stamp recipients belong to the category of marginal consumers.

excessive calories (more than twice the Recommended Daily Allowance) than equivalent cash recipients.⁹

The second link between FSP participation and obesity is due to the administrative practice of distributing food stamps once a month. FSP participants might use food stamps as needed and overeat soon after receiving them and restrict their eating later in the month when food stamps are exhausted. Wilde and Ranney (2000) find that compared with the first week, FSP households that only shop once a month reduce their energy intake by 11.6% in the fourth week. Shapiro (2005) also finds that the average calorie intake of FSP participants declines by 10%-15% over the food stamp month. Hastings and Washington (2008) examine household level scanner data and find the same cycle of food purchase for food stamp recipients. They find that this cycle is not related with food variety and food quality. A person's body will adapt to the situation by lowering the metabolism and reserve calories as fat when eating is resumed (Manore et al., 1991). These chronic ups and downs in food energy intake can gradually lead to weight gain and obesity (Kendall, et al., 1996; Townsend, et al., 2001).

Both hypotheses explain why the FSP participation might cause weight gain and obesity. The treatment effects are positively correlated to spells of participation and the amount of food stamp benefits. If food stamps do cause weight gain or obesity, longer spells of participation and bigger benefits will have bigger effects on participants than would shorter spells of participation and smaller benefits. From the NLSY79, we find that households with a longer duration of participation receive a larger amount of benefits per person. Therefore, we focus on full-time participants and only examine part-time participant as a robustness check.

⁹ She finds that the constrained (marginal) participants spend more on non-dairy beverages (primarily soda and juice) than they would if they received equivalent cash benefits.

2.2. Past Studies

Ver Ploeg and Ralston (2008) summarize the literature and report that FSP participation is associated with obesity among female adult participants but not among male adult participants. Two studies used panel data and fixed effect model (Baum, 2007;Gibson, 2003)¹⁰. Both studies find that FSP participation increases the probability of being obese and BMI only for females. The validity of the fixed-effect model depends on an assumption that the unobserved individual characteristics, affecting both BMI and FSP participation, do not change through time. However, this assumption is very strong and does not hold. For example, FSP participants drop out the program very often, which may indicate that the selection bias may not be controlled for by fixed effects. Other studies use instrumental variables to deal with endogeneity of FSP participation (Baum, 2007; Chen, et al., 2005; Meyerhoefer and Pylypchuk, 2008). Baum (2007) uses households' characteristics that determine program eligibility and state eligibility criteria as instruments. Chen et al. (2005) use non-food expenditures,¹¹ residential area, and unemployment status as instruments. Meyerhoefer and Pylypchuk (2008) use state FSP outreach expenditures, the state electronic fingerprinting requirement, and state three-month recertification as instruments. Even though their basic findings are the same, Chen et al. (2005) use cross-section data and find a much larger effect of FSP participation on obesity and BMI than other two studies that use panel data.

In 1996, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) denied benefits for legal immigrants until they become U.S. citizens. States responded differently to the PWRORA and implemented different alternative state program

¹⁰ The results in Gibson (2003) are not technically correct because she creates the income variable by replace the missing values with zeros.

¹¹ Non-food expenditure is clearly endogenous because total income equals the sum of food expenditures and non-food expenditures.

to cushion the sudden shock for legal adult immigrants. Kaushal (2007) estimates a difference-in-difference model to determine the treatment effects of FSP participation on overweight and obesity, and concludes that the FSP is not a cause of obesity and overweight among adult legal immigrants. Although Kaushal's analysis is more convincing than other studies in the literature, the results for legal adult immigrants may not be generalized to native citizens.

The treatment group in the literature is not precisely defined. Each of these papers uses a binary variable for FSP participation. The variable equals 1 if the respondent answers "yes" to the survey question "Did you receive any food stamp benefits in the past calendar year?" Using a dummy for FSP participation is potentially problematic for two reasons. First, FSP participation intensity varies substantially in terms of spell and benefit level. For example, one would expect the effect of eleven-month participation and \$3,000 in benefits to be much larger than that of one-month participation and \$50 in benefits if there is truly a causal link between FSP participation and obesity. Second, short-term participation, e.g. one or two months, is hardly a convincing cause of the participants' weight gain, since substantial weight gain needs a longer period of time on average.

3. Data

The NLSY79, collected by Bureau of Labor Statistics, is used in this paper. The data set is a national representative sample of 12,686 young men and women who were 14-22 years old when the data collection began in 1979. The NLSY79 consists of three sub-samples. The cross-sectional sample of 6,111 respondents represents non-institutionalized youth. A supplemental sample of 5,295 respondents oversamples Hispanic, African-American and economically disadvantaged people. We do not use the third subsample, which consists of

1,280 military respondents. Individuals were interviewed annually from 1979 to 1994, and biennially from 1994 to the present.

The NLSY79 has several attractive features for our study. First, the longitudinal structure of the data allows us to track, over time, program participation status and changes in BMI of the same individuals. Second, the dataset includes detailed information about FSP participation so that we could distinguish between full-time participants and part-time participants. Third, the data set is a unique source of wealth and income information. Since the eligibility of FSP depends on both assets and incomes of households, we are able to construct the eligible nonparticipants group precisely. Fourth, the supplement sample increases the sample size of the treatment group and the comparison group. Accuracy of matching estimators increases as the sample size increases. Finally, the data set contains a rich set of demographic characteristics which are potentially correlated with both FSP participation and BMI.

Respondents' body weights were reported in pounds in every interview except 1983, 1984, 1987 and 1991 interviews. Respondents' heights were reported in inches in 1981, 1982 and 1985. In subsequent years, no further height information is reported. We use height reported in 1985 to calculate BMI because all respondents were adults (20 to 28 years old) at the time. We construct our sample using waves from 1985 to 1994 which are before the PRWORA made adjustments to eligibility criteria and benefit levels of the program. The longitudinal structure of the NLSY79 allows us to determine precisely how much respondents' BMIs changed over time and how long they participated in the FSP within certain periods. The information about wealth and income is used to identify eligible nonparticipants. To avoid comparing FSP participants with non-eligible people who have more resources and face no food insecurity problems, we restrict our comparison group to

eligible nonparticipants. Later, we use non-eligible people to form a comparison group as a robustness check.

4. Empirical Strategy

Since the FSP is a national program and varies little across states in terms of benefits and eligibility, the best way to estimate the treatment effects of receiving food stamp benefits on obesity is by comparing FSP participants with eligible nonparticipants. We first utilize the information on wealth, income and program participation in the data set to construct the treatment and comparison groups carefully. Then, we employ the DID PSM estimator, which does not rely on specific parametric assumptions on outcome, to control for selection into the FSP.

4.1. Defining the Treatment Group and the Comparison Group

We first examine program participation information of all FSP participants. The data set collects retrospective information about welfare program participation for all survey years. Respondents were asked to recall past FSP participation month during the interview. The FSP participants are defined as full-time (part-time) participants if they enroll in the program all the time (sometime). We find that the majority of FSP participants are actually part-time participants, a large proportion of whom enrolled in the program for a short period of time and receive a small amount of benefit. For example, full-time participants receive \$2,077 per household (\$639 per person) on average in survey year 1985-1986, as compared to \$990 per household (\$323 per person) for part-time participants. According to the discussion in section 2, the value of food stamp benefits and the spell of participation would be positively related to the treatment effects of the FSP on obesity and BMI. Therefore, full-time participants should more likely be affected by the FSP than part-time participants. The FSP is less likely

to be responsible for the obesity problem if we do not find a significant positive effect on its full-time participants. Pooling different types of participants makes it harder to interpret the estimates because of the presence of multiple treatments. Our treatment group is defined as full-time FSP participants. In other words, the treatment group is composed of those who are enrolled in the program all the time. Later, as a robustness check, we use part-time participants as a treated group and re-estimate the treatment effects.

Our comparison group consists of eligible nonparticipants. All eligible households have to meet income and asset tests.¹² Information on incomes and assets has been regularly collected in the NLSY79. We use both income and asset tests to determine the eligibility of nonparticipants. First, net household income has to be less than 100% of the federal poverty level. Second, households may have up to \$2000 in countable resources (e.g. checking/savings account, cash, stock/bonds). Table 1 presents the pre-intervention characteristics of the treatment group and the comparison group. There is one important distinction in terms of pre-treatment BMI: FSP participants have higher BMIs than eligible nonparticipants even before the treatment. For example, the average BMI of FSP participants is 27.16 which is 2.85 units higher than that of the comparison group for sample of one-year participation. The differences are quite large and statistically significant at 1% level. This indicates that the treatment effects might pick up the pre-treatment differences if econometric methods do not properly control for selection into the program.

4.2. Difference-in-Difference Propensity Score Matching

Let Y_1 measure the weight if an individual participates in FSP and Y_0 measure the weight if the individual does not participate in FSP. The treatment effect is the difference

¹² There are some exemptions and special cases. For more details, see Food and Nutrition Service website: http://www.fns.usda.gov/FSP/applicant_recipients/eligibility.htm.

between two outcomes: $Y_1 - Y_0$. However, this difference is not observable due to a missing data problem: being in the program (out of the program) reveals Y_1 (Y_0) but conceals the other potential outcome. In a randomized experiment, in which some applicants receive treatment, while similar applicants are denied treatment, a difference-in-difference estimator would give unbiased estimates of the treatment effect of the program on outcome. However, randomly denying applications for food stamps is neither legal nor ethical.

The basic idea of PSM is to impute counterfactual outcome for program participants using non-treated individuals with the same propensity score.¹³ Intuitively, comparing two individuals with the same propensity score, one is treated and the other is not, is like comparing those two individuals in a randomized experiment. Matching has several advantages over other nonexperimental evaluation techniques. First, matching does not impose functional form restriction in estimating the conditional expectation of the outcome variable, which is usually not justified neither by economic theory nor by the data used (Dehejia and Wahba, 1998; Smith and Todd, 2005). Second, matching could impose a common support requirement. The poor overlap in support between the treated and the comparison group raises questions about the robustness of parametric methods relying on functional form to extrapolate outside the common support (Caliendo and Kopeinig, 2008). Third, the dimensionality of the observable characteristics is high in our study. Propensity score matching is particularly useful under such circumstances because they provide a weighting mechanism that produces unbiased estimates (Dehejia and Wahba, 1998).

Theoretically, all PSM estimators should yield the same results asymptotically. As the sample size goes to infinity, different matching algorithms become closer to comparing only

¹³ See Abadie and Imbens (2002), Hahn (1998), Heckman, Ichimura and Todd (1997;1998b), Heckman et al. (1998b), and Hirano, Imbens and Ridder (2003) for detailed discussion of propensity score matching techniques.

exact matches (Smith, 2000). However, the choice of the matching estimator can be important for small samples (Heckman, et al., 1997). The cross-sectional matching estimators may not get rid of certain systematic differences between participant and nonparticipant outcomes even after conditional on observables. Such differences violate the identification conditions required for matching estimators. We take advantage of the panel structure of our data to implement the DID PSM estimator. The DID PSM estimator defined by Heckman, Ichimura and Todd (1997) and Heckman et al. (1998a) allows for temporally invariant differences in outcomes between participants and nonparticipants. This type of estimator matches on the basis of differences in outcomes and uses the same weighting functions as cross-sectional matching estimators. Let t and t' denote time periods before and after program participation. The DID PSM estimator requires:

$$E(Y_{0t'} - Y_{0t} | P, T = 1) = E(Y_{0t'} - Y_{0t} | P, T = 0), \quad (1)$$

where T is an indicator of program participation and $P = \Pr(T = 1 | X)$. Let i denote the individual. The sample average treatment effect for treated (SATT) is given by:

$$SATT_{DDM} = \frac{1}{N_1} \sum_{i|T_i=1} \left\{ (Y_{1it'} - \hat{Y}_{1it}) - \sum_{j|T_j=0} w_{ij} (Y_{0it'} - \hat{Y}_{0it}) \right\}, \quad (2)$$

where $N_1 = \sum_i T_i$ and the weights w_{ij} depend on matching algorithm employed.

We apply two widely used matching algorithms to our data set: nearest neighbor matching (NNM) and local linear matching (LLM). The NNM estimator is the most straightforward matching estimator. Every treated unit is compared with one or more units

from the comparison group that are most similar in terms of propensity score.¹⁴ It defines the set of matches with replacement¹⁵ as:

$$C_i^0(M) = \{l = 1, \dots, N \mid T_l = 0, |P_i - P_l| \leq d_i(M)\}, \quad (3)$$

where M indicates the number of matches (neighbors), P_i indicates the propensity score for individual i , and $d_i(M)$ is the distance from individual i to the M^{th} nearest match in the comparison group. We implicitly define $d_i(M)$ as:

$$\sum_{l:T_l=0} 1\{|P_i - P_l| < d_i(M)\} < M, \quad (4)$$

and

$$\sum_{l:T_l=0} 1\{|P_i - P_l| \leq d_i(M)\} \geq M, \quad (5)$$

where $1\{\cdot\}$ is the indicator function, which equals to 1 when the value in brackets is true, and zero otherwise. We implement this method using one and five nearest neighbors and with replacement. We impose the common support restriction and each match is weighted equally.

The LLM is a generalized version of kernel matching. It uses kernel-weighted average over multiple persons in the comparison group as the counterfactual outcome of the treated observation. Fan (1992) shows that local linear matching converges faster and that it is more robust to different densities of data than kernel matching. The weight of local linear matching is given by:

¹⁴ In nearest neighbor matching, the estimator could be biased if the distance between “best” matches are actually far away. The caliper matching estimator puts a tolerance level on maximum propensity score distance so that it can avoid bad matches. We find that NNM matching estimators are very similar to caliper matching estimators for our data set.

¹⁵ Replacement means that untreated units can be used more than once as the matches for the treated units.

$$w_{ij} = \frac{G_{ij} \sum_{l \in C_i^0} G_{il} (P_l - P_i)^2 - [G_{il} (P_l - P_i)] \left[\sum_{l \in C_i^0} G_{il} (P_l - P_i) \right]}{\sum_{j \in C_i^0} \left[G_{ij} \sum_{l \in C_i^0} G_{il} (P_l - P_i)^2 \right] - \left[\sum_{l \in C_i^0} G_{il} (P_l - P_i) \right]^2}, \quad (6)$$

where $G_{ij} = G\left(\frac{P_j - P_i}{h}\right)$ and h is the bandwidth. We use the Epanechnikov distribution as the kernel function.¹⁶

To test the FSP-obesity hypothesis, we need to estimate the standard errors for the estimators. However, calculating analytical standard errors can be cumbersome. Bootstrapping is often used to get standard errors for matching estimators (e.g. Black and Smith, 2004; Heckman, et al., 1997; Sianesi, 2004). Each bootstrap sample is a random sampling with replacement from the original data set. We draw 500 bootstrap samples and estimate 500 average treatment effects for the treated. The distribution of these means approximates the sampling distribution (and thus the standard error) of the population mean.

5. Results

5.1. The Propensity Score and the Balancing Test

In this paper, the set of baseline matching covariates includes pre-treatment BMI, age, ethnicity, income per person,¹⁷ education level, marital status, number of children, mother's education level, urban/rural residence and a geographical dummy of living in the west.¹⁸ We

¹⁶ We also experimented with different distributions such as tricube and normal kernel. The choice of kernel function has very little effect on the performance of LLM estimator.

¹⁷ The total net family income variable in the NLSY 79 includes all the income received by the household (e.g. salaries from work, Food Stamp benefits, AFDC payments). The income variable used in matching excludes the food stamp benefits from the total net family income because we are comparing pre-treatment resources between participants and eligible nonparticipants. The income per person variable is calculated by dividing the total net family income (excluding the food stamp benefits) by the number of people in the household. For a detailed discussion of the income variables in the NLSY 79, please refer to <http://www.nlsinfo.org/nlsy79/docs/79html/79text/income.htm>.

¹⁸ No mechanical algorithm or strict rule is available to choose a set of variables that satisfy the identification condition of matching estimators. However, there are three generally agreed rules of thumb about the choice of matching covariates. First, variables that have been affected by the treatment should not be used as matching covariates. One typical example is the middle outcome during the treatment period. Second, not all relevant variables should be matching covariates if the size of the treatment or comparison group is small. Covariates that are not correlated, or weakly correlated with outcomes, and

would expect that pre-treatment BMI, age and ethnic characteristics would be correlated with the unobserved biological differences affecting BMI. Educational level would proxy the awareness of nutrition and self-portrait. We would expect per capita income, marital status and number of children to reflect the unobserved information about household's resource and intra-household allocation. We exclude food stamp benefits from the income variable because we are interested in estimating the effects of extra in-kind resources on obesity and BMI. Geographical variables also enter the equation since they might be correlated with both BMI and decisions of participation.

The propensity score is estimated using a logistic regression of the treatment indicator on the set of covariates. Table 2 reports the logistic estimates for the FSP participation. The PSM uses the propensity score to balance the pre-treatment variables between the treatment group and the comparison group. The first thing to check is the overlap of the propensity score between the treatment group and the comparison group. Lack of overlap can result in imbalance of the pre-treatment variables. For our purpose of estimating the average treatment effects for the treated, imposing common support ensures that the combination of characteristics observed in the treatment group is also observed in the comparison group. The crucial conditional independence assumption of propensity score matching is not testable. Rosenbaum and Rubin (1983) state that additional information on X should not provide new information about treatment decision conditioning on the propensity score. The basic idea is to compare the characteristics before and after matching and check if significant differences still exist after matching. One practical way of testing the balance property of the

treatment indicator may exacerbate the common support problem and result in larger variances (Augurzky and Schmidt, 2001; Imbens, 2004). Third, after stratifying the sample into subgroups by propensity score, the differences between the treatment group and the comparison group in each subgroup should not be statistically significant. We also match on different sets of variables. Readers may find the results in the appendix.

propensity score is to compare the mean of each matching variable for the treated group with the mean of each variable in the comparison group.

Table 3 presents the mean differences between the treatment group and the comparison group before and after matching and the p-values for t-statistics. The results show a clear lack of balance for unmatched samples: 32 of 44 differences are statistically significant. The adjustment for propensity score improves the balance significantly. After adjustment for the propensity score, none of the differences between the treatment group and the comparison group are statistically significant. These low post-matching t-statistics of differences are not merely the result of the increased variance. For example, in 1985-1986 sample mean differences for BMI in 1985 between the treatment group and the comparison group is 1.81 (p-value=0). After adjusting for the propensity score, the difference is reduced to 0.27(p-value=0.65).

5.2. Matching Results

Our baseline treatment group consists of full-time FSP participants. And our baseline comparison group includes only eligible nonparticipants. The dependent variables are unconditional differences of outcomes before and after treatment. We estimate both short-term (one year) and long-term (three years) treatment effects of FSP participation on changes in obesity, overweight and BMI. We impose the common support condition¹⁹ on matching because matches off the support could result in substantial bias (Heckman, et al., 1997). The standard errors are based on bootstrap with 500 repetitions and 100% sampling.

¹⁹ Common support requires that any combination of characteristics seen among those in the treatment group may also be observed among those in the comparison group. For matching estimators, imposing common support could reduce the bias of estimators.

For NNM, we use 5 and 10 neighbors in the comparison group to match every treated individual. For LLR, we use a fixed bandwidth 0.1²⁰ and a plug-in rule of thumb bandwidth discussed in Fan and Gijbels (1996). We first estimate the treatment effects using the whole sample. Then we redo our analysis year by year.²¹ We show the results using the whole sample and 1985 as the pre-treatment base year. Results from other years are similar to those of 1985. Table 4 presents estimates of the average treatment effects for FSP participants. For the same outcome variable, the estimates are remarkably consistent across different matching algorithms. We first consider Panel A. Column 1 gives the number of observations used by matching algorithms from the treatment group and the comparison group. Very few observations are off support. Columns 2 to 5 show the estimated treatment effects on changes in obesity. Every estimate is a percent (not a percentage point) representing the difference between the probability of becoming obese for an average FSP participant and that for an average eligible nonparticipant. For example, if the initial obesity rate is 20 percent, an increase of 2 percentage points from 10 percent to 12 percent translates to 20 percent increase. Our estimates are neither statistically nor economically significant for both short-term and long-term treatments. Panel B presents estimates for changes in overweight status and BMI. The estimates for changes in overweight status are also percent. All estimates are either negative or close to zero. None of them is statically significant. The estimated treatment effects on BMI are either negative or too small to be economically significant. Again, all estimates are statistically insignificant. In summary, the matching results suggest that higher obesity of female FSP participants than eligible nonparticipants is largely due to the fact that participants initially had a bigger BMI. No evidence shows that the FSP is responsible for a higher obesity rate, overweight rate and BMI.

²⁰ We also tried many different bandwidth including 0.5, 0.4, 0.3, 0.2, 0.08, and 0.06. Compared to the other bandwidths, the 0.1 fits the data most smoothly.

²¹ Any year between 1985 and 1993 can be the base year and we estimate the treatment effects of FSP participation in the next year and in the next three years.

5.3. Robustness Check

In this section, we provide a series of robustness analysis for the matching results.²² First, we experiment with different specifications of the propensity score equation. These alternative specifications will show if the matching results are sensitive to non-linearity of the propensity score equation. Second, we are concerned about the measurement error in self-reported heights and weights in NLSY79. Following Lakdawalla and Philipson (2002), we address this issue by estimating self-reported error correction equations with data from the National Health and Nutrition Examination Survey III,²³ which has both self-reported and doctor-measured height and weight. Third, we are also concerned that some people in our comparison group were FSP participants before 1985. Including the previous participants in the comparison group could be problematic if the FSP alters the individuals' eating habits, which in turn affect BMI. Fourth, we are also concerned about the fact that most treated people were long-time FSP participants and their BMI might have been affected by the program even before the period that we study in this paper. Fifth, we are also concerned with the assumption PSM makes about treatment assignment. The covariate matching (CVM) estimator is not based on assumptions about either treatment assignment or the relationship between covariates and outcomes. Finally, we use part-time participants as the treatment group and investigate whether or not FSP contributes to obesity among female part-time participants. None of the robustness check results is statistically significant. The results show no evidence that FSP participation causes obesity, overweight and higher BMI.

6. Concluding Remarks

²² For the sake of brevity, results of robustness checks are not displayed. Readers may find them in the appendix.

²³ The National Health and Nutrition Examination Survey III was conducted during the period from 1988-1994, which is most closely related to NLSY79 in terms of timing.

The exceptional high obesity rate among low-income women has caught the attention of policy makers and the public. Since the purpose of FSP is to increase food consumption of low-income Americans, economists and nutrition scientists begin to hypothesize that the FSP contributes to the obesity problem among low-income women. However, estimating the treatment effects of the FSP on obesity and BMI is a difficult research question for the following reasons. First, there is no controlled random experiment since it is illegal to deny one's application for food stamps randomly. In addition, the FSP eligibility and benefit levels do not vary across regions. Researchers have to find ways to deal with selection bias related to FSP participation in the observational studies. Second, among participants, the duration of participation and the benefit level vary significantly across households. An accurate estimation requires a good measurement of participation. Third, BMI does not change overnight but over a certain period of time the length of which depends on personal characteristics. There are few panel data sets collecting individuals' BMI other than the NLSY79.

This paper uses a difference-in-difference propensity score matching technique and the NLSY79 to investigate whether participation in the FSP causes obesity, overweight and higher BMI for its female participants. The NLSY79 is the best available data set for this paper because: (1) The data set includes rich information of body weight over time. We exploit the panel structure to examine the changes in BMI and control for pre-treatment BMI. (2) The survey collects detailed information about FSP participation. The part-time participants could be distinguished from the full-time participants who are most likely to be affected by the program. (3) The data set has collected wealth and income information which can be used to determine eligibility of the FSP. (4) In addition to BMI and FSP participation, there are many important personal characteristics and environmental variables that are correlated with both FSP participation and BMI in the data set.

This paper distinguishes full-time participants from part-time participants and focuses on full-time participants rather than treating all FSP participants homogenously. We separately examine those who were in the program for some period of time but not the entire time. Difference-in-difference matching techniques are employed to control for selection into the program and pre-treatment differences in BMI. This econometric technique can also avoid the possible mistakes of imposing specific functional form on outcome. Three measures of weight are examined in this paper: obesity, overweight, and BMI. The estimates are neither statistically nor economically significant. The results hold for various robustness checks and for both short-term and long-term participation. Our analysis shows that the FSP is not responsible for obesity, overweight or higher BMI among its female participants. One important question left unaddressed in this paper is how to examine food allocation within a household without treating it as a “black box”. Unfortunately, intra-household food consumption information is not available in the NLSY79. As data become available, hopefully our future research will examine caloric distribution among household members.

References

- Abadie, A and Imbens, Guido W.** Simple and Bias-Corrected Matching Estimators for Average Treatment Effects, NBER Working Paper no. 283, 2002.
- Augurzky, B. and Schmidt, C. M.** The Propensity Score: A Means to an End, IZA, Discussion Paper No. 271 2001.
- Averett, S and Korenman, S.** "The Economic Reality of the Beauty Myth." *The Journal of Human Resources*, 1996, 31(2), pp. 304-30.
- Baum, Charles.** "The Effects of Food Stamps on Obesity," *Contractor and Cooperator Report No. 34.* Economic Research Service, USDA, 2007.
- Black, D. A. and Smith, J. A.** "How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*, 2004, 121(1-2), pp. 99-124.
- Caliendo, M. and Kopeinig, S.** "Some Practical Guidance for the Implementation of Propensity Score Matchings." 2008, 22(1), pp. 31-72.
- Cawley, John.** "Body Weight and Women's Labor Market Outcomes," NBER working paper no. 7841, 2000.
- Chen, Zhuo; Yen, Steven T. and Eastwood, David B.** "Effects of Food Stamp Participation on Body Weight and Obesity." *American Journal of Agricultural Economics*, 2005, 87(5), pp. 1167-73.
- Dehejia, R. H. and Wahba, S.** "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," NBER working paper no. 6586, 1998.
- Fan, J.** "Design-Adaptive Nonparametric Regression." *Journal of the American Statistical Association*, 1992, 87(420), pp. 998-1004.
- Fan, J. and Gijbels, I.** *Local Polynomial Modelling and Its Applications.* Chapman & Hall, 1996.
- Fraker, TM; Martini, AP and Ohls, JC.** "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations." *The Journal of Human Resources*, 1995, 30(4), pp. 633-49.
- Gibson, Diane.** "Food Stamp Program Participation Is Positively Related to Obesity in Low Income Women." *Journal of Nutrition*, 2003, 133(7), pp. 2225-31.
- Hahn, J.** "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica*, 1998, 66(2), pp. 315-31.
- Hastings, J. S. and Washington, E. L.** "The First of the Month Effect: Consumer Behavior and Store Responses," NBER Working Paper No. 14578 2008.
- Heckman, J.; Ichimura, H. and Todd, P.** "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*, 1997, 64(4), pp. 605-54.
- _____. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies*, 1998a, 65(2), pp. 261-94.
- Heckman, James; Ichimura, Hidehiko; Smith, Jeffrey and Todd, Petra.** "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 1998b, 66(5), pp. 1017-98.
- Hirano, K.; Imbens, G. W. and Ridder, G.** "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 2003, 71(4), pp. 1161-89.
- Imbens, G. W.** "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics*, 2004, 86(1), pp. 4-29.
- Kaushal, N.** "Do Food Stamps Cause Obesity? Evidence from Immigrant Experience." *Journal of Health Economics*, 2007, 26(5), pp. 968-91.

Kendall, Anne; Olson, Christine M. and Frongillo, Jr., Edward A. "Relationship of Hunger and Food Insecurity to Food Availability and Consumption." *Journal of the American Dietetic Association*, 1996, 96(10), pp. 1019-24.

Lakdawalla, D and Philipson, T. The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination, NBER working paper no. 8946
2002.

Manore, M. M.; Berry, T. E.; Skinner, J. S. and Carroll, S. S. "Energy Expenditure at Rest and During Exercise in Nonobese Female Cyclical Dieters and in Nondietsing Control Subjects." *The American journal of clinical nutrition*, 1991, 54(1), pp. 41-6.

Meyerhoefer, Chad D. and Pylypchuk, Yuriy "Does Participation in the Food Stamp Program Increase the Prevalence of Obesity and Health Care Spending?," 2008, 287-305.

Pagan, J. A. and Davila, A. "Obesity, Occupational Attainment, and Earnings: Consequences of Obesity." *Social science quarterly*, 1997, 78(3), pp. 756-70.

Rosenbaum, P. R. and Rubin, D. B. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 1983, 70(1), pp. 41.

Shapiro, Jesse M. "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle." *Journal of Public Economics*, 2005, 89(2-3), pp. 303-25.

Sianesi, B. "An Evaluation of the Active Labour Market Programmes in Sweden." *Review of Economics and Statistics*, 2004, 86(1), pp. 133-55.

Smith, J and Todd, P. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 2005, 125(1-2), pp. 305-53.

Smith, J. "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies." *Swiss Journal of Economics and Statistics*, 2000, 136(3), pp. 1-22.

Townsend, Marilyn S.; Peerson, Janet; Love, Bradley; Achterberg, Cheryl and Murphy, Suzanne P. "Food Insecurity Is Positively Related to Overweight in Women." *J. Nutr.*, 2001, 131(6), pp. 1738-45.

Ver Ploeg, M. and Ralston, K. "Food Stamps and Obesity: What Do We Know?," *Economic Information Bulletin No. 34*. Economic Research Service, USDA, 2008.

Whitmore, D. "What Are Food Stamps Worth?," Industrial Relations Section, Princeton University, 2002.

Wilde, Parke E. and Ranney, Christine K. "The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework." *American Journal of Agricultural Economics*, 2000, 82(1), pp. 200.

Table 1: Summary Statistics

Variable name	One-Year Participation ^a		Three-Year Participation ^b		1985-1986		1985-1988	
	Treatment Comparison		Treatment Comparison		Treatment Comparison		Treatment Comparison	
	group	group	group	group	group	group	group	group
BMI(t) ^c	27.16	24.31	26.71	24.10	25.09	23.29	25.69	23.27
	[6.91]	[5.49]	[6.72]	[5.51]	[6.38]	[4.90]	[6.49]	[4.93]
Age	28.56	27.05	26.84	25.35	24.06	22.99	24.06	23.10
	[3.32]	[3.78]	[2.77]	[3.23]	[2.02]	[2.34]	[1.93]	[2.43]
African-American(=1)	0.56	0.35	0.62	0.35	0.47	0.36	0.56	0.35
	[0.50]	[0.48]	[0.49]	[0.48]	[0.50]	[0.48]	[0.50]	[0.48]
Hispanic(=1)	0.2	0.18	0.17	0.16	0.19	0.18	0.17	0.15
	[0.40]	[0.38]	[0.37]	[0.36]	[0.39]	[0.38]	[0.38]	[0.36]
Income per Person(Log)	7.14	6.43	6.99	6.53	7.19	6.75	7.06	6.84
	[1.09]	[2.47]	[1.42]	[2.43]	[1.03]	[2.05]	[1.04]	[2.02]
Married(=1)	0.18	0.26	0.15	0.23	0.15	0.2	0.15	0.18
	[0.38]	[0.44]	[0.36]	[0.42]	[0.36]	[0.40]	[0.35]	[0.38]
Number of Children	2.43	1.23	2.4	0.94	2.02	0.8	2.15	0.71
	[1.35]	[1.25]	[1.32]	[1.16]	[1.19]	[1.08]	[1.25]	[1.04]
Education ^d	1.16	1.51	1.09	1.66	1.13	1.78	1.10	1.98
	[0.47]	[0.94]	[0.35]	[1.09]	[0.46]	[1.21]	[0.37]	[1.30]
Mother's Education ^d	1.33	1.57	1.30	1.68	1.32	1.62	1.21	1.69
	[0.58]	[0.80]	[0.56]	[0.91]	[0.57]	[0.87]	[0.45]	[0.91]
Urban(=1)	0.76	0.76	0.75	0.77	0.77	0.76	0.79	0.76
	[0.42]	[0.43]	[0.43]	[0.42]	[0.42]	[0.43]	[0.41]	[0.43]
West(=1)	0.17	0.2	0.14	0.21	0.14	0.18	0.13	0.18
	[0.37]	[0.40]	[0.34]	[0.41]	[0.35]	[0.38]	[0.34]	[0.38]
Observations	1444	1260	738	803	215	350	124	251

Notes:

Standard deviations are in brackets. The treatment group includes females who enrolled in the FSP all the time. The comparison group consists of all eligible nonparticipants.

^a. This is a stacked sample including subsample of 1985-1986, 1988-1989, 1989-1990, 1992-1993 and 1993-1994.

^b. This is a stacked sample including subsample of 1985-1988, 1986-1989, 1989-1992, and 1990-1993.

^c. Body Mass Index before the treatment.

^d. Education has four levels: (1) less than high school; (2) high school graduate or equivalent, (3) some college or two year college degree, (4) college graduate or more.

Table 2: Estimates of Propensity Score Equations

Variable name	One-Year Participation	Three-Year Participation	1985-1986	1985-1988
BMI(t)	0.0495** [6.30]	0.0410** [3.60]	0.0438* [2.29]	0.0544* [2.10]
Age	0.0543** [4.00]	0.0331 [1.41]	0.0977 [1.90]	-0.0045 [-0.063]
African-American(=1)	0.7598** [6.88]	0.9403** [6.01]	-0.0411 [-0.16]	0.5349 [1.61]
Hispanic(=1)	0.4374** [3.04]	0.5844** [2.75]	0.0482 [0.15]	0.1685 [0.36]
Income per Person(Log)	0.2681** [8.45]	0.1792** [4.84]	0.3549** [3.82]	0.3089* [2.47]
Married(=1)	-1.0042** [-8.40]	-1.3436** [-7.38]	-1.6416** [-5.18]	-1.8942** [-4.30]
Number of Children	0.7134** [16.9]	0.9593** [14.2]	0.9544** [8.25]	1.1470** [6.86]
Education	-0.5389** [-6.65]	-1.0126** [-7.17]	-0.6584** [-3.73]	-0.8680** [-3.09]
Mother's Education	-0.1355 [-1.75]	-0.0836 [-0.77]	-0.1001 [-0.54]	-0.4123 [-1.42]
Urban(=1)	-0.0241 [-0.21]	0.1135 [0.69]	-0.1485 [-0.55]	0.3913 [1.06]
West(=1)	-0.0962 [-0.74]	-0.3949* [-2.05]	-0.0548 [-0.17]	-0.1918 [-0.43]
Constant	-5.0684** [-10.8]	-3.6852** [-5.34]	-6.1826** [-4.45]	-4.1298* [-2.21]
Observations	2704	1541	565	375

Notes:

Dependent variable is the treatment status. 1 indicates being treated, 0 otherwise.

z-statistics are in brackets. * and ** indicate statistical significance at 5% and 1% level.

The treatment group includes females who enrolled in the FSP all the time. The comparison group consists of all eligible nonparticipants.

Table 3: Balancing Test

		One-Year Participation		Three-Year Participation		1985-1986		1985-1988	
		Difference ^a p> t		Difference p> t		Difference p> t		Difference p> t	
BMI(t)	Unmatched	2.86	0	2.6	0	1.81	0	2.41	0
	Matched	0.3	0.25	0.57	0.16	0.27	0.65	0.84	0.28
Age	Unmatched	1.52	0	1.51	0	1.07	0	0.96	0
	Matched	-0.13	0.33	0.13	0.46	-0.02	0.93	-0.30	0.26
African-American(=1)	Unmatched	0.21	0	0.27	0	0.11	0.01	0.20	0
	Matched	0.02	0.24	0.01	0.86	-0.04	0.37	-0.07	0.29
Hispanic(=1)	Unmatched	0.02	0.11	0	0.86	0.01	0.69	0.02	0.58
	Matched	0.02	0.23	0.01	0.74	0.04	0.28	0.01	0.83
Income per Person(Log)	Unmatched	0.71	0	0.29	0.02	0.44	0	0.22	0.26
	Matched	-0.07	0.08	0	0.97	0.01	0.93	0.05	0.80
Married(=1)	Unmatched	-0.09	0	-0.09	0	-0.05	0.15	-0.03	0.41
	Matched	-0.03	0.07	-0.03	0.22	-0.06	0.13	0.01	0.86
Number of Children	Unmatched	1.19	0	1.46	0	1.22	0	1.44	0
	Matched	-0.08	0.13	-0.13	0.08	-0.11	0.34	-0.04	0.81
Education	Unmatched	-0.34	0	-0.58	0	-0.65	0	-0.88	0
	Matched	-0.01	0.69	-0.01	0.58	0.01	0.89	-0.02	0.69
Mother's Education	Unmatched	-0.25	0	-0.37	0	-0.31	0	-0.48	0
	Matched	0	0.98	-0.06	0.09	0	0.95	-0.06	0.35
Urban(=1)	Unmatched	0	0.83	-0.02	0.5	0	0.90	0.03	0.47
	Matched	-0.01	0.75	0.01	0.82	0.03	0.45	-0.06	0.25
West(=1)	Unmatched	-0.03	0.02	-0.07	0	-0.04	0.27	-0.05	0.22
	Matched	0.01	0.33	-0.04	0.11	0	0.93	0.02	0.55

Notes:

^a. Mean differences of variables equal means of the treatment group minus those of the comparison group. The differences that are statistically significant are in bold.

All tests are based on Local Linear Matching with bandwidth=0.1. Results from other specifications are similar.

T-statistics are calculated as $t_{(\bar{X}_{Treated} - \bar{X}_{Control})} = \frac{\bar{X}_{Treated} - \bar{X}_{Comparison}}{\sqrt{\frac{\sigma_{Treated}^2}{n_1} + \frac{\sigma_{Comparison}^2}{n_2}}}$, where n_1 and n_2 are the numbers of

observations for the treatment and comparison group on the support.

Table 4: Estimates of Treatment Effects for Full-time FSP Participants

Panel A: Outcome = Obesity

	# of Treated/Comparison on the Support	Outcome = Obesity			
		Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
	(1)	(2)	(3)	(4)	(5)
One-Year Participation	1428/1260	-0.0151 [0.0112]	-0.0109 [0.0101]	-0.0115 [0.0114]	-0.013 [0.0109]
Three-Year Participation	695/803	0.0256 [0.0214]	0.0269 [0.0328]	0.0222 [0.0263]	0.0318 [0.0231]
1985-1986	204/350	-0.0514* [0.0259]	-0.0441 [0.0267]	-0.0461 [0.0296]	-0.051 [0.0293]
1985-1988	119/251	-0.0147 [0.0479]	0.0278 [0.0456]	0.0104 [0.0491]	0.0209 [0.0492]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Specification of propensity score function: $Prob(T=1)=F(\text{pre-treatment BMI, age, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west})$.

Panel B: Outcome = Overweight/BMI

	Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
One-Year Participation	0.0029 [0.0180]	-0.0013 [0.0201]	-0.0015 [0.0240]	-0.0032 [0.0216]	0.1797 [0.1318]	0.1705 [0.1506]	0.165 [0.1558]	0.2558 [0.1543]
Three-Year Participation	0.0157 [0.0312]	0.0098 [0.0321]	0.0069 [0.0362]	-0.0096 [0.0363]	0.0123 [0.2894]	0.0357 [0.3017]	0.0674 [0.3030]	-0.0026 [0.2764]
1985-1986	-0.0447 [0.0379]	-0.0382 [0.0441]	-0.0284 [0.0519]	-0.0181 [0.0485]	0.0977 [0.3048]	0.0967 [0.3182]	0.1329 [0.3522]	0.1297 [0.3361]
1985-1988	-0.0255 [0.0793]	-0.0239 [0.0874]	-0.0609 [0.0851]	-0.0626 [0.0807]	-0.2978 [0.8198]	-0.5164 [0.6606]	-0.5301 [0.6514]	-0.7975 [0.6171]

Notes:

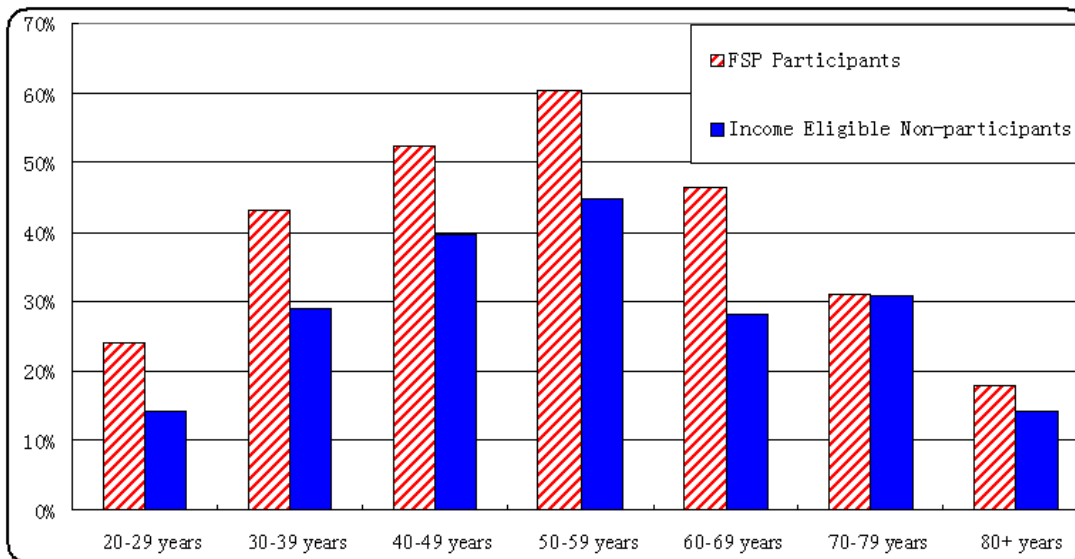
Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Specification of propensity score function: $Prob(T=1)=F(\text{pre-treatment BMI, age, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west})$.

Figure 1: Obesity/Overweight Rates (Percentage) of Adult Female of FSP participants and Income Eligible Non-participants

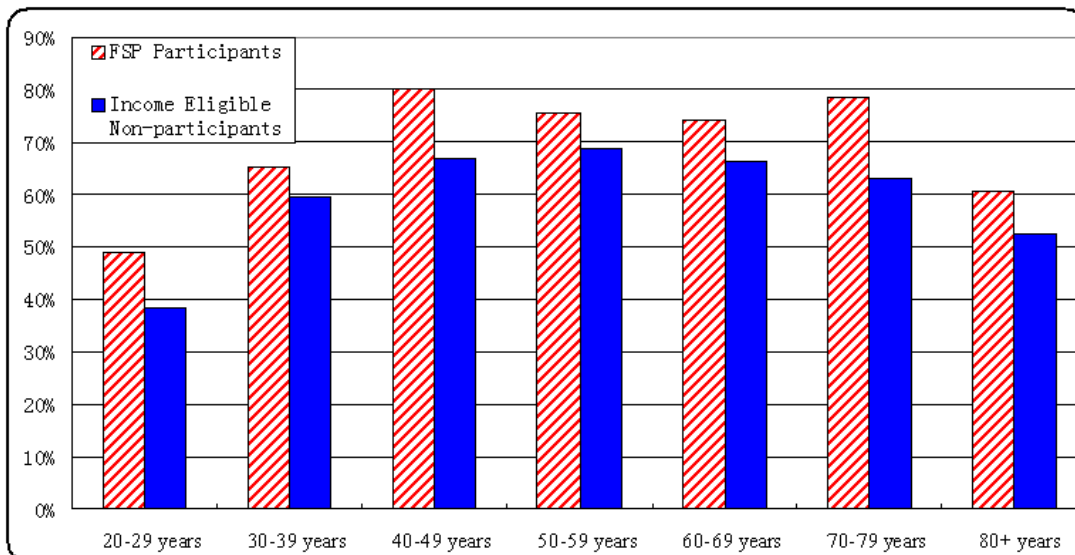
A: Obesity Rates



X-axis: Age Groups

Y-axis: Obesity Rates

B: Overweight Rates



X-axis: Age Groups

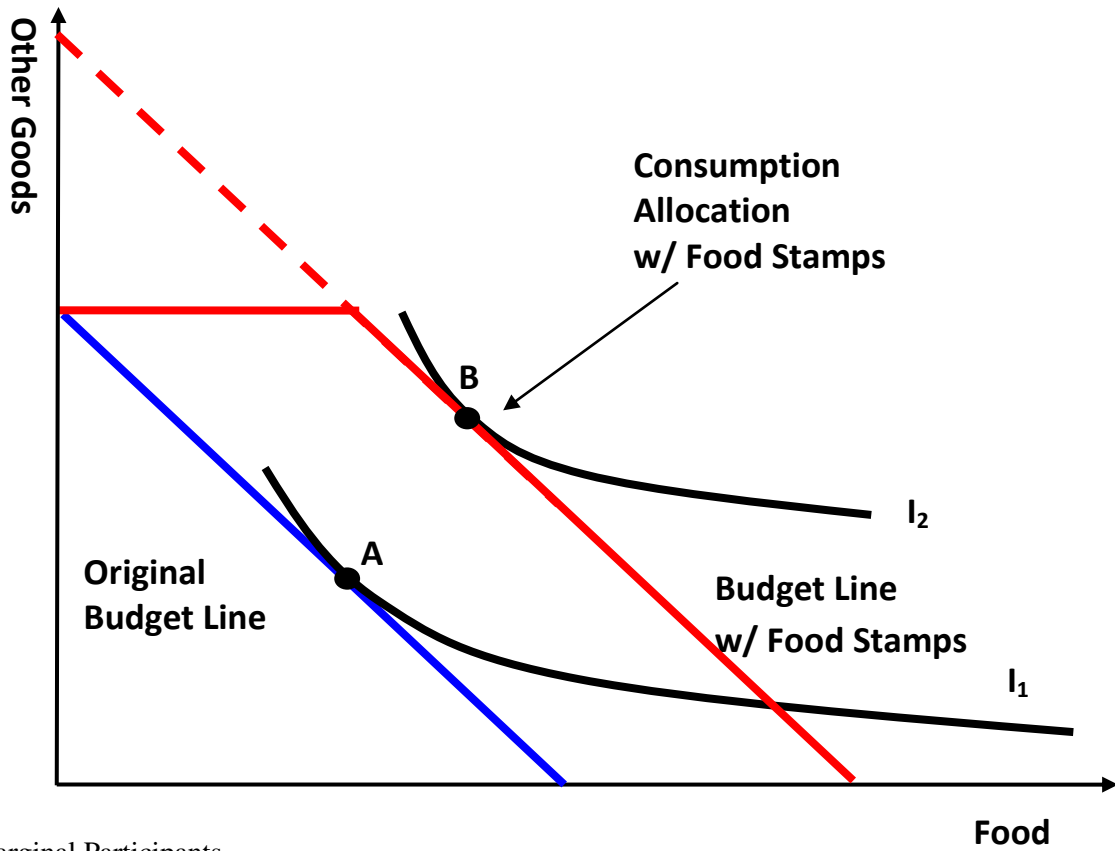
Y-axis: Overweight Rates

Note: BMI is calculated using doctor measured weights and heights.

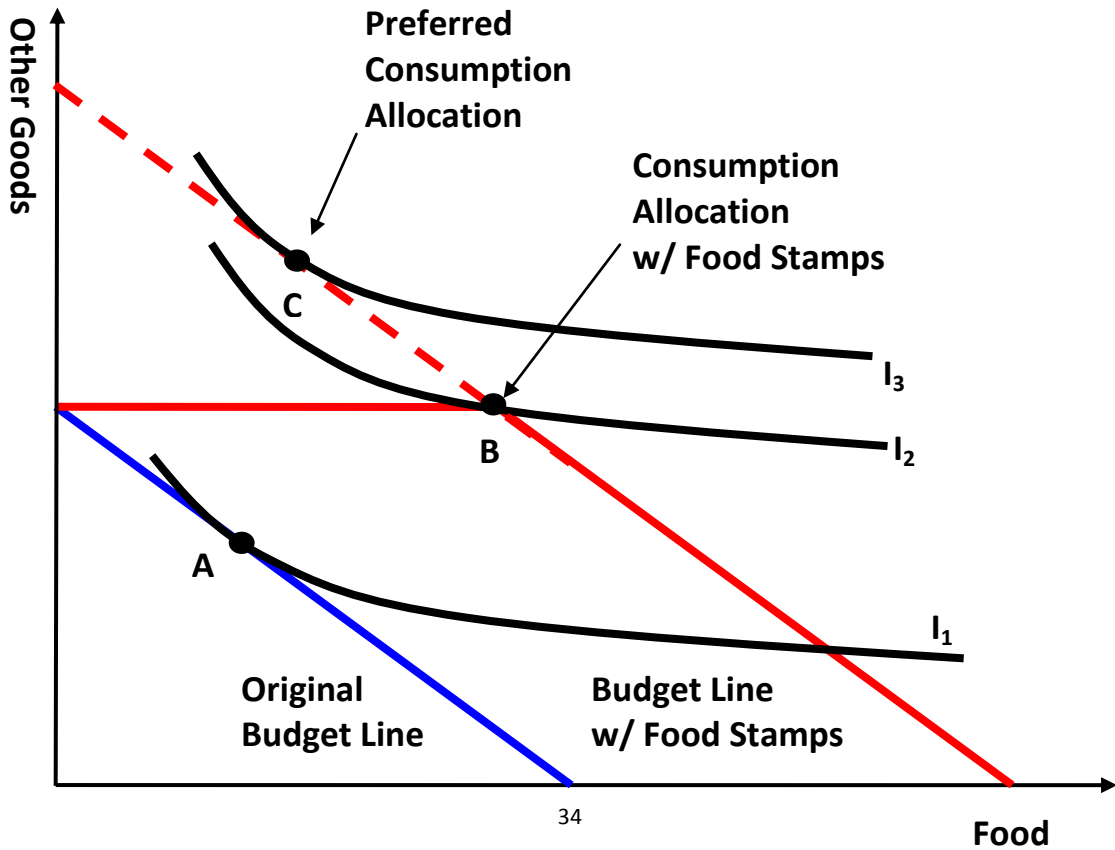
Data source: National Health and Nutrition Examination Survey III 88-94

Figure 2: The Effect of Food Stamp Program on Food Consumption

(a) Inframarginal Participants



(b) Marginal Participants



Appendix A: Background of the Food Stamp Program

The Food Stamp Program (FSP) was designed to promote food consumption and nutrition intake of low income Americans by providing in-kind benefits that can be redeemed in authorized stores.¹ It is the only food program that does not have demographic restrictions other than financial requirements.² The intent of the FSP was to kill two birds with one stone,³ and was first implemented in late 1930s. The food stamps were not distributed freely to eligible people; rather, participants had to buy orange stamps in order to obtain the free blue stamps. For each dollar of orange stamp purchased, they received a 50-cent blue stamp. Blue stamps could only be used to buy the food determined by the USDA to be surplus while the orange stamp could be used to buy any food. An important purpose of the program was to promote the consumption of American agricultural products surplus. The program ended in 1943, when the unmarketable agricultural surplus and large scale unemployment no longer existed.

In 1961, the FSP was revived as a pilot program. The pilot program retained the requirement that the food stamps be purchased, but eliminated the concept of special stamps for surplus foods. A USDA spokesman indicated that the emphasis would be on increasing the consumption of perishables. The program was made permanent by Food Stamp Act under Johnson administration in 1964. The major provisions stated that “recipients purchase their food stamps, paying an amount commensurate with their normal expenditures for food and receiving an amount of food stamps representing an opportunity more nearly to obtain a low-cost nutritionally adequate diet”. In accordance with P.L. 93-86, the program began operating nationwide on July 1, 1974.

Currently, the FSP is the largest federal food intervention program that is based only on financial need.⁴ The program operates in partnership with the states. The federal government sets national program standards for eligibility and benefits, funds all benefit costs and approximately half of state administrative expense, and monitors program implementation. The program uses a very sophisticated system to monitor eligibility of recipients and give the proper amount of benefits. Food stamps can be certified for up to two-year period. However, recipients are required to report changes in household size and income every month.⁵ Food stamp benefits to an eligible household depends on household’s size, its net monthly income, and

¹ Alcoholic drinks or prepared food could not be purchased with food stamps.

² Other food assistance and nutrition programs can only be used by certain demographic groups. For example, the WIC program is restricted to women with infants or children. AFDC is for poor families with children.

³ The program's first administrator, Milo Perkins said, "We got a picture of a gorge, with farm surpluses on one cliff and under-nourished city folks with outstretched hands on the other. We set out to find a practical way to build a bridge across that chasm."

⁴ The FSP is unlike all other food assistance and nutrition programs which are only applicable to only certain demographic group. For example, WIC program is for woman, infant and children. AFDC is for poor families with children.

⁵ The basic food stamp beneficiary unit is the “household”. Generally, individuals living together constitute a single food stamp household if they customarily purchase food and prepare meals in common.

maximum monthly benefit levels. The amount of benefits is negatively related with the net family income.

⁶ Families with lower income and larger family size receive higher benefits. The financial requirements include a gross income limit of 130% of the poverty level,⁷ a net income limit of poverty level, and a countable asset limit of \$2,000. The maximum monthly food stamp allotments are adjusted by the United States Department of Agriculture (USDA) annually on October 1st to account for changes in the cost of Thrifty Food Plan (TFP). The thresholds and maximum allotments for fiscal year 2009-2010 are in table 1.

The program caseload has fluctuated in the past several decades. The caseload fluctuations are largely affected by national economic situation and changes of relevant policies. The caseload rose steadily in 1970s and 1980s and reached its peak of 29 million in 1994. The caseload decreased in the second half of 1990s and increased again after 2000. The FSP served more than 31.6 million people in 2008.⁸ The most recent policy affecting the program caseload is the Personal Responsibility and Work Opportunities Reconciliation Act of 1996 (PRWORA). The Act places a time limit on food stamp receipt of three out of 36 months for able-bodied adults without dependents who are not working at least 20 hours a week or participating in a work program.

The FSP is not mutually exclusive with other federal welfare programs. The food stamp recipients can also receive public assistance income, such as Aid to Families with Dependent Children (AFDC) or Temporary Aid to Needy Families (TANF), and Supplementary Security Income (SSI). During the 1980s and 1990s, around 40% FSP participants received AFDC income and around 20% of them received SSI income.

⁶ See Fact Sheet on Resources, Income, and Benefits for detail:

http://www.fns.usda.gov/fsp/applicant_recipients/fs_Res_Ben_Elig.htm

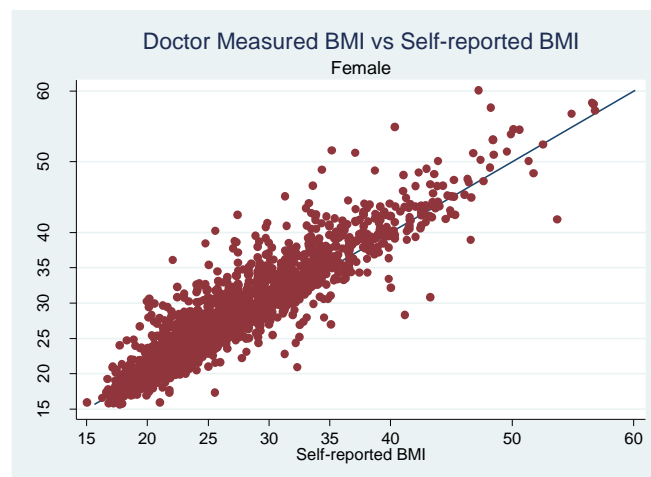
⁷ Poverty level is different for different households since the household sizes are different.

⁸ See details on <http://www.fns.usda.gov/pd/annual.htm>, Food and Nutrition Service, USDA.

Appendix B: Correction of Self-Reported Height and Weight

One of the potential problems, which could bias my estimates, is measurement error in self-reported height and weight. Although height and weight were reported with small errors on the average, self-reported height and weight suffer from significant errors for important subgroups i.e. underweight people, obese people.

National Health and Nutrition Examination Survey III (NHANES III) is conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). The sample of this survey is national representative. Self-reported height and weight are recorded to the nearest pound and inch in the home survey. The respondents were asked the following questions: “How tall are you without shoes – inches” and “How much do you weigh w/out clothes -lbs”. Medical examination data include doctor measured height and weight information. It is so far the best source to study the measurement error of self-reported height and weight for NLSY 79 cohort.



In order to determine how the self-reported BMI is correlated with the true outcome, we first present the scatter plot of self-reported BMI and doctor-measured BMI using the sample including only people between 20 and 37.⁹ The plot suggests that doctor measured BMI is a concave function of self-reported BMI.

Reporting errors are supposed to be correlated with some other demographic characteristics except gender such as age, ethnicity, race, education and if living in urban area. The final correction equation for female is:

$$\begin{aligned} \text{True_BMI} = & -2.6697 + 1.2699\text{Self_BMI} - 0.0036\text{Self_BMI}^2 - 0.0348\text{Age} \\ & (0.8391) \quad (0.0610) \quad (0.0010) \quad (0.0103) \end{aligned}$$

We will use these correction equations and variables of NLSY79 to predict the true BMI for the NLSY79 cohort.

⁹ I am using 1985-1994 data for my analysis. The NLSY79 cohort was 20-28 when their heights were reported in 1985. They were 29-37 in 1994. Pregnant women are dropped from the sample.

Appendix C: Robustness Checks

We provide a series of robustness analysis for the matching results. Five confounding factors are discussed and examined carefully. The first experiments with the propensity score equation. The second focuses on measurement error of self-reported BMI. The third uses a different definition of the comparison group. The fourth analysis refines the treatment group by dropping those who had full-time participation experience before our starting point (1985). Finally, we estimate the treatment effects using covariate matching instead of propensity score matching.

C.1. Different Specifications of the Propensity Score Equation

Estimating the propensity score is the crucial step in implementing matching. The matching estimator is sensitive to choices of the propensity score specification (e.g. Heckman et al., 1997; Smith and Todd, 2005). Our results would be less likely to be statistically convincing if they are sensitive to the propensity score specification. Thus, two alternative specifications are used for sensitivity analysis. The first specification adds the interactions between ethnicity and pre-treatment BMI to the original propensity score equation. The second specification adds the second-order terms of non-binary explanatory variables to the original propensity score equation. These alternative specifications will show if the matching results are sensitive to non-linearity of the propensity score equation.

Table 2 presents the estimates with different propensity score specifications. None of the 72 estimates is statistically or economically significant. Consistent with

the baseline specification, both alternative specifications show little evidence that FSP participation has positively affected any of the three outcomes for both short-term and long-term treatment.

C.2. Correct for Self-reported BMI

The third concern is the measurement error in self-reported heights and weights in NLSY79. Self-reported measurement errors are common in survey data. Researchers have found that self reported weights are directly related to the respondent's overweight status and heights are generally over-reported (Rowland, 1990; Stunkard and Albaum, 1981). They also found that bias and unreliability increase positively with the magnitude of overweight and age, with overweight females more likely to underreport weight than overweight males. Following Lakdawalla and Philipson (2002), we address this issue by estimating self-reported error correction equations with data from NHANES III¹⁰, which has both self-reported and doctor-measured height and weight. Table 3 presents the results using the predicted outcomes rather than self-reported outcomes. None is statistically or economically significant. The results show that FSP participation has no significant positive effects on obesity/overweight/BMI even we adjust the possible underestimation of obesity/overweight/BMI.

C.3. An Alternative Comparison Group

¹⁰ NHANES III was conducted during the period from 1988-1994, which is most closely related to NLSY79 in terms of timing.

We are also concerned about our definition of the comparison group. For example, individuals in the comparison group could be FSP participants before 1985. Including the previous participants in the comparison group could be problematic if the FSP alters the individuals' eating habits, which in turn affect BMI. For example, the treatment effects would be underestimated if the FSP causes a binge eating habit and some individuals in the comparison group picked up the habit when they were previously enrolled in the FSP. So the first alternative comparison group consists of individuals who had not participated in the FSP since 1978¹¹ and lived in poverty in 1985. We are also interested in comparing FSP participants with all other people including high-income Americans.

In Table 4 we re-estimate the treatment effects on all outcomes using two alternative comparison groups and baseline propensity score specification. Compared to the baseline comparison group, the alternative comparison group yields similar treatment effects on all three outcomes. None out of 36 estimates is statistically significant. Our results indicate that DID propensity score matching is robust to various definitions of the comparison group.

C.4. Refining the Treatment Group

We are also concerned about the fact that most treated people were long-time FSP participants and their BMI might have been affected by the program even before the period that we study in this paper. For example, if FSP participants gained weight and

¹¹ Strictly speaking, the data set does not contain the full history of individuals' FSP participation. The earliest history dates back to 1978.

became obese before 1985 and their BMIs had been stable ever since, the treatment effects are going to be underestimated. To get clean estimates of treatment effect on full-time participants, we drop those who were full-time participants previously from the treatment group. Table 5 presents estimates using the baseline comparison group and baseline propensity score equation. None of estimates is statistically and economically significant.

C.5. Covariate Matching

The PSM imputes the missing potential outcomes using the estimated propensity score. We are also concerned with the assumption PSM makes about treatment assignment. The covariate matching (CVM) estimator is not based on assumptions about either treatment assignment or the relationship between covariates and outcomes. We employ CVM as our last robustness check. In CVM, every treated unit is matched to a number of units in the control group based on the distance measured by the vector norm $\|\cdot\|$. Let $\|x\|_V = (x'Vx)^{1/2}$ be the vector norm with positive definite matrix $V^{1/2}$, the CVM defines $\|z - x\|_V$ as the distance between the vector x and z , where x and z represent the covariates for a treated unit and a potential match. Let $d_M(i)$ be the distance from unit i to the M^{th} nearest match with the opposite treatment. Consider the set of observed covariates for unit i to be X_i , the set unit i will match with is:

$$\Psi_M(i) = \{l = 1, \dots, N \mid T_l = 1 - T_i, \|X_l - X_i\|_V \leq d_M(i)\}, \quad (1)$$

¹² We use the diagonal matrix, of which the diagonal elements are the inverses of the variances of X_i (the element of the set of covariates), as our weighting matrix V . The weighting matrix V accounts for the difference in the scale of the covariates.

and $d_M(i)$ is defined as

$$\sum_{l:T_l=1-T_i} 1\{\|X_l - X_i\|_V \leq d_M(i)\} = M, \quad (2)$$

where $1\{\cdot\}$ is the indicator function, which is equal to 1 when the value in brackets is true and zero otherwise.

The simple matching estimator will be biased in finite samples when the matching is not exact. Abadie and Imbens (2002) develop a bias-corrected matching estimator adjusting the difference within the matches for the differences in their covariate values. Although theoretically matching on multidimensional covariates can lead to substantial bias, the matching approach combined with bias adjustment often leads to estimates with little remaining bias.

We re-estimate the short-term and long-term treatment effects using all three comparison groups: baseline comparison group and two alternative comparison groups. The set of covariates include pre-treatment BMI, age, education, ethnicity dummies, family size, per capita income, marital status dummy, urban dummy, and mother's education. The CVM results are presented in Table 6. None of 54 estimates is statistically significant. The results show no evidence that FSP participation causes obesity, overweight or higher BMI.

C.6. Part-time Participants

To give a complete picture of how participants are affected by FSP, we also investigate the treatment effects of FSP on part-time participants. We apply the baseline control group and the propensity score specification and estimate the

treatment effects of FSP on part-time participants. The estimates are presented in table 7. None of the estimates is statistically significant. We also experiment with different propensity score specifications and control groups for all periods. Our results show that the FSP is not responsible for obesity, overweight and higher BMI among part-time participants.

Table 1: Household Size, Net income Limit and the Max FSP benefits (2009-2010)

Household Size	Gross Monthly Income (130% of Poverty Level)	Net Income Limit (Poverty Level)	Max FSP Benefits(\$)
1 person	1,174	903	200
2 persons	1,579	1,215	367
3 persons	1,984	1,526	526
4 persons	2,389	1,838	668
5 persons	2,794	2,150	793
6 persons	3,200	2,461	952
7 persons	3,605	2,773	1,052
8 persons	4,010	3,085	1,202
Each additional person	+406	+312	150

Data Source: USDA, Food and Nutrition Service, Supplemental Nutrition Assistance Program

Table 2: Estimates Using Different Specifications of Propensity Score Equation

	Outcome = Obesity				Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
One-Year Participation												
Specification-1	-0.0289	-0.0239	-0.0259	-0.0245	-0.0067	-0.0075	-0.0017	-0.0172	0.1353	0.1185	0.1018	0.0794
	[0.0205]	[0.0183]	[0.0204]	[0.0191]	[0.0299]	[0.0294]	[0.0340]	[0.0326]	[0.1934]	[0.1785]	[0.2099]	[0.1985]
Specification-2	-0.0114	-0.0084	-0.0114	-0.0109	0.0012	0.0008	0.0029	-0.003	0.1923	0.1888	0.2783	0.2607
	[0.0110]	[0.0099]	[0.0113]	[0.0104]	[0.0178]	[0.0207]	[0.0225]	[0.0216]	[0.1353]	[0.1568]	[0.1673]	[0.1496]
Three-Year Participation												
Specification-1	0.027	0.0289	0.0461	0.0506	0.0347	0.0223	0.0016	0.0033	0.0266	0.2407	0.1916	0.1979
	[0.0282]	[0.0349]	[0.0316]	[0.0305]	[0.0368]	[0.0387]	[0.0447]	[0.0410]	[0.2938]	[0.3008]	[0.3135]	[0.3083]
Specification-2	0.0356	0.0285	0.0217	0.037	0.0127	0.0079	0.0192	0.001	0.0379	0.0605	0.0289	0.0467
	[0.0206]	[0.0377]	[0.0308]	[0.0263]	[0.0331]	[0.0377]	[0.0424]	[0.0464]	[0.3035]	[0.3489]	[0.3355]	[0.3248]
1985-1986												
Specification-1	-0.0581*	-0.0508	-0.058	-0.058	-0.049	-0.0468	-0.0518	-0.057	0.0195	-0.0085	0.05	-0.0618
	[0.0261]	[0.0264]	[0.0302]	[0.0282]	[0.0409]	[0.0418]	[0.0467]	[0.0444]	[0.2849]	[0.3123]	[0.3473]	[0.3315]
Specification-2	-0.034	-0.0427	-0.047	-0.044	-0.056	-0.0613	-0.05	-0.0355	0.3128	0.3901	0.3197	0.2839
	[0.0252]	[0.0268]	[0.0281]	[0.0271]	[0.0444]	[0.0520]	[0.0559]	[0.0536]	[0.3481]	[0.3858]	[0.4092]	[0.3965]
1985-1998												
Specification-1	-0.0297	0.0019	0.0088	0.0167	-0.0259	-0.022	-0.0439	-0.093	-0.6602	-0.3788	-0.1779	-0.36
	[0.0509]	[0.0552]	[0.0543]	[0.0509]	[0.0774]	[0.4586]	[0.0992]	[0.0907]	[0.7960]	[0.8065]	[0.7185]	[0.6529]
Specification-2	0.0012	0.0059	0.0226	0.013	-0.0385	-0.0491	-0.0522	-0.0765	-0.2672	-0.5381	-0.3218	-0.564
	[0.0487]	[0.0534]	[0.0533]	[0.0507]	[0.0832]	[0.0878]	[0.0968]	[0.0892]	[0.8378]	[0.6187]	[0.6379]	[0.5848]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Propensity score specifications:

Specification 1: $Prob(T=1)=F$ (pre-treatment BMI(two periods), age, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west, Interactions between pre-treatment BMIs and ethnicity).

Specification 2: $Prob(T=1)=F$ (pre-treatment BMI, age, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west, BMI_1985², age², educ², income per person², fsize², motedu²).

Table 3: Estimates Using Corrected BMI

	Outcome = Obesity				Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
One-Year Participation	-0.003 [0.0155]	0.0125 [0.0177]	0.0106 [0.0197]	0.0173 [0.0184]	0.0008 [0.0168]	-0.006 [0.0181]	-0.0102 [0.0224]	-0.0154 [0.0206]	0.1854 [0.1380]	0.1804 [0.1523]	0.1335 [0.1634]	0.2692 [0.1573]
Three-Year Participation	0.0171 [0.0254]	0.0396 [0.0422]	0.038 [0.0366]	0.0322 [0.0324]	-0.031 [0.0330]	-0.0343 [0.0359]	-0.0282 [0.0427]	-0.0383 [0.0435]	-0.0256 [0.3206]	0.0217 [0.3200]	0.0357 [0.3143]	-0.0058 [0.2938]
1985-1986	0.0054 [0.0264]	0.0067 [0.0276]	0.0117 [0.0314]	0.0034 [0.0295]	-0.0043 [0.0383]	0.0122 [0.0415]	0.0234 [0.0479]	0.0132 [0.0433]	0.0752 [0.3176]	0.0342 [0.3389]	0.0147 [0.3771]	0.033 [0.3618]
1985-1988	0.0027 [0.0583]	-0.0002 [0.1631]	0.035 [0.0576]	0.0275 [0.0596]	0.016 [0.0754]	0.0129 [0.4137]	0.0717 [0.1006]	-0.0125 [0.0862]	-0.2399 [0.8635]	-0.1666 [14.0801]	-0.0842 [0.8827]	-0.8695 [0.7854]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Specification of propensity score function: $Prob(T=1)=F$ (per-treatment BMI, age, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west).

Table 4: Estimates Using Alternative Comparison Groups

	Outcome = Obesity				Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
Alternative Comparison Group 1: Individuals who never joined FSP and live in poverty at time t												
One-Year Participation	-0.0137	-0.0113	-0.0117	-0.0132	0.0463	0.0159	0.0216	0.021	0.2053	0.2705	0.2285	0.22
	[0.0154]	[0.0129]	[0.0153]	[0.0140]	[0.0239]	[0.0357]	[0.0375]	[0.0368]	[0.1536]	[0.1704]	[0.1724]	[0.1737]
Three-Year Participation	0.0149	-0.0162	0.022	0.0051	-0.0265	-0.1216	-0.1352	-0.0889	-0.2452	-0.5804	-0.3694	-0.4867
	[0.0264]	[0.0532]	[0.0410]	[0.0399]	[0.0431]	[0.0657]	[0.0737]	[0.0707]	[0.4018]	[0.4537]	[0.4561]	[0.4627]
1985-1986	-0.0254	-0.012	-0.0109	-0.0124	-0.0471	-0.0529	-0.0607	-0.0861	-0.2932	-0.1948	-0.3798	-0.4576
	[0.0216]	[0.0175]	[0.0187]	[0.0181]	[0.0523]	[0.0524]	[0.0602]	[0.0617]	[0.2908]	[0.3282]	[0.3190]	[0.2948]
1985-1988	-0.0215	-0.0031	-0.0035	0.0183	-0.1063	-0.1403	-0.1774	-0.1991	-1.0927	-1.2681	-1.2588*	-1.1702*
	[0.0741]	[0.0915]	[0.0835]	[0.0684]	[0.0945]	[0.8636]	[0.1214]	[0.1114]	[0.9620]	[0.7447]	[0.5856]	[0.5508]
Alternative Comparison Group 2: Individuals who never joined FSP												
One Year of Participation	0.0104	0.0192	0.0131	0.0161	0.016	0.0032	0.0144	0.0095	0.2033	0.2576	0.3465	0.2584
	[0.0108]	[0.0120]	[0.0119]	[0.0121]	[0.0166]	[0.0174]	[0.0199]	[0.0199]	[0.1187]	[0.1784]	[0.1943]	[0.1943]
Three Years of Participation	0.0186	0.0164	0.0126	0.0228	-0.0164	-0.0155	-0.0198	-0.0055	-0.0788	-0.1514	-0.2104	-0.0324
	[0.0252]	[0.0224]	[0.0273]	[0.0255]	[0.0299]	[0.0306]	[0.0318]	[0.0309]	[0.2471]	[0.2158]	[0.2357]	[0.2254]
1985-1986	-0.0238	-0.0165	-0.0126	-0.0188	-0.0309	-0.0243	-0.0338	-0.0469	-0.2838	-0.2952	-0.3107	-0.3225
	[0.0244]	[0.0225]	[0.0259]	[0.0281]	[0.0304]	[0.0402]	[0.0385]	[0.0372]	[0.2482]	[4.6288]	[0.2638]	[0.2553]
1985-1988	-0.0475	0.0049	0	0.0129	-0.0559	-0.0823	-0.0903	-0.0847	-1.0409	-1.0957	-1.136*	-1.1049*
	[0.0499]	[0.0698]	[0.0484]	[0.0433]	[0.0623]	[0.1051]	[0.0769]	[0.0718]	[0.5281]	[0.5806]	[0.4682]	[0.4287]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Alternative comparison groups:

Comparison group 1 consists of individuals who had not participated in the FSP since 1978 and lived in poverty in 1985.

Comparison group 2 consists of all people who had not participated in the FSP since 1978.

Table 5: Estimates Using Refined Treatment Group

	Outcome = Obesity				Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
One-Year Participation	-0.0148	-0.0151	0.0011	-0.0075	0.0313	0.0291	0.0247	0.0312	0.2418	0.2198	0.2179	0.1277
	[0.0191]	[0.0197]	[0.0234]	[0.0215]	[0.0297]	[0.0291]	[0.0361]	[0.0337]	[0.1704]	[0.1784]	[0.2270]	[0.2069]
Three-Year Participation	0.035	0.0467	0.0353	0.0365	0.0068	0.0166	0.0118	0.0035	-0.6042	-0.6416	-0.7916	-0.6977
	[0.0457]	[0.3039]	[0.0500]	[0.0469]	[0.0567]	[0.0593]	[0.0632]	[0.0582]	[0.4126]	[0.7391]	[0.4643]	[0.4190]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Matching covariates include pre-treatment BMI, age, Black dummy, Hispanic dummy, income per person, marital status, number of children, urban dummy, and mother education.

Table 6: Estimates of Covariate Matching

	Outcome = Obesity			Outcome = Overweight			Outcome = BMI		
	Covariate Matching (N=1)	Covariate Matching (N=5)	Covariate Matching (N=10)	Covariate Matching (N=1)	Covariate Matching (N=5)	Covariate Matching (N=10)	Covariate Matching (N=1)	Covariate Matching (N=5)	Covariate Matching (N=10)
One-Year Participation	-0.022 [0.0148]	-0.0135 [0.0120]	-0.01 [0.0116]	-0.0115 [0.0182]	-0.0045 [0.0139]	0.0064 [0.0134]	-0.0646 [0.1409]	0.0459 [0.1203]	0.0991 [0.1175]
Three-Years Participation	-0.0011 [0.0284]	0.03 [0.0215]	0.0336 [0.0201]	0.0406 [0.0386]	0.0038 [0.0308]	-0.0003 [0.0293]	0.1555 [0.2510]	0.1039 [0.1960]	0.1097 [0.1849]
1985-1986	-0.0054 [0.0282]	-0.0145 [0.0220]	-0.0342 [0.0218]	-0.035 [0.0321]	-0.0743 [0.0248]	-0.071 [0.0238]	0.0739 [0.2605]	0.053 [0.2211]	-0.0573 [0.2080]
1985-1988	0.0183 [0.0443]	0.0102 [0.0370]	-0.0516 [0.0371]	0.0024 [0.0520]	-0.0144 [0.0434]	0.001 [0.0425]	-1.8179** [0.5575]	-1.1522* [0.4637]	-1.0697** [0.3941]

Notes:

Standard errors are in brackets.

* and ** indicate statistically significance at 5% and 1% level.

Matching covariates include pre-treatment, age, Black dummy, Hispanic dummy, income per person, marital status, number of children, urban dummy, and mother education.

Table 7: Estimates of Treatment Effects on Part-time FSP Participants

	Outcome = Obesity				Outcome = Overweight				Outcome = BMI			
	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)	Local Linear Matching (ROT)	Local Linear Matching (bw=0.1)	Nearest Neighbor Matching (N=5)	Nearest Neighbor Matching (N=10)
One-Year Participation	0.0139 [0.0120]	0.0143 [0.0121]	0.0168 [0.0134]	0.0136 [0.0130]	-0.0066 [0.0189]	-0.0101 [0.0190]	-0.0107 [0.0218]	-0.0116 [0.0206]	0.1926 [0.1177]	0.1993 [0.1159]	0.1946 [0.1331]	0.2087 [0.1259]
Three-Years Participation	0.0123 [0.0157]	0.014 [0.0182]	0.0108 [0.0187]	0.0122 [0.0173]	0.0005 [0.0231]	-0.0148 [0.0246]	-0.0097 [0.0263]	-0.0151 [0.0256]	0.0021 [0.1685]	-0.0276 [0.1731]	-0.0726 [0.1778]	-0.0856 [0.1762]
1985-1986	-0.0005 [0.0226]	0.0012 [0.0312]	-0.0043 [0.0341]	0.0043 [0.0322]	-0.0241 [0.0359]	0.0025 [0.0504]	-0.0113 [0.0551]	0.0035 [0.0515]	0.2806 [0.2851]	0.2352 [0.3172]	0.0997 [0.3841]	0.1777 [0.3432]
1985-1988	-0.0073 [0.0361]	0.0067 [0.1136]	0.034 [0.0528]	0.0292 [0.0483]	-0.0069 [0.0445]	-0.0273 [0.0601]	-0.066 [0.0662]	-0.0358 [0.0637]	-0.4193 [0.4020]	-0.4948 [0.9863]	-0.5255 [0.4848]	-0.4199 [0.4345]

Notes:

Standard errors in brackets are calculated with 500 bootstrap repetitions with 100% sampling.

* and ** indicate statistical significance at 5% and 1% level.

Specification of propensity score function: $Prob(T=1)=F(\text{pre-treatment BMI, age, age squared, education, Black, Hispanic, income per person, married, number of children, education, mother education, urban, west})$.

Reference

- Abadie, A and Imbens, Guido W.** Simple and Bias-Corrected Matching Estimators for Average Treatment Effects, NBER Working Paper no. 283, 2002.
- Heckman, J.; Ichimura, H. and Todd, P.** "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*, 1997, 64(4), pp. 605-54.
- Lakdawalla, D and Philipson, T.** The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination, NBER working paper no. 8946 2002.
- Rowland, ML.** "Self-Reported Weight and Height." *Am J Clin Nutr*, 1990, 52(6), pp. 1125-33.
- Smith, J and Todd, P.** "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 2005, 125(1-2), pp. 305-53.
- Stunkard, AJ and Albaum, JM.** "The Accuracy of Self-Reported Weights." *Am J Clin Nutr*, 1981, 34(8), pp. 1593-99.