Three Essays on Money Input and Time Input in Food Poverty Measurement
and Healthy Eating Index

Yanliang Yang

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
In
Economics, Agricultural and Life Sciences

George C. Davis, Co-Chair
Wen You, Co-Chair
Suqin Ge
Srijan Sengupta

May 20, 2019
Blacksburg, Virginia

Keywords: Money Input, Time Input, Food Poverty Measurement, and Healthy Eating Index

Funding for the third chapter was provided in part, by the University of Kentucky Center for Poverty Research (funded through USDA ERS and FNS) and the Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture. Funding for the whole dissertation is partially provided by the College of Agriculture and Life Sciences Graduate Teaching Scholar (GTS) Program, Virginal Tech.

Three Essays on Money Input and Time Input in Food Poverty Measurement
and Healthy Eating Index

Yanliang Yang

ABSTRACT

A healthy diet is related to a low risk of chronic diseases. A large body of research is devoted to improving social welfare by promoting healthy eating. This dissertation addresses the relation of food and health by analyzing the money and time inputs in food, the food poverty measurement, and a corresponding health outcome.

The second chapter extends the current food poverty measure in headcount and proposes a set of Foster, Greer, and Thorbecke (FGT) indices, which is commonly used in development literature, in food poverty to allow for a more comprehensive understanding in food poverty evaluation. The counter-factual analysis on removing the American Recovery and Reinvestment Act (ARRA) component from the food expenditure shows that the original metrics underestimate the reduction to food expenditure poverty associated with ARRA, whereas the FGT indices indicate a slightly larger impact of ARRA in alleviating food poverty.

The third chapter uses the same FGT indices in food poverty measurement but focuses on the sensitivity of these measurements to a different spatial and temporal food price. We use linear regression to estimate the local level of food poverty thresholds. The results show the spatial and temporal-specific thresholds are higher than the national threshold. The West region shows the most severe poverty situation, indicating the importance of considering spatial and temporal variations in measuring food expenditure poverty. The decompositions of food expenditures show that both the Supplemental Nutrition Assistance Program (SNAP) benefits and money spent on protein play an essential role in reducing food expenditure poverty.

The fourth chapter combines the two datasets used in the previous two chapters to investigate the connection between the resources (money and time) devoted to food and a corresponding health outcome (Healthy Eating Index, HEI). Two-Sample-2-Stage-Least-Square (TS2SLS) model is used to account for the two different datasets in predicting the time spent on food-related activities. After obtaining the time input, a Three-Stage-Least-Square (3SLS) model shows the time input improves the HEI for Non-SNAP households, who are more constrained by time. The decomposition of the impact of education on the HEI shows the indirect impact account for 22% of the total impact. This analysis breaks down the impact of the characteristics on HEI through different channels, thus offers more comprehensive policy recommendations.

Three Essays on Money Input and Time Input in Food Poverty Measurement
and Healthy Eating Index

Yanliang Yang

General Audience Abstract

A healthy diet is related to a low risk of chronic diseases. A large body of research is devoted to improving social welfare by promoting healthy eating. This dissertation is a series of studies on food and health regarding the money and time input on food, the food poverty measurement, and the corresponding health outcome.

The second chapter extends the current food poverty measure in headcount and proposes a set of distributional metrics: depth and severity, which measures how far away households are away from the targeted threshold and how severe the food poverty is respectively. These distributional metrics allow for a more comprehensive understanding of food poverty evaluation. We also analyzed the change of the metrics when removing part of the food expenditure funding source. The analysis shows the original metrics tend to underestimate the reduction to food expenditure poverty and indicates a slightly larger impact of removed funding source in alleviating food poverty.

The third chapter uses the same distributional food poverty metrics, but focuses on the sensitivity of these measurements to different spatial and temporal food prices. We use linear regression in estimating the local food poverty thresholds. The results show the spatial and temporal-specific thresholds are higher than the national threshold. The West region shows the most severe poverty situation, indicating the importance of considering spatial and temporal variations in measuring food expenditure poverty.

The forth chapter combines the two datasets used in the previous two chapters to investigate the connection between the resources (money and time) spent on food and a corresponding health outcome. A special econometrics model is used to predict the time spent on food-related activities with two datasets. After obtaining the time input, a system of equations model shows the time input improves the healthy eating for households who are more constrained by time. The decomposition of the impact of education on healthy eating shows the indirect impact account for 22% of the total impact. This analysis breaks down the impact of the characteristics on HEI through different channels, thus offers more comprehensive policy recommendations.

ACKNOWLEDGMENT

This dissertation would not be possible if without the support from my co-chairs, my committee members, my family, and friends. It is my great honor to express my appreciations to those who have influenced me academically and personally during my Ph.D. studies.

George C. Davis was my mentor since my first day at Virginia Tech and later became my co-chair. He has been the major influencer in my academic journey in Economics. I learned a lot from his courses and working with him on projects. My understanding of economics research was greatly reshaped in his Food and Health Microeconomics course when he systematically introduced the gaps between the theoretical assumptions and empirical models. I was deeply amazed by the theoretical elegance of economics, especially in household production. That is when I found my passion for Food and Health Economics. The opportunity of working with Dr. Davis closely on various research projects is another blessing for me. His rigor in research and persistence in the face of difficulties was always inspiring for me to be a better researcher and communicator.

Wen You is another co-chair for my dissertation, from whom I learned the econometrics skills systematically. The learning in Dr. You's course was always challenging yet rewarding. Working with Dr. You in the research projects was also exciting. In our weekly meeting, her comments always enlighten me with new ways of looking at the questions. Dr. You not only provided me with support regarding my academic development, but also helped me during my family difficulties. Her encouragement and empathy helped me in going through the toughest moments as an international student studying alone far away from my home country. I am indebted for her support throughout my Ph.D. learning at Virginia Tech.

I also want to extend my appreciation to my dissertation committee members (Suqin Ge and Srijia Sengupta) for their great support and invaluable advice. I am grateful to Dr. Ge, an expert in labor economics, for her suggestions about the household production model used in this dissertation. I am equally thankful to Dr. Sengupta for offering insights about the prediction model and estimation used in this dissertation.

Besides my academic supporters, I would like to thank my friends in Blacksburg who opened their heart to me and helped me in numerous ways. Special thank you to Mina, Kate, Peti, Ulrike, Mengjiao, Miao, Ruoding and Xin. Some of you have started a new chapter in different places and I will cherish those good old days forever.

Last but not least, a special thank you to my family for their countless love for the past years. My parents and my brother(Peiliang) always believe in me and encourage me to pursue my dreams. My beloved husband, Guanghua, has always been my cheerleader academically and personally, and without whom, I would not have had the courage to embark on this journey in the first place. To all of the warm love and endless support, I dedicate this dissertation.

Table of Contents

1 Introduction

A healthy diet is related to low risk of chronic diseases. A large body of research is devoted to improving social welfare by promoting healthy eating. This dissertation is a series of studies on food and health regarding three main topics: 1) food poverty measurement impacted by both money and time spent on food eating; 2) the food poverty measurement impacted by different spatial and temporal thresholds; 3) the corresponding health outcome measured by the Healthy Eating Index (HEI) as a result of food purchasing using money and time resources.

The second chapter extends the current food poverty measure in headcount and proposes a distributional measure (Foster, Greer, and Thorbecke (FGT) indices) in food poverty to allow for a more comprehensive understanding in food poverty evaluation. As an example to illustrate the use of this new set of measurements, it offers a policy evaluation of the American Recovery and Reinvestment Act (ARRA) by using counterfactual analysis assuming the ARRA component of the SNAP benefit were removed. The third chapter uses the same FGT food poverty measurements, but focuses on the sensitivity of the measurement of different spatial and temporal food prices. A decomposition of the money expenditure components and segmentation by household demographics are also provided. Chapter four combines the datasets used in the previous two chapters to investigate the connection between the money and time input devoted to food and a corresponding health outcome: the Healthy Eating Index (HEI). A special econometrics model to take account of two different datasets is adapted in the prediction of time in food production. The decomposition of the impact of demographic characteristics on the HEI into direct and indirect effects is also derived from the theoretical framework and illustrated in the empirical analysis.

In the second chapter, the FGT indices are applied to the Supplemental Nutrition Assistance Program (SNAP) participants. The SNAP is the largest nutrition program in the United States, accounting for about 80% of the USDA budget. Recently the adequacy of these benefits has become a concern (see Caswell and Yatkine 2013), but by definition, adequacy implies some goal or target. According to section 2 of 7 of the US Code 2011, the purpose of the SNAP is to "permit low-income households to obtain a more nutritious diet through normal channels of trade by increasing food purchasing power for all eligible households who apply for participation" (Supplemental Nutrition Assistance Program 7 USC 2011). The purpose of this chapter is to extend the most common measure for evaluating this explicitly stated intermediate goal to provide a more comprehensive picture of the adequacy of SNAP benefits. The extended metrics are used to answer the following question: Did the SNAP component of the American Recovery and Reinvestment Act (ARRA) improve the food expenditure poverty situation for SNAP participants as was intended?

We find that the less comprehensive measures are not closely tied to the purpose of the SNAP and tend to underestimate the reduction to food expenditure poverty associated with the ARRA, whereas the more comprehensive metrics presented here are more closely tied to the purpose of the SNAP and indicate a slightly larger impact.

While the second chapter on food poverty metrics relies on an important concept: food poverty threshold, the third chapter dives into this threshold. There are three specific objectives. Firstly, we want to determine how sensitive these poverty indexes are to national versus spatial and temporal-specific food prices and threshold estimates. Secondly, we measure the attribution of different types of food expenditures (e.g. Food-Away-From-Home (FAFH) vs. Food-At-Home (FAH), personal funded FAH vs. SNAP funded FAH, or among major food groups) to the poverty indexes. Finally, we explore the additive decomposability property of these poverty indexes to determine the contribution to the overall poverty index associated with various policy-relevant data partitions, such as household labor force participation, household composition, food security status and various spatial categorizations.

The Thrifty Food Plan (TFP) is the minimal amount of money required to meet the nutrition standard, which serves as the food expenditure poverty threshold. The results show the spatial and temporal-specific TFPs are higher than the national TFP, indicating the importance of considering spatial and temporal variations in measuring food expenditure poverty. The decompositions of food expenditures by funding sources show that SNAP benefits play an essential role in reducing food expenditure poverty. The food group decomposition results show spending on protein is the most significant source in alleviating food expenditure poverty. Finally, the household partitions by regions show large heterogeneity of poverty indexes across regions, with the West region showing the most severe poverty situation mainly due to a higher regional temporal-specific TFP threshold.

The last chapter analyzes the connection between money and time resources spent on food acquisition and a health outcome measured by the Healthy Eating Index (HEI). Diets of high quality are associated with a low risk of noncommunicable diseases. Policies in promoting a healthy eating focus on improving HEI, a measurement on diet quality. Two of the most important economic factors that impact HEI are money and time. For example, Venn and Strazdins (2017) found both money and time restrict healthy food choice. Despite the obvious economic connection between money, time, and HEI, little is known empirically about this relationship. Observational data shows money and time are substitutes: Food-Away-From-Home (FAFH) saves time in food preparation but is often more expensive on a per serving basis than

Food-At-Home (FAH). For empirical analysis, almost all analyses are based on just the relationship between money (i.e., income and prices) and HEI, while completely ignoring time. This is understandable because no single dataset contains all three of these elements. Ignoring the time input in the analysis leads to several problems. First, when the food decision is made jointly with the market goods and time input constraint in reality, the theoretical framework with time input constraint missing in the decision-making process will be incomplete. Secondly, the decision on market goods and time inputs are joint decisions and thus correlated. Thus, models considering only market goods will lead to empirical biased estimation due to the omitted variable problem. Finally, the policy recommendation on food diet to improve the public health based on market goods only analysis will be misleading.

This chapter follows the household production theory (HPT). A household obtains utility from the FAH and FAFH, produced by market goods (the ingredients) and the labor/time input, conditional on other factors. The econometric techniques based on instrumental variable (IV) techniques are also adapted by merging disparate datasets that contain information on healthy eating, money, and time expenditures in estimating a healthy eating production function. The inclusion of time input in the model reduces the threat of the omitted variable problem and provides theoretical guidance in identification strategies. It also enables the pathway analysis of understanding the direct and indirect effects of other factors through their effect on money and time allocations and thus on HEI.

The result shows the impact of time input on healthy eating index varies by household status. The models based on Non-SNAP households indicate 1) statistically significant impact of FAH time input on the HEI; 2) the time input and money input on FAH are complements to each other. There is no statistically significant effect of FAFH time input on HEI and no effect of time on SNAP households either. The decomposition of the effect of education on HEI shows 20% and 3% of the effect passes through the expenditure on FAH and FAFH indirectly, compared to 77% of the direct impact on HEI.

2 Measuring Food Expenditure Poverty in SNAP Populations: Some Extensions with an Application to the American Recovery and Reinvestment Act

2.1 Introduction

The Supplemental Nutrition Assistance Program (SNAP) is the largest US nutrition program, accounting for 80% of the USDA budget (Monke 2013). Recently, questions have arisen about the adequacy of SNAP benefits (Caswell and Yatkine 2013). But, adequacy, by definition, implies some goal or target. The US Code 2011 states the purpose of SNAP is to "permit low-income households to obtain a more nutritious diet through normal channels of trade by increasing food purchasing power for all eligible households who apply for participation" (SNAP 7 USC 2011).

Though general welfare may be the final goal, in the language of economic policy theory (e.g., Acocella, Di Bartolomeo, Hughes-Hallett 2012), "increasing food purchasing power" is an intermediate goal or target. Ease in measurement, observability, and monitoring, and greater uncertainties associated with the structure and number of determinants of final outcomes are just some of the reasons why an intermediate target may be favored over a final target (e.g., Holbrook and Shapiro 1970 in a macro context).

The purpose of this article is to extend the most common measure of this explicitly stated intermediate goal to provide a more comprehensive picture of the adequacy of SNAP benefits. This extension is important because the effectiveness of any anti-poverty program depends on the stated goal and how accurately the chosen metrics reflect that goal. The advantages of the extended metrics are demonstrated by answering the question: Did the SNAP component of the American Recovery and Reinvestment Act (ARRA) improve the food expenditure poverty situation for SNAP participants as was intended? We find that the current less comprehensive measures are not closely tied to the purpose of the SNAP and tend to underestimate the effectiveness of the ARRA, whereas the more comprehensive metrics presented here are more closely tied to the purpose of the SNAP and indicate a slightly larger impact.

The next section presents the most common measure of purchasing power required for a nutritious diet and discusses some of its limitations. The following section demonstrates how this measure is embedded and extended with metrics from the poverty literature that address these limitations. Recent literature, based on Becker's (1965) seminal household production theory, has also identified the exclusion of labor cost as another important limitation and we also demonstrate how this literature is also extended with these poverty metrics. All the metrics are then applied to evaluating the impact of the ARRA and we close with conclusions and limitations.

2.2 Normalized Food Expenditures

SNAP benefits are derived from the USDA Thrifty Food Plan (TFP), which is considered the minimum monetary cost required to meet the US dietary guidelines (see Wilde and Llobrera 2009 for a good overview). As quoted, the US Code 2011 goal is about "purchasing power" relative to a "nutritious diet" so the ratio of actual food expenditures to the TFP is a simple, intuitive, and ubiquitous evaluation metric.[1]

Let M_i^a denote actual food expenditures and M_i^{TFP} the recommended TFP food expenditure amount for household *i*. The TFP normalized money expenditures are then

$$(1) \quad NME_i = M_i^a \div M_i^{TFP} \qquad : Normalized\ Money\ Expenditures$$

[1] This ratio goes by various names (e.g. Needs Standard, Hoynes, McGranaham, and Schanzenbach 2016; TFP Adjusted Food Expenditures, Nord and Prell 2011; Money Expenditure Ratio Davis and You 2011; Ratio of Actual Expenditures to TFP Stewart and Blisard 2006; Standardized Cost Horning and Fulkerson 2014; Food Spending Relative to the TFP, Coleman-Jensen et al. 2017, Tiehen, Newman, Kirlin 2017)

SNAP benefits are not intended to cover all food expenditures. Participants are assumed to spend about 30% of their own money on food, so the appropriate numerator is food expenditures from all sources. If $NME_i > 1$, the household is spending more than enough to reach the nutrition target. Otherwise it is not. Of course, the *NME* does not tell us anything about household diet quality or more distant outcomes, such as food security or childhood obesity, but these are not the focus of the US Code 2011.

Some central tendency measure of the normalized food expenditure is the most common measure of resource adequacy found in the literature. The Annual Food Security Report of USDA (e.g., Coleman-Jensen, et al. 2017) reports the normalized food expenditure every year and it can be found in numerous reports and journal articles (e.g., Davis and You 2011; Horning and Fulkerson 2014; Hoynes, McGranaham, and Schanzenbach 2016; Katare and Kim 2017; Nord 2013; Nord and Prell 2011; Rose 2007; Stewart and Blisard 2006; Tiehen, Newman, and Kirlin 2017). For SNAP eligible or participating households, the mode across studies is usually a little less than 1.0, implying SNAP households are not spending enough to reach the TFP target but that interpretation is misleading.

While the *NME* does provide useful information about expenditures, it suffers three limitations. First, it only provides information on expenditures, not households. The ultimate subject of interest in US Code 2011 is "low-income households" not money. Even for a sample restricted to SNAP participants, the mean or median *NME* tells us nothing about how many low-income households are above or below the TFP threshold. Second, the *NME* does not give a clear indication of how far low income households may be below the TFP threshold. Finally, the *NME* does not give any indication of the concentration of households below the TFP threshold.

Fortunately, these limitations are easily addressed by connecting the *NME* to a well-known poverty metric.

2.3 A Poverty Index Extension of the Normalized Money Expenditure

The fundamental question implied by the US Code 2011 is: how are individual households doing relative to some minimum standard or target? This is a poverty question (See Ziliak 2006 for a good overview of poverty). As normally defined, poverty is to be below some minimum income level. The implicit assumption is that if income, which is an intermediate target and an input in an indirect utility function, is above some threshold so too will be the final output or target (utility). Of course, this poverty concept can be applied more generally to any case where a household is below some minimum resource (input) threshold (Citro and Michael 1995). Thus the TFP defines a minimum food expenditure threshold and standard poverty metrics can be used to define what we will call *food expenditure poverty*: food expenditures below the TFP threshold.

The poverty index developed by Foster, Greer, and Thorbecke (FGT: 1984) has very appealing theoretical properties, is very easy to implement, and is a staple in the poverty literature. With the notable exception of Jolliffe, et al. (2005), and similar studies by Tiehen, Jolliffe, and Gundersen (2012) and Tiehen, Jolliffe, and Smeeding (2016), the SNAP has not been viewed through the lens of the FGT index. These authors considered how SNAP benefits affected income poverty, so their relevant threshold was the income poverty level. In line with US Code 2011, our interest is in food expenditure poverty, so our relevant threshold is the TFP food expenditures.

The FGT poverty index is

(2) $P_\alpha = N^{-1}\sum_{i=1}^{N} I\left(z_i > y_i\right)\left(\frac{z_i - y_i}{z_i}\right)^{\alpha}$,

where y_i denotes the value of the variable of interest and z_i the threshold value defining 'poverty' for the i$^{\text{th}}$ household. The i subscript on z indicates the threshold may vary by household. The indicator function $I(z_i > y_i) = 1$ if the household is below the poverty threshold and is zero otherwise. The term $(z_i - y_i) \div z_i = 1 - (y_i \div z_i) = g_i$ is the normalized poverty gap g_i. In the present application, $y_i = M_i^a$ and $z_i = M_i^{TFP}$, so the indicator function only counts those households below the TFP and the normalized poverty gap is $g_i = 1 - (M_i^a \div M_i^{TFP}) = 1 - NME_i$, which shows how the normalized gap is related to the NME_i. The normalized gap is expressed in percentage terms below the threshold, so if $g_i = 0.25$, the household is 25% *below* the threshold. The parameter α defines the poverty measure of interest: $P_{\alpha=0}$ gives the percentage of households below the poverty threshold – the poverty rate or *prevalence*, $P_{\alpha=1}$ gives the per capita household distance from the poverty threshold in percentage terms or *depth*, and $P_{\alpha=2}$ gives degree of skewness in depth or *severity*.

The FGT index addresses all three of the *NME* limitations mentioned above. The mean *NME*, that is usually reported, is related to the poverty index as

(3) $\overline{NME} = (P_0 - P_1) + N^{-1}\sum_{i \in N_A} NME_i$

where N_A is the sample above the threshold. The mean *NME* conflates information on prevalence and depth (the first term) and also includes information for those above the threshold (the last term) and so could increase even as the prevalence and depth of poverty did not change or actually increased. Alternatively, the FGT poverty index, via the indicator function, limits the sample to those individuals who are of most concern, those below the threshold or who are in poverty. Furthermore, by setting $\alpha = 0,1$, and 2 the focus is on the number of households in poverty (prevalence) and how far the households in poverty are from reaching the threshold (depth) and to what degree they are in poverty (severity) (e.g., Ravallion 1994).

2.4 Is the Denominator Wrong? The "Full Cost" of the TFP

The normalized money expenditure also suffers another limitation. It underestimates the cost of a nutritious diet because the TFP only estimates the cost of *one* input (groceries). But a nutritious diet also requires labor: labor in meal planning, travel to the store, shopping, preparing the food for assembly, and cooking. Ignoring the labor cost in food production leads to an underestimation of the full cost of a diet, which in turn leads to an overestimation of the effectiveness of SNAP and an underestimation of food expenditure poverty. This is a direct application of household production economics and the "full cost" of production (Becker 1965).

There is now a rather substantial literature on the role that time plays in nutrition related outcomes, ranging from lower diet quality being associated with lower time in food preparation (e.g., Jabs and Devine 2006; Monsiavais, et al. 2014), to mealtime planning and food insecurity (Fiese, et al 2016) and healthy child weight (Fiese, et al. 2012). Marshall and Pires (2017) and Hilbert, et al. (2014) find that travel costs, which is directly proportional to time, is more important in determining grocery choices and diet quality than food prices. In a comprehensive review of the literature of numerous nutrition and health outcomes related to home cooking, Mills et al. (2017) find that one of the main factors affecting food preparation is time availability and employment.

In the present context, recent research finds labor (time) in food production is more important than money in reaching the TFP nutrition target (e.g., Davis and You 2011; Raschke 2012; Rose 2007). The underlying logic is straightforward and can be illustrated with a standard two input isocost isoquant (household) production diagram (figure 1). The vertical axis represents household (food) money expenditures *M*. The horizontal axis household (food) time expenditures or labor *T*. There is an implicit target nutrition isoquant that is consistent with the TFP (dashed N_{TFP}) and $TFP \equiv (M^{TFP}, T^{TFP})$ denotes a money−time input combination that lies on this isoquant.

Given there are two required inputs for reaching a nutrition target, a household could be 'time poor and money rich,' or vice versa (Davis and Serrano 2016 ch. 6). Any of the households C through G are 'money rich but time poor', but household H is 'money poor and time rich.' So the reason for not reaching the nutrition target could be different for different households. In this context, poverty can be multidimensional and, as Atkinson (2003 p. 51) states, "There is widespread agreement that deprivation is multidimensional. It is not enough to look only at income poverty; we have to also look at other attributes." One of the *atheoretical* multidimensional poverty measures could be implemented, such as the intersection method, union method, or counting method (see Alkire et al. 2015 for overview) but in the present context, there are two fundamental questions: (i) how much is the "full" cost of the TFP? and (ii) how far is the household from this full cost TFP target? These are standard compensation questions, so a well-established theoretical measure can be utilized.

Davis and You (2011) demonstrate that the cost difference approach (Malchup 1957), in conjunction with the market substitute approach to valuing time in food production (Gronau 1986), will give an isocost line consistent with the TFP nutrition level. The cost difference approach answers the question: what is the appropriate compensation isocost line showing different input combinations that lead to the same "full" cost that includes all inputs?[2] Regarding valuing of time, there are generally two approaches: the opportunity cost approach and the market substitute approach. The opportunity cost approach values the input at how the individual values the input (i.e. a point on their demand curve or willingness to pay) whereas the market substitute approach values the input how the market values the input. As with any good, there is a difference in the individual's willingness to pay and how the market values the good, but similar to all income product accounting, the TFP is based on market value prices of groceries, not individual's willingness to pay for groceries. Given the purpose of the exercise is to determine the "full" market value of all inputs (groceries and labor), and for internal consistency with the TFP, the market substitute approach is appropriate (see Chiswick 1982; Hawrylyshyn 1976 for more discussion of this point in general).

The cost difference isocost line (money-time threshold) is given by the equation

$$(4)\quad MT_i^{TFP} = M_i^{TFP} - p(T_i^a - T_i^{TFP}): \quad \textit{Money-Time Threshold}$$

where p is the market value of time in food production and T_i^a is actual food production time. Equation (4) gives the amount of money the household needs to reach the "full" cost of the TFP (groceries + labor) once their labor in food production is taken into account.[3] So given a household's time allocation to food production, the money-time threshold becomes the relevant "full cost" nutritional expenditure for normalizing food expenditures.

[2] Similar to compensating variation or equivalent variation, the cost difference compensation is about the amount of resources needed to *possibly* reach the targeted isoquant. There is no guarantee that households will allocate these resources in the optimal way to reach the targeted level. That is a different question.

[3] The intuition of the cost difference approach is easily seen by rewriting (4) as $MT + pT^a = M^{TFP} + pT^{TFP}$. The right side gives the "full cost" to reach the TFP nutrition target, including labor. With a given time allocation valued at the market substitute rate pT^a, solving for MT gives the amount required to have the equivalent expenditures to the full cost TFP consistent expenditures.

$$(5) \quad NMTE_i = M_i^a \div MT_i^{TFP} \qquad : Normalized\ Money\ Time\ Expenditures$$

The normalized money time expenditures (*NMTE*) in equation (5) will nest the normalized money expenditures (*NME*) from equation (1) as follows

$$(6) \quad NMTE_i = \frac{1}{[1 + p(T_i^{TFP} - T_i^a)/M_i^{TFP}]} \times NME_i \, .$$

With a positive price of the labor input ($p > 0$) and a time requirement to reach the TFP greater than actual time $(T_i^{TFP} - T_i^a) > 0$, the partial normalized money expenditures (*NME*) will always overestimate the full cost normalized money time expenditures (*NMTE*), implying the *NME* will always overestimate SNAP benefit adequacy (Davis and You 2011). Davis and You (2011) are the only ones to have calculated the *NME* and *NMTE* and estimate these to be about 1.35 and 0.60, respectively, for single headed households, demonstrating the importance of taking into account labor costs.

Pulling all this together, by defining y_i and z_i appropriately in the FGT index equation (2), the prevalence, depth, and severity can be generated in three dimensions: (i) money expenditure only, (ii) time expenditure only, and (iii) 'full' expenditure. None of the mentioned literature has used the FGT or in these three dimensions. The importance of measuring prevalence, depth, and severity in all three dimensions can be seen in figure 1. For example, the prevalence rates for money only, time only, and money-time are 37.5% (=3/8), 87.5% (=7/8), and 62.5% (=5/8). Regarding depth, household A is further below the money only and money-time threshold than B, whereas both are at the same depth from the time only threshold. An increase in SNAP benefits that increased spending from point A to point B will *not* change any of the poverty rates and one may conclude the policy was ineffective. However, the policy certainly got the households closer to the money and money-time thresholds, which is consistent with the SNAP goal stated in the introduction of obtaining “a more nutritious diet...by *increasing* [emphasis added] food purchasing power.” Finally, the severity of poverty is greatest in the time only dimension, followed by the money-time dimension and then the money only dimension. If point B represented 1,000 households and point A represented say 10,000 households, the severity of poverty in the population would be much worse than if these numbers were reversed. The poverty literature has long recognized depth and severity provide a much more comprehensive picture of poverty and policy effectiveness than simple prevalence and all these metrics are easy to implement using existing data already being reported in the literature.

2.5 An Application to the Effect of the ARRA on Food Expenditure Poverty

The idea behind an increase in SNAP benefits, such as occurred with the American Recovery and Reinvestment Act (ARRA), is that actual money (food) expenditures will increase – and also possibly time in food production – thereby decreasing all of the three poverty measures in money/time or both dimensions. The ARRA provides a good case for highlighting the strengths of the more comprehensive measures because the ARRA did not increase SNAP benefits by a large amount. Thus we would expect the current measures that are not very sensitive to distance will likely show little impact whereas the metrics proposed here would provide a more complete picture and show a larger impact.

2.5.1 Data and the ARRA Premium

Equation (2) is used to measure of prevalence (P_0), depth (P_1), and severity (P_2) for (i) only a money threshold, (ii) only a time threshold, and (iii) the theoretically based cost–difference money-time threshold. Each of these measures is evaluated with and without the ARRA premium amount to determine (i) how much the ARRA affected all three measures and (ii) if it affected some measures more than others.

The ARRA was signed into law in February of 2009. Effective on April 1, 2009, the ARRA raised the maximum SNAP allotment by an average of 13.6% which ended on October 31, 2013(USDA/ERS/2015). However, because the maximum allotments varied by household composition and the level of benefits received vary by household specific deductions, all households did not receive the same dollar increase in benefits.

For this analysis, the data requirements are household food money *and* time expenditure data, household composition, and SNAP benefit data. These requirements limit dataset options. We follow Davis and You (2011) and use the Food Security Supplement (FSS: USDA/ERS/CPS-FSS 2009-2011) and the American Time Use Survey (ATUS: USDL/BLS 2009-2012), which can be matched because they are both supplements and subsamples of the Current Population Survey (CPS) and contain household compositional information. We focus on single-headed households who are most susceptible to resource constraints (e.g., Casey and Maldonado 2012; Meyer and Abdul-Malak 2015) and because the ATUS collects time diary information from a single individual in the household. Given the focus of the study, we also limit the analysis to SNAP participating households. Finally, the ARRA increment varies over Alaska, Hawaii, Guam and Virgin Islands but is the same across the 48 contiguous states and DC, so we focus only on the 48 contiguous states and DC. The time period is 2009 to 2011.

Money and Time Thresholds

The weekly money expenditure threshold M_i^{TFP} comes from the June Thrifty Food Plan (TFP) Official USDA Food Plans table (USDA Food Plans: Cost of Food 2009-2011) with household composition adjustments made according to the table footnotes. For the weekly time expenditure threshold T_i^{TFP}, we use the median 13.13 hours per week estimate from Davis and You (2011), which is based on 1,000 USDA weekly meal plans satisfying the TFP. As in Davis and You (2011), we use the annual median hourly wage of Cooks, Private Household (code: 35-2013) from the Bureau of Labor Statistics, Occupational Employment Statistics, Occupational Employment Wages from 2009 to 2011(BLS 2009-2011) for the market substitute price, *p*.

Money and Time Expenditures with ARRA Premium

For actual money expenditures M_i^a, we use the "usual" weekly food expenditures reported in the FSS. During the time period under consideration, actual food money expenditures M_i^a would *include* the effect of the ARRA on food expenditures. The actual weekly time expenditures come from the American Time Use Survey (USDL/BLS/ATUS 2009-2012) and are for Food and Drink Preparation (ATUS code 020101), Food Presentation (ATUS Code 020202), Kitchen and Food Clean-up (ATUS Code 020203), Grocery Shopping (ATUS Code 070101), and Travel Related to Food and Drink Preparation, Clean-up, and Presentation (ATUS Code 180202). The ATUS is a daily time diary but because the highest frequency of M_i^{TFP} and M_i^a is weekly, a weekly household time estimate is obtained using a nonparametric Horovitz and Thompson (1952) estimator as described in Davis and You (2011). Similar to the actual money expenditures, we use the superscript *a* to indicate actual time expenditures, T_i^a.

Money expenditure and time expenditure without ARRA

We need estimates of M_i^a *without* the ARRA. The actual money expenditure M_i^a will depend on the amount of SNAP benefit received and the marginal propensity to spend on food out of the received SNAP benefit (*MPS*). The general benefits formula *without* the ARRA premium (see Caswell and Yatkine 2013 Box 2-2. p. 2-6) is,

$$(7) \quad SNAP_i^o = M_i^{TFP} - Deductions_i$$

After April 2009, the SNAP benefits were increased by a fixed $ARRA_i$ amount, depending on the household composition, but the deduction formulas were not affected. So the SNAP benefit *with* the ARRA premium is

(8) $SNAP_i^w = (M_i^{TFP} + ARRA_i) - \ Deductions_i$.

The difference in SNAP benefits ($\Delta SNAP_i$) is then

(9) $\Delta SNAP = SNAP_i^w - SNAP_i^o = ARRA_i$.

The Food Nutrition Service of USDA reports the maximum benefit *with the* ARRA or $(M_i^{TFP} + ARRA_i)$ (USDA/FNS 2009-2011). The Center for Nutrition Policy and Promotion at USDA reports the value of M_i^{TFP} (USDA Food Plans: Cost of Food 2009-2011). Taking the difference between the two reported figures yields the value of $ARRA_i$ per household. The per capita amount is about $5.00 per week.

As is well documented, the marginal propensity to spend out of SNAP benefits (*MPS*) is normally less than one (Beatty and Tuttle 2014; Breunig and Dasgupta 2005; Fox, Hamilton, and Lin 2004; Fraker 1990; Hoynes and Schazenbach 2009; Tuttle 2016). Thus the reduction in food expenditures associated with removing the *ARRA* would be

(10) $\Delta M_i = -MPS \times ARRA_i$.

The *MPS* falls in the interval of [0.17, 0.47] for most studies, though Fox, Hamilton, and Lin (2004) have a high estimate of 0.86 and Tuttle (2016) reports estimates in the 0.39 to 0.62 range. Given this uncertainty, we consider three cases: MPS_1 = 0.17, 0.47, 1.00.

With respect to changes in time expenditures, only Beatty, Nanney and Tuttle (2014) have looked at the relationship between SNAP benefit levels and food production time. For single-headed households, which we are considering here, SNAP benefits had no statistically significant effect on any meal preparation or grocery shopping time. Consequently, in this main text we focus on the case where the time allotted to food production does not change as a result of the ARRA. In the supplementary appendix results are provided where the time allocated to food production could change by –5% and +5%. We will briefly allude to these findings in the discussion as well.

In summary, the poverty index formulas *with* and *without* the ARRA premium are

$$(11)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[M_i^{TFP} > (M_i^a + \Delta M_i)\right]\left(\frac{M_i^{TFP} - (M_i^a + \Delta M_i)}{M_i^{TFP}}\right)^\alpha$$

$$(12)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[T_i^{TFP} > T_i^a\right]\left(\frac{T_i^{TFP} - T_i^a}{T_i^{TFP}}\right)^\alpha$$

$$(13)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[MT_i^{TFP} > (M_i^a + \Delta M_i)\right]\left(\frac{MT_i^{TFP} - (M_i^a + \Delta M_i)}{MT_i^{TFP}}\right)^\alpha.$$

The *with* ARRA premium formula occurs when $\Delta M_i = 0$ and the j superscript is w (i.e., P_α^w). The *without* ARRA premium formula occurs when $\Delta M_i \neq 0$ for the values discussed and the j superscript is o (i.e., P_α^o). Prevalence, depth, and severity are associated with $\alpha = 0$, 1, and 2, respectively.

2.5.2 Results

We present the components of the normalized poverty gap from equation (2) (table 1) and the poverty metrics with and without the ARRA (table 2). As in Jolliffe, et al. (2005), table 2 contains the percentage change in each poverty metric associated with the ARRA premium, which is calculated as $(P_\alpha^w - P_\alpha^o) \div P_\alpha^w$. Their general variance calculation approach is followed for each metric and the percentage change. However, in contrast to Jolliffe, et al. (2005), whose focus was overall income, given our focus is food expenditures, we use per capita food expenditures in the sorting step. We use the four regions (Northeast, Midwest, South and West) as the synthetic strata.[4]

2.5.2.1 Components

The money expenditure gap $(M_i^{TFP} - M_i^a)$ in table 1 shows that on average the TFP target expenditures are \$5.49 above what households are actually spending, ignoring the value of labor. The average normalized money expenditure gap is – 0.03, so on average households are spending 3% *more than enough* to reach the TFP target. The higher median of 0.08 indicates the distribution is skewed left so the average may be a little misleading. Half of the households are spending *at least* 8% *less than* the TFP target. The 95th percentile intervals indicate there households above and below the target.

The time components paint a more severe picture. On average, the TFP time threshold is about 8.50 hours per week higher than actual time expenditures (i.e., time expenditure gap $T_i^{TFP} - T_i^a$). The median is similar. The 95th percentile (5.90 to 9.27) indicates virtually all households fall short of the required TFP time. The average normalized time expenditure gap indicates that households are 65% below the TFP time threshold.

As shown earlier, the required money-time threshold will be greater than when labor is ignored and this is confirmed. The required money-time expenditure threshold averages about \$105 higher than the actual money expenditures (i.e., the money-time expenditure gap $MT_i^{TFP} - M_i^a$). Given the 95th percentile overlaps zero (– 87.79 to 219.83), some households are spending more than required to meet the TFP threshold. However, the normalized money-time expenditure gap indicates households are only spending on average about 50% of the amount

[4] We considered another more complicated sorting approach using household size and money expenditures and the variances estimates are virtually identical to what we report here.

required to meet the TFP target that takes into account labor cost. All within the 95th percentile interval are below the labor inclusive threshold.

2.5.2.2 Poverty metrics with and without the ARRA

Did the ARRA premium improve the poverty metrics? Column three in table 2 gives the prevalence, depth, and severity measures *with* the ARRA premium and demonstrates that focusing solely on prevalence and monetary resources gives a distorted picture of the degree of poverty.

In terms of *prevalence*, about 60% of the sample is below the money expenditure threshold, but 100% are below the time expenditure and about 93% are below the money-time expenditure thresholds. Ignoring labor cost, 40% of the sample is above the money expenditure threshold. With labor cost included, only 7% are above the money-time threshold. Regarding *depth*, the money expenditure depth of 21.8 indicates that the average 'poor' household (i.e., those below the threshold) falls about 22% below the threshold. For time expenditure, the average 'poor' household falls about 65% short of the time threshold and for the money-time threshold the average household falls about 53% short of the money-time threshold. *Severity* places a greater weight on a larger normalized poverty gap and, like prevalence and depth, a smaller severity number is preferred. Severity appears smallest in the money only dimension (11.7) and is the worst in the time only dimension (42.2), but as expected falls between these extremes in the money-time dimension (34.0). Regardless of the poverty measure used, ignoring labor cost underestimates food expenditure poverty.

As indicated, the effect of the change in SNAP benefits on total food expenditures depends on the *MPS*. Columns four through six in table 2 give the poverty measures *without* the ARRA premium for values of the *MPS* of 0.17, 0.47, and 1.00, assuming there is no change in the time allocation. As the *MPS* increases, the *removal* of the ARRA premium will lead to a *greater* decrease in food expenditures and *greater* poverty measures. The absolute changes in the poverty measures (not shown) are quite small however and this is due to the fact that the average ARRA premium is about $12 per week or about $5.00 per capita for our sample. Thus the issue is *not* that the ARRA did not have the desired effect, but rather that the "dose" was perhaps too small to make much of a difference in absolute terms. However, columns seven through nine reveal the ARRA premium was more impactful in percentage terms than in absolute terms.

The last three columns in table 2 demonstrate the importance of going beyond just the prevalence rate and monetary resources. Regardless of the *MPS* value, removing the ARRA leads to a larger increase in the severity measure, followed by a larger increase in the depth measure, and finally an increase in the prevalence measure for the money only expenditure threshold. A similar pattern emerges for the money-time poverty metrics. The fact that there is only statistical significance when the $MPS = 1$ provides further evidence that dose is the issue, in that a larger *MPS* is effectively equivalent to a larger dose of SNAP benefits being used on food expenditures. The practical translation is the ARRA did more for improving depth and severity than it did prevalence. Just focusing on prevalence underestimates the positive impact of the ARRA.

Interestingly, an important ordinal finding is that across all measures the percentage change in severity is greater than the percentage change in depth, which is greater than the percentage change in prevalence. As demonstrate in the supplementary appendix, under some rather mild conditions, this ordinal ranking is actually an analytical relationship. This indicates the effectiveness of a policy will be understated if only the prevalence rate is considered, regardless of what threshold is considered. Furthermore, this implies the choice of the measurement matters. Again, as stated in the US Code 2011, the purpose of SNAP is “to permit low-income households to obtain a more nutritious diet through normal channels of trade by increasing food purchasing power for all eligible households who apply for participation.” This directive is better measured by depth or even severity than a normalized food expenditure or prevalence. And as shown, the ARRA was more effective at improving depth and severity than prevalence.

2.6 Conclusions and Limitations

In evaluating the effectiveness of a poverty policy, measurement matters. The chosen metric(s) should capture the intent of the policy and include the most important resources for reaching the policy target. This chapter extends the literature on measuring SNAP benefit adequacy as called for in the IOM report (Caswell and Yaktine 2013) by using the FGT poverty index to capture the prevalence, depth, and severity, and by incorporating labor (time) cost into the analysis. Previous analyses, even those including time cost, have only considered the prevalence rate. Consistent with this previous research, if time cost is ignored there is an overly optimistic evaluation of the effectiveness of SNAP benefits that extends to depth and severity. In terms of the impact of the ARRA, it had a much larger positive impact on the percentage change in depth and severity, than prevalence. One could argue the issue was dose level, not systematic ineffectiveness. We believe that depth and severity are more appropriate for measuring SNAP benefit adequacy because they are more in line with the language of the policy intent than the commonly encountered normalized money expenditure or prevalence rate.

As with all analyses, there are limitations and future research needs. Though the ATUS is a drastic improvement in time use data, there are still some outstanding measurement issues (e.g., accounting for intra-household time substitution). This is one of the reasons we limited our analysis to single-headed households. The ‘time deficit’ between actual and TFP consistent time expenditures is likely to be smaller in dual headed households. Furthermore, as Davis and You (2011) have discussed, much more work is needed on the amount of time required to meet the TFP target.

Also, much of the SNAP literature focuses on estimating the effects of SNAP participation (or benefit levels) on some more distant nutrition related outcomes, such as diet quality, food security, child health outcomes, along with moderators or mediators (e.g., education level, employment) via statistical modeling (see Bartfield et al. 2016 for a good overview). The research reported here focuses on the directly stated intermediate target of the US Code 2011 and should be viewed as complementary not competitive with these endeavors. Some rather straightforward mathematics, such as found in the structural equation modeling literature, can demonstrate that (in)significance in an intermediate target implies nothing about (in)significance in a more distant target and vice-versa. This is an area in need of a lot more research, figuring out the causal relationships between intermediate and final targets.

Though important, none of these remaining limitations or future directions change the main conclusion: measurement matters in evaluating the SNAP benefit adequacy and the extensions presented here are very easy to implement with existing data and overcome several existing limitations.

2.7 References

Acocella, N. G. Di Bartolomeno, and A. Hughes-Hallett 2012. *The Theory of Economic Policy in a Strategic Context*. Cambridge University Press. Cambridge, UK.

Alkire, S., J. Foster, S. Seth, J. Roche, and M. Santos. 2015. *Multidimensional Poverty Measurement and Analysis*. Oxford University Press, USA.

Atkinson, A. B. 2003. "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches." *Journal of Economic Inequality*, 1(1): 51-65.

Bartfield, J., C. Gundersen, T. Smeeding, and J. Ziliak (eds). 2016. *SNAP Matters: How Food Stamps Affect Health and Well Being*. Stanford University Press. Stanford, CA.

Beatty, T.K., M.S. Nanney, and C. Tuttle. 2014. "Time to Eat? The Relationship Between Food Security and Food-Related Time Use." *Public Health Nutrition*, 17(01):66-72.

Beatty, T. K and C. Tuttle. 2014. "Expenditure Response to Increases in In-Kind Transfers:Evidence from the Supplemental Nutrition Assistance Program." *American Journal of Agricultural Economics*, 97(2):390-404.

Becker, G.S. 1965. "A Theory of the Allocation of Time." *The Economic Journal*, 75(299):493-517.

Breunig, R., and I. Dasgupta. 2005. "Do Intra-household Effects Generate the Food Stamp Cash-Out Puzzle?" *American Journal of Agricultural Economics*, 87 (3): 552–68.

Bureau of Labor Statistics. 2009-2011. Occupational Employment Statistics. Washington DC. (accessed Feb 29, 2016)

Casey, T. and L. Maldonado. 2012. *Worst Off: Single-parent Families in the United*

States. A Cross-National Comparison of Single Parenthood in the US and Sixteen Other High-Income Countries. New York: Legal Momentum, the Women's Legal Defense and Education Fund.

Caswell, J.A., and A.L. Yaktine, eds. 2013. *Supplemental Nutrition Assistance Program: Examining the Evidence to Define Benefit Adequacy*. Washington DC: National Academies Press.

Chiswick, C.U. 1982. "The Value of a Housewife's Time." *The Journal of Human Resources*, 17(3): 413-425.

Citro, C.F. and R.T. Michael. eds., 1995. *Measuring Poverty: A New Approach*. Washington DC: National Academies Press.

Coleman-Jensen, A., M.P. Rabbitt, C. Gregory, and A. Singh. 2017. *Household Food Security in the United States in 2016.* 194. Washington DC: U.S. Department of Agriculture, Economic Research Service.

Davis, G.C. and E. L. Serrano. 2016. *Food and Nutrition Economics: Fundamentals for Health Sciences*. Oxford University Press, 2016.

Davis, G.C., and W. You. 2011. "Not Enough Money or Not Enough Time to Satisfy the Thrifty Food Plan? A Cost Difference Approach for Estimating a Money–Time Threshold." *Food Policy*, 36(2):101-107.

Fiese, B. H., A. Hammons, and D. Grigsby-Toussaint. 2012. "Family Mealtimes: A Contextual Approach to Understanding Childhood Obesity." *Economics & Human Biology*, 10(4): 365-374.

Fiese, B. H., C. Gundersen, B. Koester, and B. Jones. 2016. "Family Chaos and Lack of Mealtime Planning is Associated with Food Insecurity in Low Income Households."

Economics & Human Biology, 21: 147-155.

Foster, J., J. Greer, and E. Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica*, May 1: 761-766.

Fox, M.K., W. Hamilton, and B.H. Lin. 2004. "Effects of Food Assistance And Nutrition Programs on Nutrition And Health." *Executive Summary of the Literature Review, Food Assistance and Nutrition Research Report* 19(4).

Fraker, T. 1990. *The Effects of Food Stamps on Food Consumption: A Review of the Literature.* Washington, DC: U.S. Department of Agriculture, Food and Nutrition Service.

Gronau, R. 1986. "Home Production-A Survey." In Ashenfelter, O., Layard, R. eds. *Handbook of labor economics*, vol. 1. New York: Elsevier, pp.273-304.

Hawrylyshyn, O. 1976. "The Value of Household Services: A Survey of Empirical Estimates." *Review of Income and Wealth*, 22 (2): 101-103.

Hilbert, N., J. Evans-Cowley, J. Reece, C. Rogers, W. Ake, and C. Hoy. 2016. "Mapping the Cost of a Balanced Diet, as a Function of Travel Time and Food Price." *Journal of Agriculture, Food Systems, and Community Development*, 5(1): 105-127.

Holbrook, R., and H. Shapiro. 1970. "The Choice of Optimal Intermediate Economic Targets." *American Economic Review*, 60(2): 40-46.

Horning, M.L. and J.A. Fulkerson. 2015. "A Systematic Review on the Affordability of a Healthful Diet for Families in the United States." *Public Health Nursing*, 32(1): 68-80.

Horvitz, D.G. and D.J. Thompson. 1952. "A Generalization of Sampling Without

Replacement from a Finite Universe." *Journal of the American Statistical Association* 47(260):663-685.

Hoynes, H., L. McGranaham, and D. Schanzenbach. 2016. "SNAP and Food Consumption" in *SNAP Matters: How Food Stamps Affect Health and Well Being*, Judith Bartfield, Craig Gundersen, Timothy M. Smeeding and James Ziliak (eds), Stanford University Press. Stanford, CA.

Hoynes, H.W., 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): 109–39.

Jabs, J., and C. Devine. 2006. "Time Scarcity and Food Choices: An Overview." *Appetite* 47(2): 196-204.

Jolliffe, D., C. Gundersen, L. Tiehen, and J. Winicki. 2005. "Food Stamp Benefits and Child Poverty." *American Journal of Agricultural Economics* 87(3):569-581.

Katare, B. and J. Kim. 2017. "Effects of the 2013 SNAP Benefit Cut on Food Security." *Applied Economic Perspectives and Policy*: ppx025.

Malchup, F. 1957. "Professor Hicks' Revision of Demand Theory." *American Economic Review*, 47(1):119–135.

Marshall, G., and T. Pires. 2017. "Measuring the Impact of Travel Costs on Grocery Shopping." *The Economic Journal* (2017). Accepted Author Manuscript doi:10.1111/ecoj.12523.

Meyer, M.H. and Y. Abdul-Malak. 2015. "Single-Headed Family Economic Vulnerability and Reliance on Social Programs." *Public Policy & Aging Report* 25(3):102-106.

Mills, S., M. White, H. Brown, W. Wrieden, D. Kwasnicka, J. Halligan, S. Robalino, and J. Adams. 2017. "Health and Social Determinants and Outcomes of Home Cooking: A Systematic Review of Observational Studies." *Appetite*, 111: 116-134.

Monke, J. 2013. *Budget issues shaping a Farm Bill in 2013. Congressional Research Service Report*, *42484*. https://www.fas.org/sgp/crs/misc/R42484.pdf

Monsivais, P., A. Aggarwal, and A. Drewnowski. 2014. "Time Spent on Home Food Preparation and Indicators of Healthy Eating." *American Journal of Preventive Medicine*, 47(6): 796-802.

Nord, M. 2013. "Effects of the Decline in the Real Value of SNAP Benefits from 2009 to 2011." Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No.151.

Nord, M., and M.A. Prell. 2011. Food Security Improved Following the 2009 ARRA Increase in SNAP Benefits. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No.116.

Raschke, C. 2012. "Food stamps and the time cost of food preparation." *Review of Economics of the Household*, 10(2): 259-275.

Ravallion, M. 1994. *Poverty comparisons*, vol. 56. England Oxfordshire: Taylor & Francis.

Rose, D. 2007. "Food Stamps, The Thrifty Food Plan, and Meal Preparation: The Importance of The Time Dimension For US Nutrition Policy." *Journal of Nutrition Education and Behavior*, 39(4):226-232.

Stewart, Hayden, and Noel Blisard. 2006. "The Thrifty Food Plan and Low-income

Households in the United States: What Food Groups are Being Neglected?" *Food Policy*, 31(5): 469-482.

Supplemental Nutrition Assistance Program. Title 7 *U.S. Code*, Sec. 2011. http://uscode.house.gov/browse/prelim@title7/chapter51&edition=prelim

Tiehen, L., C. Newman, and J. Kirlin. 2017. "The Food-Spending Patterns of Households Participating in the Supplemental Nutrition Assistance Program: Findings From USDA's FoodAPS." USDA Economic Research Service. Economic Information Bulletin Number 176. August.

Tiehen, L., D. Jolliffe, and C. Gunderson. 2012 *Alleviating Poverty in the United States: The Critical Role of SNAP Benefits*. USDA, Economic Research Service. Economic Research Report Number 132. April.

Tiehen, L., D. Jolliffe, and T. Smeeding. 2016. "The Effect of SNAP on Poverty" in *SNAP Matters: How Food Stamps Affect Health and Well Being*, Judith Bartfield, Craig Gundersen, Timothy M. Smeeding and James Ziliak (eds), Stanford University Press. Stanford, CA.

Tuttle, C. 2016. "The Stimulus Act of 2009 and Its Effect on Food-at-Home Spending by SNAP Participants." *Economic Research Service. Economic Research Report Number 213*. United States Department of Agriculture. August.

U.S. Department of Labor, Bureau of Labor Statistics. 2009-2012. *American Time Use Survey.* Washington DC. (accessed Feb 29, 2016)

U.S. Department of Agriculture. Economic Research Service, 2009-2011. *Food Security in the United States: Current Population Survey Food Security Supplement (CPS-FSS)*. Washington DC. (accessed Feb 29, 2016).

U.S. Department of Agriculture, Economic Research Service, 2015. *ARRA.* Washington DC. (accessed Feb 29, 2016).

U.S. Department of Agriculture, Food and Nutrition Service. (2017). SNAP data. http://www.fns.usda.gov/pd/supplemental- nutrition-assistance-program-snap.

U.S. Department of Agriculture, Food and Nutrition Service. 2009-2011. *Supplemental Nutrition Assistance Program (SNAP) Cost of Living Adjustment (COLA) Information.* Washington DC. . (accessed Feb 29, 2016).

U.S. Department of Agriculture. 2009-2011. *USDA Food Plans: Cost of Food.* Washington DC. (accessed Feb 29, 2016).

Waehrer, G. and Deb, P. 2012. *Food stamp effects on food preparation and consumption patterns*. Food and Nutrition Assistance Research Program.

Wilde, P., and J. Llobrera. 2009. "Using the Thrifty Food Plan to Assess the Cost of a Nutritious Diet." *Journal of Consumer Affairs* 43(2): 274-304.

Ziliak, J. P. 2006. "Understanding Poverty Rates and Gaps: Concepts, Trends, and Challenges." *Foundations and Trends® in Microeconomics* 1(3): 127-199.

2.8 Figures

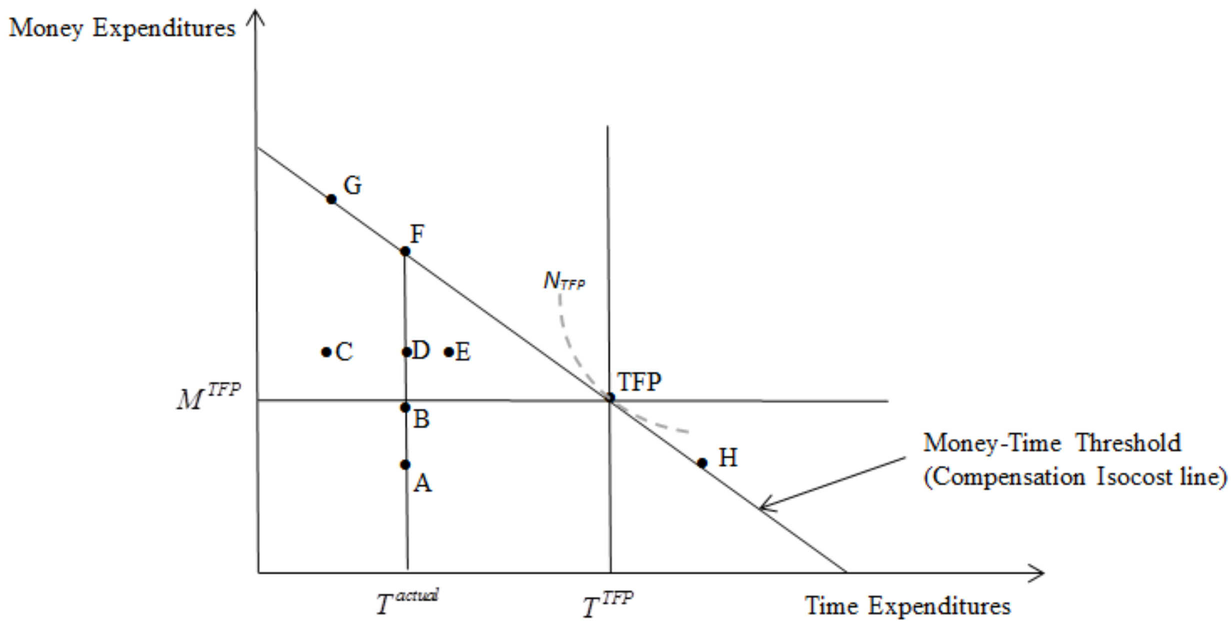


Figure 1. Alternative thresholds and measurement differences

2.9 Tables

(This page is left empty)

Table 1. Summary Statistics for Metric Components

Variables	Units	Mean	Median	S.D	95% Percentile	
Actual money expenditure (M^a)	$/week	109.15	91.00	80.36	9.33	319.58
Money expenditure threshold (M^{TFP})	$/week	114.64	109.10	59.79	40.44	229.49
Money expenditure gap ($M^{TFP} – M^a$)	$/week	5.49	9.10	72.33	–184.71	116.98
Normalized money gap ($1– M^a / M^{TFP}$)		–0.03	0.08	0.71	–2.35	0.89
Actual time expenditure (T^a)	h/week	4.64	4.37	0.84	3.86	7.23
Time expenditure threshold (T^{TFP})	h/week	13.13	13.13	0.00	13.13	13.13
Time expenditure gap ($T^{TFP} – T^a$)	h/week	8.49	8.76	0.84	5.90	9.27
Normalized time gap ($1 – T^a / T^{TFP}$)		0.65	0.67	0.06	0.45	0.71
Actual money expenditure (M^a)	$/week	109.15	91.00	80.36	9.33	319.58
Money-Time expenditure threshold (MT^{TFP})	$/week	213.74	206.93	61.63	122.39	329.48
Money-Time expenditure gap ($MT^{TFP} – M^a$)	$/week	104.59	107.90	73.95	–87.79	219.83
Normalized money-time gap ($1 – M^a / MT^{TFP}$)		0.50	0.55	0.33	0.39	0.95

Note: N = 692

Table 2. FTP Indices with ARRA and Without ARRA Assuming No Household Time Allocation Change Associated with ARRA.[a]

Indices	Expenditure Type	With ARRA	Without ARRA			Percent Poverty Increase Without ARRA			
			MPS=0.17	MPS=0.47	MPS=1	MPS=0.17	MPS=0.47	MPS=1	
Prevalence(P_0)	Money	59.3	60.1	62.4	67.1	1.36%	5.22%	13.19%	*
		(0.03)	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.08)	
	Time	100.0	100.0	100.0	100.0	0.00%	0.00%	0.00%	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Money–Time	92.8	93.4	93.8	94.4	0.61%	1.07%	1.74%	
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	
Depth(P_1)	Money	21.8	23.0	25.3	29.5	5.64%	15.89%	35.12%	**
		(0.02)	(0.02)	(0.02)	(0.02)	(0.11)	(0.12)	(0.13)	
	Time	64.7	64.7	64.7	64.7	0.00%	0.00%	0.00%	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Money–Time	52.7	53.6	55.3	58.4	1.82%	5.05%	10.81%	**
		(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.05)	
Severity(P_2)	Money	11.7	12.7	14.4	18.0	7.88%	22.90%	53.13%	**
		(0.01)	(0.01)	(0.01)	(0.02)	(0.16)	(0.17)	(0.21)	
	Time	42.2	42.2	42.2	42.2	0.00%	0.00%	0.00%	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
	Money–Time	34.0	35.1	37.0	40.6	3.15%	8.87%	19.47%	***
		(0.01)	(0.01)	(0.01)	(0.02)	(0.06)	(0.06)	(0.07)	

a. Supplementary appendix has results assuming time allocation change associated with ARRA.

2.10 Appendix

Measuring Food Expenditure Poverty in SNAP Populations: Some Extensions with an Application to the American Recovery and Reinvestment Act

This appendix provides information on two issues touched on in the chapter but not included due to space considerations: (i) the potential change in the poverty metrics if household time allocation is also allowed to change with the ARRA and (ii) the ordinal ranking of the poverty measures.

2.10.1 Poverty Metrics With and Without the ARRA and a Time Change

With respect to changes in time expenditures, standard economic analysis implies if food and time expenditures are normal inputs, movements along the expansion path due to a decrease in income cost may also lead to a decrease in time devoted to food production. We are aware of only one study where the relationship between SNAP benefit levels and time allocated to food production (an intensive margin) has been examined. Beatty, Nanney, and Tuttle (2014) find that for married households a 1% increase in SNAP benefit levels leads to a 7% decline in minutes in meal preparation time. Alternatively, for single-headed households, which is what we are analyzing, SNAP benefits had no statistically significant effect on any meal preparation or grocery shopping time. However, there are three studies investigating the relationship between SNAP *participation* and time in food production (the extensive margin): Beatty, Nanney, and Tuttle (2014), Roy, Millimet, and Tchernis (2012), and Waehrer and Deb (2012). The common finding is that the direction of the relationship between SNAP and time in food production tends to be household composition and employment dependent. As is well known in discrete/continuous modeling, a variable may not have the same effect sign or magnitude in the extensive and intensive margins. Given the limited evidence on the effect of SNAP benefits on food production time, we first ignore any time adjustment with a change in SNAP benefits and then consider a five percent decrease and increase, so $\Delta T_i = 0, -5\%, +5\%$.

In summary, the poverty index formulas *with* and *without* the ARRA premium are

$$(A.1)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[M_i^{TFP} > (M_i^a + \Delta M_i)\right]\left(\frac{M_i^{TFP} - (M_i^a + \Delta M_i)}{M_i^{TFP}}\right)^\alpha$$

$$(A.2)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[T_i^{TFP} > (T_i^a + \Delta T_i)\right]\left(\frac{T_i^{TFP} - (T_i^a + \Delta T_i)}{T_i^{TFP}}\right)^\alpha$$

$$(A.3)\quad P_\alpha^j = N^{-1}\sum_{i=1}^{N} I\left[MT_i^o > (M_i^a + \Delta M_i)\right]\left(\frac{MT_i^o - (M_i^a + \Delta M_i)}{MT_i^o}\right)^\alpha$$

where $MT_i^o = M_i^{TFP} - p(T_i^a + \Delta T_i - T_i^{TFP})$. The *with* ARRA premium formula occurs when ΔM_i = 0 and ΔT_i = 0 and the j superscript is w (i.e., P_α^w). The *without* ARRA premium formula occurs when $\Delta M_i \neq 0$ and $\Delta T_i \neq 0$ for the values discussed and the j superscript is o (i.e., P_α^o). Prevalence, depth, and severity are associated with α = 0, 1, and 2, respectively.

Table A.1 repeats the analysis presented in table 2 of the main text, but now with two time changes. Only the metrics that change (time only and money-time) are shown. A five percent decline in actual time per week would be on average about 14 minutes per week less in food production. Given that 100 percent of the households were below the time required to be consistent with the TFP before the counterfactual decrease in time, the prevalence rate will not change. However, consistent with the figure and discussion, the depth and severity measures are worse when actual time expenditures decline (2.73% and 5.39% increases, respectively). More importantly, the money-time poverty metrics are greater when time decreases regardless of the MPS value: For example, when MPS = 0.47, the prevalence percent change increased from

1.07% (table 2) to 1.58% (table 3), depth from 5.05% to 6.04%, and severity from 8.87% to 10.34%. Interestingly however the only statistically significant change in depth occurs when the MPS $\geq$ 0.47. This underscores the fact that while the ARRA premium was effective at reducing poverty, the dose was insufficient to make a statistically significant difference when the MPS is less than 0.47.

For completeness, we also include the case where time actually increases by five percent and the results are as expected and are consistent with the figure (e.g., point E): all poverty metrics improve (relative to time decreasing) if more time is spent in food production, even if actual money expenditures decline. This case cannot be ruled out theoretically, but it implies a backward bending expansion path in money-time space. One possible argument for this case could be that lower money expenditures are associated with fewer purchases of more expensive pre-prepared items, thus requiring more labor. This again highlights the need to take into account labor in the poverty analysis, as ignoring this component in this case would tend to overstate the level of food expenditure poverty.

2.10.2 Ordinal Ranking of Poverty Metrics?

An important ordinal finding is that across all measures the percentage change in severity is greater than the percentage change in depth, which is greater than the percentage change in prevalence. This indicates the effectiveness of a policy will be understated if only the prevalence rate is considered, regardless of what threshold is considered. This observation warrants further consideration, especially given Jolliffe, et al. (2005) found similar results in a completely different application (see their table 2). Are these ordinal rankings an analytical result or just an empirical coincidence? Jolliffe et al. (2005) do not explore this question but it is important for drawing policy implications.

First from the FGT poverty index formula in the chapter (i.e. equation 2) the indicator function $I(z_i > y_i)$ only counts those below the poverty threshold, so in the standard case of $y_i > 0$, the normalized gap $(z_i - y_i) \div z_i \in (0, 1)$. Consequently, the weight $[(z_i - y_i) \div z_i]^{\alpha} \in (0, 1)$ is decreasing as α increases. So prevalence is greater than depth, which will be greater than severity (i.e. $P_0 > P_1 > P_2$). Yes, the interpretations are different, but this ordinal ranking is an analytical result.

Next consider the percentage change. All poverty measures will be greater as y_i decreases so $(P_{\alpha}^{w} - P_{\alpha}^{0}) < 0$, where the *w* superscript indicates *with* the *ARRA* and the *o* superscript is *without* the *ARRA*. The following conditions must be satisfied for the percentage change in the higher order measure to be a greater negative number than the lower order measure:

$$(A.4)\quad \frac{(P_{\alpha+1}^{w} - P_{\alpha+1}^{0})}{P_{\alpha+1}^{w}} < \frac{(P_{\alpha}^{w} - P_{\alpha}^{0})}{P_{\alpha}^{w}} \quad \Leftrightarrow \quad \frac{P_{\alpha+1}^{0}}{P_{\alpha+1}^{w}} > \frac{P_{\alpha}^{0}}{P_{\alpha}^{w}} \qquad \alpha = 0,1.$$

If this right hand side condition is satisfied empirically, then the left hand side will be satisfied automatically (i.e. analytically). In our application the right hand side condition is always satisfied and so the percentage change in severity is always greater than the percentage change in depth, which is always greater than the percentage change in prevalence.

2.10.3 Tables

(This page is left empty)

Table A.1. FGT Indices without ARRA---Adjustment on Both Money Expenditure and Time Expenditure

	Indices	Expenditure Type	Without ARRA			Percent Poverty Increase Without ARRA		
			MPC=0.17	MPC=0.47	MPC=1	MPC=0.17	MPC=0.47	MPC=1
	Prevalence(P_0)	Time	100.00	100.00	100.00	0.00%	0.00%	0.00%
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
		Money–Time	93.90	94.30	94.80	1.14%	1.58%	2.19%
			(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
	Depth(P_1)	Time	66.40	66.40	66.40	0.03 ***	0.03 ***	0.03 ***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
ΔT=–5%		Money–Time	54.10	55.80	58.80	2.83%	6.04%	11.73% ***
			(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
	Severity(P_2)	Time	44.50	44.50	44.50	5.39% ***	5.39% ***	5.39% ***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
		Money–Time	35.60	37.50	41.10	4.64%	10.34% *	20.90% ***
			(0.01)	(0.01)	(0.02)	(0.06)	(0.06)	(0.07)

Note: Each index is multiplied by 100. For each index P_i , i=0,1,2, a percent decrease is calculated as $(P_{i,w} - P_{i,w/o})/P_{i,w}$ but when this number is negative it is the percent increase, hence the heading title and the negative sign is dropped. If there is a negative sign this means the poverty measure decreased by this amount. Standard errors are in parentheses. The estimated change is superscripted with asterisks *, **, or *** if the p-value is less than 0.1, 0.05, or 0.001, respectively.

Table A.1. (Cont.) FGT Indices without ARRA---Adjustment on Both Money Expenditure and Time Expenditure

	Indices	Expenditure Type	Without ARRA			Percent Poverty Increase Without ARRA		
			MPC=0.17	MPC=0.47	MPC=1	MPC=0.17	MPC=0.47	MPC=1
	Prevalence(P_0)	Time	100.00	100.00	100.00	0.00%	0.00%	0.00%
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
		Money–Time	92.80	93.50	94.40	0.00%	0.74%	1.74%
			(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
	Depth(P_1)	Time	62.90	62.90	62.90	–2.73% ***	–2.73% ***	–2.73% ***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
ΔT=+5%		Money–Time	53.10	54.80	57.90	0.79%	4.06%	9.87% ***
			(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
	Severity(P_2)	Time	40.00	40.00	40.00	–5.23% ***	–5.23% ***	–5.23% ***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
		Money–Time	34.50	36.50	40.10	1.64%	7.37% *	18.03% ***
			(0.01)	(0.01)	(0.02)	(0.06)	(0.06)	(0.07)

Note: Each index is multiplied by 100. For each index P_i , i=0,1,2, a percent decrease is calculated as $(P_{i,w} - P_{i,w/o})/P_{i,w}$ but when this number is negative it is the percent increase, hence the heading title and the negative sign is dropped. If there is a negative sign this means the poverty measure decreased by this amount. Standard errors are in parentheses. The estimated change is superscripted with asterisks *, **, or *** if the p-value is less than 0.1, 0.05, or 0.001, respectively.

3 Food Acquisitions, the Thrifty Food Plan, and Benefit Adequacy for SNAP Participants

3.1 Background and Motivation

The USDA Thrifty Food Plan (TFP) is an estimate of the minimum food expenditure needed to reach a nutritious diet and it provides guidance for determining the maximum benefits of USDA Supplemental Nutrition Assistance Program (SNAP). The TFP is normally used as a threshold in measuring food expenditure poverty. A common measure of SNAP effectiveness is the ratio of actual household food expenditure to the TFP expenditure threshold (e.g., Davis and You 2011; Horning and Fulkerson 2014; Hoynes, McGranaham, and Schanzenbach 2016; Katare and Kim 2017; Nord 2013; Nord and Prell 2011; Rose 2007; Stewart and Blisard 2006; Tiehen, Newman, and Kirlin 2017; Yang, Davis, and You 2018). This normalized money expenditure is generally greater than one suggesting individuals are reaching the TFP target. Though useful, the normalized money expenditure is limited in several ways (Yang, Davis, and You 2018).

First, the normalized money expenditure focuses on expenditures, not households and consequently does not provide information on the number of households that are below the threshold (prevalence), how far they are below the threshold (depth), and the concentration of those below the threshold (severity). Measures that are more comprehensive can be imported from the general poverty literature for measuring the distance from the TFP.

Second, the denominator of the TFP used in the normalized money expenditure ratio is based on national prices, not the local prices. However, food prices vary spatially and temporally, so the national TFP, and any related poverty metrics using this baseline, may be too high in some areas and too low in others (e.g., Nord and Hopwood 2007; Todd, Leibtag, and Penberthy 2011; Bronchetti, Christensen, and Hansen 2015).

Finally, the numerator in the normalized food expenditure is the total household food expenditure. In the current literature, it is only considered in aggregate. Therefore, there is a paucity in the literature about the contribution of different food expenditure types to reach the TFP. The total food expenditures can be decomposed along several dimensions, and such decompositions can provide useful information in understanding what categories are contributing the most to food expenditures and thus to reducing food expenditure poverty. The only published work we know that has considered this issue is Stewart and Blisard (2006). Stewart and Blisard (2006) considered food at home food expenditures and decomposed those expenditures into different food groups. However, while useful, this is a partial view since total nutrition is based on total intake from Food-At-Home (FAH) and Food-Away-From-Home (FAFH), so decomposing food expenditures into FAH expenditures and FAFH expenditures would be more informative to understanding the degree of importance of FAH in terms of households' food expenditure poverty levels. You, et al. (2009) estimate the TFP threshold needs to be 7% higher if FAFH is included. Furthermore, SNAP benefits are also not intended to cover all FAH expenditures as households are assumed able to spend 30% of their own income on FAH (USDA FNS 2017). Consequently, the decomposition of FAH expenditures into personal-funded-FAH and SNAP-funded-FAH expenditures would reveal the contribution actual SNAP purchases are making to food expenditure poverty measures[5].

[5] The *actual* SNAP-FAH purchases considered in this report should be distinguished from the *hypothetical* 'SNAP benefits + 30% of adjusted income FAH purchases' being considered by Bronchetti, Christensen and Hansen (2015) in their current UKCPR grant. They are answering the question *could* the household reach the TFP with the SNAP benefits + 30% of their adjusted income? We are interested in answering the question: how much *do* actual SNAP-FAH expenditures contribute to reducing the poverty indexes?

3.1.1 Poverty Indexes

To tackle the above three limitations, we use the Foster, Greer, and Thorbecke indexes (FGT poverty indexes, 1984), a well-established poverty measurement in the development literature, to measure food expenditure poverty.

$$P_\alpha = \frac{1}{H}\sum_{h=1}^{H} I(z_h > y_h^{total})(\frac{z_h - y_h^{total}}{z_h})^\alpha \quad (3.1.1)$$

Here y_h^{total} is the variable of interest for household h and z_h is the corresponding threshold. Since our research interest is on food expenditure poverty, the y_h^{total} in this chapter is the household's weekly food expenditure. The z_h is the weekly TFP threshold. The $I(\cdot)$ is the indicator function that equals one if the food expenditure is below the TFP threshold and equals zero, otherwise. The gap between the food expenditure and the threshold is represented by $z_h - y_h^{total}$ with the normalized gap as $\frac{z_h - y_h^{total}}{z_h}$. The total number of households in the population is H. The parameter α defines the poverty index of interest. When $\alpha = 0$, P_0 gives the percentage of households below the poverty threshold – the poverty rate or *prevalence*. When $\alpha = 1$, P_1 gives the per capita household distance from the poverty threshold in percentage terms or *depth*. When $\alpha = 2$, P_2 gives an indication of the degree of skewness in the household per capita distance from the poverty threshold or *severity*.

3.1.2 The Possible Thresholds (z_h)

The TFP threshold (z_h) in the denominator of equation (3.1.1) contains three layers of variation among households: spatial, temporal, and household composition. The national TFP table assumes no spatial price variation (e.g., across regions). However, significant spatial variation in prices has been documented in the literature on food expenditure (Bronchetti, Christensen, and Hansen 2015; Todd, Leibtag, and Penberthy 2011; Nord and Hopwood 2007;

Jolliffe 2006a; 2006b; 2003; Jolliffe, Datt, and Sharma 2004; Jolliffe et al. 2005; Andrews et al. 2001). Failure to account for spatial price variations may generate inaccurate and misleading policy analysis. Furthermore, the national TFP table assumes no price variation between weeks in the same month, or the same year, which ignores temporal price fluctuations. Apart from spatial and temporal variations, the nutrition requirements (quantities) vary by individuals in the households (i.e., household composition), but the national TFP does take into account this variation.

Because of the spatial-temporal structure of the data, there are seven possible z_h = TFP thresholds that could be constructed. This can be denoted by using different subscripts and superscripts: $TFP_{s,t}^{d,e}$. The subscript *s* is the level of disaggregation in the spatial dimension and the maximum value depends on the level of spatial aggregation: N (nation), R(region), D(division), S(state) and L(local). The subscript *t* is the level of disaggregation in the temporal dimension and the maximum value depends on the level of temporal aggregation: Y(year), M(month) and W(week). The superscript *d* indicates the data source, which is either from the Thrifty Food Plan or FoodAPS-GC data (Gundersen et al., 2016). The superscript *e* represents the estimation approach if the FoodAPS-GC data is used (e.g., model or match).

As shown in the following figure 3.1.1, the level of spatial disaggregation increases from left to right. The nation TFP is the spatially most aggregated threshold so is on the left side of the x-axis. The local TFP is the spatially most disaggregated threshold, which can be household-specific depending on the definition of the local neighborhood. The data in FoodAPS is best suited for weekly threshold construction.

(Insert figure 3.1.1 here)

Clearly from equation (3.1.1), as the z_h (= TFP) measure changes, then too could the poverty indexes.

3.2 Total Food Expenditure Decompositions

As mentioned, the decomposition of total food expenditures y_h^{total} in the numerator will provide useful information. The general decomposition is achieved simply by decomposing y_h^{total} into *K* different components ($y_h^{total} = \sum_{k=1}^{K} y_h^k$) and substituting into equation (3.1.1):

$$P_\alpha = \frac{1}{H}\sum_{h=1}^{H} I(z_h > \sum_{k=1}^{K} y_h^k)(\frac{z_h - \sum_{k=1}^{K} y_h^k}{z_h})^\alpha \qquad (3.2.1)$$

The $\sum_{k=1}^{K} y_h^k$ is the total 7-day food expenditure for household *h* aggregated from different expenditure partitions $k, k = 1,2, \dots, K$. An equivalent representation of (3.2.1) is more intuitive and useful.

$$\begin{aligned} P_\alpha &= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > \sum_{k=1}^{K} y_h^k\right)\left(\frac{z_h - \sum_{k=1}^{K} y_h^k}{z_h}\right)^\alpha \\ &= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > \sum_{k=1}^{K} y_h^k\right)\left(1 - \frac{\sum_{k=1}^{K} y_h^k}{z_h}\right)^\alpha \\ &= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > \sum_{k=1}^{K} y_h^k\right)\left(1 - \sum_{k=1}^{K} \frac{y_h^k}{y_h^{total}} * \frac{y_h^{total}}{z_h}\right)^\alpha \\ &= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > \sum_{k=1}^{K} y_h^k\right)\left(1 - \sum_{k=1}^{K} \mathrm{S}_h^k * R_h^{total}\right)^\alpha \qquad (3.2.2) \end{aligned}$$

The $\mathrm{S}_h^k = \frac{y_h^k}{y_h^{total}}$ in equation (*3.2.2*) is the *k* component share of total food expenditures and can be thought of as the weight the *k* group contributes to the total food expenditure. The $R_h^{total} = \frac{y_h^{total}}{z_h}$ is the ratio of total food expenditure to the TFP threshold, which is the normalized money expenditure ratio as currently used in the literature.

Different ways of grouping/partitioning not only can quantify the shares of each component of the total food expenditure but also offer insights regarding the contribution of each component to the food expenditure poverty indexes.[6]

3.2.1 Decomposition by Food-at-Home (FAH) and Food-Away-from-Home (FAFH)

Food expenditure is normally considered in aggregate, combining both expenditures on FAH and FAFH. Yet, nutrition is obtained from both FAFH and FAH. It is important to decompose the total food expenditure into FAFH and FAH components to understand the contribution of each to the food expenditure poverty indexes. The total food expenditure can be written as

$$y_h^{total} = y_h^{FAFH} + y_h^{FAH} \tag{3.2.3}$$

The corresponding FGT food expenditure poverty indexes are decomposed into the following:

$$\begin{aligned} P_\alpha &= \frac{1}{H}\sum_{h=1}^{H} I(z_h > (y_h^{FAFH} + y_h^{FAH}))\left(1 - \frac{y_h^{FAFH}}{y_h^{total}} * \frac{y_h^{total}}{z_h} - \frac{y_h^{FAH}}{y_h^{total}} * \frac{y_h^{total}}{z_h}\right)^\alpha \\ &= \frac{1}{H}\sum_{h=1}^{H} I(z_h > (y_h^{FAFH} + y_h^{FAH}))\left(1 - \mathrm{S}_h^{FAFH} * R_h^{total} - \mathrm{S}_h^{FAH} * R_h^{total}\right)^\alpha \end{aligned} \tag{3.2.4}$$

The R_h^{total} is the ratio of total money expenditure to the TFP threshold, or the normalized money expenditure ratio. The $\mathrm{S}_h^{FAFH} = \frac{y_h^{FAFH}}{y_h^{total}}$ is the FAFH expenditure share and $\mathrm{S}_h^{FAH} = \frac{y_h^{FAH}}{y_h^{total}}$ the FAH expenditure share of total food expenditure, respectively. This decomposition can be used to determine the contribution of FAFH and FAH to reducing food expenditure poverty.

3.2.2 Decomposition of FAH by Funding Source

SNAP is designed as a hunger safety net for inframarginal households. As the biggest welfare program in the United States, it provides benefits to those eligible households to purchase ingredients for food preparation at home (i.e., FAH). However, the SNAP benefit is not designed to cover all FAH expenditures for eligible households. According to the SNAP design, SNAP

[6] Due to the existence of the index function, the contribution of each part is non-linear.

households are supposed to spend 30% of their own net income on food (USDA FNS 2017). Thus it will be useful to further decompose the FAH expenditures into those that are funded from personal funds and those that are funded from SNAP funds.

Let $y_h^{FAH_P}$ denote personal funded FAH expenditures and $y_h^{FAH_S}$ SNAP funded FAH expenditure and so

$$y_h^{total} = y_h^{FAFH} + y_h^{FAH} = y_h^{FAFH} + y_h^{FAH_P} + y_h^{FAH_S} \qquad (3.2.5)$$

The corresponding FGT indexes become:

$$P_\alpha = \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > y_h^{FAFH} + y_h^{FAH_P} + y_h^{FAH_S}\right)\begin{pmatrix} 1 - \frac{y_h^{FAFH}}{y_h^{total}} * \frac{y_h^{total}}{z_h} \\ -\frac{y_h^{FAH_P}}{y_h^{total}} * \frac{y_h^{total}}{z_h} \\ -\frac{y_h^{FAH_S}}{y_h^{total}} * \frac{y_h^{total}}{z_h} \end{pmatrix}^{\alpha}$$

$$= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > y_h^{FAFH} + y_h^{FAH_P} + y_h^{FAH_S}\right)\begin{pmatrix} 1 - \mathrm{S}_h^{FAFH} * R_h^{total} \\ -\mathrm{S}_h^{FAH_P} * R_h^{total} \\ -\mathrm{S}_h^{FAH_S} * R_h^{total} \end{pmatrix}^{\alpha} \qquad (3.2.6)$$

The $\mathrm{S}_h^{FAH_P} = \frac{y_h^{FAH_P}}{y_h^{total}}$ is the share of personally funded FAH expenditures (i.e., household's non-SNAP funds) and $\mathrm{S}_h^{FAH_S} = \frac{y_h^{FAH_S}}{y_h^{total}}$ the share of SNAP funded FAH expenditures of the total expenditures, respectively. This further decomposition of FAH expenditure into personal-funded expenditures and SNAP-funded expenditures allows us to analyze the contribution of SNAP expenditures to reducing the poverty indexes.

3.2.3 Decomposition of FAH by Food Groups

Many food and nutrition-related policies and recommendations are based on food groups. Take *MyPlate* for example, it offers nutrition recommendations in five food groups: fruits, vegetables, grains, protein, and milk. Therefore, decomposition of FAH expenditures into food groups expenditures will enable us to analyze the contribution of each food group to the normalized gap and thus to the food expenditure poverty indexes. This will offer insights for policies related to food groups.

The FAH can be partitioned into G food group expenditures ($y_h^{FAH_1}, y_h^{FAH_2}, \ldots, y_h^{FAH_G}$):

$$y_h^{total} = y_h^{FAFH} + y_h^{FAH} = y_h^{FAFH} + \sum_{g=1}^{G} y_h^{FAH_g}, g = 1,2, \ldots, G \tag{3.2.7}$$

The corresponding FGT indexes become:

$$P_\alpha = \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > (y_h^{FAFH} + \sum_{g=1}^{G} y_h^{FAH_g})\right)\left(1 - \frac{y_h^{FAFH}}{y_h^{total}} * \frac{y_h^{total}}{z_h} - \sum_{g=1}^{G} \frac{y_h^{FAH_g}}{y_h^{total}} * \frac{y_h^{total}}{z_h}\right)^\alpha$$

$$= \frac{1}{H}\sum_{h=1}^{H} I\left(z_h > (y_h^{FAFH} + \sum_{g=1}^{G} y_h^{FAH_g})\right)\left(1 - \mathrm{S}_h^{FAFH} * R_h^{total} - \sum_{g=1}^{G} \mathrm{S}^{FAH_g} * R_h^{total}\right)^\alpha \tag{3.2.8}$$

The $\mathrm{S}^{FAH_g} = \frac{y_h^{FAH_g}}{y_h^{total}}$ is the food group g = 1,2, …,G share of total expenditures. As above, this allows us to construct a measure of the poverty reduction contribution of each FAH food group for each FGT poverty index.

3.2.4 Decomposition by Household Characteristics

One important property of the FGT index is additive decomposability (also called subgroup monotonicity or subgroup consistency) with population share weights (Foster, Greer, and Thorbecke 1984; Foster and Shorrocks 1991; Foster, Greer, and Thorbecke 2010). If the total number of households *H* can be broken down into *M* subgroups, with $H_1, H_2 \ldots, H_M$ in each group ($H = H_1 + H_2 + \cdots + H_M$), the FGT indexes can be decomposed as:

$$P_\alpha = \sum_{m=1}^{M} \frac{H_m}{H} * \; P_\alpha^m \qquad (3.2.9)$$

The $\frac{H_m}{H}$ is the share of households in each subgroup m and the P_α^m is the sub-FGT index calculated for each subgroup m. This decomposition helps to identify the different subgroups contribution to the total food expenditure poverty. Households can be broken down by different subgroups based on household characteristics, such as household labor force participation, household size, adult food security status, spatial categorizations. This chapter will examine all of these decompositions.

3.2.5 Counterfactual Contribution Analysis

Using the above decompositions, we conduct counterfactual contribution analyses to examine the change in the FGT indexes when the expenditure of one component is removed, *ceteris paribus*. This removal will impact the poverty indexes in two parts. The first part is in the indicator function, which simply counts the number of individuals below the TFP threshold, so the removal of one expenditure component will likely result in more households below the TFP threshold. For the second part, the normalized gap, the removal of one expenditure component will result in a larger normalized gap. Both of these effects will, therefore, lead to an *increase* in all FGT poverty indexes (i.e., prevalence, depth, and severity). Therefore, the *without* FGT indexes will be larger than the *with* FGT indexes, so taking the difference between the 'without' and the 'with' will give a simple measure of the contribution to poverty reduction of the component.

3.3 Overview of Datasets Availability

There are three data sources used in the analysis: the USDA's National Household Food Acquisition and Purchase Survey (FoodAPS) (USDA-ERS, 2016), the FoodAPS-Geographic Component (FooodAPS-GC) (Gundersen et al. 2016), and the USDA Food Plans tables (USDA-

CNPP, 2012). In this section, we will introduce the key characteristics of these datasets for a better understanding of the methods used in variable creation in the next section.

3.3.1 FoodAPS

The FoodAPS is the first nationally representative survey of American households to collect data about household food purchases and acquisitions from April 2012 to January 2013. It contains comprehensive information for each food acquisition, including FAH and FAFH, during a 7-day survey period using survey books, interviews, scanning data, and receipts. According to the FoodAPS's user guide, FAH is defined as food and drinks brought home and used to prepare meals for consumption at home or elsewhere. For example, food used to make a sandwich that you bring to work. Alternatively, FAFH is the foods and drinks that are obtained and consumed away from home and prepared foods that are brought home or delivered (e.g., pizza). The main difference between FAH and FAFH is that the former requires *preparation at home* while the latter does not.[7] The food expenditures for FAH and FAFH are documented at both event level and item level.

Besides the food expenditures, the household geographic information, and individual household member's characteristics, such as gender and age, are also collected. Also, the geographic information of the stores where the household FAH acquisitions took place is also recorded. In total, 4,826 national representative households were surveyed and low-income households were over sampled.

[7] See page 10 of FoodAPS User Guide. https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey.aspx

3.3.2 FoodAPS-GC

The FoodAPS-GC data provides information on the local food price for the households (Gundersen et al., 2016). It contains a TFP-like weekly food cost index ($TFP_{q,t}$) for a standard household of four using food prices of store q for calendar week t from January 2012 to January 2013. A standard household of four consists of a couple of age 19-50 and two children of age 6-8 and 9-11. Spatial information of the store is also available in FoodAPS-GC data. This TFP cost index is used in calculating spatial and temporal-specific TFP thresholds.

3.3.3 USDA TFP Food Plan Table

The third data source comes from the national USDA TFP Food Plan Table. The Thrifty Food Plan (TFP) cost is the cheapest food plan to reach the desired nutrition target. Table 3.3.1 shows a typical TFP table. It contains both weekly and monthly costs. To be consistent with the 7-days FoodAPS survey, we focus on the weekly cost. For the analysis of food expenditure poverty, the weekly cost Thrifty Plan (second column) provides our TFP threshold values.

(Insert Table 3. 3.1 here)

The TFP is age-gender-specific, with 15 different age-gender composition groups. It also provides a weekly cost for a standard household of four, which contains the same individual composition as the FoodAPS-GC data described above. The TFP cost of households with a different household size or age-gender composition can be calculated based on the individual TFP cost in the table and then adjusted by the economies of scale factors provided in footnote 3 of the USDA TFP cost table (as appears in table 3.3.1 of this chapter).

Temporally, the June TFP cost is used as the maximum SNAP allotment level across the fiscal year starting on October 1 of each year. The June TFP cost is updated each month by adjusting for monthly food price inflation. In comparison to the annual TFP, the monthly updated TFP accounts for the temporal price variation. As a result, there are two TFP thresholds: the annual TFP based on June TFP table and the monthly updated TFP. However, both TFP thresholds assume the same cost for all regions across the United States.

A more refined TFP threshold would be spatially and temporally sensitive. The purpose of the next section is to construct a spatially and temporally sensitive TFP table, similar to the TFP weekly cost provided by the table 3.3.1. Such a TFP will provide a more accurate threshold for food expenditure poverty research, which in turn will offer more accurate information for policy analysis. To obtain this more refined TFP threshold, there are two potential challenges. (1) How to obtain a spatial and temporal TFP for a standard household of four with given age-gender composition? (2) How to assign this standard spatial and temporal TFP to each household with varying age-gender composition and household size? The following section gives our approach to these two challenges.

3.4 The Estimation of a Spatially and Temporally Sensitive TFP Thresholds

The FoodAPS main data set surveyed 4,826 households in 50 primary sampling units (PSUs) of 27 states. The FoodAPS-GC data, which is supporting data for FoodAPS main data set, provides a nutritious food basket low cost for stores within the 50 PSUs, as well as within counties adjacent to the PSUs. It covers each calendar week in 2012, a total of T = 53 weeks from January 2012 to December 2012. This low basket cost is calculated based on the 10th percentile of the price for each TFP category-like food purchased in the store and the quantity recommended for a standard household of four. There is a TFP-like cost for stores in 35 states for 53 weeks. To be consistent with the FoodAPS main dataset, the TFP cost of the 27 states surveyed in the FoodAPS main data is used, which leads to 229,420 observations for 5,328 stores during 53 calendar weeks. Keep in mind, not every store contains price information for each calendar week, so it is not a balanced panel dataset.

Based on the data availability, we use a modeling (i.e., regression) approach to estimate spatial-temporal-specific specific TFPs. The construction of the TFPs contains two steps:

(1) estimate a spatial-temporal-specific TFP threshold;

(2) assign this TFP threshold to households to obtain the household composition level specific TFP threshold.

3.4.1 A Spatial-Temporal TFP model

In the first step we utilize a regression model with spatial and temporal dummy variables to model the TFP-like cost index of the store q for week t ($TFP_{q,t}$) in the FoodAPS-GC dataset. The spatial dummy variables s =1, 2, …, S capture the spatial disaggregation identifying the location of the store. Different spatial categorizations lead to different values of S. The temporal dummy variables t = 1, 2, …, T capture the temporal category based on the time when the cost index is collected. The estimated spatial- temporal-specific TFPs are for a standard household of four.

Depending on the granularity on the spatial disaggregation, the spatial category can be defined in several ways:

a. S = 4 regions: 1-Northeast, 2-Midwest, 3-South, and 4-West ($s = 1, 2, 3, 4$).
b. S = 9 divisions (1-New England, 2-Middle Atlantic,3- East North Central, 4-West North Central, 5-South Atlantic, 6-East South Central, 7-West South Central, 8-Mountain, and 9-Pacific) ($s = 1, 2, \dots, 9$).
c. S = 27 surveyed states ($s = 1, 2, \dots, 27$).[8]

[8] Due to the disclosure risk, we are not able to release the result for states.

The region and division definitions are based on the Census Bureau's regions and divisions with State FIPS Codes.[9] For the same reason, depending on the granularity on the spatial disaggregation, the time category can be defined as either weekly (T = 53) or monthly (T = 12). The general notation is $TFP_{s,t}^{d}$: the subscript *s* is the level of disaggregation in the spatial dimension: N (nation), R(region), D(division), S(state), and q(store). The subscript *t* is the level of disaggregation at the temporal dimension: Y(year), M(month), and W(week). The superscript *d* indicates the data source, which is either from the USDA reported table or the FoodAPS-GC data.

The FoodAPS-GC data provides a TFP-like cost ($TFP_{q,w}^{GC}$) for a standard household of four using the food prices of store q for calendar week w. However, one of the objectives of this chapter is to determine the differences in the FGT indexes caused by different TFPs used (national TFP vs. spatial-temporal-specific TFPs). Therefore, it is more relevant in this study to understand the deviation of the spatial-temporal-specific TFP from the national TFP and so we use the dependent variable, $\Delta TFP_{q,W}^{GC} \ = (TFP_{q,W}^{GC} - TFP_{N,Y}^{USDA})$, which is a vector of deviations of the TFP cost at the store-week level from the national TFP yearly level.[10] The regional week-specific TFP cost is then estimated from the following model:

---

[9] Source: https://www2.census.gov/geo/docs/maps-data/maps/reg_div.txt

[10] As mentioned, the $TFP_{N,Y}^{USDA}$ is the yearly TFP cost using TFP of June released by USDA for a household of four. USDA updates the SNAP maximal allotment for October 1 to September 30 of next year to account for price inflation based on the TFP cost of June of the current year. For this reason, the $TFP_{N,Y}^{USDA}$ for a calendar week before October 1, 2012 would be the June TFP 2011 constant ($141.20). After October 1, 2012, the $TFP_{N,Y}^{USDA}$ would be the June TFP 2012 constant ($144.90). For the calendar week overlapping October 1, 2012 (i.e., Sep 30 to Oct 6), the day-weighted TFP is used: $\$141.2\times\left(\frac{1}{7}\right) + \$144.9 \times \left(\frac{6}{7}\right)$.

$$\Delta TFP^{GC}_{q,W} = \alpha_0 + \sum_{s=2}^{S} \alpha_s DR_s + \sum_{t=2}^{T} \beta_t DW_t + \sum_{s=2}^{S}\sum_{t=2}^{T} \eta_{s,t} DR_s \cdot DW_t + \varepsilon \tag{3.4.1}$$

The DR_s is a region dummy variable for region $s, s = 1, 2, \ldots, S$. The elements equal one if the store *q* is located in the corresponding region and equals zero otherwise. The DW_t is the calendar week dummy variables for week *t* ($t = 1, 2, \ldots, T$) and equals one if the week belongs to the week *t* and equals zero otherwise. The FoodAPS-GC data contains a total of 53 weeks. The term, $DR_s \cdot DW_t$, is the interaction term of the regional dummy and weekly dummy and captures the regional and week specific price interaction effect. The first calendar week (*t* = 1) and the first region (*s* = 1) will serve as the base. The α_0 is the intercept, α_s the slope parameters for the regional variables, β_t the slope parameters for the calendar week and $\eta_{s,t}$ the slope parameters for interaction between regional dummy and calendar dummy.

The deviation of regional weekly TFP cost ($\widehat{\Delta TFP}^{GC}_{s,t}$) can be estimated based on the estimated parameters from the quation (3.4.1):

$$\widehat{\Delta TFP}^{GC}_{s,t} = \begin{cases} \widehat{\alpha_0}, & s = 1;\ t = 1 \\ \widehat{\alpha_0} + \widehat{\alpha_s}, & s = 2, 3, \ldots., S; t = 1 \\ \widehat{\alpha_0} + \widehat{\beta_t}, & s = 1;\ t = 2, 3, \ldots, T \\ \widehat{\alpha_0} + \widehat{\alpha_s} + \widehat{\beta_t} + \widehat{\eta_{s,t}}, & s = 2, 3, \ldots, S;\ t = 2, 3, .., T \end{cases} \tag{3.4.2}$$

The intercept $\hat{\alpha}_0$ is the estimated ΔTFP variable for week one of region one and is the base. For region *s* for week one, the estimate is $\widehat{\alpha_0} + \widehat{\alpha_s}$. For region one and other week the estimate is $\widehat{\alpha_0} + \widehat{\beta_t}$. The estimate of other *s* and week *t* is $\widehat{\alpha_0} + \widehat{\alpha_s} + \widehat{\beta_t} + \widehat{\eta_{s,t}}$. The national level TFP is added back to the predicted deviation to obtain the regional-week-specific TFP cost $\widehat{TFP}^{GC,model}_{s,t} = \widehat{\Delta TFP}^{GC}_{s,t} + TFP^{USDA}_{N,Y}$. This then gives a 4 $(regions)$×53$(weeks)$ matrix of estimated regional-week-specific TFPs for a household of four.

For simplicity, we only present the case when the surveyed households are divided into four regions ($s = 1, 2, 3, 4, S = 4$) and the temporal category divided by week ($t = 1, 2, \ldots, T$; $T = 53$). The estimation process of divisions, states, or monthly is essentially the same with a slight adjustment in the dimensions.

3.4.2 Assigning the regional week-specific TFP to households

To convert the $\widehat{TFP}_{s,t}^{GC}$ standard household of four estimates (i.e., a couple of age 19-50 and two children of age 6-8 and 9-11) to the appropriate household composition TFP requires two further steps. First, disaggregating the estimated TFP cost for this standard household of four into a region-week age-gender *individual* specific estimates in the household. Second, adding up all the individual estimates in the household and then applying the appropriate economies of scales adjustment according to the footnotes provided by USDA to obtain household level spatial - week-composition TFP cost.

The USDA monthly reported TFP cost table contains the individual TFP cost of 15 specific age-gender compositions along with the TFP cost for a standard household of four for the month. This allows us to estimate weighting parameters ($\boldsymbol{\theta} = (\theta_{t,c}) \in \mathcal{R}^{56\times15}$) of individuals of different age-gender compositions from the standard household of four. For those calendar weeks belonging to the same calendar month, the same weighting parameters are used:

$$\theta_{t,\mathrm{c}} = \frac{TFP_{N,Month}^{c}}{TFP_{N,Month}} \; for\ week\ t \in month\ m; c = 1,2, \ldots ,15 \qquad (3.4.3)$$

The $TFP_{N,Month}^{c}$ is the weekly *individual* TFP cost from the monthly TFP table for the age-gender composition *c*. The $TFP_{N,Month}$ is the weekly standard household of four TFP cost from the monthly USDA TFP table. The $\theta_{t,\mathrm{c}}$ is the weight parameter for composition *c* in week *t*. For example, the calendar week 6 (starting on February 5th and ending on February 11th) belongs to February 2012, and for the household of four the weekly total TFP is \$144.30. For children of one-year-old, the individual weekly TFP cost is \$21.20, which is 0.147 of the total TFP, so the weighting parameter for the first composition (children of one-year-old) is $\theta_{6,1} = 0.147$. The weight parameter θ allows us to disaggregate the region week-specific TFP for a household of four into a region (*s*), week (*t*), and composition (c) specific TFP for the individual using the following equation:

$$\widehat{TFP}_{s,t,c}^{GC} = \widehat{TFP}_{s,t}^{GC} * \theta_{t,c} \; for \; s = 1,2,3,4; t = 1,2,\ldots,56; c = 1,2,\ldots,15 \qquad (3.4.4)$$

Continuing the above example, suppose our estimate of the $\widehat{TFP}_{s,t}^{GC}$ is \$150.00, then for children of one-year-old in calendar week 6, the individual cost estimate would be \$22.05 (i.e., 14.7% of the estimated cost). This procedure provides estimates for all individual age and gender combinations needed to construct the appropriate TFP for various household compositions.[11]

Finally, to get the household aggregate TFP, the individual TFPs described above are added together and then adjusted by the economies of scales factor (φ) based on household size given in the footnote of the USDA Food Plan Tables. The final equation is

---

[11] For calendar week that overlaps two months, a day-weighted method is used. This method weights the weighting parameters for these two months based on the proportion of week days belonging to each month. To match with the FoodAPS main data, a total of 13 USDA TFP monthly tables (from January 2012 to January 2013) are used for a total of 56 calendar weeks. The FoodAPS survey ends on 23 Jan 2013. There are 3 more weeks than calendar weeks provided in the FoodAPS-GC data. For the week of 54, 55 and 56, the estimated $TFP_{s,t}$ for week 1, 2 and 3 are used as proxies respectively, in other words, $TFP_{s,54} = TFP_{s,1}, TFP_{s,55} = TFP_{s,2}, TFP_{s,56} = TFP_{s,3}$. One thing to note is that the week defined in the region-week-composition-specific TFP is a calendar week starting on Sunday. However, a household survey week is a 7-day period with varying starting day. For a household with a survey week not starting on Sunday, we again use a day-weighted $TFP_{s,w,c}$ of the overlapping calendar weeks. For example, a survey starting on April 2nd and ending on April 8th 2012 has six days belongs to the 14th calendar week and one day belongs to the 15th calendar week. The $TFP_{r,w,c}$ used to assign to individuals in this household is: $TFP_{s,t,c}^{i} = TFP_{s,14,c} \times \left(\frac{6}{7}\right) + TFP_{s,15,c} \times \left(\frac{1}{7}\right)$. So $TFP_{s,t,c}^{i}$ means TFP threshold for individual *i* with age-gender composition c from household surveyed at region *s*.

$$TFP_r^h = \left(\sum_{c\in\boldsymbol{h}} TFP_{s,t,c}\right) * \varphi \tag{3.4.5}$$

The economies of scale factor is provided by the Center of Nutrition and Policy and Promotion(CNPP) USDA. It is 1.20 (120%) for household size 1; 1.10 (110%) for size 2; (1.05) 105% for size 3; 1.00 (100%) for size 4; 0.95 (95%) for size 5 and 6; 0.90 (90%) for size 7 and above.

To illustrate how spatial and temporal price variations impact the FGT poverty measures, in the following analysis, we focus on $TFP_{N,Y}^{USDA}$ and $TFP_{R,W}^{GC,Model}$, with $TFP_{N,Y}^{USDA}$ serving as the benchmark and $TFP_{R,W}^{GC,Model}$ containing both spatial and temporal variations. Based on the analysis, these two estimates represent the extremes of the seven we considered. The $TFP_{N,Y}^{USDA}$ gives the lowest estimates and $TFP_{R,W}^{GC,Model}$ the highest, so provide the lower and upper boundaries for our analysis.

3.5 Food Expenditure Decomposition

The FoodAPS data contains food expenditures for FAH and FAFH, both at the item level and event level. This section discusses how the household total FAH and FAFH expenditures are constructed, as well as the funding source and food group decompositions.

3.5.1 Food-at-Home (FAH)

The FoodAPS contains FAH expenditure from both the event level and item level. An event indicates expenditures for a specific visit to a specific location. However, the expenditure information is aggregated at the event level, thus contains the cost of non-food items and bottle deposits. There is no way to separate those. Therefore, we use the expenditure information at the item level instead, which are then aggregated by event ID into event level food expenditure and by household ID into household level FAH expenditure. After dropping households with missing expenditures and data abnormalities, there are 3,656 'clean' households, 75.76% of the full household sample, with complete FAH expenditure. [12] The detailed decision tree on data construction of FAH can be found in Appendix I. One thing worth mentioning is that the item expenditure does not include state and local food sales tax. We collected the food sales tax rates for the FoodAPS survey states for 2012 and applied the tax rates to the aggregated event level non-SNAP-funded expenditures we generated from item level, as SNAP expenditures are not subject to taxes.[13]

3.5.2 Food-Away-From-Home (FAFH)

The FoodAPS contains FAFH expenditure from both the item level and event level. The total FAFH expenditure can be obtained from the event level and we can single out guest expenditures. However, we are not able to conduct food group break downs for FAFH. The main reason is that a combo meal cost at the item level data only reflects the main meal component cost. For example, a combo meal may contain chicken strips, cole slaw, hushpuppy, and diet soda. The cost of chicken strips and diet soda is reported, but not the cost of cole slaw and hushpuppy. This leads to a biased estimation of the FAFH expenditure by food groups. The household level cost is the summation of total FAFH event cost by household. The detailed decision tree on data construction of FAFH can be found in Appendix II.

After dropping households with missing FAFH expenditure and data abnormalities, there are 3,549 households. These 'clean' households constitute 73.54% of the full household. Based on the data availability of TFP thresholds, FAH and FAFH expenditures, and other demographic

[12] Here the clean household means households with no missing information,

[13] The expenditure of food paid by SNAP is not taxable, so the sales tax only applies to the non-SNAP expenditure, that is the personal expenditure ($y_h^{FAH_P}$). https://www.fns.usda.gov/snap/retailer-sales-tax-notice

variables (i.e., age and gender) , we are left with 3,380 households with 956 SNAP participants and 2,424 Non-SNAP participants.

3.6 Results

We first present and discuss the aggregate results by considering the differences in the poverty measures by using $TFP_{N,Y}^{USDA}$ vs $TFP_{R,W}^{GC}$ before decomposing the expenditures by funding type or food group. We then proceed to the decomposition analysis by first looking at the contribution of FAH and FAFH to the poverty measures using the regional estimated $TFP_{R,W}^{GC}$. We then further break down the FAH funding source into personal and SNAP. We conclude by decomposing the effects of different food groups on the poverty measures, again using the regional estimated $TFP_{R,W}^{GC}$. Based on the household characteristics, results are presented by the percentage of work-eligible individuals in the household employed (labor force participation rate), household size, adult food security status, and other spatial sub-groups. To highlight the spatial impact on FGT indexes, we focus on the regions sub-groups. All SNAP households are grouped into four regions: Northeast, Midwest, South, and West. In the conclusions, we will describe general patterns for other partitions.

3.6.1 National Versus Regional TFP Thresholds and Poverty

Table 3.6.1 presents the summary statistics of the food expenditures, the two TFP thresholds ($TFP_{N,Y}^{USDA}$and $TFP_{R,W}^{GC}$) and the FGT indexes based on these two TFP thresholds for SNAP households by various data partitions. We also give the percentage change in the FGT indexes between these two TFP thresholds to determine how sensitive these poverty indexes are to national versus local based TFP estimates (i.e., objective one). All numbers are weighted by the Jackknife replication method.

(Insert Table 3.6.1 here)

The first row reports the FGT indexes based on all SNAP households. The mean of the total weekly food expenditure is \$109.09. The national yearly TFP threshold is \$100.35, indicating households are spending more than the required TFP cost on average. However, the regional weekly TFP threshold is \$111.95, showing households are spending *less* than the required TFP cost on average when the regional variation is considered. [14] When the national TFP is used the prevalence rate (i.e., the number of individuals below the threshold) is 50% (P_0 = 0.50) but this increases to 55% when the regional TFP is used. The depth and severity measures show a similar pattern of worsening poverty when the regional TFP is used. For example, using the national TFP, the average household is 25% below the threshold (P_1 = 0.25) but this increases to 28% when the regional TFP is used. The last three columns show all the FGT indexes increase by over 9% when moving from the national TFP to the regional TFP.

The total food expenditures show geographic differences in food spending. The households in the West region spend the most per week (\$137.15) on food while households in the Midwest spend the least (\$101.11). The TFP threshold and FGT indexes also vary across the regions with a similar pattern to the expenditure. Based on the national yearly TFP, the Midwest contains the most food expenditure poor households, with the largest food expenditure poverty prevalence of 54% (P_0 = 0.54). The deprived households in the West are the largest distance below the threshold, 25% (depth, P_1 = 0.25). The most severe food poverty situation is in the South region (with the largest severity score of P_1 = 0.16). However, based on the regional weekly-specific TFP, the households in the West region show the highest degree of food expenditure poverty: with the highest scores in all three FGT indexes (prevalence of 0.60, depth of 0.30 and severity of 0.19). The percentage change between the national yearly TFP and the regional weekly TFP

[14] The national yearly TFP and the regional weekly TFP are not based on the same stores or sampling frame.

confirms the largest difference is for the West region (the difference ranged from 15.65% to 24.31% for West). This also shows the FGT indexes calculated from national yearly TFP distorts the West region food expenditure poverty the most.

3.6.2 Contribution to Poverty Reduction by Food Source

Table 3.6.2 reports the summary statistics for the food expenditures and FGT indexes based on $TFP_{R,W}^{GC}$ by FAFH and FAH decomposition.

(Insert Table 3.6.2 here)

Based on the first row, the share of FAH is twice as large as the share of FAFH (0.68 vs. 0.32) confirming that for SNAP households FAH is the largest source of total food expenditure, which in turn will have the largest impact on the FGT indexes. The ratio of total food expenditure to the TFP threshold is 1.09; indicating households are spending 9% more than the TFP cost on average. However, as the FGT indexes in table 3.6.1 shows, there are still deprived households suffering from food expenditure poverty.

Table 3.6.2 also calculates the FGT indexes by removing FAFH and FAH, respectively. Looking back at table 3.6.1 we know the FGT index values include *all* sources, so taking the difference in the table 3.6.2 and table 3.6.1 measures can be interpreted as the contribution to the *reduction* in food expenditure poverty associated with a particular food source. For example, in row one in table 3.6.2, without FAFH the prevalence rate is 73% (P_0 = 0.73), however in table 3.6.1 with FAH the prevalent rate is 55% (P_0 = 0.55). Thus, the simple interpretation here is that adding FAFH spending reduces the prevalence rate by almost 20%. In terms of depth and severity, comparing table 3.6.1 (with) and table 3.6.2 (without) for FAFH indicates that FAFH gets the individuals on average 15% closer (depth reduction) to the TFP threshold (i.e., 0.42 – 0.27 = 0.15) and severity is decreased by 0.13. Similar logic applies to the break down by regions, where FAFH contributes the most to reducing food expenditure poverty in the South, which has the largest FAFH share.

Given the FAH focus of the SNAP, there is probably interesting to look at FAH. Including FAH, the prevalence rate (P_0) rate reduces from 95% (table 3.6.2) to 55% (table 3.6.1), a difference of 40%, which is not too surprising given that FAH makes up over 2/3 of food expenditures. Furthermore, FAH gets the average household 47% closer to the TFP threshold (i.e. P_1 = 0.74 in table 3.6.2 minus P_1 = 0.27 in table 3.6.1) and reduces the severity by 0.46 (i.e. P_2 = 0.63 in table 3.6.2 minus P_1 = 0.17 in table 3.6.1).

The decomposition by regions shows different poverty reduction patterns, depending on the measure, and this highlights the importance of distinguishing prevalence from the depth from severity. The exact same dollar increase in spending in two regions may have very different impacts on the respective prevalence rates, depths, and the severities, depending on how far the households in each region are from the TFP threshold. The prevalence rate is affected most in the South by removing FAH as it increases from 53% (table 3.6.1) to 96% (table 3.6.2) or FAH reduces the prevalence in the South by 43%. Alternatively, depth and severity are impacted most in the Midwest where FAH gets the average household 51% closer to the TFP threshold (i.e. P_1 = 0.76 in table 3.6.2 minus P_1 = 0.25 in table 3.6.1) and reduces the severity by 0.50 (i.e. P_2 = 0.65 in table 3.6.2 minus P_1 = 0.15 in table 3.6.1).

3.6.3 Contribution to Poverty Reduction by FAH Funding Source

Table 3.6.3 presents summary statistics for the food expenditures and FGT indexes based on $TFP_{R,W}^{GC,M}$ by FAFH, personal funded FAH (y^{FAH_P}) and SNAP funded FAH (y^{FAH_S}) decomposition. Compared to table 3.6.2, the only difference is that the FAH is further decomposed by the funding source.

(Insert Table 3.6.3 here)

Table 3.6.3 shows the total food expenditure shares of FAFH, personal FAH, and SNAP FAH. From the second row for all SNAP households, personal FAH expenditures and SNAP FAH expenditures account for 31% and 37% of total food expenditures, respectively. In terms of the four regions, the Northeast has the greatest share from personal spending at 39% whereas the Midwest has the greatest share from SNAP spending at 44%. The other regions have much more balanced shares between personal and SNAP in the 30% t to 35% range.

Similar to table 3.6.2, table 3.6.3 calculate the FGT indexes by removing personal FAH and SNAP FAH expenditures. Including personal FAH expenditures, the prevalence rate (P_0) rate reduces from 71% (table 3.6.3) to 55% (table 3.6.1), a difference of 16%. The average household gets 17% closer to the TFP threshold via personal FAH expenditures (i.e. $P_1 = 0.44$ in table 3.6.3 minus $P_1 = 0.27$ in table 3.6.1) and severity is reduced by 0.15 (i.e. $P_2 = 0.32$ in table 3.6.3 minus $P_1 = 0.17$ in table 3.6.1). The region decomposition shows that personal FAH expenditures have the largest poverty reduction in the Northeast and the least in the Midwest.

Turning to the SNAP FAH expenditure contributions, the prevalence rate (P_0) rate is reduced by 25%, from 80% (table 3.6.3) to 55% (table 3.6.1). Furthermore, SNAP FAH expenditures get the average household 25% closer to the TFP threshold (i.e. $P_1 = 0.$ in table 3.6.3 minus $P_1 =$ 0.27 in table 3.6.1) and reduces the severity by 0.23 (i.e. $P_2 = 0.63$ in table 3.6.3 minus $P_1 = 0.17$ in table 3.6.3). The SNAP FAH expenditure effects differ by region and measure again underscoring why it is important to consider prevalence, depth, and severity. Generally speaking, SNAP FAH expenditures tend to decrease the poverty measures in the Midwest and South the most and the Northeast the least, which is not surprising and opposite of the personal FAH expenditures.

3.6.4 Contribution to Poverty Reduction by Food Groups

Table 3.6.4 reports the summary statistics of the TFP thresholds, expenditures, and FGT poverty indexes for SNAP households by food groups based on $TFP_{R,W}^{GC,M}$ for the four regions: Northeast, Midwest, South, and West. The 10 food groups are 1) milk and dairy, 2) protein, 3) mixed dishes, 4) grains 5) snacks and sweets, 6) fruit and vegetables, 7) beverages, 8) fats and oils, condiments and sugars 9) infant formula and body food, or not in a category, 10) food code not assigned. Here we focus our discussion on four food groups

1. Milk and Dairy,
2. Protein,
4. Grains,
6. Fruits and Vegetables.

As a percentage of total food expenditures, the average shares are milk and dairy 6%, protein 17%, grains 6%, and fruits and vegetables 9%. Dividing these numbers by the FAH expenditure share will give an estimate of the share of the FAH of these foods: milk and dairy 9%, protein 25%, grains 9% and fruits and vegetables 13%. The same decomposition approach used for looking at food sources and funding sources in tables 3.6.2 and 3.6.3 can be used here as well. From the relatively small shares of total food expenditures, it is perhaps no surprise that the contribution of any specific food group to food expenditure poverty reduction is rather small. Here we will only discuss the ones that are over a difference of 0.05. In the milk and dairy category, there is only one region and one poverty measure that changes by more than 0.05 by adding milk and dairy expenditures and that is for the West where the prevalence rate decreases by 6% when milk and dairy expenditures are added. With protein constituting the largest share within the food groups, it is perhaps no surprise then that protein has the largest effect. For the total (across all regions), protein expenditures reduce the prevalence rate and the depth rate by

14% and 6% respectively. And in terms of the regions, the Midwest, South, and West all show decreases in prevalence and depth greater than 5% when protein is added. Adding fruit and vegetable expenditures only decreases the prevalence rates in the Midwest and West over 5% (6%).

3.7 Conclusions

As the least expensive food plan for healthy eating calculated by USDA, the TFP is normally considered the minimal food expenditure threshold required to reach a nutritious diet. In the analysis of food expenditure poverty for SNAP households, one common question is whether SNAP households spend enough to reach the TFP target. Falling short of this food expenditure target is known as food expenditure poverty (Yang, Davis, and You 2018). We used the Foster, Greer, and Thorbecke (FGT: 1984) indexes to give a comprehensive picture of food expenditure poverty based on the FoodAPS data that includes prevalence, depth, and severity of food expenditure poverty.

One concerning issue on the TFP and thus the poverty indexes is that the TFP is a national estimate but food costs may vary over space and time. Using the FoodAPS and FoodAPS-GC data we estimate weekly regional TFPs that are then applied to each household in the respective regions and survey weeks and note the changes in the poverty indexes. Also, we considered various food expenditure decompositions based on FAFH vs. FAH, FAFH vs. SNAP funded FAH vs. Personal funded FAH, and FAH food groups. This allowed us to examine the contribution of each expenditure component to the expenditure gaps away from the TFP threshold and also the contribution to the FGT poverty indexes. Based on the additive decomposability property of the FGT indexes, the SNAP households are also portioned into different sub-groups to analyze the sub-groups' poverty impact on total poverty indexes.

The results show that the USDA based national annual TFP threshold is lower than the regional weekly estimates here based on the FoodAPS data. The greatest difference was for the West region, which has the highest TFP estimate. These higher estimates, in turn, meant that all of the food expenditure poverty indexes were higher with the regional weekly estimates than with the national annual USDA TFP threshold. Simply stated, food expenditure poverty is underestimated when one uses the nation annual USDA TFP estimate.

The decomposition of food expenditure shows that FAH contributes the most to reducing food expenditure poverty overall and SNAP benefits plays a much more important role in reducing food expenditure poverty that personal FAH expenditures, but the contribution of SNAP benefit to reducing food expenditure poverty varies across regions. The food group decomposition results show spending on protein is the most significant source in alleviating food expenditure poverty. Finally, the household partitions by regions show large heterogeneity of poverty indexes across regions, with the West region showing the most severe poverty situation mainly due to higher regional temporal-specific TFP threshold.

3.8 Reference

Andrews, M., Kantor, L.-S., Lino, M., Ripplinger, D., 2001. Using USDA's Thrifty Food Plan to assess food availability and affordability. FoodReview 24 (2), 45–53.

Bronchetti, E.T., G. Christensen, and B. Hansen. Variation in Food Prices and SNAP Adequacy for Purchasing the Thrifty Food Plan. Presentation made at the 37th Annual Conference for the Association of Public Policy Analysis. November 14, 2015. Miami, FL.

Davis, G.C., and W. You. 2011. "Not enough money or not enough time to satisfy the Thrifty Food Plan? A Cost difference approach for estimating a money–time threshold." Food Policy 36(2):101-107.

Foster, J., Greer, J. and Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica: Journal of the Econometric Society, pp.761-766.

Foster, J., and A. Shorrocks. 1991. "Subgroup Consistent Poverty Indexes," Econometrica 59(1991):687-709.

Foster, J., Greer, J. and Thorbecke, E., 2010. The Foster–Greer–Thorbecke (FGT) poverty measures: 25 years later. The Journal of Economic Inequality,8(4), pp.491-524.

Gundersen, C., Baylis, K., Fan, L., House, L. and Dutko, P., 2016. Construction of Weekly Store-Level Food Basket Costs Documentation May 2016.

Hoynes, H., L. McGranaham, and D. Schanzenbach. 2016. "SNAP and Food Consumption" in SNAP Matters: How Food Stamps Affect Health and Well Being, Judith Bartfield, Craig Gundersen, Timothy M. Smeeding and James Ziliak (eds), Stanford University Press. Stanford, CA.

Jolliffe, D., 2003. Comparisons of metropolitan-nonmetropolitan poverty during the 1990s (No. 96). Washington, DC: US Department of Agriculture, Economic Research Service.

Jolliffe, D., 2006a. The cost of living and the geographic distribution of poverty. US Department of Agriculture, Economic Research Service.

Jolliffe, D., 2006b. Poverty, prices, and place: How sensitive is the spatial distribution of poverty to cost of living adjustments?. Economic Inquiry,44(2), pp.296-310.

Jolliffe, D., Datt, G. and Sharma, M., 2004. Robust poverty and inequality measurement in Egypt: correcting for spatial-price variation and sample design effects. Review of Development Economics, 8(4), pp.557-572.

Jolliffe, D., Gundersen, C., Tiehen, L. and Winicki, J., 2005. Food stamp benefits and child poverty. American Journal of Agricultural Economics,87(3), pp.569-581.

Katare, B. and J. Kim. 2017. “Effects of the 2013 SNAP Benefit Cut on Food Security.” Applied Economic Perspectives and Policy: ppx025.

Nord, M. 2013. “Effects of the Decline in the Real Value of SNAP Benefits from 2009 to 2011.” Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No.151.

Nord, Mark, and Heather Hopwood. 2007. “Higher Cost of Food in Some Areas May Affect Food Stamp Households’ Ability to Make Healthy Food Choices,” In Can Food Stamps Do More to Improve Food Choices? An Economic Perspective, J. Guthrie, B. Lin, M. Ver Ploeg, and E. Frazao, eds., Economic Research Services, U.S. Department of Agriculture.

Nord, M., and M.A. Prell. 2011. Food Security Improved Following the 2009 ARRA Increase in SNAP Benefits. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No.116.

Rose, D. 2007. "Food Stamps, The Thrifty Food Plan, and Meal Preparation: The Importance of The Time Dimension For US Nutrition Policy." Journal of Nutrition Education and Behavior, 39(4):226-232.

Stewart, H., and N. Blisard. 2006. "The Thrifty Food Plan and Low-Income Households in the United States: What Food Groups are Being Neglected?" Food Policy 31(5):469-482.

Supplemental Nutrition Assistance Program. Title 7 U.S. Code, Sec. 2011. http://uscode.house.gov/browse/prelim@title7/chapter51&edition=prelim

Tiehen, L., C. Newman, and J. Kirlin. 2017. "The Food-Spending Patterns of Households Participating in the Supplemental Nutrition Assistance Program: Findings From USDA's FoodAPS." USDA Economic Research Service. Economic Information Bulletin Number 176. August.

Todd, Jessica, Ephraim Leibtag, and Courtney Penberthy. 2011. "Geographic Differences in the Relative Price of Healthy Foods." Economic Information Bulletin 78, U.S. Department of Agriculture, Economic Research Service.

U.S. Department of Agriculture. 20012-2013. USDA Food Plans: Cost of Food. Washington DC. http://www.cnpp.usda.gov/USDAFoodPlansCostofFood/reports (accessed October 12, 2017).

U.S. Department of Agriculture (USDA), Center for Nutrition Policy and Promotion (CNPP). 2012. "Official USDS Food Plans: Cost of Food at Home at Four Levels, U.S. Average, June 2012."

U.S. Department of Agriculture (USDA), Economic Research Service (ERS). 2016. "National Household Food Acquisition and Purchase Survey (FoodAPS): User's Guide to Survey Design, Data Collection, and Overview of Datasets."

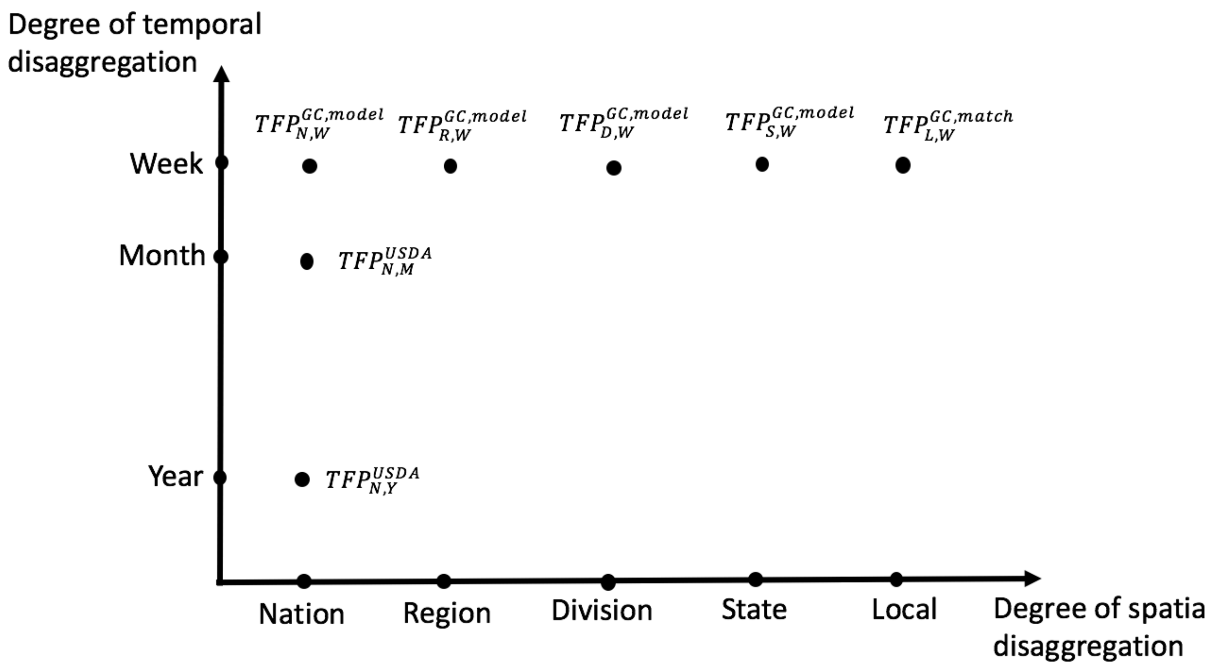
USDA FNS "Supplemental Nutrition Assistance Program (SNAP): Fact Sheet on Resources, Income, and Benefits" Assessed on Oct 2017: https://www.fns.usda.gov/snap/fact-sheet-resources-income-and-benefits.

Yang, Y., Davis, G.C. and You, W., 2018. "Measuring Food Expenditure Poverty in SNAP Populations: Some Extensions with an Application to the American Recovery and Reinvestment Act." Applied Economic Perspectives and Policy, 41(1), pp.133-152.

You, W., G. Zhang, B.M. Davy, A. Carlson, B.H. Lin. "Food Consumed Away from Home Can Be a Part of a Healthy and Affordable Diet." Journal of Nutrition. 139(2009): 1994-1999. doi: 10.3945/jn.109.107615.

3.9 Figures

Figure 3.1.1: TFP Thresholds at Different Disaggregation Level from Spatial and Temporal Dimensions.



Note: $TFP_{s,t}^{d,e}$ represents TFPs under different meaning: the subscript *s* is the level of disaggregation at the spatial dimension, takes the value of N (nation), R(region), D(division), S(state) and L(local). The subscript *t* is the level of disaggregation at the temporal dimension, takes the value of Y(year), M(month), and W(week). The superscript *d* indicates the data source, which is either from the USDA reported table or the FoodAPS-GC data. The superscript *e* represents the estimation approach if the FoodAPS-GC data is used. For example, $TFP_{R,W}^{GC,model}$ means the regional-week specific TFP estimated using modeling based on FoodAPS-GC dataset. We used two different approaches (model and match) to estimate TFP thresholds at different spatial disaggregation level.

3.10 Tables

Table 3.3.1: Official USDA Food Plans: Cost of Food at Home at Four Levels, U.S. Average, June 2012

Age-gender groups	Weekly cost[2]				Monthly cost[2]			
	Thrifty plan	Low-cost plan	Moderate-cost plan	Liberal plan	Thrifty plan	Low-cost plan	Moderate-cost plan	Liberal plan
Individuals[3]								
Child:								
1 year	21.30	28.30	32.10	39.10	92.20	122.50	139.00	169.40
2-3 years	23.30	29.30	35.60	43.30	100.90	127.10	154.10	187.60
4-5 years	24.20	30.60	37.70	46.00	104.80	132.40	163.30	199.30
6-8 years	30.90	42.50	51.50	60.80	133.90	184.30	223.00	263.40
9-11 years	35.00	46.40	59.60	69.50	151.80	201.00	258.10	301.10
Male:								
12-13 years	37.80	53.40	66.80	78.10	163.60	231.20	289.50	338.30
14-18 years	38.90	54.50	68.90	78.80	168.50	236.20	298.60	341.60
19-50 years	41.80	54.00	67.60	83.20	181.10	234.00	292.80	360.50
51-70 years	38.20	50.90	62.80	76.00	165.70	220.60	272.30	329.10
71+ years	38.50	50.70	62.40	77.20	166.90	219.80	270.40	334.40
Female:								
12-13 years	37.80	46.10	55.00	67.40	163.90	199.90	238.40	292.00
14-18 years	37.20	46.30	56.10	69.00	161.40	200.50	243.30	299.10
19-50 years	37.20	46.90	58.00	74.00	161.10	203.00	251.40	320.60
51-70 years	36.70	45.80	56.80	68.00	159.00	198.30	246.00	294.70
71+ years	35.70	45.50	56.20	67.60	154.90	197.30	243.50	292.90
Families								
Family of 2:[4]								
19-50 years	86.90	110.90	138.10	172.90	376.40	480.70	598.60	749.20
51-70 years	82.40	106.30	131.60	158.40	357.20	460.80	570.10	686.20
Family of 4:								
Couple, 19-50 years and children—								
2-3 and 4-5 years	126.50	160.70	198.80	246.50	547.90	696.50	861.60	1068.00
6-8 and 9-11 years	144.90	189.80	236.60	287.50	627.90	822.30	1025.30	1245.60

[1]The Food Plans represent a nutritious diet at four different cost levels. The nutritional bases of the Food Plans are the 1997-2005 Dietary Reference Intakes, 2005 Dietary Guidelines for Americans, and 2005 MyPyramid food intake recommendations. In addition to cost, differences among plans are in specific foods and quantities of foods. Another basis of the Food Plans is that all meals and snacks are prepared at home. For specific foods and quantities of foods in the Food Plans, see *Thrifty Food Plan, 2006* (2007) and *The Low-Cost, Moderate-Cost, and Liberal Food Plans, 2007* (2007). All four Food Plans are based on 2001-02 data and updated to current dollars by using the Consumer Price Index for specific food items.
[2]All costs are rounded to nearest 10 cents.
[3]The costs given are for individuals in 4-person families. For individuals in other size families, the following adjustments are suggested: 1-person—add 20 percent; 2-person—add 10 percent; 3-person—add 5 percent; 4-person—no adjustment; 5- or 6-person—subtract 5 percent; 7- (or more) person—subtract 10 percent. To calculate overall household food costs, (1) adjust food costs for each person in household and then (2) sum these adjusted food costs.
[4]Ten percent added for family size adjustment.

Table 3.6.1. Summary Statistics of Expenditures, TFP Thresholds, and FGT Indices

			TFP Threshold		Expenditure	FGT Indices by TFP_{NY}^{USDA}			FGT Indices by TFP_{RW}^{GC}			% Change		
		N	TFP_{NY}^{USDA}	TFP_{RW}^{GC}	y^{total}	P_0	P_1	P_2	P_0	P_1	P_2	P_0	P_1	P_2
All		**956**	**100.35**	**111.95**	**109.09**	**0.50**	**0.24**	**0.15**	**0.55**	**0.27**	**0.17**	**9.12%**	**12.75%**	**13.31%**
4 Regions	Northeast	158	96.22	106.08	113.15	0.50	0.23	0.15	0.52	0.26	0.17	4.78%	11.28%	10.49%
	Midwest	173	95.86	105.22	101.11	0.54	0.22	0.13	0.57	0.25	0.15	5.57%	13.19%	14.06%
	South	416	97.91	106.81	103.37	0.48	0.24	0.16	0.53	0.27	0.18	10.68%	9.30%	9.77%
	West	209	118.96	143.70	137.15	0.52	0.25	0.15	0.60	0.30	0.19	15.65%	22.39%	24.31%
Regional CBSA	In CBSA	871	103.56	115.76	110.95	0.51	0.24	0.15	0.56	0.27	0.18	9.91%	12.65%	13.47%
	Not CBSA	85	80.70	88.66	97.75	0.47	0.20	0.12	0.49	0.22	0.14	3.89%	13.55%	12.06%
Regional Rural	Urban	689	103.75	116.23	109.49	0.54	0.26	0.17	0.58	0.30	0.19	8.62%	11.85%	13.27%
	Rural	267	92.70	102.32	108.21	0.43	0.17	0.11	0.47	0.20	0.12	10.56%	15.86%	13.45%
Labor Force Participation Rate	(0%-25%)	439	89.53	99.60	92.66	0.52	0.25	0.16	0.55	0.28	0.18	4.88%	11.74%	12.75%
	(25%-50%)	332	129.33	143.79	143.07	0.50	0.24	0.16	0.58	0.27	0.18	15.58%	12.84%	12.27%
	(50%-75%)	80	138.99	157.77	130.57	0.68	0.22	0.11	0.74	0.28	0.14	8.51%	26.32%	30.07%
	(75%-100%)	105	68.11	76.60	91.98	0.38	0.20	0.13	0.43	0.22	0.15	12.63%	10.50%	11.95%
Household Size	1	177	44.69	49.22	67.11	0.43	0.19	0.11	0.45	0.21	0.13	4.86%	12.90%	14.66%
	2	205	80.34	88.83	105.02	0.40	0.22	0.15	0.47	0.24	0.16	18.86%	10.25%	10.84%
	3	175	110.98	124.55	108.48	0.59	0.26	0.16	0.66	0.30	0.18	12.37%	14.75%	15.43%
	4	182	138.06	155.11	139.70	0.58	0.29	0.21	0.63	0.33	0.23	7.73%	12.23%	9.99%
	5	113	157.96	177.70	166.46	0.57	0.23	0.13	0.62	0.26	0.15	8.53%	17.27%	20.39%
	6	56	195.54	218.12	168.19	0.66	0.29	0.16	0.70	0.33	0.19	7.18%	13.21%	18.16%
	>=7	28	216.96	239.37	154.17	0.65	0.38	0.27	0.68	0.41	0.30	4.49%	7.76%	8.44%
Adult Food Security Status	High Food Security	300	101.28	113.37	115.67	0.43	0.19	0.12	0.48	0.22	0.14	11.36%	14.62%	14.82%
	Marginal Food Security	241	109.05	121.28	113.98	0.53	0.24	0.15	0.59	0.27	0.17	11.27%	14.02%	13.82%
	Low Food Security	249	100.41	112.10	104.37	0.60	0.28	0.18	0.62	0.31	0.20	4.12%	11.85%	12.84%
	Very Low Food Security	166	88.67	98.56	97.70	0.49	0.26	0.17	0.55	0.28	0.18	10.73%	10.13%	11.43%

Note: All numbers are weighted by Jackknife replication method. TFP_{NY}^{USDA} is the national yearly TFP obtained from the USDA Food Cost table. TFP_{RW}^{GCM} is the regional weekly TFP estimated from modeling from FoodAPS-GC data. y^{total} is the total food expenditure. Core Based Statistical Area (CBSA) consist of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core. P_0 is the prevalence. P_1 is the depth and P_2 is the severity.

Table 3.6.2. Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAH and FAFH Decomposition

					Expenditures		Share of Expenditures	
		N	TFP_{RW}^{GC}	y^{total}	y^{FAFH}	y^{FAH}	S^{FAFH}	S^{FAH}
All		**956**	**111.95**	**109.09**	**31.93**	**77.16**	**0.32**	**0.68**
4 Regions	Northeast	158	106.08	113.15	30.77	82.38	0.30	0.70
	Midwest	173	105.22	101.11	26.96	74.16	0.29	0.71
	South	416	106.81	103.37	33.60	69.77	0.35	0.65
	West	209	143.70	137.15	37.91	99.24	0.29	0.71
Regional_CBSA	In CBSA	871	115.76	110.95	34.27	76.68	0.33	0.67
	Not CBSA	85	88.66	97.75	17.64	80.11	0.22	0.78
Regional_Rural	Urban	689	116.23	109.49	36.00	73.48	0.35	0.65
	Rural	267	102.32	108.21	22.75	85.46	0.25	0.75
Labor Force Participation Rate	(0%-25%)	439	99.60	92.66	22.92	69.74	0.28	0.72
	(25%-50%)	332	143.79	143.07	45.07	98.00	0.37	0.63
	(50%-75%)	80	157.77	130.57	50.22	80.35	0.41	0.59
	(75%-100%)	105	76.60	91.98	29.47	62.51	0.31	0.69
Household Size	1	177	49.22	67.11	15.83	51.28	0.26	0.74
	2	205	88.83	105.02	36.92	68.10	0.32	0.68
	3	175	124.55	108.48	29.97	78.51	0.34	0.66
	4	182	155.11	139.70	43.75	95.94	0.40	0.60
	5	113	177.70	166.46	45.32	121.14	0.31	0.69
	6	56	218.12	168.19	52.61	115.59	0.33	0.67
	>=7	28	239.37	154.17	37.83	116.34	0.35	0.65
Adult Food Security Status	High Food Security	300	113.37	115.67	29.69	85.98	0.28	0.72
	Marginal Food Security	241	121.28	113.98	41.71	72.27	0.40	0.60
	Low Food Security	249	112.10	104.37	32.12	72.25	0.32	0.68
	Very Low Food Security	166	98.56	97.70	24.56	73.14	0.30	0.70

Table 3.6.2(Cont.) Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAH and FAFH Decomposition

			Expenditures. over TFP Threshold			FGT indices w/o FAFH			FGT indices w/o FAH		
		N	R^{total}	R^{FAFH}	R^{FAH}	P_0	P_1	P_2	P_0	P_1	P_2
All		**956**	**1.09**	**0.31**	**0.78**	**0.73**	**0.42**	**0.30**	**0.95**	**0.74**	**0.63**
4 Regions	Northeast	158	1.23	0.37	0.86	0.64	0.37	0.27	0.90	0.73	0.63
	Midwest	173	1.08	0.25	0.83	0.73	0.39	0.28	0.97	0.76	0.65
	South	416	1.08	0.33	0.75	0.74	0.44	0.32	0.96	0.73	0.60
	West	209	1.05	0.32	0.73	0.77	0.44	0.31	0.91	0.74	0.64
Regional_CBSA	In CBSA	871	1.07	0.33	0.75	0.75	0.43	0.31	0.95	0.73	0.61
	Not CBSA	85	1.22	0.20	1.02	0.61	0.32	0.23	0.98	0.81	0.72
Regional_Rural	Urban	689	1.05	0.35	0.71	0.78	0.46	0.34	0.94	0.72	0.61
	Rural	267	1.19	0.23	0.96	0.63	0.31	0.21	0.99	0.78	0.67
Labor Force Participation Rate	(0%-25%)	439	1.07	0.24	0.83	0.71	0.40	0.29	0.97	0.78	0.68
	(25%-50%)	332	1.08	0.37	0.71	0.76	0.44	0.32	0.96	0.72	0.59
	(50%-75%)	80	0.84	0.31	0.53	0.89	0.53	0.37	1.00	0.69	0.53
	(75%-100%)	105	1.31	0.43	0.88	0.69	0.39	0.27	0.86	0.66	0.58
Household Size	1	177	1.36	0.32	1.04	0.59	0.33	0.23	0.91	0.75	0.65
	2	205	1.21	0.44	0.77	0.75	0.40	0.29	0.92	0.70	0.60
	3	175	0.88	0.24	0.63	0.83	0.47	0.34	1.00	0.76	0.62
	4	182	0.91	0.28	0.63	0.80	0.52	0.39	0.98	0.72	0.60
	5	113	0.94	0.26	0.69	0.73	0.41	0.29	0.97	0.76	0.65
	6	56	0.78	0.25	0.53	0.91	0.50	0.34	0.99	0.75	0.63
	>=7	28	0.65	0.16	0.49	0.82	0.53	0.41	1.00	0.84	0.73
Adult Food Security Status	High Food Security	300	1.19	0.28	0.90	0.69	0.35	0.24	0.95	0.76	0.65
	Marginal Food Security	241	1.04	0.40	0.64	0.76	0.48	0.35	0.92	0.65	0.53
	Low Food Security	249	1.00	0.32	0.69	0.78	0.47	0.35	0.97	0.77	0.66
	Very Low Food Security	166	1.11	0.24	0.86	0.71	0.41	0.30	0.96	0.77	0.66

Table 3.6.2(Cont.) Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAH and FAFH Decomposition

Note: All numbers are weighted by Jackknife replication method. TFP_{RW}^{GCM} is the regional weekly TFP estimated from modeling from FoodAPS-GC data. y^{total} is the total food expenditure. y^{FAFH} is the FAFH expenditure. y^{FAH} is the FAH expenditure. S^{FAFH} is the share of FAFH expenditure over the total food expenditure. S^{FAH} is the share of FAH expenditure over the total food expenditure. $R_{\square}^{total}$ is the ratio of total food expenditure over the TFP $_{RW}$. $R_{\square}^{FAFH}$ is the ratio of FAFH food expenditure over the TFP $_{RW}$. $R_{\square}^{FAH}$ is the ratio of FAH food expenditure over the TFP $_{RW}$. Core Based Statistical Area (CBSA) consist of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core. P_0 is the prevalence. P_1 is the depth and P_2 is the severity.

Table 3.6.3. Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAFH and y^{FAH_P} and y^{FAH_S} Decomposition

					Expenditures			Share of Expenditures		
		N	TFP_{RW}^{GCM}	y^{total}	y^{FAFH}	y^{FAH_P}	y^{FAH_S}	S^{FAFH}	S^{FAH_P}	S^{FAH_S}
All		**956**	**111.95**	**109.09**	**31.93**	**30.23**	**46.80**	**0.32**	**0.31**	**0.37**
4 Regions	Northeast	158	106.08	113.15	30.77	42.22	40.16	0.30	0.39	0.31
	Midwest	173	105.22	101.11	26.96	21.86	52.20	0.29	0.27	0.44
	South	416	106.81	103.37	33.60	26.65	42.86	0.35	0.30	0.35
	West	209	143.70	137.15	37.91	47.00	52.24	0.29	0.36	0.35
Regional_CBSA	In CBSA	871	115.76	110.95	34.27	31.06	45.50	0.33	0.31	0.35
	Not CBSA	85	88.66	97.75	17.64	25.16	54.74	0.22	0.31	0.47
Regional_Rural	Urban	689	116.23	109.49	36.00	29.98	43.43	0.35	0.31	0.34
	Rural	267	102.32	108.21	22.75	30.78	54.37	0.25	0.32	0.43
Labor Force Participation Rate	(0%-25%)	439	99.60	92.66	22.92	27.25	42.34	0.28	0.35	0.37
	(25%-50%)	332	143.79	143.07	45.07	35.02	62.82	0.37	0.26	0.37
	(50%-75%)	80	157.77	130.57	50.22	36.42	43.75	0.41	0.30	0.29
	(75%-100%)	105	76.60	91.98	29.47	28.58	33.84	0.31	0.29	0.39
Household Size	1	177	49.22	67.11	15.83	21.81	29.41	0.26	0.35	0.39
	2	205	88.83	105.02	36.92	25.72	42.22	0.32	0.30	0.38
	3	175	124.55	108.48	29.97	31.15	47.16	0.34	0.29	0.37
	4	182	155.11	139.70	43.75	37.26	58.56	0.40	0.30	0.30
	5	113	177.70	166.46	45.32	34.91	86.12	0.31	0.23	0.47
	6	56	218.12	168.19	52.61	50.74	64.39	0.33	0.31	0.36
	>=7	28	239.37	154.17	37.83	38.09	78.14	0.35	0.25	0.40
Adult Food Security Status	High Food Security	300	113.37	115.67	29.69	31.35	54.50	0.28	0.31	0.41
	Marginal Food Security	241	121.28	113.98	41.71	32.82	39.27	0.40	0.32	0.29
	Low Food Security	249	112.10	104.37	32.12	33.17	38.92	0.32	0.32	0.35
	Very Low Food Security	166	98.56	97.70	24.56	21.47	51.59	0.30	0.30	0.40

Table 3.6.3(Cont.) Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAFH and y^{FAH_P} and y^{FAH_S} Decomposition

		N	FGT indices w/o y^{FAH_P}			FGT indices w/o y^{FAH_S}		
			P_0	P_1	P_2	P_0	P_1	P_2
All		**956**	**0.71**	**0.44**	**0.32**	**0.80**	**0.52**	**0.40**
4 Regions	Northeast	158	0.72	0.49	0.39	0.69	0.46	0.36
	Midwest	173	0.66	0.39	0.28	0.84	0.56	0.44
	South	416	0.73	0.43	0.31	0.81	0.51	0.40
	West	209	0.76	0.51	0.38	0.77	0.51	0.39
Regional_CBSA	In CBSA	871	0.74	0.45	0.33	0.80	0.52	0.40
	Not CBSA	85	0.57	0.37	0.29	0.81	0.53	0.41
Regional_Rural	Urban	689	0.75	0.46	0.34	0.79	0.53	0.41
	Rural	267	0.63	0.38	0.29	0.81	0.52	0.40
Labor Force Participation Rate	(0%-25%)	439	0.74	0.47	0.36	0.80	0.54	0.43
	(25%-50%)	332	0.71	0.42	0.31	0.83	0.53	0.40
	(50%-75%)	80	0.87	0.47	0.31	0.92	0.48	0.32
	(75%-100%)	105	0.57	0.35	0.26	0.70	0.47	0.38
Household Size	1	177	0.64	0.41	0.31	0.70	0.48	0.39
	2	205	0.66	0.38	0.29	0.80	0.49	0.36
	3	175	0.78	0.46	0.33	0.88	0.57	0.43
	4	182	0.79	0.49	0.37	0.85	0.53	0.40
	5	113	0.71	0.38	0.26	0.86	0.62	0.49
	6	56	0.86	0.51	0.35	0.79	0.56	0.45
	>=7	28	0.77	0.54	0.43	0.98	0.69	0.55
Adult Food Security Status	High Food Security	300	0.65	0.40	0.30	0.78	0.52	0.41
	Marginal Food Security	241	0.77	0.45	0.33	0.74	0.44	0.33
	Low Food Security	249	0.80	0.48	0.35	0.81	0.55	0.43
	Very Low Food Security	166	0.67	0.42	0.33	0.89	0.59	0.45

Table 3.6.3(Cont.) Summary Statistics of Expenditures and FGT Indices on Regional Weekly TFP by FAFH and y^{FAH_P} and y^{FAH_S} Decomposition

Note: All numbers are weighted by Jackknife replication method. TFP_{RW}^{GCM} is the regional weekly TFP estimated from modeling from FoodAPS-GC data. y^{total} is the total food expenditure. y^{FAFH} is the FAFH expenditure. y^{FAH_P} is the FAH expenditure paid by personal. y^{FAH_S} is the FAH expenditure paid by SNAP. Core Based Statistical Area (CBSA) consist of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core. P_0 is the prevalence. P_1 is the depth and P_2 is the severity.

Table 3.6.4. Summary Statistics of TFP Threshold, Expenditures and FGT Poverty Index for SNAP Households by Food Groups Using Regional Weekly TFP Threshold

			Total	Northeast	Midwest	South	West
		N	956	158	173	416	209
		$TFP_{R,W}^{GC,Model}$	111.95	106.08	105.22	106.81	143.7
		y^{total}	109.37	113.12	101.34	103.88	137.11
		y^{FAFH}	31.93	30.77	26.96	33.6	37.91
		y^{FAH}	77.44	82.36	74.38	70.28	99.2
		y^{FAH_1}	6.04	6.08	5.95	5.21	8.45
		y^{FAH_2}	20.78	20.22	17.89	21.88	23.85
	Panel A1: Summary Statistics	y^{FAH_3}	6.36	4.88	8.12	4.66	8.73
		y^{FAH_4}	6.45	5.33	6.03	5.97	9.42
		y^{FAH_5}	9.79	11.19	10.8	8.04	11.51
		y^{FAH_6}	8.42	9.97	6.47	7.55	13.4
		y^{FAH_7}	10.18	12.58	10.02	8.71	12.68
		y^{FAH_8}	6.43	6.37	6.17	5.85	8.59
		y^{FAH_9}	0.7	1.28	0.66	0.68	0.39
		$y^{FAH_{10}}$	2.29	4.45	2.27	1.73	2.18
		S^{FAFH}	0.32	0.3	0.29	0.35	0.29
		S^{FAH}	0.68	0.7	0.71	0.65	0.71
		S^{FAH_1}	0.06	0.06	0.05	0.06	0.07
		S^{FAH_2}	0.17	0.15	0.15	0.19	0.18
	Panel B1: Share of Sub Expenditures over Total	S^{FAH_3}	0.05	0.05	0.07	0.04	0.05
Mean		S^{FAH_4}	0.06	0.04	0.06	0.06	0.06
		S^{FAH_5}	0.08	0.1	0.09	0.08	0.09
		S^{FAH_6}	0.08	0.09	0.07	0.07	0.09
		S^{FAH_7}	0.1	0.12	0.11	0.09	0.09
		S^{FAH_8}	0.05	0.05	0.05	0.05	0.06
		S^{FAH_9}	0	0.01	0	0	0
		$S^{FAH_{10}}$	0.02	0.03	0.04	0.01	0.02
		R^{total}	1.1	1.23	1.08	1.09	1.05
		R^{FAFH}	0.31	0.37	0.25	0.33	0.32
		R^{FAH}	0.79	0.86	0.83	0.76	0.73
		R^{FAH_1}	0.06	0.06	0.07	0.06	0.06
		R^{FAH_2}	0.22	0.22	0.2	0.25	0.18
	Panel C1: Ratio of Sub Exp. over TFP Threshold	R^{FAH_3}	0.06	0.06	0.09	0.05	0.06
		R^{FAH_4}	0.06	0.05	0.07	0.06	0.06
		R^{FAH_5}	0.09	0.12	0.11	0.08	0.08
		R^{FAH_6}	0.09	0.1	0.08	0.09	0.1
		R^{FAH_7}	0.11	0.14	0.12	0.09	0.1
		R^{FAH_8}	0.07	0.07	0.07	0.07	0.06
		R^{FAH_9}	0.01	0.01	0	0	0
		$R^{FAH_{10}}$	0.02	0.04	0.03	0.02	0.01

Table 3.6.4(Cont.) Summary Statistics of TFP Threshold, Expenditures and FGT Poverty Index for SNAP Households by Food Groups Using Regional Weekly TFP Threshold

			Total	Northeast	Midwest	South	West
FGT Indices	Panel E0: FGT indices_Full	P_0	0.55	0.52	0.57	0.52	0.6
		P_1	0.27	0.26	0.25	0.27	0.3
		P_2	0.17	0.17	0.15	0.18	0.19
	Panel E1: FGT indices w/o y^{FAFH}	P_0	0.73	0.64	0.73	0.74	0.77
		P_1	0.42	0.37	0.39	0.44	0.45
		P_2	0.3	0.27	0.28	0.32	0.31
	Panel E2: FGT indices w/o y^{FAH}	P_0	0.95	0.9	0.97	0.96	0.91
		P_1	0.74	0.73	0.76	0.73	0.74
		P_2	0.63	0.63	0.65	0.6	0.64
	Panel E3: FGT indices w/o y^{FAH_1}	P_0	0.58	0.54	0.6	0.54	0.66
		P_1	0.28	0.27	0.27	0.28	0.32
		P_2	0.18	0.18	0.16	0.19	0.21
	Panel E4: FGT indices w/o y^{FAH_2}	P_0	0.69	0.62	0.66	0.71	0.72
		P_1	0.33	0.31	0.31	0.33	0.37
		P_2	0.21	0.2	0.19	0.22	0.24
	Panel E5: FGT indices w/o y^{FAH_3}	P_0	0.61	0.59	0.65	0.56	0.67
		P_1	0.28	0.27	0.28	0.28	0.32
		P_2	0.18	0.18	0.17	0.19	0.2
	Panel E6: FGT indices w/o y^{FAH_4}	P_0	0.58	0.55	0.6	0.55	0.64
		P_1	0.28	0.27	0.28	0.28	0.32
		P_2	0.18	0.18	0.16	0.19	0.21
	Panel E7: FGT indices w/o y^{FAH_5}	P_0	0.6	0.56	0.63	0.57	0.66
		P_1	0.3	0.29	0.29	0.29	0.34
		P_2	0.19	0.19	0.17	0.19	0.22
	Panel E8: FGT indices w/o y^{FAH_6}	P_0	0.61	0.63	0.61	0.58	0.69
		P_1	0.29	0.29	0.28	0.29	0.33
		P_2	0.19	0.18	0.18	0.19	0.22
	Panel E9: FGT indices w/o y^{FAH_7}	P_0	0.62	0.58	0.61	0.61	0.68
		P_1	0.3	0.29	0.3	0.29	0.34
		P_2	0.2	0.2	0.18	0.19	0.22
	Panel E10: FGT indices w/o y^{FAH_8}	P_0	0.58	0.56	0.59	0.55	0.67
		P_1	0.28	0.27	0.27	0.28	0.32
		P_2	0.18	0.18	0.16	0.18	0.2
	Panel E11: FGT indices w/o y^{FAH_9}	P_0	0.55	0.52	0.57	0.53	0.6
		P_1	0.27	0.26	0.25	0.27	0.3
		P_2	0.17	0.17	0.15	0.18	0.19
	Panel E12: FGT indices w/o $y^{FAH_{10}}$	P_0	0.56	0.52	0.59	0.53	0.6
		P_1	0.28	0.27	0.27	0.27	0.31
		P_2	0.18	0.18	0.17	0.18	0.2

Note: All numbers are weighted by Jackknife replication method.
$TFP_{R,W}^{GC,Model}$ is the regional weekly TFP estimated from modeling from FoodAPS-GC data.
y^{total} is the total food expenditure. y^{FAFH} is the FAFH expenditure. y^{FAH} is the FAH expenditure. y^{FAH_g} is the FAH expenditure for food group g, g=1,2..,10.
S^{FAFH} is the share of FAFH expenditure over total expenditure. S^{FAH} is the share of FAH expenditure over total expenditure. S^{FAH_g} is the share of FAH expenditure for food group g over total expenditure. R^{FAFH} is the share of FAFH expenditure over TFP threshold. S^{FAH} is the share of FAH expenditure over TFP threshold. S^{FAH_g} is the share of FAH expenditure for food group g over TFP threshold, g=1,2..,10.
The 10 food groups are 1) milk and dairy, 2) protein, 3) mixed dishes, 4) grains 5) snacks and sweets, 6) fruit and vegetables, 7) beverages, 8) fats and oils, condiments and sugars 9) infant formula and body food, or not in a category, 10) food code not assigned.
P_0 is the prevalence. P_1 is the depth and P_2 is the severity.

3.11 Appendix A: Decision Tree of FAH and FAFH expenditure

This appendix explains the decision tree on the construction of the FAH and FAFH expenditures.

3.11.1 Decision Tree for FAH Expenditure

The FAH expenditure data at FoodAPS contains some anomalies due to missing observations. Those missing expenditures can be partially uncovered by other information, such as free event indicator and ERS imputation. Figure 3.11.1-3.11.3 below are three figures summarizing the anomaly and the decisions on the construction of FAH expenditure on item level, event level, and household level respectively. Numbers in regular font represent the summary based on the item level data. Numbers in *italic font* represent the summary based on the event level data. Numbers in **bold font** represent the summary based on the household level data. For example, the figure about item level anomaly is presented in regular font, event level anomaly in *italic font*, and household level anomaly in **bold font**. Within the figure, the three percentage in parenthesis summarize the percent of the case impacted at item level, *event level* and **household level**, thus highlighted in different fonts. Problematic cases noted by different numbers. Problems appear in the item level will impact the event level and later household level summary, which is explained in the notes. In each figure, there are two shapes. The rectangular represents different cases or subcases. The oval represents the solutions adapted after this case. Cases without any oval shapes after them are dropped in the construction of a clean household dataset.

(Insert Figure 3.11.1 here)

Figure 3.11.1 analyzes the anomalies of the original item cost and offers suggestions in tackling each anomaly cases to get a clean FAH_{hei}. Each case is flagged based on the availability of original item cost, free event indicator at item level data and event-level data, and the availability of ERS imputation.

Case 1.1 is the case when an item expenditure is missing for the event that is non-free but not imputed by ERS. This case accounts for 1.99% of all items at item level data, impacted 5.79% of the events at event level data, and 14.94% of households at the household level data. This case is flagged. It will later be dropped in the construction of clean sub household sample. Case 1.2 is the case when item expenditure is missing for non-free event, but the cost is later imputed by ERS. The ERS imputation value is used as the item cost. Case 1.3 is the case when the item expenditure is missing, but the event is marked as free by the household. The item cost is identified as zero for this case. Case 1.4 is the case when item expenditure is available. The original item expenditure is used.

It is worth noting, for figure 3.11.1 the summation of the percent at item level equals to 100%. However, the summation of the percentage at event level will exceed 100% because there are events with items belonging to different cases. Such events are counted more than once and so is the case for percentages at the household level. After these adjustments, the item level expenditure is aggregated into event level expenditure analyzed in Figure 3.11.2 ($\mathrm{FAH_{he\cdot}} = \sum_{i=1}^{I_{he}} \mathrm{FAH_{hei}}$).

(Insert Figure 3.11.2 here)

Figure 3.11.2 analyzes the anomaly of the aggregated event level expenditure ($\mathrm{FAH_{he\cdot}}$). Case 2.1 and 2.2 represent the cases that an event is recorded at the event level, but not recorded at the item level. That is, the household indicates such FAH acquisition, but no item is recorded at the item level due to the missing of receipt and Blue Page information. For those cases, free event indicator provides additional information for the event expenditure. If the event is free, the event expenditure is set to be zero (case 2.2). If the event is not free, there is no way to uncover the cost, thus it is flagged as well. It will be later dropped in construction of clean sub household sample (case 2.1).

Case 2.3 and 2.4 summarize the case when the item of the event is recorded at the item level, but the expenditure information is missing, because of case 1.1 in figure 3.11.1.

Case 2.5 and 2.6 summarize the events that all item cost is available at the item level. The event cost is obtained by aggregation from the item cost. For SNAP households, there is also information about payments made by SNAP EBT for this event. Because the payment by SNAP EBT is not taxable, the food and beverage sales tax only apply to payments made through personal funding. This taxable personal payment is constructed by subtracting the SNAP payment amount from the aggregated event level cost. Such subtraction may lead to a negative value[15]. Case 2.5 identified such cases and is later dropped in the construction of a clean sub household sample.

The percentage numbers in parenthesis are the summary at the item level, event level, and household level, respectively. The item level summary is not available at event level analysis, thus represented by N/A. The second number is the portion of this case in the whole FAH events. The summation over all cases will be 100%. The third number is the percentage of impacted households by this case. A household with more than one event may be impacted by more than one case. Thus the summation of this number would exceed 100%.

After these adjustments, the event level expenditure is aggregated into household level expenditure analyzed in Figure 3.11.3 ($\mathrm{FAH}_{\mathrm{h}\cdot\cdot} = \sum_{e=1}^{E_h} \mathrm{FAH}_{\mathrm{he}\cdot}$).

---

[15] We use $-1, instead of $0, as the cutoff point of negative case. The distribution of the negative taxable personal payment is negatively skewed with a mean of -$1.25 and median of -$3.23*e-07. More than 99% of the observation clustered within the internal of (-$1, $0) and close towards the left of $0 due to the rounding errors. More importantly, events with taxable personal payment between ($-1, $0) will be replaced by $0. This leads to an overestimation of the FAH expenditure, making our conclusion on the severeness of food expenditure poverty even more convincing.

(Insert Figure 3.11.3 here)

Figure 3.11.3 analyzes the anomalies of the aggregated household level expenditure ($FAH_{h..}$). Case 3.1 are households with no FAH event level record. Those household are considered as not making any FAH purchases during the survey week. The reasons can be the household happens to make FAH purchase before or after the survey week, or the household ate FAFH for the entire survey week. As a result, the FAH expenditure is assigned zero. Those household constitute part of the clean sub household sample.

Case 3.2 are households with missing/problematic event level cost due to the case 2.1, 2.3, 2.4 and 2.5 in figure 3.11.2. It is important to distinguish case 3.1 and case 3.2. The former means the household does not have any FAH purchases during the survey week, thus assigned zero cost. The latter means the household has FAH purchases during the survey week, but the event expenditure information is problematic, thus dropped.

Case 3.3 are households with all events recorded and the event expenditure available, thus constitute the second part of a clean sub household sample.

Case 3.2 account for 24.24% of the households (the union of problematic households in case 2.1, 2.3, 2.4, and 2.5 in figure 3.11.2). Those households are dropped, leaving the rest in the clean sub household sample. Those rest 3655 households (75.76% of the total households) are the subsample used in the analysis of FAFH expenditure.

3.11.2 Decision Tree for FAFH Expenditure

Figure 3.11.4 and 3.11.5 summarize the data anomaly and solutions for each case of FAFH at the event level and household level, respectively. The analysis is based on the previously defined clean sub-sample households by the FAH analysis. The font and shape pattern is similar to figure

3.11.1 to 3.11.3 for FAH. Because FAFH is aggregated from the event level, the analysis starts from the event level cost, instead of the item level cost. So the $\mathrm{FAFH_{he\cdot}} = \mathrm{TOTALPAID}$ is the cost at the event level, instead of aggregated from the item level cost. The household level cost is the summation of total FAFH event cost ($\mathrm{FAFH_{h\cdot\cdot}} = \sum_{e=1}^{E_h} \mathrm{FAFH_{he\cdot}}$).

(Insert Figure 3.11.4 here)

Figure 3.11.4 summarizes three cases for FAFH expenditure. Case 4.1 is the case when the event expenditure is missing for the non-free event. This is flagged and highlighted in red. This case will later be dropped out of the clean household. This case accounts for 0.35% of the total FAFH events and impacted 2.22% of the households.

Case 4.2 is the case with event expenditure missing, but the case is indicated as free events by the households. There is no such case within the predefined clean subsample (This one is kept because this can be a problem for a broader definition of the clean subset). If there were such a case, the event cost would be replaced as zero.

Case 4.3 is the case with event expenditure available. It accounts for 72.59% of the total FAFH events and impacted 67.61% of households. One weird thing about this case is that SNAP payment is used for such FAFH purchase for some events when SNAP payment is only allowed for the purchase of FAH. So, this case is further investigated based on SNAP information and payment information.

Case 4.3.1 was the event when the household identified SNAP EBT card as the *only* payment method. The event expenditure is considered paid by SNAP. As a result, the SNAP amount is the $\mathrm{FAFH_{he\cdot}}$, while the personal payment is zero. Case 4.3.2 was the event when the household identified SNAP EBT card as *one of* the payments used. Because no other information about

SNAP amount is available, the SNAP amount is assumed to be zero. Case 4.3.3 is the event when no SNAP EBT card is used. All payment is considered as a personal payment.

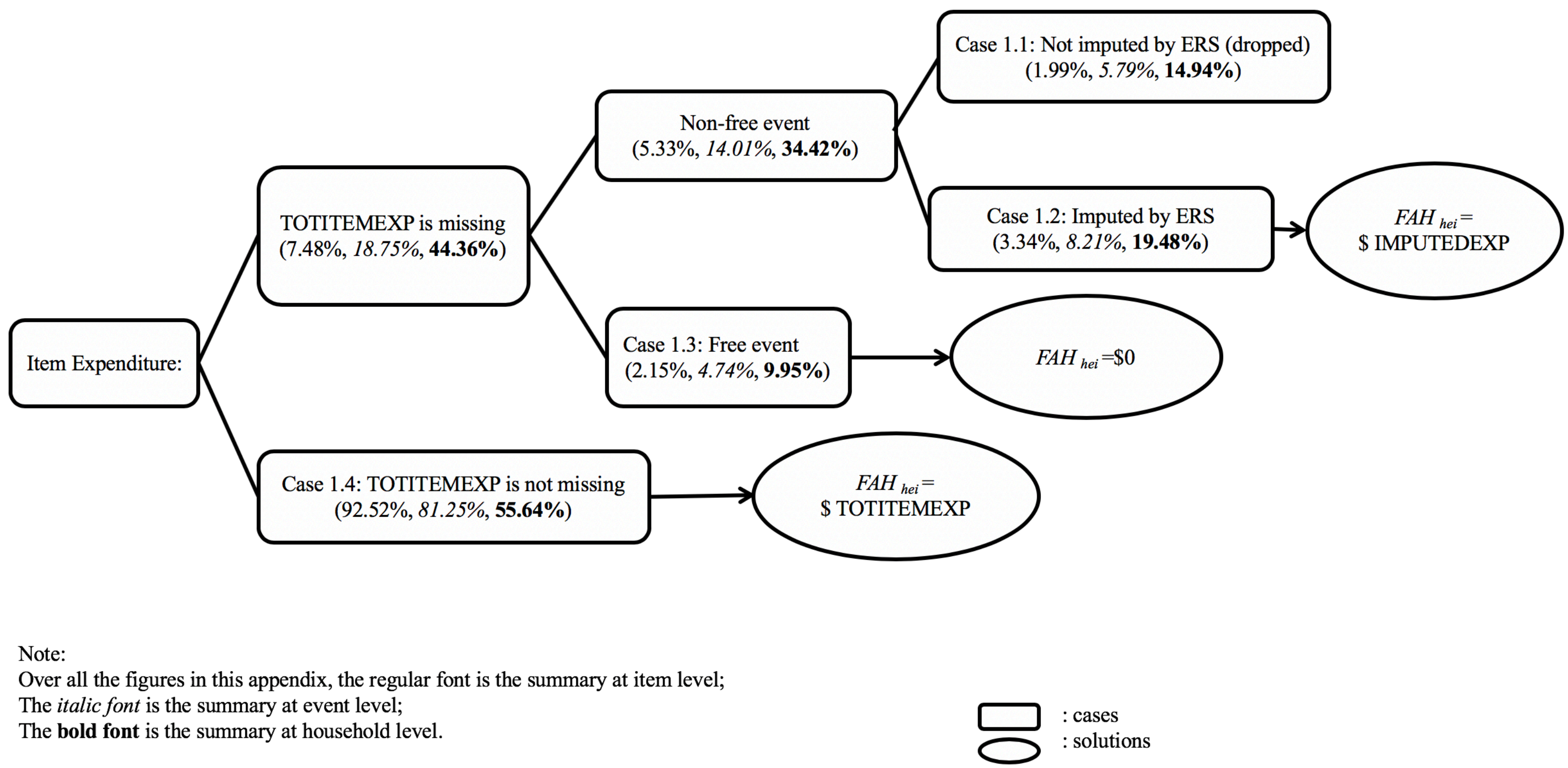
It is worth noting that the percentage number at event level does not sum into 100% because the analysis is based on the events of clean sub households only.

(Insert Figure 3.11.5 here)

Figure 3.11.5 analyzes the anomaly of the aggregated household level expenditure ($FAFH_{h..}$). Case 5.1 are households with no FAFH event level record. Those household are considered not making any FAFH purchases during the survey week. The reasons can be the household happens ate FAH for the entire survey week. As a result, the FAFH expenditure is assigned zero. There are 392 households (4.25% of the total sample size).

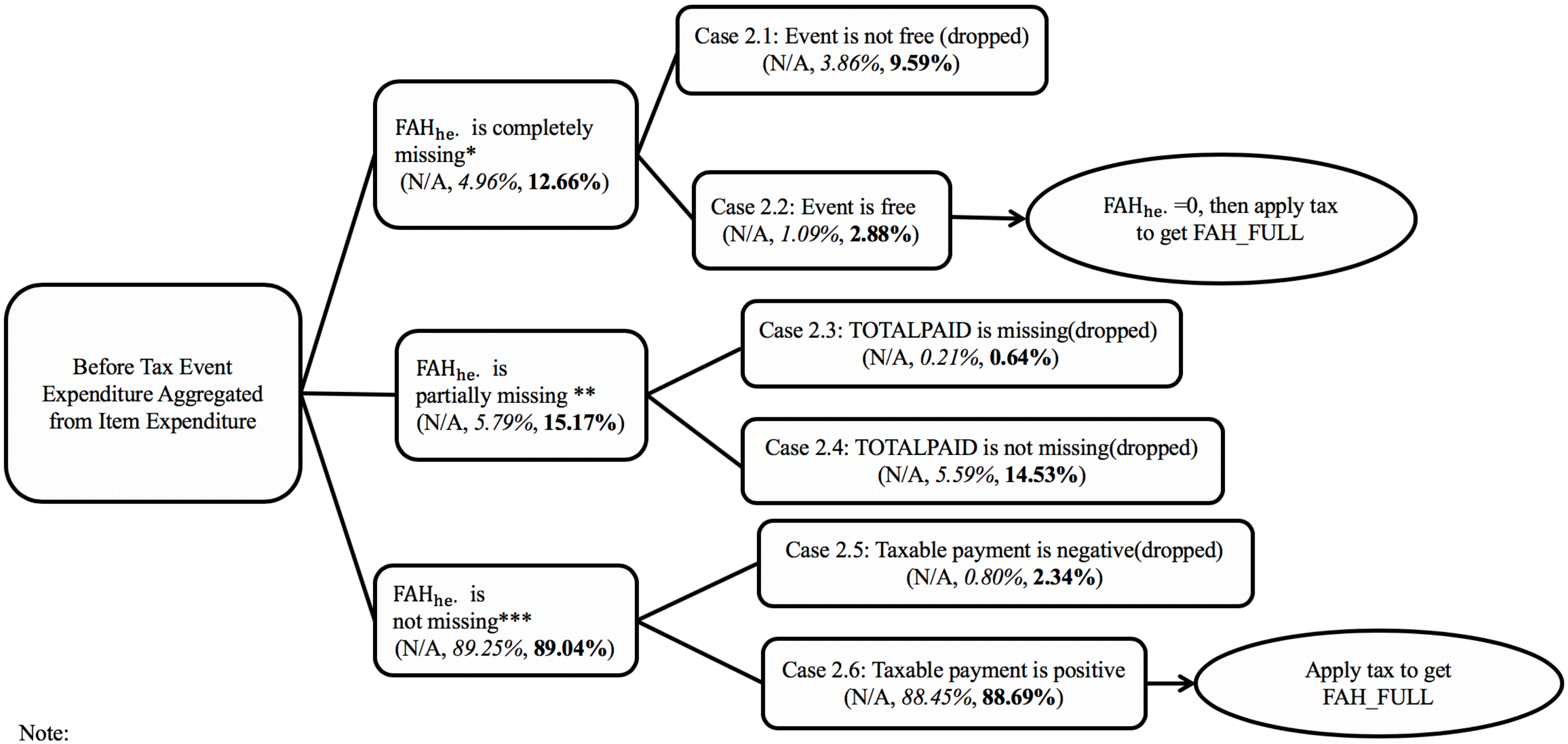
Case 5.2 are households with missing event level cost due to the case 4.1 in figure 3.11.4. Case 5.3 are households with all events recorded and the event expenditure available. Case 5.2 account for 2.22% of the households. Those households are dropped, leaving the rest in the clean sub household sample. Those rest 3549 households (73.54% of the total household) are the sub-sample used in the following analysis.

Figure 3.11.1: Anomaly and Decision of FAH Item Level Expenditure in Obtaining FAH_{hei}



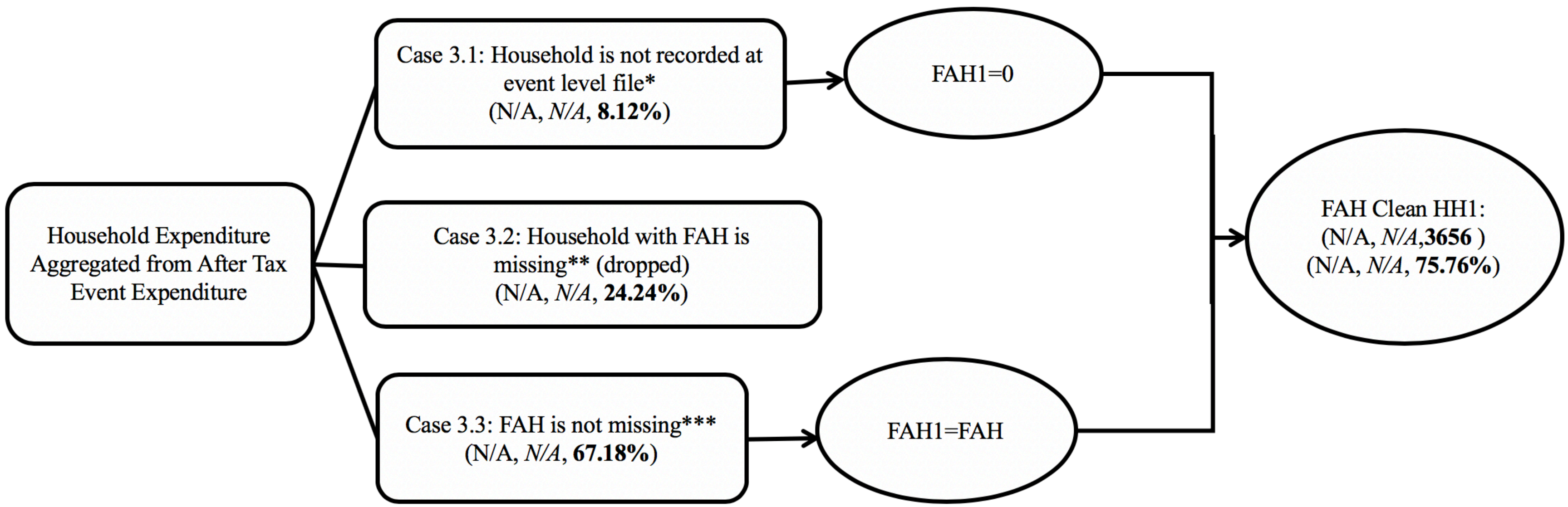
Note:
Over all the figures in this appendix, the regular font is the summary at item level;
The *italic font* is the summary at event level;
The **bold font** is the summary at household level.

Figure 3.11.2: Anomaly and Decision of FAH Event Level Expenditure in Obtaining *FAH* $_{he.}$



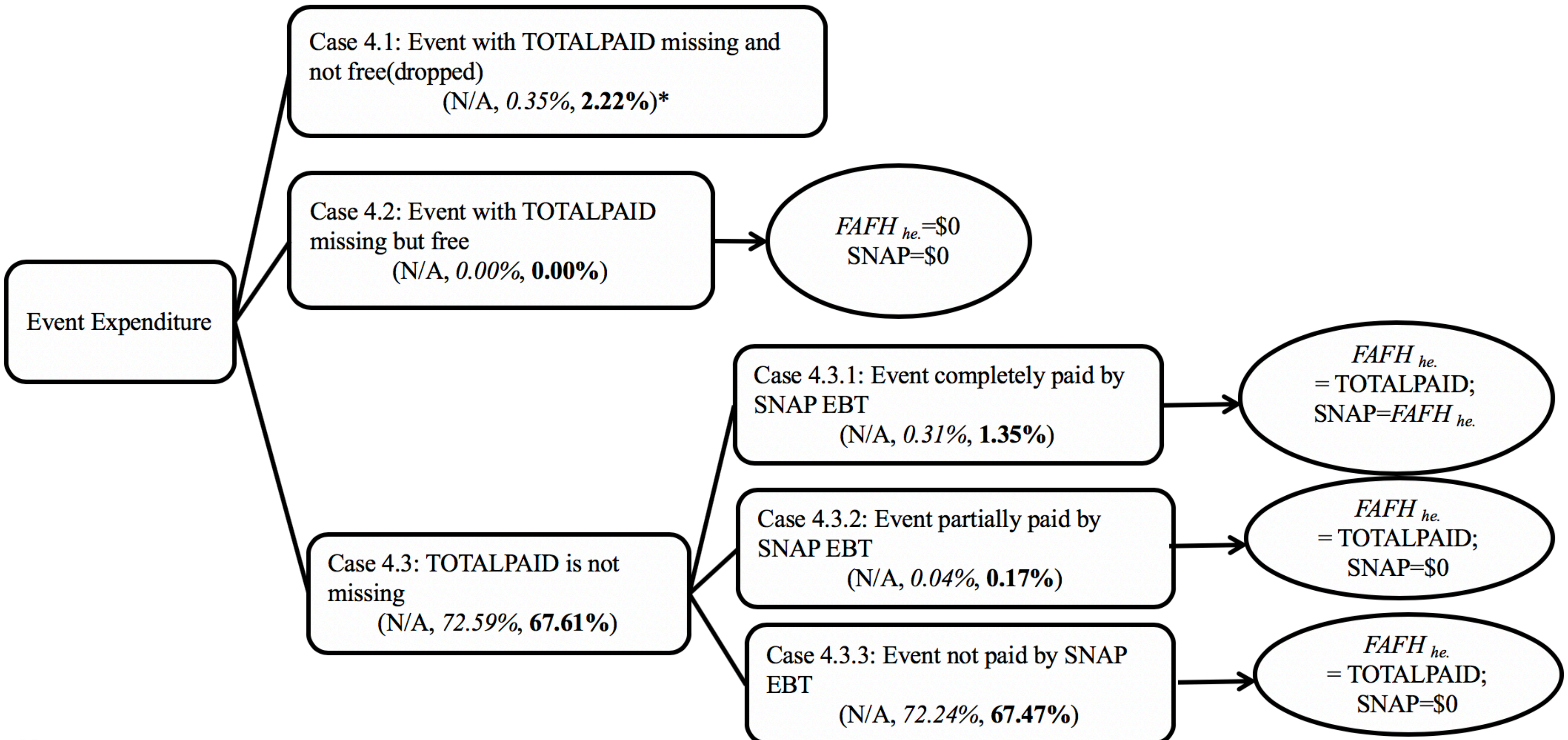
Note:
*: This is the event with no record at the item level, but a record at the event level. No show in figure 3.11.1.
**: This is the event with both recorded at the item and event level, but with some item cost missing due to case 1.1 in figure 3.11.1.
***: This is the clean events aggregated from case 1.2, 1.3, 1.4 in figure 3.11.1.

Figure 3.11.3: Anomaly and Decision of FAH Household Level Expenditure in Obtaining FAH_{h}.



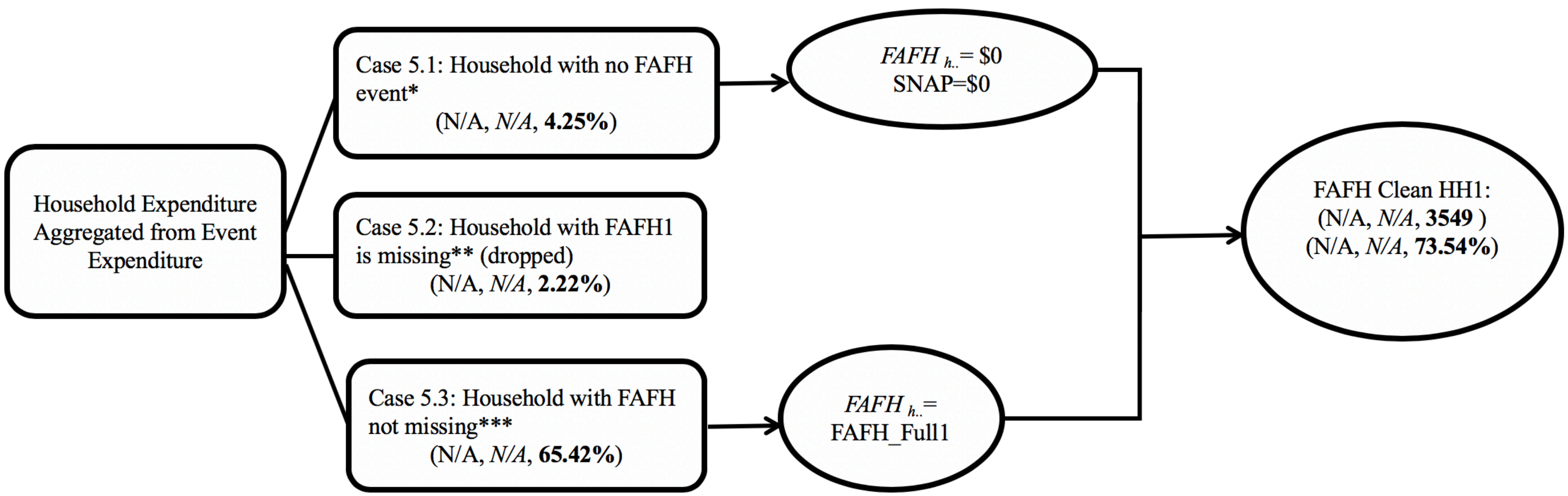
Note
*: This is the household with no record at event level. No show in figure 3.11.2.
**: This is the household with record at event level, but with some event cost missing due to case 2.1, 2.3, 2.4 and 2.5 in figure 3.11.2.
***: This is the clean household aggregated from case 2.2 and 2.4 in figure 3.11.2.

Figure 3.11.4: Anomaly and Decision of FAFH Event Level Expenditure in Obtaining $FAFH_{he.}$



Note:
*: All percentage is based on the full sample size of 39120 FAFH events and 4826 households.

Figure 3.11.5: Anomaly and Decision of FAFH Household Level Expenditure in Obtaining $FAFH_{h..}$



Note:
*: This is the household with no record at event level. No show in figure 3.11.4.
**: This is the household with record at event level, but with some event cost missing due to case 4.1 in figure 3.11.4.
***: This is the clean household aggregated from case 4.2 and 4.3 in figure 3.11.4.

4 Money, Time, and the Healthy Eating Index: How Are They Related? A Structural Analysis with Disparate Datasets

4.1 Introduction

A healthy diet is associated with a lower risk of chronic diseases (Schwingshackl, and Hoffmann, 2015; Schwingshacki, Bogensberger, and Hoffmann, 2018). As a measurement of diet quality, the Healthy Eating Index (HEI) measures how well the diet is in alignment with the most recent Dietary Guidelines for the American. It measures the adequacy of nine food components (Total Fruit, Whole Fruit, Total Vegetables, Greens and Beans, Whole Grains, Dairy, Total Protein foods, Seafood, and Plant Proteins and Fatty Acids) and the moderation of three other food components (Refined Grains, Sodium and Empty Calories) of the diet. Diet quality can be assessed on meals/food prepared at home(M^H) and meals/food prepared away from home (M^A) using the HEI.[16] Here the superscript H representing meals/food at **H**ome and A representing meals/food **A**way from home. Similar to Gronau (2017), the meals (M^H, M^A) are produced from the market inputs of the food (q^H, q^A) and the corresponding time inputs (T^H, T^A) given meal preparation related characteristics (Z^H, Z^A) as controls to capture the meal production efficiency, which can be at the individual or household level. Mathematically, the HEI is determined by the following relationship:

$$\begin{aligned} HEI = F(M^H, M^A) &= F(M^H(q^H, T^H;\ Z^H), M^A(q^A, T^A;\ Z^A)) \\ &= H(q^H, q^A, T^H, T^A;\ Z^H, Z^A) \qquad (4.1.1) \end{aligned}$$

[16] It is important to distinguish the meals/food from the ingredients. The meal/food is the outcome of household production with ingredients as one of the inputs. For example, a frozen pizza in the super market is an ingredient. But if you purchase this pizza, put it into microwaves and place it onto the plate for dinner, this hot pizza is a meal, because it is a production result of the ingredient input (frozen pizza purchased in the super market) and you time input (your trip to the super market, time to microwave). Though out this chapter, the word meal and food are used interchangeably.

Here the $M^H(.)$ and $M^A(.)$ are production functions, capturing the meal/food production technology for meals/food prepared at home (H=FAH) and meals/food prepared away from home (A=FAFH), respectively. The $F(.)$ is the HEI function showing how the HEI score is obtained from the meals produced (both M^H, M^A). As a result, the composite function $H(.) = F(M^H(.),\ M^A(.))$ is an algorithm transformation or an aggregator function that summarizes the contribution of the market input and time input of meals toward the HEI. It captures a HEI score based on the market input, time input, and characteristics. One thing to highlight about this function $H(.)$ is that it captures the relationship between the inputs and characteristics. But, it does not show any individual/household taste or preferences for the inputs. In other words, the (Z^H, Z^A) are characteristics regarding food production efficiency, not the preferences related characteristics.[17] For simplicity, we will refer to this as the HEI direct function to distinguish from other HEI functions discussed in the following section.

For the cross-sectional data, the price of market input for food at home (p^H) and for food away from home (p^A) are considered fixed. So the quantity consumed (q^H, q^A) can be replaced by the corresponding money expenditure (E^H, E^A). The HEI direct HEI function then becomes

$$HEI = H(E^H, E^A, T^H, T^A;\ Z^H, Z^A) \qquad (4.1.2)$$

where the $E^H = p^H * q^H, E^A = p^A * q^A$.

One problem with the empirical research that estimates the equation (4.1.2) is that it suffers from endogeneity problem due to the correlation between the food expenditures/time inputs and the error term, which may contain other eating health-related choice variables (Rozenweig and Schultz 1983). For example, people with more health literacy may spend more money on food, so health literacy is related positively to money expenditure. But even given the same money

[17] The preference related characteristics are captured by the utility function, which will be detailed later.

expenditure and time input, healthier people may choose healthier food, thus obtain a higher HEI score. As a result, the unobservable health literacy will also impact the HEI score given the money and time input. Stated more generally, there will be moderating variables in this function, besides the Z^H and Z^A.

When such moderating variables are not observed, it leads to endogeneity problem of input variables (E^H, E^A, T^H, T^A) in the estimation. A commonly used solution to the endogeneity problem is the Instrumental Variables (IV) approach to find a (set of) instrument variable(s) that is 1) relevant to the endogenous variable and 2) exogenous to the dependent variable. However, the main critique of the IV approach is that the choice of instrumental variables is ad hoc, with no theoretical guidance for the choices of the IVs. In this chapter, a theoretical framework is provided to give support to the choice of IVs in the HEI production function estimation. The empirical estimation strategy on the irrelevance condition regarding the special dataset is also provided. Altogether, this structure and approach provide a theory-guided interpretation of different model specifications that does not exist in current literature. This, in turn, helps in organizing the empirical literature as shown later.

The next section offers a theoretical model and the empirical explanation for estimating the HEI function, followed by a literature review of past empirical analysis guided by the theoretical model in section 3. Section 4 introduces the two unique datasets used in this chapter, with one dataset providing the HEI and money input and another providing time input. Section 5 explains the estimation strategy regarding the two datasets. The construction and definition of the variables are presented in section 6 with the time prediction result presented in section 7. The final estimation on HEI is presented in section 8 followed by the conclusion and limitations in section 9.

4.2 Theoretical Model

Food choice is considered a household decision, especially for a household with more than one household member. Originating from the essential work of Becker (1965), economists have developed various approaches in capturing household's decisions. The three common approaches are the unitary model, the collective model, and non-cooperative model. The unitary model (Barten and Bohm, 1982; Deaton and Muellbauer, 1980; Muellbauer, 1974; and Pollak and Wachter, 1975) treats a household similar to an individual, deriving utility from commodities produced with money and time input. One assumption of this model is the income-pooling assumption that the source of income (who makes the money) does not matter, rather it is just the total household income that impacts the overall distribution of consumption among household members. In other words, the individual with high income does not have bargaining power over the allocation of resources among household members. On the other hand, the collective model, developed by Chiappori (1988, 1992), recognizes the intra-household interaction and captures the individual bargaining power weighted by the individual's income. This model assumes Pareto efficient of intra-household allocation. The non-cooperative model (Leuthold, 1968; Ashworth and Ulph, 1981; You and Davis, 2010) assumes household members maximize their utility, subject to their own constraints, taken account into other individual's responses. This model does not promise a Pareto efficient allocation outcome.

Various empirical works have tested the assumptions of different models and in support of different models in different scenarios with no solid conclusion settled. According to Vermulen (2002), the unitary model is a special case of the collective model under the following three circumstances. First, when individuals in the household contain the same preferences, the collective model collapses into a unitary model. It uses individual preferences to represent the household preference, with a smaller constant scale that does not change the optimization solution. Second, when the household's decision is dictated by a benevolent dictator and other individuals have zero bargaining power, the whole household behaves as one individual. The collective model is reduced into the unitary model as well in this case. Lastly, when the household behaves as if it were a single individual with latently separable preferences, that is, no consumption on public goods or externalities in consumption.

In this chapter, we use the unitary model for three main reasons. First, in the household food preparation analysis, the decision is normally made by the main preparer, especially for the single-headed household with one adult as the main preparer. This satisfies the second circumstance mentioned above. Second, the main focus of this chapter is the single-headed household, where all household members share the income of the household head. The income pooling assumption is a valid one to hold. Furthermore, although the latter two models capture more interesting interactions between the household's members, they are more demanding on data. The individual data required for every household member is not always available for researchers. The unitary model requires mainly household level observations and is more empirically feasible.

4.2.1 Unitary Model

Following household production theory (e.g, Becker, 1965, Muellbauer, 1974, Pollak and Wachter, 1975), the household is assumed to derive utility from four broader groups of commodities: Meals prepared at home (M^H), meals prepared away from home (M^A), other market goods (q^o) and leisure (L), conditional on individual/household characteristics (Z) (i.e., U = U(M^H, M^A, q^o, L; Z^U). The other goods (q^o) is the consumption of all other commodities besides food related goods. For simplicity, it is represented by a scalar. The leisure (L) is as scalar as well, measured by the time input only.

The general household utility maximization problem based on the unitary model is represented as the following.

$$\max_{\{q^H, q^A, T^H, T^A, q^o, \mathrm{L}\}} \mathrm{U}(M^H, M^A, q^o, \mathrm{L}; Z^U) \qquad (4.2.1)$$

s.t.

$$M^H = M^H(q^H, T^H;\ Z^H) \qquad (4.2.2)$$

$$M^A = M^A(q^A, T^A;\ Z^A) \qquad (4.2.3)$$

$$p^H * q^H + p^A * q^A + p^o * q^o + \mathrm{w} * (T^H + T^A + \mathrm{L}) = \mathrm{w} * T + I \qquad (4.2.4)$$

Here the equation (4.2.1) is the utility function. The Z^U is the individual and household characteristics related to preferences. The (4.2.2) and (4.2.3) are the meal production functions for FAH(*H*) and FAFH(*A*). It is important to distinguish this Z^U and Z^H, Z^A in the production function. The Z^U and (Z^H, Z^A) contain overlaps, such as age, gender, education, and health status, which have impacts on both the preferences and food productions. However, they each share a unique set of variables as well. Take geographic location, for example. It belongs only to Z^U, but not (Z^H, Z^A). People from different regions may have different taste preferences on food, but, it is hard to imagine people from different regions will have different food production functions/efficiency. On the other hand, there are other characteristics that only belong to (Z^H, Z^A), such as health status. The health status will impact the food cooking functions/efficiency but not necessarily the personal preferences in the short run. The p^H, p^A and p^o are prices of market input for food at home and food away from home and other commodities. The hourly wage rate and the non-labor income is represented by *w* and *I* respectively. The non-labor income may include earnings from investment, unemployment insurance, pension, and social welfare transfer program (SNAP program). The *T* is the total time available for the household. Equation (4.2.4) is the full-budget constraint, which pulls the money budget and time budget together. It implicitly assumes a free transition between working time, leisure time, and food production time. The $\{q^H, q^A, T^H, T^A, q^o, \mathrm{L}\}$ are choice variables. The rest of the variables $\{p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}\}$ are the non-choice variables.

Solving this problem yields the input demand equations for $\{q^H, q^A, T^H, T^A, q^o, L\}$ as functions of the non-choice variables: price, individual or household characteristics, wage rate, total time, non-labor income: $\{p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}\}$. Since $\{q^H, q^A, T^H, T^A\}$ is the main focus of this chapter, we will focus on them only. The system of the demand equations is shown as the following.

$$\begin{cases} q^H = q^H(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}) \\ q^A = q^A(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}) \\ T^H = T^H(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}) \\ T^A = T^A(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I}) \end{cases} \tag{4.2.5}$$

By multiplying the quantity with price, the dependent variable in the first two equations of system (4.2.5) can be transformed into expenditures demand function ($E^H = p^H * q^H = E^H(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I})$, $E^A = p^A * q^A = E^A(p^H, p^A, p^o, Z^H, Z^A, Z^u, w, \mathrm{T}, \mathrm{I})$). Also, in the cross-sectional setting the price vector (p^H, p^A, p^o) is often assumed fixed for all households, which can then be dropped in the system of equations due to the lack of variations. Also, the total time (T) is constant across households, thus dropped as well. Rewriting the system of input demands above and combining with the production function in (4.1.2), we get the following recursive system:

$$
\begin{cases}
HEI = H(E^H, E^A, T^H, T^A, Z^H, Z^A) \\
\quad E^H = E^H(Z^H, Z^A, Z^u, w, \mathrm{I}) \\
\quad E^A = E^A(Z^H, Z^A, Z^u, w, \mathrm{I}) \\
\quad T^H = T^H(Z^H, Z^A, Z^u, w, \mathrm{I}) \\
\quad T^A = T^A(Z^H, Z^A, Z^u, w, \mathrm{I})
\end{cases}
\tag{4.2.6}
$$

As mentioned in the introduction, if the first equation is estimated directly, there will be an endogeneity problem with (E^H, E^A, T^H, T^A). However, combining this HEI function with the second to the fifth function as a system offers theoretical guidance on the selection of instrument variables of the endogenous variables. Here the set of variables ($Z^u, w,$ I) can be a valid set of instrumental variables for the estimation of HEI in the first equation in the system. The set of IVs appears in the second to the fifth equations of the system (4.2.6), which indicates the relevance of the endogenous variables. By not appearing in the first equation of the system (4.2.6), the exclusion identification restriction is satisfied. That is, given the money input and time input, the utility-specific characteristics (Z^u) such as geographic information, and the wage rate (w) and income (I) do not impact the HEI. The above system offers a theory based guidance to the choice of IVs empirically in the estimation of HEI production function.

4.2.2 Structures of HEI Production Function

Based on the theoretical analysis above, this section summarizes three forms of HEI production functions used in the empirical analysis: direct, indirect, and hybrid HEI function. Variables of interest in different HEI functions will have different impacts on the HEI, through preference, or/and meal production function or/and budget constraints. Consequently, the interpretation of a marginal effect will and should differ depending on what structure is estimated. This will be discussed in detail in the following sub-section.

4.2.2.1 Direct HEI Function

The direct HEI is the estimation based on the first equation of the (4.2.6) only:

$$HEI = H_1(E^H, E^A, T^H, T^A, Z^H, Z^A) \qquad (4.2.7)$$

The main explanatory variables are food expenditures (E^H, E^A), total time input on food (T^H, T^A) and production related characteristics (Z^H, Z^A). The Z^H and Z^A are the characteristics that impact home cooking and away from home consumption including (e.g., age, gender, race/ethnicity, number of kids, and education). The subscript 1 in H_1 is used to help distinguish the direct HEI function (4.2.7) from later variations of the HEI functions.

The derivative with respect to one of the characteristics of this function shows the *direct* marginal impact of the characteristic on HEI, *ceteris paribus*; holding the level of the inputs (E^H, E^A, T^H, T^A) constant. Take education for example. The derivative with respect to education is represented as the following.

$$\frac{\partial HEI}{\partial edu} = \frac{\partial H_1}{\partial edu} \qquad (4.2.8)$$

The equation (4.2.8) shows the partial impact of education on HEI conditional on the four input variables and other characteristics. However, this is *not* the *total* impact of education on HEI. As shown in the second to fifth equations in the system of (4.2.6), education will also impact the money input and time input, thus impact the HEI *indirectly*. More specifically, the derivation presented in equation (4.2.8) is the *direct* impact of education on HEI based on this direct HEI production function, controlling for the four endogenous variables: money expenditure and time inputs (E^H, E^A, T^H, T^A).

4.2.2.2 Indirect HEI Function

When *all* demand functions for the endogenous variables (E^H, E^A, T^H, T^A) are substituted into the direct HEI function this yields the indirect HEI function (e.g., Diewert 1982). In this case, the first equation of (4.2.6) becomes:

$$HEI = H_1\begin{pmatrix} E^H(Z^H, Z^A, Z^u, w, \mathrm{I}), E^A(Z^H, Z^A, Z^u, w, \mathrm{I}), \\ T^H(Z^H, Z^A, Z^u, w, \mathrm{I}), T^A(Z^H, Z^A, Z^u, w, \mathrm{I});\ Z^H, Z^A \end{pmatrix} \qquad (4.2.9)$$

or simply

$$HEI = H_2(Z^H, Z^A, Z^u, w, \mathrm{I}) \qquad (4.2.10)$$

As shown in the above equation (4.2.9), the HEI function is a composite function with endogenous variables as functions of exogenous variables. It is simplified into the equation (4.2.10), called the *indirect* HEI function. According to Rosenzweig and Schultz (1983), this *indirect* HEI function combines health production technology, preference, and budget constraints. Following the similar logic in You and Davis (2010), the direct HEI function (the first equation of the system 2.6) can be treated as part of the structural equation of the HEI. It shows a pure production technology effect. On the other hand, the *indirect* HEI function is a reduced form function of the HEI. It is a mixture of production technology, preference, and budget constraints.

Due to this difference of mechanism in generating these two HEI functions, the interpretation of the partial effect varies greatly from these two functions. The *direct* HEI function uncovers the direct impact of education on HEI production, holding the endogenous inputs constant. The *indirect* HEI function captures the total effect (both direct and indirect) of education on HEI because it does not hold the endogenous variables constant (i.e., control for them directly). Because of the composite function in equation (4.2.9), the interpretation of the derivative of the characteristics of this *indirect* HEI function differs from the interpretation of the derivative from the *direct* HEI function. Take the education, for example, which appears in five different places in equation (4.2.9). Theoretically, it shows that the education impacts the HEI from five directions: one directly from the production-related characteristics and four indirectly through changes in the endogenous expenditure and time input variables.

This distinction between the *direct* and *indirect* impact has huge empirical interpretation and corresponding policy implications. From equation (4.2.9), one can find that education (edu) not only impacts the HEI production function directly ($edu \in (Z^H, Z^A)$), but also impacts indirectly through the choices of the inputs levels E^H, E^A, T^H and T^A. The derivative of education in equation (4.2.9) can be decomposed into the following five parts (given the wage rate fixed) based on the chain rule effect in equation (4.2.11):

$$\frac{\partial HEI}{\partial edu} = \frac{\partial H_2}{\partial edu} = \frac{\partial H_1}{\partial E^H} * \frac{\partial E^H}{\partial edu} + \frac{\partial H_1}{\partial E^A} * \frac{\partial E^A}{\partial edu} + \frac{\partial H_1}{\partial T^H} * \frac{\partial T^H}{\partial edu} + \frac{\partial H}{\partial T^A} * \frac{\partial T^A}{\partial edu} + \frac{\partial H_1}{\partial edu} \qquad (4.2.11)$$

This partial derivative reveals in detail how the change of education impact the HEI through HEI production technology and the input demands.[18] The first and second parts are the indirect impact of education through the channels of FAH and FAFH expenditures. The third and fourth parts are the indirect impact of education through the channel of FAH and FAFH time inputs. Each of these indirect impacts contains two parts in specific. Take $\frac{\partial H_1}{\partial E^H} * \frac{\partial E^H}{\partial edu}$ for example, the $\frac{\partial H_1}{\partial E^H}$ is the change of input on HEI through the production technology. The $\frac{\partial E^H}{\partial edu}$ is the impact of education on the FAH expenditures, taking into account optimization subject to both the technology constraint and budget constraint. The last part is the direct impact of education on the food production function, through health production technology. This last part is the same to equation (4.2.8).

Compared to the equation (4.2.8) with only the direct effect, equation (4.2.11) contains four more indirect effects of education on HEI through the inputs channel. From these two equations,

[18] The education is considered as a continues variable here. In the empirical part, the education is treated as a dummy variable due to the data availability. The general idea to illustrate the difference between direct and indirect impact of education on HEI still holds here.

it is understandable that empirical research estimating a direct HEI production function obtains different partial effects from research estimating indirect HEI production function.

It is possible to compare the sign or magnitude of the partial derivation of the direct and indirect HEI function depending on the variable and the targeting population under discussion. Take education for example. The difference between the partial derivative from the indirect function to the direct HEI function can be expressed as the following:

$$\frac{\partial H_2}{\partial edu}-\frac{\partial H_1}{\partial edu}=\frac{\partial H_1}{\partial E^H}*\frac{\partial E^H}{\partial edu}+\frac{\partial H_1}{\partial E^A}*\frac{\partial E^A}{\partial edu}+\frac{\partial H_1}{\partial T^H}*\frac{\partial T^H}{\partial edu}+\frac{\partial H}{\partial T^A}*\frac{\partial T^A}{\partial edu} \qquad (4.2.12)$$

This difference is exactly the indirect impact of education on HEI through input demands. Take the first part $\frac{\partial H_1}{\partial E^H}*\frac{\partial E^H}{\partial edu}$ for example. The first term ($\frac{\partial H_1}{\partial E^H}$) is the direct impact of FAH expenditure on HEI. In general, this would be positive ($\frac{\partial H_1}{\partial E^H}>0$), as more expenditure on FAH indicates better diet quality. The second term ($\frac{\partial E^H}{\partial edu}$) is the impact of education on money expenditure on FAH through preference. If, for example, one would prefer to spend more money on FAH as education increases, this term will be positive as well. The same analysis can be conducted for each component of the equation (4.2.12). If all components turn out positive, in the empirical analysis, we would expect to find the partial derivative of the indirect HEI function to be higher than the partial derivation of the direct HEI function.

However, it is not always easy to justify a sign for each term. Take a different characteristic AGE, for example. One can derive the same partial derivation of age on HEI to obtain similar difference function of equation (4.2.12) concerning AGE. The sign of the second term ($\frac{\partial E^H}{\partial age}$) may be uncertain. As age increases from infant to adult and then to senior, the expenditure may be an inversed U-shape, increasing first then decreasing. In other words, depending on the population sample of focus and the magnitude of each part in equation (4.2.12), the difference between the partial derivative of AGE based on the indirect HEI function and the direct HEI function can be larger or smaller or equal to zero.

4.2.2.3 Hybrid HEI function:

The *hybrid* HEI function is an HEI function that includes both endogenous input demand variables and exogenous determinants of the input demand variables. This *hybrid* function is constructed by substituting out a subset of the endogenous input demand variables with their demand function. Depending on the endogenous variables substituted, one can get different combinations of explanatory variables in the hybrid HEI function. The more the endogenous variables are substituted, the more the function is closer to the indirect HEI function. Again, according to Rosenzweig and Schultz (1983), this *hybrid* HEI function also conflates information on HEI production technology, preference, and budget constraints.

For example, if the two input demand functions (E^H, E^A) are substituted into the direct HEI function (equation 2.9). It becomes:

$$HEI = H_1(\, E^H(Z^H, Z^A, Z^u, w, \mathrm{I}), E^A(Z^H, Z^A, Z^u, w, \mathrm{I}), T^H, T^A;\; Z^H, Z^A) \qquad (4.2.13)$$

$$HEI = H_3(T^H, T^A, Z^H, Z^A, Z^u, w, \mathrm{I}) \qquad (4.2.14)$$

Here T^H, T^A are the endogenous variables, and $(Z^H, Z^A, Z^u, w, \mathrm{I})$ are the exogenous variables. The equation (4.2.14) is the simplified version of equation (4.2.13). When a researcher estimates such hybrid HEI function, the interpretation of the partial derivative of the characteristics also differs from the direct and indirect HEI function. Based on the chain rule effect on equation (4.2.13), the partial derivative of education is:

$$\frac{\partial HEI}{\partial edu} = \frac{\partial H_3}{\partial edu} = \frac{\partial H_1}{\partial E^H} * \frac{\partial E^H}{\partial edu} + \frac{\partial H_1}{\partial E^A} * \frac{\partial E^A}{\partial edu} + \frac{\partial H_1}{\partial edu} \qquad (4.2.15)$$

The first and second term of the equation (4.2.15) is the indirect impacts of education on HEI through the money input in FAH and FAHE, respectively. The third term is the impact of education on HEI through the production function directly, which is the part in equation (4.2.8). The marginal effect of education on HEI based on such hybrid HEI function offers the combined effects of direct and indirect effect, given the time input on FAH and FAFH fixed. It is important to notice that this equation (4.2.15) is exactly the three parts of equation (4.2.12). As a result, the partial derivative of the hybrid HEI function reveals part of the partial derivation from the indirect HEI function. For the same reason discussed in the last part of chapter 4.2.2.2, the sign and magnitude of this partial derivative vary from the partial derivative from the direct and indirect HEI function.

4.2.2.4 Summary of partial derivatives on different production functions.

Based on the analysis above, a general understanding of the partial derivatives on different production function is summarized in a general case represented in the following function.

When the variable of interest y is a function of the endogenous variable (x) and characteristics (z) ($y = f(x, z)$), where x is a function of z and other factors w ($x = x(z, w)$). As a result, the function of y can be represented by the following three forms:

$$y = f_1(x, z) \qquad (4.2.16)$$

$$y = f_1(x(z, w), z) = f_2(w, z) \qquad (4.2.17)$$

Both x and z can be a set of vectors with more than one variables. For the above case of HEI as the dependent variable, the endogenous variables (x) are money expenditures and time inputs. The characteristics (z) can be the education or the other factor such as age. When different explanatory variables are picked along with z, the function reveals a different effect of z on y. These effects of z on y can be summarized by the following figure 4.2.1.

(Insert figure 4.2.1 here)

When the function (4.2.16) is estimated, the partial effect of z on y is the effect through channel 1 in figure 1, represented by an arrow with a solid line. This is the direct impact of z on y, given the x fixed. When the function (4.2.17) is estimated, according to the chain rule, the partial effect of z on y is $\frac{\partial y}{\partial z} = \frac{\partial f_1}{\partial z} + \frac{\partial f_1}{\partial x} * \frac{\partial x}{\partial z}$. It is a combination of the direct effect (through channel 1) and the indirect effect (through channel 2 and 3). The direct effect is represented by an arrow with solid line and the indirect effect is represented by a dashed line.

The function of (4.2.16) and (4.2.17) are the direct and indirect HEI function, respectively. When part of the x is included, and the rest part of the x is not included as the explanatory variables, we get the hybrid HEI function. The omission of one of the endogenous variable, for example x_1, leads to one indirect effect of z on y, picked up by the $\frac{\partial y}{\partial z}$ through $\frac{\partial f_1}{\partial x_1} * \frac{\partial x_1}{\partial z}$. The more the endogenous variable is omitted, the more the indirect effect picked up by the $\frac{\partial y}{\partial z}$.

This chapter will focus on the estimation of the direct HEI function, instead of an indirect or hybrid production. It focuses on the pure production technological effect of the right-hand-side variables on the HEI, separating from the preferences and constraint effects (Rosensweig and Schultz 1983).

4.3 Theory-Guided Literature Review

There are numerous empirical estimates of the impact of different variables on the diet quality (HEI) (Beydoun and Wang, 2008; Drewnowski et al., 2016; Gibbs et al., 2017; Guo, Warden, Paeratakul and Bray, 2004; Kuczmarski et al., 2016; Manios et al., 2009; McNaughton, Ball, Crawford and Mishara, 2008; Mullie, Clarys, Hulens and Vansant, 2010; Rehm, Monsivais, Drewnowski, 2015; Schroder, Marrugat and Covas, 2006; Zoellner et al., 2011). Most papers do not use the theory to guide their empirical model specifications, which are, therefore, ad hoc. This makes it difficult to know exactly what they are trying to estimate (direct, indirect, or hybrid), other than the effect of a variable. With the above theory, we can identify different categories of the HEI functions, which provides a good framework for readers to understand the effect that is estimated. We summarize the literature according to this framework, as shown in table 4.3.1 below. This categorization helps in synchronizing the interpretation of the empirical results found in the literature. As the derivative in the theoretical part (chapter 4.2.2) indicates, the partial derivative of the same variable on different HEI functions summarizes different effects from different channels. There is no reason to believe that the partial derivatives from those different production functions will be the same in either the sign or the magnitude even with the same data sets.

(Insert Table 4.3.1 here)

4.3.1 Direct HEI function Category

The main difference between the direct HEI function and indirect one is that the direct HEI function contains the endogenous variables (i.e., choice variables) including both food expenditure and time input while the indirect production contains none. Due to the lack of sufficient data, none of the papers so far have estimated the direct HEI function yet. By merging the FoodAPS and ATUS-EHM datasets, we can estimate this direct HEI function in this chapter with a closer look at the direct effect of the inputs and those characteristics on the HEI through the production technology, separating from the preferences and budget constraints. A closer comparison of the estimated HEI function with the indirect HEI function and the hybrid HEI function will be offered in the following sub-section.

4.3.2 Indirect HEI function Category

Before the discussion of indirect HEI function, there are certain criteria to be relaxed. First, the indirect HEI function should contain no endogenous variables. However, it is hard to justify a strict exogenous variable except for age, gender, and races. When the scope of the question changes, variables may change from the endogenous variables into the exogenous variables. For example, education is an endogenous variable in the dynamic utility maximization problem with the school decision as the choice variable in a certain age range. However, if the individuals of interest is an adult who has finished schooling or the research time frame is short or static, it is reasonable to say that education is an exogenous or predetermined variable and thus we have conditional demand functions similar to the one in Pollak (1971). Second, an ideal indirect HEI function, at least coming from the theoretical structures presented here, should include the wage rate (w) as well. When wage rate is not available, the common proxy used in the literature for wage rate are household income (Drewnowski et al., 2016; McNaughton et al., 2008; Zoellner et al., 2011;)[19], property value/tax (Drewnowski et al., 2016), income category (Mullie, Clarys, Huens and Vansant, 2010), Poverty to Income Ratio (Kuczmarski et al., 2016 and Rehm, Monsivais and Drenowski, 2015), employment status (Maniso el at., 2009), and household income per capita (Beydoun and Wang, 2008).

Several of those studies estimated what is considered in our framework as an indirect HEI function (Beydoun and Wang, 2008; Drewnowski et al., 2016; Maniso el at., 2009; Mullie et al., 2010) with exogenous variables and no endogenous variables. These papers show a positive connection between the wage rate proxies and the HEI. Maniso et al. (2009) estimated the HEI-2005 for children aged 2-5 in Greece and found that the older, male baby living in the rural/small towns with higher maternal education enjoys higher HEI than younger female baby living in the larger urban area with lower maternal education.

[19] Income would show up if the labor decision is fixed and thus the earned income is simply the multiplication of the wage rate and the working time ($w*T_w$).

Although the papers included education and age as the control covariates, not all of them disclose the estimate for education and age in the paper. With papers that do disclose the parameter of education and age, both have a positive correlation to the HEI. Mullie et al., (2010) found an individual with a college, or higher degree will have a higher HEI score of 1.59 to 2.27, depending on the individual's obesity level. Drewnowskl et al. (2015) found that a college or higher education will have a higher HEI score of 4.76, comparing to a high schooler and other less educated individuals. Manios et al. (2009) show that the baby of one-year-old enjoys an average of 2.05 higher in HEI score.

4.3.3 Hybrid HEI function Category

The hybrid HEI function contains both the exogenous and endogenous variables. The main endogenous variables treated in the empirical analysis are dieting and skipping breakfast (Woodruff et al., 2008), smoking (Kuczmarski et al., 2016), drinking (Guo et al., 2004), SNAP participation (Zeollner et al., 2011), property (home or/and vehicle) ownership status (Drewnowski et al., 2016; Scharadin, 2017). Similar to the indirect HEI function, the price was considered fixed for cross-sectional data, and the wage rate was approximated by other income related measurement such as household income (McNaughton et al., 2008; Scharadin, 2017; Zoellner et al., 2011) and poverty-to-income ratio (Kuczmarski et al., 2016).

This hybrid HEI function allows more types of variables, including both endogenous variables and exogenous variables. Due to its flexibility in empirical specifications, most of the empirical work falls into this category (Kuczmarski et al., 2016; Guo et al., 2004; McNaughton et al., 2008; Scharadin, 2017; Zeollner et al., 2011). There are three main differences to the empirical finding of the indirect HEI function estimation in the literature. First, the hybrid HEI function reveals the impact of endogenous choice variables to the HEI. People who smoke have a lower

HEI (Kuczmarski et al., 2016), so do people with a higher BMI (Guo et al., 2004). Participation in SNAP increases diet quality (Zeollner et al., 2011). Secondly, for the papers with parameter information revealed, controlling for those endogenous choice variables leads to no significant impact of wage proxies on the HEI (Kuczmarski et al., 2016; Zeollner et al., 2011) with one exception (Scharadin, 2017). This chapter found a different effect for different proxies included in one model (while residency ownership and income is negatively related to HEI, the vehicle ownership is positively related to HEI). This is consistent with the finding in Case, Lubotsky, and Paxson (2002) that the parent's income effect on the kids' health is attenuated when health-related behavior factor is included. Those kids related behavior includes the regular bedtime and the use of a seat belt. The parent related behavior contains BMI, regular doctor visit, and smoking. Finally, the gender effect on HEI flips to negative, compared to the positive impact in the indirect HEI function. In other words, the female tends to enjoy higher HEI than the male. This change in the effect of gender is consistent with the theoretical discussion above that the partial effect of characteristics on HEI may change between different production function because the different indirect effects are included through different channels.

4.3.4 Interpretation of the Empirical Results Under These Production Categories.

The discussion above offers general guidance in the interpretation of the differences in estimates among different empirical works. When facing the difference in the magnitude or even the sign of the estimates of the same variable in different kinds of literature, researchers have confined their differences to differences in the sample population, or estimation methods, or data. While this is true, this offers very little insight into how the variable of interest impacts the dependent variables. Based on the above theoretical analysis and the corresponding categorization of the literature, this chapter offers a more systematic way in interpreting the difference in estimates, that is, through the breakdown of the partial derivatives into direct and indirect impacts.

Take the wage-proxy variables for example. Wage-proxy variables are included in both the indirect and hybrid HEI functions. The proxies used are income categories (Mullie et al., 2010), Z-score of SES index with combination of income and education (Manios et al., 2009), household income (Drewnowski el at., 2015, McNaughton et al., 2008, and Zoellner et al., 2011), or log of annual household income (Scharadin, 2017). Though not all papers show the effect of such wage proxies, the papers that do reveal the effect show some interesting findings. For the indirect HEI function, the marginal impact of wage-proxy on HEI are all significantly positive (Mullie et al., 2010, Manios et al., 2009, Beydoun and Wang, 2008, Drewnowski el at., 2015). However, for hybrid HEI function, where behavioral/choice variables appear, the previous significantly positive effect of wage-proxy on HEI disappears into no significant effect for all studies (Kuczmarski et al., 2016, Scharadin, 2017, and Zoellner et al., 2011). In other words, the wage-proxy on HEI is negligible when another choice variables are included. This implies the impact of wage-proxy on HEI comes more from the indirect effect through other choice variables.

Another example is the impact of education on HEI. The empirical estimation shows people with college or higher education tend to have higher HEI score than people with lower education. The difference in HEI score is 4.76 in indirect HEI function (Drewnowski et al., 2015), 3.80 in hybrid HEI function (Scharadin, 2017, used HEI-2010 for FAH only) and 2.92 in mincer-hybrid HEI function (Gibbs et al., 2017).[20] All numbers are statistically significant. Scharadin (2017) used endogenous variables, including the time spent on FAH activities, primary/secondary

[20] Here in the mincer-bybrid HEI function is a special form of hybrid HEI function. It considers wage as an endogeneous variable, but is substituted into the equation by other characteristics (w=w(z)).

childcare, smoking decision, and SNAP participation. Instrumental variables were used to predict these four endogenous variables at the first stage to ensure a consistent estimator at the second stage estimation on HEI. Gibbs et al. (2017) controlled for the health condition (such as BMI and diabetes condition), where no instrumental variable is used. The result of this model is potentially biased when the health condition is correlated with the error term. Drewnowski et al., (2015) controlled none, which implicitly estimated the total effect of education, including the effect of those endogenous variables to HEI through the channel of the education. Thus the indirect HEI function reflects a total impact of education on HEI of 4.76. By controlling those endogenous variables, the impact of education reduced for the hybrid HEI functions.

4.4 Datasets

This chapter will briefly introduce the datasets available for this part for a better understanding of the estimation approaches used later. A detailed explanation of the data cleaning and variable constructions is presented in section 6.

4.4.1 The National Household Food Acquisition and Purchase Survey (FoodAPS)

The FoodAPS is a nationally representative survey of American households. It collects unique and comprehensive data about household food purchases and acquisitions for a total of 4,826 households. It is a nationally representative survey on household from Apr. 2012 to Jan. 2013. If food/drinks brought home and used to prepare meals for consumption at home or elsewhere, it is defined as FAH, otherwise FAFH. The main difference between the FAH and FAFH is that if additional meal preparation is needed or not. The detailed expenditure on FAH and FAFH expenditures for seven consecutive days for a household are collected by survey books, interviews, scanning data, and receipts. The expenditure for FAH is recorded at the household level, while the expenditure for FAFH is recorded at the individual level. The primary respondent of the survey is identified as the primary food shopper or meal planner. The main variables available for this chapter are household HEI, money expenditure on FAH and FAFH, HEI production related demographic variables and income ($HEI, E^H, E^A, Z^H, Z^A, Z^u, I$). Though some variables (E^A, Z^H, Z^A, Z^u, I) are available at both the individual level and household level, we will focus the analysis on household level because the HEI and E^H are only available at the household level.

4.4.2 Eating and Health Module of American Time User Survey (ATUS-EHM)

The American Time User Survey (ATUS) is also a nationally representative U.S. time diary survey conducted annually since 2003. It collects the 24-hour recall diary on time used in activities for a randomly select one individual (age>=15) of the selected household. The Eating and Health Module (EHM) is a supplement to the ATUS collected in 2006, 2007, 2008, 2014 and 2015. It contains an additional 38 questions regarding individual eating, drinking, food preparation, physical health, and income information. The survey also asked whether the surveyed individual is the main shopper, or main food preparer, or not. The separation of the main shopper, main food preparer, or not helps in understanding the heterogeneity in the behavior pattern between individual of different roles. The main variables for this chapter are individual and household characteristics, time spent on different food-related events. The time information is available at the individual-event level. For the consistency to the FoodAPS dataset, after the prediction of individual time in FoodAPS, aggregation is needed to transform this individual-event level time information into household level data. More details regarding this transformation will be introduced in chapter 4.5.2.2 in the prediction of time information for

FoodAPS individuals from these ATUS-EHM individuals. To increase the sample size, we use the ATUS-EHM for all five years and treat them as a cross-sectional dataset. [21]

4.5 Two Sample Instrument Variable Estimator

It would be ideal to have a dataset that contains the detailed input data of both money and time (E^H, E^A, T^H, T^A). However, such informative dataset does not exist. Take the previous datasets for example. The FoodAPS contains the food expenditure information (E^H, E^A), but no time information (T^H, T^A). The ATUS-EHM contains the time information (T^H, T^A) but not money expenditure(E^H, E^A). Due to this restriction, a lot of research analyzes the change in HEI based on food expenditure information, ignoring the time input (Beydoun and Wange, 2008, Rehm, Monsivals, and Drewnowski, 2015, and Schroder, Marrugat, and Covas, 2006). As a result, the empirical function estimated becomes this hybrid function:

$$HEI = H(E^H, E^A; Z^H, Z^A) + \varepsilon \qquad (4.5.1)$$

The lack of time input information leads to an omitted variable problem in the context of direct HEI, because the decision of money expenditure and time input are made simultaneously with a household's food choice, which leads to the correlation between the time input and money input. Such omission of time input leads to endogeneity problems of the money expenditure, thus result in the inconsistency of estimated effects of money expenditure.

In this paper, we will use the direct health production function with both money input and time input information.

$$HEI = H(E^H, E^A, T^H, T^A; Z^H, Z^A) + \varepsilon \qquad (4.5.2)$$

[21] The summary statistic on the time spent on food related activity shows no significant change from the year 2006 to 2015. This supports our decision on treat them as cross section data.

As documented in the theoretical part, the HEI is based on the food-related decision in solving the utility maximization problem, the decision on food money input and time input are correlated with other non-meal related decisions, such as the unobservable health literacy, thus leads to endogeneity problem as well.

As mentioned before, a common solution to the endogeneity problem is the use of Instrument Variables (IVs). One can use the IV estimator when endogenous variables are exactly identified, or Two-stage-least-squares (2SLS) estimator in the case of over-identification. However, both estimators require the endogenous variable, the dependent variable, and the instrument variables in the same dataset. It is always a luxury for empirical researchers to obtain a sample set that meets such a requirement.

For the cases when researchers have two independent samples: one containing the dependent variable and some IVs and another containing the endogenous variables and some overlapping IVs, the two-sample two-stage least square estimator (TS2SLS) is developed exactly for such two-sample case. This TS2SLS estimator is a consistent estimator, as long as these two datasets are independent to each other and are drawn from the same population (Inoue and Solon, 2010, assumption a).

One clear advantage of TS2SLS approach in two sample case over the one sample case is the justification of exogeneity of instruments. The two fundamental conditions of a good instrument are the relevance and exogeneity conditions. The relevance condition indicates the instrument should be related to the endogenous RHS variables. The exogeneity condition requires the instrument should not be directly related to the error term. In other words, the instrument is only related to the dependent variable through the RHS endogenous variable. In the two sample cases, because the IVs comes from another independent dataset, the exogeneity condition is satisfied automatically.

The following part introduces the general TS2SLS in a single-equation framework. After that, we will extend the TS2SLS framework to fit the estimation framework of this paper.

4.5.1 General Framework of TS2SLS

The general single-equation linear framework proposed by Augrist and Krueger (1992) is as follows. Suppose we want to know the relationship between a variable Y and X and we think X is likely correlated with the error term (e.g. endogenous).[22] Now we have data on Y from dataset one and we have data on X from dataset two. We have data on Z from both datasets one and two. If we use subscripts to denote datasets, then the data we have is as follows:

$$(Y_1, \quad Z_1) \text{ for dataset 1}$$

$$(\quad X_2, Z_2) \text{ for dataset 2}$$

The possible models are then:

$$Y_1 = X_1\beta + \varepsilon_1 \qquad (4.5.3)$$

$$X_1 = Z_1\pi + \eta_1 \qquad (4.5.4)$$

$$X_2 = Z_2\pi + \eta_2 \qquad (4.5.5)$$

The (Y_1, Z_1) are the dependent variable and instrument variable, only available in dataset 1 and (X_2, Z_2) are the endogenous variable and the same instrument variables, but only available in dataset 2. Here the X_1 is a set of endogenous variables, but not observable in dataset 1, with Z_1 serves as its instrumental variables. The equations (4.5.3) and (4.5.4) indicates the relationship

[22] In a more general case, the X_1 can also contains exogenous variables which serves as its own instrument variable contained in Z_1.

between Y_1, X_1 and Z_1 , if X_1 were observed. However, if π can be estimated from (4.5.5), then X_1 can be estimated from (4.5.4) because Z_1 is known.

Following this logic, Inoue and Solon (2010) proposed the TS2SLS estimator by the following two steps:

- Create a proxy (prediction) for the unobserved regressor of the dataset 1: $\hat{X}_{1,2} = Z_1 * \hat{\pi}_2$, where $\hat{\pi}_2 = \pi(Z_2, X_2) = (Z_2'Z_2)^{-1}Z_2'X_2$.
- Estimate the TS2SLS estimates using $\hat{\beta}^{TS2SLS} = (\hat{X}_{1,2}'\hat{X}_{1,2})^{-1}\hat{X}_{1,2}'Y_1$.

For the predicted variable $\hat{X}_{i,j}$ contains two subscripts (i, j) with i means the data source of variables and j means the data source of the parameters. Here the $\hat{X}_{1,2}$ indicates the predicted value is based on the variable Z_1 from the dataset 1 and the parameter $\hat{\pi}_2$ estimated from the dataset 2 using $(Z_2'Z_2)^{-1}Z_2'X_2$.

Pacini and Windmeijer (2016) extended this framework to a more general framework with two different types of endogenous variables, denoted by superscript 1 and 2 ($X^1$and X^2). The type 1 endogenous variable (X^1) exists only in dataset 1, represented by X_1^1. The type 2 endogenous variable X^2 exists only in dataset 2, represented by X_2^2. In other words, the observed data structure is:

$$(Y_1, X_1^1, \qquad Z_1) \text{ for dataset 1}$$

$$(\qquad X_2^2,\ Z_2) \text{ for dataset 2}$$

Comparing with the Augrist and Krueger (1992), this extended framework contains a mix of endogenous variables from different datasets. Similar to the approach proposed by Inoue and Solon (2010), the predicted endogenous variables in the first step are given by:

$$\hat{X}_{1,.} = \left(\hat{X}_{1,1}^1, \hat{X}_{1,2}^1\right) = Z_1\hat{\Pi}$$

Where $\widehat{\Pi} = (\hat{\pi}_1, \hat{\pi}_2)$, and the estimated parameters are: $\hat{\pi}_1 = \pi_1(Z_1, X_1^1) = (Z_1{}'Z_1)^{-1}Z_1{}'X_1^1$, and $\hat{\pi}_2 = \pi_2(Z_2, X_2^2\ ) = (Z_2{}'Z_2)^{-1}Z_2{}'X_2^2$. Again, the TS2SLS estimator is given by:

$$\hat{\beta}^{TS2SLS} = \left(\hat{X}_{1,.}{}'\hat{X}_{1,.}\right)^{-1}\hat{X}_{1,.}{}'Y_1$$

$$= (\widehat{\Pi}'Z_1{}'Z_1\widehat{\Pi})^{-1}\widehat{\Pi}'Z_1{}'Y_1 \qquad (4.5.6)$$

4.5.2 TS2SLS framework extended for this paper

In this chapter, due to our special data structure and information availability, we estimated the TS2SLS estimator similar to Inoue and Solon (2010) by following Pacini and Windwijer (2016). One difference between our model to these two papers should be kept in mind: our prediction function is non-linear because there is censoring at zero and there will be some aggregation over functions as well.

As discussed in the theoretical part, the endogenous variables of interest are $(E^H,\ E^A, T^H,\ T^A)$. These two set of endogenous variables $(E^H,\ E^A)$ and $(T^H,\ T^A)$ will be tackled differently due to different dataset availability. To simplify the notation, we denote the demographic variable (I, Z^u, Z^H, Z^A, D) as (Z, Z^i, D), where Z is a vector of household-level variables (eg., region, number of kids and household head's characteristics, such as age, gender, race, education, and income), Z^i is a vector of individual-level variables(eg., age, gender, race, education) and D is the day of the week dummy variables.[23] The day of the week, dummy D, is used to capture the heterogeneity in the time use pattern in the time prediction modeling. The distinguishing between (Z, Z^i, D) is important because the $(Z^i, D\ )$will be used in the prediction

[23] By adding this D dummy variables, we are capturing the different behavior pattern by the day of the week. One can also capture the heterogeneity of the daily pattern by focusing on the sub-dataset of the day. The model here focuses only on the change of the intercepts by the different day of the week. Using the sub-dataset reduces sample size for each model, thus reduce the efficiency of the data, therefore is not adopted here.

of individual time input and then aggregated into household time input (T^H, T^A), but not used in the prediction of the household-week expenditure (E^H, E^A).

Denote the FoodAPS as the first dataset, with subscript 1 and the EHM-ATUS as the second dataset, with subscript 2. The (E^H, E^A) is available in the first dataset, thus denoted as (E_1^H, E_1^A). The (T^H, T^A) is available in the second dataset, thus denoted as (T_2^H, T_2^A). The prediction for each of them is documented below.

4.5.2.1 Prediction of expenditure ($\mathbf{E_1^H}, \mathbf{E_1^A}$)

For the endogenous variables (E_1^H, E_1^A), it can be predicted by the traditional 2SLS using the following function:

$$\hat{E}_1^f = Z_1({Z_1}'Z_1)^{-1}{Z_1}'E_1^f \quad for\ f = H, A \tag{4.5.7}$$

This is equivalent to the first stage of 2SLS, where the endogenous variable is predicted by the instrument variables.

4.5.2.2 Prediction of time ($\mathbf{T^H}, \mathbf{T^A}$)

The prediction of the time variables is more complicated. The FoodAPS does not contain observed time input. There are three main challenges. Firstly, the behavior pattern varies between the responsibilities of the individuals in the household (main food preparer vs. the non-main food preparer). Secondly, two of the main food-related activities (grocery shopping vs. meal preparation) requires separate prediction models due to the difference in data availability. For FoodAPS, grocery shopping is documented if such activity happens. However, there is no such information on meal preparation. As a result, different models are adapted for the different amounts of information related to the engagement in the activities. Thirdly, the data availability for EHM-ATUS and the FoodAPS are at a different disaggregation level. The EHM-ATUS documented time spent on different activities at the individual level, while the expenditure and HEI information available at the FoodAPS is at the household level.

The following approaches are used to tackle each of the above challenges. First, separate models are built for main food preparer and non-main food preparer. Second, different functions are used for different activities (i.e., grocery shopping vs. meal preparation) to make the best use of information. Finally, the predicted individual time is aggregated into household time based on household compositions.

The FoodAPS distinguishes the main food preparer vs. the non-main food preparer. It also contains information on whether grocery shopping or FAFH event takes place in the survey data. But no information is available on whether meal preparation happens. The ATUS-EHM contains the time spent on four categories of role-specific activities ($c = 1,2,3,4$) for FAH and one category of FAFH event ($c = \mathrm{A}$). These categories are grocery shopping for main preparer ($c = 1$), grocery shopping for non-main preparer($c = 2$), meal preparation for main preparer($c = 3$), meal preparation for non-main preparer ($c = 4$) and FAFH for all individuals ($c = \mathrm{A}$).

The time variables available in ATUS-EHM are at the individual-category-day level ($t_2^{i,c}$), while the predicted FAH and FAFH time used in the HEI function is at the household-week level ($\hat{T}_{1,2}^H, \hat{T}_{1,2}^A$). The prediction of time on FAH and FAFH will be documented here.

Based on the above category definitions, the prediction function of FAH time will be the summation of the predicted four individual categories at the household-week level:

$$\hat{T}_{1,2}^H = \hat{T}_{1,2}^1 + \hat{T}_{1,2}^2 + \hat{T}_{1,2}^3 + \hat{T}_{1,2}^4 \tag{4.5.8}$$

Here the $\hat{T}_{1,2}^c$ are four FAH sub-categories respectively, with the superscript c representing different categories ($c = 1, 2, 3, 4$) and subscript 1, 2 indicating parameter data source and variable data source (1 is FoodAPS and 2 is ATUS-EHM).

For each predicted sub-categorical household-week level time ($\hat{T}_{1,2}^{c}$), it is aggregated from the predicted individual-day ($\hat{t}_{1,2}^{i,c,d}$) with i representing the household member belongs to the household and d indicating the date of the survey week when such event category happens:

$$\hat{T}_{1,2}^{c} = \sum_{i} \sum_{d} \hat{t}_{1,2}^{i,c,d} \tag{4.5.9}$$

The predicted individual-day time ($\hat{t}_{1,2}^{i,c,d}$) comes from the general prediction function $g^{c}(.)$:

$$\hat{t}_{1,2}^{i,c,d} = g^{c}\big(Z_1, Z_1^i, D_1 = d;\ \hat{\pi}_2^c\big) = g^{c}\big(Z_1, Z_1^i, D_1 = d;\ \hat{\pi}_2^c(Z_2, Z_2^i, D_2)\big) \tag{4.5.10}$$

The prediction function is based on variables from the first dataset and parameters estimated using the second dataset. The day dummy d takes the value of 1 to 7, indicating the different day of the week when the category of event c happens.

There is one additional challenge in the time prediction on FAFH. When more than one household member attends the FAFH, the event is only reported by one of the household members. So the prediction function is:

$$\hat{T}_{1,2}^{A} = \sum_{i} \sum_{d} \lambda_1^{A,d} \hat{t}_{1,2}^{i,A,d} \tag{4.5.11}$$

Here, $\lambda_1^{A,d}$ is the multiplication factor accounting for more than one household members in the event. For example, when it is documented that three of the household members went to the same FAFH event, then the multiplication factor is 3, assuming each household individual spent the same amount of time on the same event.

The individual prediction function is given the same as FAH as:

$$\hat{t}_{1,2}^{i,A,d} = g^{A}\big(Z_1, Z_1^i, D_1 = d;\ \hat{\pi}_2^A\big) = g^{A}\big(Z_1, Z_1^i, D_1 = d;\ \hat{\pi}_2^A(Z_2, Z_2^i, D_2)\big) \tag{4.5.12}$$

After the predicted value $\hat{X}_1 = (\hat{E}^H_{1,1},\ \hat{E}^A_{1,1}, \hat{T}^H_{1,2},\ \hat{T}^A_{1,2}\ )$, the TS2SLS estimator is calculated based on step 2 of Inoue and Solon (2010).

To sum up, the household-week level time is aggregated from the individual-day level time, which is estimated form the individual prediction functions $g^c(.)$. The c takes the value of 1, 2, 3, 4 and A, which represents meal preparation of main food preparer (c=1), meal preparation of non-main food preparer (c=2), shopping of the main food preparer (c=3), shopping of the non-main food preparer (c=4) and eating on FAFH (c=A), respectively. There is a total of five different individual prediction functions to take account of the role difference and information difference.

4.5.2.3 Specification of Individual Prediction function($g^c(.), c = 1, 2, 3, 4, A.$)

Now we are left with the last problem, how to set the individual-day time prediction function$g^c(.)$ for $c = 1,2,3,4, A$. As mentioned above, the FoodAPS does not offer information on whether the meal preparation happens for a particular date, but it offers information on whether the grocery shopping or FAFH event happens. Due to this difference, the prediction function of meal preparation is different from grocery shopping and FAFH event.

For meal preparation ($c = 1,2$), a two-part model (2PM) is used. The 2PM, commonly used in health econometrics, is a model to estimate outcome variable in a two-part context: whether an action is taken (censoring mechanism); and if so, how much is obtained (outcome mechanism) (Cameron and Trivedi, 2005). The 2PM allows the censoring mechanism and outcome to be modeled using separate processes. It contains the first part of probit selection model capturing the probability of preparing a meal on the survey date and the second part of an exponential model capturing the time spent on food preparation given the decision to prepare the food.

The traditional 2PM estimator assumes homoskedastic lognormal of the error terms or a homoskedastic retransformation. An alternative model, a modified two-part model(M2PM) is used here for a heteroskedasticity-robust covariance estimator (Mullahy 1998, and You and Davis 2018). The M2PM estimator is a special version of Non-Linear Least Square (NLLS) estimator, where the orthogonality condition is based on the combination of the error term of the first part and second part, instead of the second part along in the 2PM (Davidson and MacKinnon, 1993). The Duan's smearing factor is used here to adjust the transformation of the exponential function using in the second part.

In the ATUS dataset, individuals i may have zero time spent on activity c on day d ($t_2^{i,c} = 0$). Therefore, the individual prediction modeling is as follows:

$$E\left(t_2^{i,c}\right) = Pr(t_2^{i,c} > 0) * E\left(t_2^{i,c} \middle| t_2^{i,c} > 0\right) \tag{4.5.13}$$

Where:

$$Pr(t_2^{i,c} > 0) = Pr\left(t_2^{i,c} > 0 \middle| Z_2, Z_2^i, D_2; \mu_2^c\right) = \phi\left(Z_2, Z_2^i, D_2; \mu_2^c\right) \tag{4.5.14}$$

$$E\left(t_2^{i,c} \middle| t_2^{i,c} > 0\right) = E\left(t_2^{i,c} \middle| t_2^{i,c} > 0, Z_2, Z_2^i, D_2; \eta_2^c\right) = exp\ (Z_2, Z_2^i, D_2; \eta_2^c) \tag{4.5.15}$$

Equation (4.5.13) is the unconditional expectation of the time spent on event c for individual i. It is the product of the probability of the event happens (equation 4.5.14) and the conditional amount of time spent given the individual decided to do the event on that day (equation 4.5.15). Again, as a reminder, the subscribe 2 indicates the variables are coming from the second dataset ATUS-EHM.

Plugging in the equations (4.5.14) and (4.5.15) into (4.5.13), we get the estimation model for meal preparation for main food preparer and non-main food preparer are:

$$E\left(t_2^{i,c}\right) = \phi\left(Z_2, Z_2^i, D_2; \mu_2^c\right) * exp\left(Z_2, Z_2^i, D_2; \eta_2^c\right) \tag{4.5.16}$$

The parameter (μ_2^c, η_2^c) can be estimated accordingly:

$$\hat{\mu}_2^c = \mu^c\left(t_2^{i,c}, Z_2, Z_2^i, D_2\right), c = 1, 2 \qquad (4.5.17)$$

$$\hat{\eta}_2^c = \eta^c\left(t_2^{i,c}, Z_2, Z_2^i, D_2\right), c = 1, 2 \qquad (4.5.18)$$

Based on the estimated parameters $(\hat{\mu}_2^c,\ \hat{\eta}_2^c)$ using variables from the ATUS-EHM (the second dataset), the meal preparation time for FoodAPS individual can be predicted using:

$$\begin{aligned}\hat{t}_{1,2}^{i,c,d} &= g^c\left(Z_1, Z_1^i, D_1 = d;\ \hat{\pi}_2^c(Z_2, Z_2^i, D_2)\right) \\ &= \phi\left(Z_1, Z_1^i, D_1 = d; \hat{\mu}_2^c\right) * exp\left(Z_1, Z_1^i, D_1 = d; \hat{\eta}_2^c\right), \qquad c = 1,2\end{aligned} \qquad (4.5.19)$$

The intuition is to estimate the parameter using the second dataset and replace the household and individual demographics and event information using the variables in the first dataset.

For the grocery shopping and FAFH event (c=3, 4, A), there is additional information in FoodAPS on whether the activities happen. This, in the context of equation (4.5.14), we know the probability is either 1 or 0. A piecewise function can be adapted to incorporate this additional information.

$$\hat{t}_{1,2}^{i,c,d} = \begin{cases} \exp\left(Z_1, Z_1^i, D_1 = d; \hat{\eta}_2^c\right) & event\ observed \\ 0 & event\ not\ observed \end{cases} \qquad (4.5.20)$$

Here we only need to estimate the parameter$(\hat{\eta}_2^c)$ on the exponential function from the ATUS-EHM dataset from the following equation:

$$E\left(t_2^{i,c} \middle| t_2^{i,c} > 0\right) = \exp\left(Z_2, Z_2^i, D_2; \eta_2^c\right), c = 3,4, A \qquad (4.5.21)$$

To sum it up, the prediction function $g^c(.)$ Takes the following form:

$$g^c(.) = \begin{cases} \phi\left(Z_1, Z_1^i, D_1 = d; \hat{\mu}_2^c\right) * exp\left(Z_1, Z_1^i, D_1 = d; \hat{\eta}_2^c\right) & c = 1,2 \\ \begin{cases} exp\left(Z_1, Z_1^i, D_1 = d; \hat{\eta}_2^c\right) & event\ observed \\ 0 & event\ not\ observed \end{cases} & c = 3,4, A \end{cases} \quad (4.5.22)$$

The parameter $\hat{\pi}_2^c$ is a collection of the parameters estimated from the second dataset.

$$\hat{\pi}_2^c = \begin{cases} (\hat{\mu}_2^c, \hat{\eta}_2^c) & c = 1,2 \\ \hat{\eta}_2^c & c = 3,4, A \end{cases} \quad (4.5.23)$$

4.5.3 Schematic of Algorithm for Estimation

The overall estimation process can be depicted schematically in the following figure 4.5.1.

(Insert Figure 4.5.1 here)

4.5.4 Variance Estimated by Bootstrapping

Both Inoue and Solon (2010) and Pacini and Windwijer (2016) proposed the analytical form on estimating the standard error of the TS2SLS estimator. However, both of them are based on the linear prediction function in the first step. For the non-linear prediction function in our case, the bootstrapping method is used following Bjorklund and Jantti (1997). For each bootstrap sample, a random sample (n=2000) are drawn with replacement from the ATUS dataset, to obtain the $\hat{\pi}_2^c$ and the predicted $\hat{X}_1$. The TS2SLS estimate is calculated based on this $\hat{X}_1$. After repeating these steps by B=1,000 times, the standard error of $\hat{\beta}_{TS2SLS}$ is calculated based on the standard deviation of the bootstrap TS2SLS estimates.

4.6 Variables and Sample Size

4.6.1 Variable Definition and Adjustment

Since we are using two datasets: FoodAPS and ATUS-EHM, some adjustments of variables are required to keep the definition consistent between these two datasets. Table 4.6.1 below listed the definition of the variables in each of the datasets, respectively. The final decisions on which definition is adapted is also listed with certain assumptions if applies.

<Table 4.6.1 here>

The detailed explanation of the variables is summarized below:

HEI: This is the healthy eating index that measures the health level of the food the household purchased over the surveyed week based on the 2010 Dietary Guideline of America. It measures the energy density consumption of 12 food components, 9 of which assess adequacy (1) total fruit, 2) whole fruit, 3) total vegetables, 4) greens and beans, 5) whole grains, 6) diary, 7) total protein foods, 8) seafood and plant proteins, 9) Fatty Acids.) and 3 of which assess moderation (1) refined grains, 2) sodium, 3) empty calories). A score is assigned to each food components depending on the amount consumed proportional to the optimal level. Each component is assigned to a maximum score of 5, 10, or 20, depending on the importance of each component. The maximal possible total score is 100.

This HEI is calculated using code offered by USDA-ERS. [24] It contains three main steps: 1) match the food items into 12 food components based on the coding groups and nutrient database. 2) construct the gram weight of the food items based on reported package size or weight. Imputation of the weight is used when weight information is missing. For FAH, the inedible portions of the food is excluded. 3) calculate the household week HEI index of the food item purchased from the survey week based on the HEI 2010 scoring standard.

It is worth noting that the original HEI (2010) is established in measure food intake, the HEI used here is the measure of food purchase due to the lack of information on the final food intake.

[24] Mancino, L., Todd, J.E. and Scharadin, B., 2018. USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups.

However, this discrepancy does not impact our estimation structure because all other key variables, such as money expenditure and time input are measured at the purchase and preparation domain (see detail later) as well. Therefore, the definition here is consistent with the definition to the other variables.

ATUS-EHM Time: For the prediction of the time use variable, the dependent variable is only available in ATUS-EHM dataset. The construction of times related to FAH activities follows the definition of You and Davis (2011), which is further divided into activities related to meal preparation and activities related to grocery shopping. Activities related to meal preparation ($t_2^{i,c}, c = 1, 2$) contains Food and Drink Preparation (ATUS Code 020201), Food Presentation (ATUS Code 020202), Kitchen and Food Clean-up (ATUS Code 020203). Grocery shopping ($t_2^{i,c}, c = 3,4$) includes: Grocery Shopping (ATUS Code 070101), and Travel Related to Food and Drink Preparation, Clean-up, and Presentation (ATUS Code 180202). Depending on the roles of the individual on the shopping activity and meal preparation in the household, the individual is categorized into a main/non-main shopper and a main/non-main meal preparer. The time for FAFH ($t_2^{i,A}$) includes travel related to purchasing food (non-grocery) (ATUS Code 180703), time spent on purchasing food (non-grocery)[25] (ATUS Code 070103) and waiting associated with eating and drinking[26] (ATUS Code 110201). One interesting observation of this time variable is the rough estimation in reported time spent on FAH and FAFH. The survey reply is an estimation based on 5 minutes unit. For example, people are more likely to reply 5 minutes or 10 minutes, instead of 4 minutes, 6 minutes, 9 minutes, or 11 minutes. As a result, the

[25] This includes buying fast food, paying check for a meal/drink/snack, paying for fast food at drive-through, paying for meal at restaurant, paying the pizza delivery person, pick-up take-out food, placing order, talking to fast food cashier, talking to the waiter

[26] This includes waiting for a table, for the food to be delivered, for the check, for the pizza delivery person, to place an order.

histogram of the time variables will experience spikes at the multiple of 5, such as 5 minutes, 10 minutes, etc. To avoid the skewness caused by outliers, we dropped the ATUS-EHM households with time input above 99% of the each of the three event categories.

Eating and Drinking time is not included for two reasons: Firstly, in this paper, we focus on the production side of meals. More specifically, what matters most is the production/preparation time between the FAH meal and FAFH meal, not the eating time. In other words, the eating time, given the specific individual, is considered fixed, thus does not impact food decision (whether to make the meal at home or purchase from outside). Secondly, the HEI constructed in FoodAPS is calculated based on food purchased, rather than the actual food consumed, for the consistency of these variables, the actual consumption time (eating/drinking time) is not included in the FAH and FAFH meal time.

Food Expenditure: The FoodAPS contains FAH expenditure (E_1^H) from both the event level and item level. An event indicates expenditures for a specific visit to a specific location. However, the expenditure information at the event level contains the cost of non-food items and bottle deposits, and there is no way to separate these. Therefore, we use the expenditure information at the item level instead, which are then aggregated by event ID into event level expenditure and by household ID into household level FAH expenditure. After dropping households with missing expenditures and data abnormalities, there are 3,656 ‘clean’ households or 75.76% of the full household sample with complete FAH expenditure.[27] The item expenditure does not include state and local food sales tax, so the food sales tax rates for the FoodAPS survey states for 2012 were collected and applied to the aggregated event level non-

[27] Here the ‘clean’ household means households with no missing information and no data abnormality.

SNAP-funded expenditures that were generated from item level, as SNAP-funded expenditures are not subject to taxes.[28]

In terms of FAFH (E_1^A), the FoodAPS contains both the item level and event level expenditures. The total FAFH expenditure can be obtained from the event level and we can separate guest expenditures, assuming equal sharing of the expenditure among the event participants. The household level cost is the summation of total FAFH event cost by household.

D_meal and D_grocery: For FoodAPS, the individual who is the main meal planner *or* the food shopper of the household is identified as the primary respondent in the household[29]. For ATUS-EHM, the main meal preparer and main grocery shopper are defined based on two questions. Question 1: Are you the main food preparer? Question 2: Are you the main grocery shopper? "Yes" to question 1 is the main preparer for the meal and "No" is the non-main preparer. Other responds including "split equally", "refused", "don't know" and "not in universe" are dropped. "Yes" to question 2 is the main grocery shopper and "No" is the non-main shopper. Other responds including "split equally", "refused", "don't know" and "not in universe" are dropped. In ATUS-EHM, an individual can be the main grocery shopper, but not the main food preparer.

However, in FoodAPS, we only know if an individual is a primary respondent or not. There is no additional information to distinguish further if the individual is the main meal planner but not food shopper, or vice versa. In this chapter, we treat this primary respondent of the FoodAPS as the main meal planner and the main shopper, assuming the primary respondent is both the main

[28] The expenditure of food paid by SNAP is not taxable, so the sales tax only applies to the non-SNAP expenditure, that is the personal expenditure ($y_h^{FAH_P}$). https://www.fns.usda.gov/snap/retailer-sales-tax-notice
[29] FoodAPS User Guide, page 7.

meal preparer and the main shopper simultaneously. The non-primary respondent is considered non-main meal planner and non-main grocery shopper.

This assumption determines the sample size used in time estimation and estimation using ATUS-EHM. In the empirical analysis of estimating individual-daily time spend on main food preparer and non-main food preparer, represented by time prediction function $g^c(.)(c = 1$ for main food preparer and $c = 2$ for non-main food preparer), the subsample is determined by the D_meal. If n_2^1 of the respondent answers "Yes" to question 1 in ATUS-EMH, then the information of these individuals are used in estimating $g^1(.)$. The information of the rest individual responding "No" to question 1 is used in estimating $g^2(.)$. For the same reason, in the estimation of individual-daily time spend on grocery shopping, represented by time prediction function $g^c(.)$ ($\mathrm{c} = 3$ for main grocery shopper and $\mathrm{c} = 4$ for non-main grocery shopper), the sample size is determined by D_grocery. If n_2^2 of the respondent answers "Yes" to question 2, then the information of those individuals are used in estimating $g^3(.)$ and the observation of the rest individuals are used in estimating $g^4(.)$.

AGE: The FoodAPS reported the individual age at the time of the survey. But in ATUS-EHM, top-coding is applied. Individuals with age between 80 to 84 are coded as 80, while individuals with age at 85 or over are coded as 85. For consistency, the ATUS-EHM top code is adopted.

SEX: Both datasets indicates 1 as male and 2 as female, with small portion reporting "refuse" at the gender column. The final coding set male as the benchmark and assigns 1 for female and 0 for male. Household with one individual of "refuse" answer is dropped.

White/Black/Asian: Both datasets contain multiple races. These three dummy variables are identified to represent the white, black, and Asian group. By setting these three race dummies, the reference race group is the other race groups other than these three, including American Indian, Native Hawaiian, or other races.

USCITIZEN: There are two issues related to this variable. First, the ATUS-EHM citizenship variable is pulled from CPS data, which is 2-5 months earlier. Though it is not the most up-to-date variable, given only small proportion who changes citizenship within 2-5 months, we use this variable to indicate the individual citizenship status at the time of the ATUS survey. Second, this citizenship question is asked for all CPS individuals, but in the FoodAPS this question is only asked for individuals whose birth state is not in the U.S. For the construction of this variable in FoodAPS, we assume individuals whose birth state is in the U.S. is a US citizen. The variable is not included later due to the small variance.

EDU1-5: This is a group of 5 dummy variables indicating different highest education level at the time the survey. Each variable captures the education level of interest: less than high school for EDU1, high school for EDU2, some college for EDU3, college for EDU4 and higher than college for EDU5. The variable takes the value of 1 if the individual's highest education is the education level of interest. EDU1 is dropped in the prediction model to set the highest education less than high school as the base case. Again, education information is collected from the CPS. Assumption of no change of final education level is assumed here.

EMPLOY_IND: This is a dummy variable capture the individual employment status of last week. It equals to 1 if the individual is employed, 0 otherwise (unemployed, looking for job, not in the labor force, or valid skip with age less than 16).

EARNING: It is the individual monthly earning in dollars. The FoodAPS contains the individual monthly income. The ATUS contains individual weekly earning. There are two adjustments made for a consistent variable between two datasets. First, the weekly earning in

ATUS is transformed into monthly earning using weekly earing * 52 / 12. Second, because the weekly earning in ATUS is also censored at the value of $2884.61, the monthly earning in FoodAPS is also censored accordingly to make the maximal value consistent.

HEALTH1-5: This is a group of 5 dummy variables indicating different self-reported health status at the time of the survey with HEALTH1 for excellent health, HEALTH2 for very good health, HEALTH3 for good health, HEALTH4 for fair health, HEALTH5 for poor health. HEATLTH5 is dropped in the prediction model to set the poor level of health as the base case.

BMI: Both datasets calculated the individual BMI based on self-reported height and weight. Household with one individual of missing BMI is dropped.

WEEKDAY1-7: The ATUS-EHM is a one day survey. The WEEKDAY is a set of seven dummy variables indicating the date of the week of the survey day. The i-th variable is turned on if the survey day is the i-th day of the week. However, the WEEKDAY in FoodAPS is more complicated due to the special data structure for different events. For the FAH grocery shopping event and FAFH event, FoodAPS individual reported every event during the seven-day survey. Given this, the conditional predicted time is used if the event happens, other the predicted time should be zero. When the event happens on a particular day, the WEEKDAY of seven dummy variables is generated. For example, if an individual reported FAFH event for every day of the survey week, then there are seven sets of WEEKDAY variables with each of them containing seven dummy variables. In other words, the time for FAFH will be predicted seven times for each day of the event. When two FAFH events are recorded at the same day, the prediction is made only once for that day, given the ATUS time is the time spend on one particular day (not a particular event). For the days, when no such event is reported, the WEEKDAY is coded as missing, with the final predicted time recorded as zero. For the FAH meal preparation event, there is no explicit report regarding if such event happens, the unconditional prediction time based on 2PM is used for each day of the survey week. As a result, there are seven WEEKDAY variables for each of the survey day.

HHSIZE: This is the household size variable. For FoodAPS, this variable is derived from the number of individuals, excluding guests, surveyed during the week within the household. For ATUS-EHM, this variable is directly reported.

REGION1-4: This is a set of 4 dummy variables with 1 representing Northeast, 2 representing Midwest, 3 for South and 4 for West. The region information is derived from the state information based on household residency. D_REGION1 is dropped in the model to set Northeast as the base case.

METRO: The definition of METRO varies slightly in FoodAPS and ATUS-EHM. The FoodAPS defines the METRO equals to 1 if the household resides in a Census Core Based Statistical Area (CBSA) with population size larger than 10,000, while the ATUS defines the METRO takes 1 if the household resides in CBSA with population size larger than 100,000. Despite this slight difference in definition, we still include this METRO variable in Z for two reasons. First, excluding this METRO variable means the food-related behavior pattern is the same between households in the metro area and the non-metro area. This is a dangerous assumption to make given the FAH related event such as grocery shopping, or FAFH event varies between these two groups. Secondly, this METRO variable assumes the household resides in CBSA between population between (10,000, 100,000) to have a similar pattern with another household resides on CBSA with population more than 100,000. This is a relatively weaker assumption compared to the previous assumption of no difference between households from the metro and non-metro area.

HHINCOME1-16: It is a set of 16 dummy variables on the 16 different household annual income intervals. The original FoodAPS household income is a continuous variable, while the ATUS is a categorical variable for family income. The ATUS coding method is adopted. There is one assumption used here: for the ATUS-EHM, the family income equals to household income. This is true under two cases: 1) there is no non-family household member in the ATUS surveyed house. 2) if the non-family household member exists, their income is zero. In the model, the households with income below 5000 (HHINCOME1) is dropped and set as the base case.

It is important to include the previous individual earning even when the household income is included for three main reasons. First, the household income is a household level characteristic, which captures the household's overall income (including teenager's income). The EARN here is an individual level characteristic, which impacts individual food-related behaviors, the HHINCOME is a categorical variable, while the EARN is a continuous variable, which contains more information to improve the estimation efficiency. Third, the EARN does not include the income for the individual who is self-employed, which is included in HHINCOME. These two variables complement each other in offering a complete picture of income.

HHTYPE: This is the variable indicating the ownership of the housing unit. The FoodAPS reports three types of ownership (1-rent, 2-own, 3-others or free). The ATUS-EHM codes the ownership as 1-owned, 2-rented, 3-occupied without payment of case rent. This is drawn from the CPS. To match this dataset, the final definition used is 1-own the house, 0-otherwise. The assumption of no change of housing status in ATUS-EHM after the CPS is assumed as well.

KIDm_n: This is a set of ordinal variables indicated the number of kids with age ranging from [m, n]. This is constructed based on the household member's age information.

SNAP: The ATUS offers a self-reported SNAP status in the past 30 days. However, for the FoodAPS, the SNAP status is confirmed by the administrative match, using the household address and the administrative record of SNAP participant's address. Households observed using SNAP EBT (electronic benefit transfer) card during the 7-day survey period is also defined as a SNAP household. To match the SNAP status of FoodAPS and ATUS-EHM, which is taken 3-5 months after the ATUS survey, it is assumed that no change of SNAP status during the 3-5 months gap between ATUS and ATUS-EHM survey.

4.6.2 Sample size

To simplify the labor division within the household, we focus on the single-headed household. Because the ATUS-EHM surveyed individuals with age 15+, the time prediction of FoodAPS only applies to individuals with age 15+ as well. For all individuals with age less than 15, the time is assumed as 0.

After dropping households with missing FAH and FAFH expenditure and expenditure related data abnormalities, there are 3,549 households (see detail for the construction of the clean subsample in chapter 3.11). These 'clean' households constitute 73.54% of the survey population.

4.7 Prediction Model

The result summaries of these prediction models on the three types of activities (FAFH, FAH: grocery shopping, and FAH: meal preparation) of the ATUS-EHM individuals are presented here in tables 4.7.1, 4.7.2, and 4.7.3, respectively. For the FAH activity, it is more often one individual doing it on behalf of the whole household, so each table is separated into two panels by the roles: if the individual is the main shopper or not for grocery shopping, main meal preparer or not for meal preparation activity. For FAFH, there is no point in separating the individual by roles. Thus, table 4.7.1 is shorter than table 4.7.2 and 4.7.3. Since the time spent on different food activities shows different weekday and weekend patterns, it is important to summarize the time based on different days of the week. The total ATUS-EHM individual-day observations used are 15,527 for grocery shopping (table 4.7.2) and meal preparation (table 4.7.3), and 15,526 for FAFH activity (table 4.7.1), with one individual dropped as an outlier.

<Table 4.7.1 here>

<Table 4.7.2 here>

<Table 4.7.3 here>

All three tables share a similar structure. Take table 4.7.1 for example, it summaries the original time spent on FAFH and the predicted time on the two-part model by each day of the week. It is further separated into two sub-panels: the ATUS-EHM individual and all individuals over the week. The ATUS-EHM individual summarizes the reported time for the surveyed individual on that day. Because each individual is surveyed only on one day, the number of observations varies for each day. The unconditional panel summarizes the average time for every individual, including those with no reported activity of interest. The conditional panel focuses on individuals with positive time spent on FAFH. Thus, the number of observations is smaller than the unconditional panel. It is the number of individuals engaged in FAFH activities.

The left panel of all individual over the week summarizes the predicted time for every surveyed individual for all seven days of the week. Even if an individual is surveyed on Money only, the time is also predicted for Tuesday to Sunday as well. As a result, the number of observations each day is the total number of observation of the original ATUS-EHM individuals. The idea of predicting all individuals over the 7-days week is to make this table more comparable to the prediction results of FoodAPS individuals, which are surveyed over the 7-days of the interview week. The conditional part summarizes the predicted results from the second part of 2PM (the exponential function of equation 4.5.15). This number is available for all predicted individuals, thus it is averaged over all individuals, leading to 100% proportion of non-zeros in the conditional panel.

On average, for the ATUS individuals the engagement rate is 11.86%. Given the decision to engage in FAFH activity, individuals spent around 30 minutes on average. From the daily pattern, people are more likely to spend time on FAFH on Thursday, Friday, Tuesday, and Saturday than the rest of the days (higher proportion of non-zeros). However, given the decision to eat outside, people trend to spend more time on Saturday (31.36 minutes) but least time on Tuesday (23.44 minutes).

The predicted unconditional and conditional times are similar to the actual unconditional and conditional times for FAFH activity. The average predicted unconditional and conditional time are of the same magnitude of the ATUS individuals. The daily pattern of the prediction also shows when eating outside, people trend to spend more time on Saturday (31.94 minutes) and least time on Tuesday (25.33 minutes). This indicates the 2PM is doing a good job for in-sample prediction on FAFH.

Table 4.7.2 summarizes the time spent on one FAH activity, grocery shopping, for the main shopper and non-main shoppers. For the ATUS observations, on average, the main shopper spent two times more than the non-main shopper (the unconditional time: 7.29 mins/day vs. 2.25 mins/day). The engagement rate is 18.1% on average for the main shopper, which is much higher than the 5.48% for the non-main shopper. Although the average conditional time between the main shopper and non-main shopper are quite similar (around 40 minutes per day), the non-main shopper enjoys higher variation over the week, ranging from the lowest 17 minutes on Tuesday to the highest of 85 minutes on Wednesday, while the conditional time for main shoppers are quite consistent over the week. For both the main shoppers, the predicted unconditional and conditional times are similar to the actual unconditional and conditional times for grocery shopping. This indicates the 2PM is doing a good job for in-sample prediction on grocery shopping.

The prediction on grocery shopping for non-meal shoppers seems quite large. This is due to the relatively small proportion of non-zero times (less than 6% on average) available for the 2PM. Take Wednesday for example. There is only one non-main shopper, upon which the prediction function on the conditional part is build. This estimation lacks the statistical power, thus offers no insight.[30] Since there are only three grocery purchasing events by the non-main shoppers in FoodAPS, which is a tiny portion comparing to the overall shopping events. It is presented here mainly for the table consistency.

Table 4.7.3 presents the time spent on another FAH activity, meal preparation, for main meal preparers and non-main preparers. On average, the main preparer spent 31 minutes on meal preparation. The engagement rate is quite high. Around 62% of the individuals participated in this meal preparation. The non-main preparers are less active in meal preparation as expected. They spent one-third of the time, around 11 minutes, with an average engagement rate of 28%. Again, the predicted unconditional and conditional times are similar to the original time. This indicates the 2PM is doing a good job for in-sample prediction on meal preparation.

The table 4.7.4, 4.7.5, and 4.7.6 summarize the out of sample prediction on FoodAPS individuals. The out-sample predicted conditional times is similar to the in-sample prediction on the ATUS-EHM individuals, while the predicted unconditional time is larger than the

[30] The Wednesday is left here mainly for the sake of table consistency.

unconditional time in ATUS-EHM individuals. This difference is due to the greater engagement rate (proportion of non-zero over all individuals) in the FoodAPS group. Take FAFH activity in table 4.7.4, for example, on average, around 41% of the FoodAPS individuals reported FAFH activities each day, 2.5 times larger than the 12% in the ATUS-EHM groups. This leads to a larger unconditional time spent on FAFH overall FoodAPS individuals. The summary statistics on the predicted time in each activity in tables 4.7.4, 4.7.5, and 4.7.6 for FoodAPS individuals are reasonable, compared to the summary statistics in tables 4.7.1, 4.7.2 and 4.7.3 for ATUS-EHM individuals. The predicted individual-day times are then aggregated into the household-week level, which will be used in the following analysis in HEI function.

<Table 4.7.4 here>

<Table 4.7.5 here>

<Table 4.7.6 here>

4.8 HEI Function

Before the estimation of HEI function, it is important to summarize the HEI, money input and time input on food-related activities for different household groups. Table 4.8.1 separates the households based on SNAP participation and the household head employment status. It also offers three panels: the total event, the FAH event, and FAFH event.

<Table 4.8.1 here>

The total event panel summarizes the HEI, money, and time inputs for all food-related activities, including both FAH event and FAFH event. The NonSNAP households enjoy slightly higher HEI and money input than the SNAP households. The largest difference comes from the time input. The NonSNAP households spend 100 minutes less than the SNAP households. The main difference comes from the time spent on FAH ---the SNAP households spent 100 minutes more on FAH each week. This makes sense because the food eligible for SNAP EBT card purchase is less processed, which requires more preparation or cooking when cooking at home.

For the households with different household head employment status, the HEI score is similar between the employed and unemployed. However, money input and time input shows different patterns. The employed households spent more money and time on the FAFH event, while the unemployed spent more time on the FAH event.

The last row summarizes the HEI, money and time input for all households. In total, the household on average spent 422 minutes on food-related activities per week, in which three quarters (315 minutes) is spent on FAH activity (this is similar to the 295 minutes estimated by You and Davis (2018) in table 2) and one quarter is spent on FAFH activities.

Table 4.8.2 summaries the HEI, money, and time input for SNAP and Non-SNAP households based on different HEI score. As the HEI score increases from 0 to 100, for both the SNAP and Non-SNAP households the main increase comes from the increase in FAH HEI score. The HEI score of FAFH reaches and fluctuates within the range of 40 to 50. This means the increase in HEI of FAH is the main source of an increasing HEI. For the FAH event, as the HEI score increases, the total time spent increases for both SNAP and Non-SNAP households, though the increases for SNAP households is larger than the Non-SNAP households. For SNAP households, the money input on FAH also increases as HEI increases. This pattern is not found for Non-SNAP households.

<Table 4.8.2 here>

In summary, the increase in total HEI mainly comes from the increase HEI of FAH. As the HEI of FAH increases, for SNAP households the money input and time input increases simultaneously. For Non-SNAP household, the increase of HEI of FAH is accompanied with more time input, but not necessarily more money input.

Based on the HEI, money input and the predicted time input on FAH and FAFH, respectively, the following system of equation (4.8.1) is used in the empirical analysis. There is one main difference between the system (4.8.1) and the system (4.2.6) proposed in the theoretical part. The system (4.2.6) contains estimation functions for the time input, while in the system (4.8.1) there is no additional estimation function of the time input. The main reason for this difference is that the time input ($\hat{T}^H, \hat{T}^A$) are predicted based on household and individual characteristics, thus is not endogenous anymore.

$$\begin{cases} HEI = H\left(E^H, E^A, \hat{T}^H, \hat{T}^A, Z^H, Z^A; \beta\right) + \varepsilon \\ E^H = E^H(\mathrm{w}, Z^u, Z^H, Z^A) + \epsilon \\ E^A = E^A(\mathrm{w}, Z^u, Z^H, Z^A) + \zeta \end{cases} \tag{4.8.1}$$

An endogeneity test on the money expenditure for FAH and FAFH was also conducted. This test contains two steps. The first step is to obtain the residuals of the regression of money expenditure of FAH and FAFH, respectively, based on the instrumental variables. The second step is the regression of the HEI on the original money expenditures and the residuals, together with other characteristics in the HEI function. If the coefficient of the residual is significant, then there is the endogeneity problem of the corresponding money expenditure. The result shows that the money expenditure of FAH is endogenous, but not the money expenditure for the FAFH.

Table 4.8.3 summarizes the four econometrics models in estimating HEI function for all single-headed households (N=1027).[31] The OLS model estimates the first function in the system (4.8.1) alone, ignoring the endogeneity of the money input (E^H, E^A). The Two-Stage Least

[31] The full table is very big. To simplify the presentation, the table 4.8.3 presented here is the shorted versions of the full estimation table with the key variables of interest.

Square (2SLS) takes into account the endogeneity of the money input, using the second and third equation of the system as a prediction function. The Seemly Unrelated Regression (SUR) estimates all three equations in (4.8.1) as a system, taking into the account of correlations between error terms in the system. The Three-Stage Least Square (3SLS) considers both the endogeneity of money input (E^H, E^A) and the correlation between error terms. The standard error is estimated based on the bootstrapping method. For each of these four models, the results of with and without predicted time (T^H, T^A) are included. The bootstrapping method with replacement in the sample is used in generating the standard error of all estimates. Comparing 2SLS to OLS and 3SLS to SUR, the standard error is slightly larger as expected, due to more relaxed assumptions on the endogenous variables.

<Table 4.8.3 here>

For each of the four econometric models, it seems adding the time input does not increase the explanatory power of the models for all single-headed households. The model selection criteria such as adjusted R-square and BIC do not show an improvement between models with and without time input. Also, the time input variables are not significant in all four econometric models. This indicates if we put SNAP households and Non-SNAP households together, the average effect of time on HEI is not significantly different from zero.

To better understand the difference between the SNAP participants (N=310) and Non-SNAP (N=717) participants, we separate the analysis for the SNAP households from the Non-SNAP households listed in table 4.8.4 and table 4.8.5 respectively.[32] Although the coefficient of money input and time input of FAH are both significant in the SUR model, the BIC criterion is the

[32] Although table 8.3 contains SNAP dummy variable in the estimation of FAH and FAFH expenditure functions, this does not capture interaction terms of the SNAP participation with other demographic features. Splitting the samples into two by SNAP household and Non-SNAP household is considered a better approach, which allows more flexibility in the coefficients.

lowest for 3SLS. This 3SLS model will be the main focus of our discussion. Similar to the result of Venn and Strazdins (2017), the models based on Non-SNAP households (table 4.8.5) indicate a significant impact of time input on the HEI with consistent magnitude across four different models: one additional minute spent on time on FAH leads to around 0.03 increase in the HEI score. Alternatively, the additional time spent on FAFH has no impact on HEI. This indicates that time spent on FAH events is a more valuable investment in improving HEI than on FAFH events for Non-SNAP households.

<Table 4.8.4 here>

<Table 4.8.5 here>

The results for SNAP households (table 4.8.4) show time inputs on FAH have no impact on HEI when controlling for the money input and other characteristics. In other words, the policy to improve the HEI for SNAP households should focusing on other factors, rather than the time input. This different impact of time on the HEI between the SNAP households and the Non-SNAP households may be because SNAP households already spend a lot of time (390 minutes) on FAH each week on average. One additional minute on FAH for SNAP households would not make a significant impact on HEI. On the other hand, the Non-SNAP households only spent 282 minutes on FAH, 108 minutes less than SNAP households. This means the Non-SNAP households are more likely to run into a tight time budget. Thus additional minutes spent on FAH will lead to a more significant impact on HEI.

Except for the above-mentioned difference, models based on SNAP household and Non-SNAP households share similar patterns regarding the money input on FAH and FAFH: the money input on FAH has more impact on HEI. Firstly, the impact of FAH expenditure on HEI is statistically significant, while the impact of the money input on the FAFH is not. Also, the magnitude of the impact of money input on FAH is larger than the magnitude of the impact of the mony input on FAFH. For the analysis on SNAP households, in particular, the gap of the effects between the expenditure on FAH and FAFH are even larger (0.152 vs. 0.043, in table 8.4). Based on this, the policy interest in improving healthy eating should focus on promoting more money input on FAH, especially for SNAP households, given their limited food budget. This is consistent with the literature that FAH is healthier than FAFH.

One advantage of analyzing the system of equation (3SLS) is the ability to decompose the effect of demographic variables on the HEI into the direct and indirect effect and addressing the endogeneity problem of the money inputs (E^H, E^A). The estimation on the OLS model gives the total effect of demographic variables on HEI. Comparing the results from these two models offers significant insights into the demographic variables impact on HEI through different channels.

Recall the equation (4.2.14) (repeated in the equation (4.8.2) below) in the theoretical part, the impact of the characteristics can be decomposed into direct and indirect effect.

$$\frac{\partial HEI}{\partial edu} = \frac{\partial H_3}{\partial edu} = \frac{\partial H_1}{\partial E^H} * \frac{\partial E^H}{\partial edu} + \frac{\partial H_1}{\partial E^A} * \frac{\partial E^A}{\partial edu} + \frac{\partial H_1}{\partial edu} \qquad (4.8.2)$$

For the Non-SNAP households (table 4.8.5), the result in 3SLS shows the impact of a bachelor degree on HEI through three different channels: 1) an indirect impact (1.64 = 19.783* 0.083) through the FAH expenditure function ($\frac{\partial E^H}{\partial edu}$) and the HEI function ($\frac{\partial H_1}{\partial E^H}$); 2) an indirect impact (0.26 = 8.07 * 0.032) through the FAFH expenditure function ($\frac{\partial E^A}{\partial edu}$) and the HEI function ($\frac{\partial H_1}{\partial E^A}$); 3) a direct impact (6.403) through HEI function($\frac{\partial H_1}{\partial edu}$). The indirect impacts indicate that higher education not only improves HEI directly, but also leads to more monetary expenditure on FAH relative to FAFH, thus leads to an indirect increase in HEI. The total impact is 8.30 (=1.64 + 0.26 + 6.40). The indirect impact through FAH and FAFH expenditures account for 20% and 3% respectively of the total effect.[33]

This decomposition of the total impact of education on HEI into indirect and direct effects on HEI helps to better explain the effectiveness of the education and its interaction with food money expenditure programs. For the Non-SNAP households, by advocating more money expenditure on FAH to college students, we are targeting at 22% of the potential increase on HEI. It helps with precise targeting under limited healthy eating promotion budget. However, without the decomposition of the effect based on the system of estimation, one may ignore the indirect effect and use only the direct effect ($\frac{\partial H_1}{\partial edu}$) as the impact of education on HEI. This will significantly underestimate the total impact of education on HEI and ignore this impact of 22%, thus lead to misleading policy recommendations.

4.9 Conclusion and Extensions

A healthy diet is related to lower risk of chronic diseases. The HEI is normally used as a proxy in measuring the health of a diet. Researchers and policymakers are interested in improving the HEI by analyzing the various resources needed in making food decisions. Most researchers analyze only the monetary resources: does more money spend on food improve the HEI? This line of research ignores another important component in the food-related decision: time. The food decision in the home is related to household production, where households spent both money and time to prepare the final meal. When time is ignored from the equation, it leads to an incomplete understanding of the change of resources input on the HEI. From the empirical perspective, this missing time component leads to the omitted variable bias, which causes a biasness in the

[33] Similar result can be found for the bachelor degree in table 8.3 as well. The total impact of bachelor degree on HEI indicated from the 3SLS model is 7.264, with 12% coming from the indirect impact through the FAH expenditure and 7% from indirect impact through the FAFH respectively.

parameters estimated. Apart from that, the majority of the empirical literature is conducted without a theoretical framework. This makes the interpretation of the impact of characteristics on the HEI difficult and ambiguous.

This chapter proposed a more complete analysis of the impact of resource input on the HEI by adding time input into the analysis. Following the household production model (Becker, 1965), the household is maximizing the utility subject to a full constraint, which takes account of both monetary constraint and time constraint. The demand equations of money expenditure and time expenditure on food are derived from utility maximization. A system of equations is established by combining the demand equations and the HEI functions. This system of equations also helps in offering insight on decomposing the direct impact and indirect impact of characteristics on HEI through the HEI function directly or through the demand functions.

Estimating this system of equations requires data of HEI, money expenditure, time, expenditure, and household characteristics. The FoodAPS dataset contains all required information, except the time expenditure. However, the ATUS-EHM contains time expenditure on food and household characteristics. Since both datasets is nationally representative, two-sample two-stage least square estimator (TS2SLS) is adapted to prediction the time expenditure of FoodAPS households based on the time expenditure of the similar household from the ATUS-EHM households.

The result shows the time input on FAH is important for the improvement of HEI for Non-SNAP households, partly due to the fact the Non-SNAP households are more constrained by the time input. But the money and time input on FAFH does not have a significant impact on HEI. No significant impact for money and time input for SNAP household either. The decomposition of the effect of education on HEI shows an indirect effect of 20% and 3% passes through the expenditure on FAH and FAFH, comparing to 77% of the direct impact on HEI.

There are certain directions for future research to extend the current work. Firstly, this research used the unitary model and focused on the single-headed household. Future research can be extended to a dual-headed household. For the dual-headed household, the theoretical framework based on the collective model or non-cooperative model should be considered to better capture the dynamics between the two household heads.

Secondly, the time impact on HEI varies for SNAP household and Non-SNAP household. One potential reason of this difference is due to the different marginal impact of time on HEI at the different time input level. One can potentially test this proposal by adding a second order or third order of the time inputs or using quantile regression.

Finally, since the datasets that are stratified designed, like the case of FoodAPS and ATUS-EHM, the weighting methods should be taken into account in the estimation process if one is more interested in the national representative result.

4.10 Reference

Angrist, J.D. and Krueger, A.B., 1995. Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, *13*(2), pp.225-235.

Ashworth, J. and Ulph, D., 1981. Estimating labour supply with piecewise linear budget constraints. *Taxation and Labour Supply, London: Allen and Unwin*.

Barten, A.P. and Böhm, V., 1982. Consumer theory. *Handbook of mathematical economics*, *2*, pp.381-429.

Becker, G.S., 1965. A Theory of the Allocation of Time. *The economic journal*, pp.493-517.

Beydoun, M.A. and Wang, Y., 2008. How do socio-economic status, perceived economic barriers and nutritional benefits affect quality of dietary intake among US adults?. *European journal of clinical nutrition*, *62*(3), p.303.

Björklund, A. and Jäntti, M., 1997. Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, *87*(5), pp.1009-1018.

Cameron, A.C. and Trivedi, P.K., 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Case, A., Lubotsky, D. and Paxson, C., 2002. Economic status and health in childhood: The origins of the gradient. *American Economic Review*, *92*(5), pp.1308-1334.

Chiappori, P.A., 1988. Rational household labor supply. *Econometrica: Journal of the Econometric Society*, pp.63-90.

Chiappori, P.A., 1992. Collective labor supply and welfare. *Journal of political Economy*, *100*(3), pp.437-467.

Davidson, R. and MacKinnon, J.G., 1993. Estimation and inference in econometrics. *OUP Catalogue*.

Deaton, A. and Muellbauer, J., 1980. *Economics and consumer behavior*. Cambridge university press.

Diewert, W.E., 1982. Duality approaches to microeconomic theory. *Handbook of mathematical economics*, *2*, pp.535-599.

Drewnowski, A., Aggarwal, A., Cook, A., Stewart, O. and Moudon, A.V., 2016. Geographic disparities in healthy eating index scores (HEI–2005 and 2010) by residential property values: findings from Seattle obesity study (SOS). *Preventive medicine*, *83*, pp.46-55.

Gibbs, H.D., Ellerbeck, E.F., Gajewski, B., Zhang, C. and Sullivan, D.K., 2018. The nutrition literacy assessment instrument is a valid and reliable measure of nutrition literacy in adults with chronic disease. *Journal of nutrition education and behavior*, *50*(3), pp.247-257.

Gundersen, C., Baylis, K., Fan, L., House, L. and Dutko, P., 2016. Construction of Weekly Store-Level Food Basket Costs Documentation May 2016.

Gronau, R., 1977. Leisure, home production, and work--the theory of the allocation of time revisited. *Journal of political economy*, *85*(6), pp.1099-1123.

Guo, X., Warden, B.A., Paeratakul, S. and Bray, G.A., 2004. Healthy eating index and obesity. *European journal of clinical nutrition*, *58*(12), p.1580.

Inoue, A. and Solon, G., 2010. Two-sample instrumental variables estimators. *The Review of Economics and Statistics*, *92*(3), pp.557-561.

Kuczmarski, M.F., Beydoun, M.A., Cotugna, N., Daniels, L., Mason, M.A., Zonderman, A.B. and Evans, M.K., 2016. Literacy contributes to greater higher diet quality in a socioeconomically diverse urban prospective cohort. *Topics in Clinical Nutrition*, *31*(1), pp.47-58.

Leuthold, J.H., 1968. An empirical study of formula income transfers and the work decision of the poor. *Journal of Human Resources*, pp.312-323.

Mancino, L., Todd, J.E. and Scharadin, B. 2018. USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups (No. 1488-2018-5472).

McNaughton, S.A., Ball, K., Crawford, D. and Mishra, G.D., 2008. An index of diet and eating patterns is a valid measure of diet quality in an Australian population. *The Journal of nutrition*, *138*(1), pp.86-93.

Manios, Y., Kourlaba, G., Kondaki, K., Grammatikaki, E., Birbilis, M., Oikonomou, E. and Roma-Giannikou, E., 2009. Diet quality of preschoolers in Greece based on the Healthy Eating Index: the GENESIS study. *Journal of the American Dietetic Association*, *109*(4), pp.616-623.

Muellbauer, J., 1974. Household Production Theory, Quality, and the" Hedonic Technique". *The American Economic Review*, *64*(6), pp.977-994.

Mullahy, J., 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of health economics*, *17*(3), pp.247-281.

Mullie, P., Clarys, P., Hulens, M. and Vansant, G., 2010. Dietary patterns and socioeconomic position. *European journal of clinical nutrition*, *64*(3), p.231.

Pacini, D. and Windmeijer, F., 2016. Robust inference for the Two-Sample 2SLS estimator. *Economics letters*, *146*, pp.50-54.

Pollak, R.A., 1971. Conditional demand functions and the implications of separable utility. *Southern Economic Journal*, pp.423-433.

Pollak, R.A. and Wachter, M.L., 1975. The relevance of the household production function and its implications for the allocation of time. *Journal of Political Economy*, *83*(2), pp.255-278.

Rehm, C.D., Monsivais, P. and Drewnowski, A., 2015. Relation between diet cost and Healthy Eating Index 2010 scores among adults in the United States 2007–2010. *Preventive medicine*, *73*, pp.70-75.

Rosenzweig, M.R. and Schultz, T.P., 1983. Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of political economy*, *91*(5), pp.723-746.

Scharadin, B., 2017. *Three Essays on Household Committed Activities and Diet Quality*. The Pennsylvania State University.

Schröder, H., Marrugat, J. and Covas, M.I., 2006. High monetary costs of dietary patterns associated with lower body mass index: a population-based study. *International journal of obesity*, *30*(10), p.1574.

Schwingshackl, L. and Hoffmann, G., 2015. Diet quality as assessed by the Healthy Eating Index, the Alternate Healthy Eating Index, the Dietary Approaches to Stop Hypertension score, and health outcomes: a systematic review and meta-analysis of cohort studies. *Journal of the Academy of Nutrition and Dietetics*, *115*(5), pp.780-800.

Schwingshackl, L., Bogensberger, B. and Hoffmann, G., 2018. Diet quality as assessed by the healthy eating index, alternate healthy eating index, dietary approaches to stop hypertension score, and health outcomes: an updated systematic review and meta-analysis of cohort studies. *Journal of the Academy of Nutrition and Dietetics*, *118*(1), pp.74-100.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans, 2010.* 7th Edition, Washington, DC: U.S. Government Printing Of ce, December 2010.

Venn, D. and Strazdins, L., 2017. Your money or your time? How both types of scarcity matter to physical activity and healthy eating. *Social Science & Medicine*, *172*, pp.98-106.

Vermeulen, F., 2002. Collective household models: principles and main results. *Journal of Economic Surveys*, *16*(4), pp.533-564.

Woodruff, S.J., Hanning, R.M., Lambraki, I., Storey, K.E. and McCargar, L., 2008. Healthy Eating Index-C is compromised among adolescents with body weight concerns, weight loss dieting, and meal skipping. *Body Image*, *5*(4), pp.404-408.

You, W. and Davis, G.C., 2010. Household food expenditures, parental time allocation, and childhood overweight: An integrated two-stage collective model with an empirical application and test. *American Journal of Agricultural Economics*, *92*(3), pp.859-872.

You, W. and Davis, G.C., 2018. Estimating dual headed time in food production with implications for SNAP benefit adequacy. *Review of Economics of the Household*, pp.1-18.

Zoellner, J., You, W., Connell, C., Smith-Ray, R.L., Allen, K., Tucker, K.L., Davy, B.M. and Estabrooks, P., 2011. Health literacy is associated with healthy eating index scores and sugar-sweetened beverage intake: findings from the rural Lower Mississippi Delta. *Journal of the American Dietetic Association*, *111*(7), pp.1012-1020.

4.11 Figures

Figure 4.2.1: Illustration of Impact of z on y Through Direct and Indirect Channels.

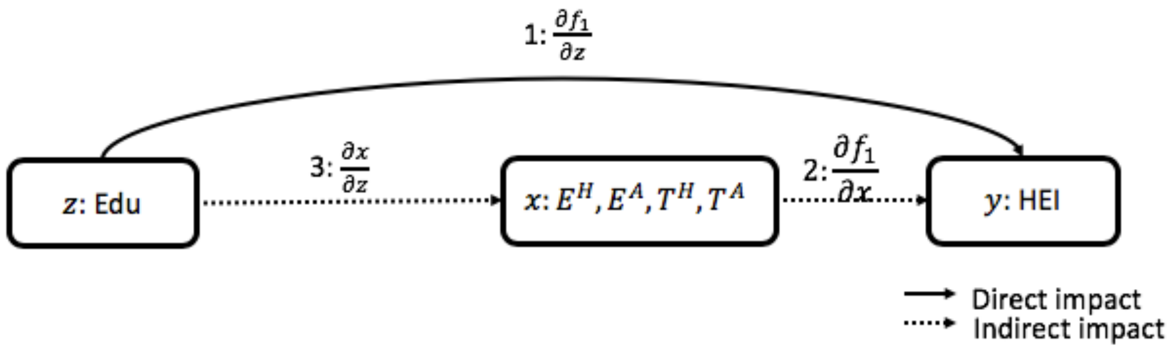
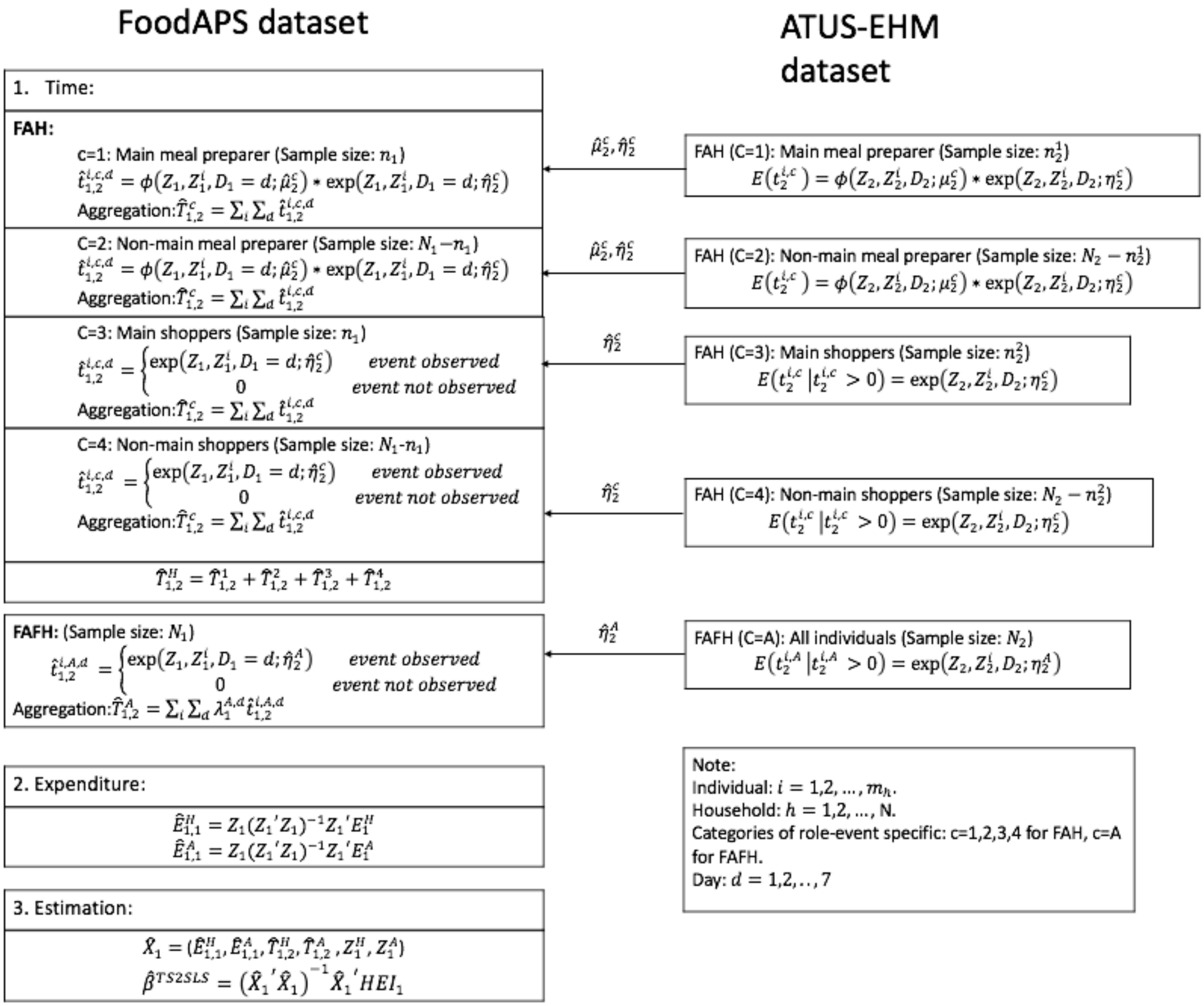


Figure 4.5.1: Time Input Estimation Schematic for FoodAPS



Note:

- c=1,2,3,4,A represents different role-event specific categories (1:meal preparation for main meal preparer; 2: meal preparation for non-main meal preparer; 3: grocery shopping for main shopper; 4: grocery shopping for non-main shopper; 5: FAFH activity for all.
- Sample size: The total sample size for ATUS-EHM is N_2. There are two survey questions asking whether an individual is 1) a main meal preparer or 2) a main grocery shopper. The sample is grouped into n_2^1 main meal preparer and $N_2 - n_2^1$ non-main preparer based on the answer to question 1. They are also grouped into n_2^2 main grocery shopper and $N_2 - n_2^2$ non-main grocery shopper based on the answer to question 2. An individual can be counted as both the main meal preparer and the main grocery shopper in the same time. For the FoodAPS dataset, we only have information to identify the main respondent, who are treated as the main meal preparer and main grocery shopper.

The total N_1 individuals are grouped into n_1 main meal preparer (grocery shopper) and $N_1 - n_1$non-main meal preparer (grocery shopper).

- $d = 1,2, \ldots ,7$ indicated the 7-days in a week. D_1 and D_2 is a vector of 1*7 indicating the day dummies when the role-event specific category c happened for individual i in dataset 1 and 2, respectively.
- The $\hat{t}_{1,2}^{i,c}$ is the predicted time for individual i of role-event specific category c. The first subscript 1 indicating the source of the variables from dataset 1 and the second subscript 2 indicating the source of parameters ($\hat{\mu}_2^c$, $\hat{\eta}_2^c$) estimated from dataset 1.

4.12 Tables

(This page is left empty)

Table 4.3.1. Summary of Production Functions Estimated in the Literature

Type	Articles	Participants	Sample Size	Method	Dependent Variable	**Characteristics Variables**	**Income related variables**	**Endogenous Variables**
	Mullie, Clarys, Hulens and Vansant (2010)	Cross-sectional design in 1852 military men	1,852	Linear Regression	HEI-2010	Age, Education	Income Category	None
	Manios et al. (2009)	Greek Children Aged 2-5	2,287	Linear Regression	HEI-2005	Age, geographic, Maternal Education	Maternal Employment	None
Indirect HEI Function	Beydoun and Wang 2008	CSFII and DHKS (1994–96)	4,356	Locally Weighted Regression	HEI-2005	Age, Gender, Region, Education	Income , taste, convenience, safety and freshness	None
	Drewnowski el at. (2015)	King County adults in 2008–09	1,116	Multivariable regression	HEI 2010	Age, Gender, Education	Home ownership status , Residential property values at tax parcel, HH income)	None

Table 4.3.1 (contd.). Summary of Production Functions Estimated in the Literature

Type	Articles	Participants	Sample Size	Method	Dependent Variable	**Characteristics Variables**	**Income related variables**	**Endogenous Variables**
Hybrid HEI Function	McNaughton, Ball, Crawford and Mishara (2008)	adult of Australian National Nutrition Survey	8,220	Linear Regression	Dietary Guideline Index (DGI)	Age	Income	Smoking, Physical activities, Health Status
	Kuczmarski et al. (2016)	Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS)	2111	Linear Regression	HEI-2010	Age, Gender, Race, Education	PIR	Smoking
	Zoellner et al. (2011)	Adults in rural Lower Mississippi Delta	376	Linear Regression	HEI-2005	Age, Gender, Race, Education, Nutrition Literacy	HH income (0)	SNAP participation
	Guo, Warden, Paeratakul and Bray 2004	NHANES III	10930	Linear Regression	HEI	Age, Gendear, Nutrition Literacy	Income	Smoking, Drinking, Physical activities, Health Status
	Scharadin (2017)	FoodAPS household	4,811	Linear Regression	HEI-FAH	Age, Race, Geographic, Education, Nutrition Literacy	owns residency (-2.14); vehicle (7.45); Ln of annual income (-0.22)	Time in FAH, Time in childcare, Smoking, SNAP participation, Health Status
	Gibbs et al. (2017)	Adults with chronic disease	429	Linear Regression	HEI-2010	Age, Education, Nutrition Literacy	None	Health Status
	Schroder, Marrugat and Covas (2006)				Althernative HEI	Age, Gender	None	Food Cost, Smoking
	Rehm, Monsivais and Drewnowski (2015)	NHENES 2007-2010	11,181	Mean of the age adjusted HEI-2010 by cost quintile	HEI-2010	Age, Gender	None	Food Cost

Table 6.1. Definition of Common Variables of FoodAPS and ATUS-EHM

Table 4.6.1 Common Variable (variable used for predictions)

Variable	Description	Data Type	Unit
Left Hand Side Variables		**Data Type**	**Unit**
HEI	FoodAPS: Healthy Eating index based on 2010 dietary guideline of America	Continuous	
$t_2^{i,c}$, C = 1,2	ATUS-EHM: Time spent on FAH meal preparation	Continuous	Mins
$t_2^{i,c}$, C = 3,4	ATUS-EHM: Time spent on FAH grocery shopping	Continuous	Mins
$t_2^{i,c}$, C = A	ATUS-EHM: Time spent on FAFH for individual i	Continuous	Mins
E_1^H	FoodAPS: Household Weekly FAH Expenditure	Continuous	$
E_1^A	FoodAPS: Household Weekly FAFH Expenditure	Continuous	$
Right Hand Side Variables (Individual Level)		**Data Type**	**Unit**
D_meal	FoodAPS: 1-if the individual is the primary respondent, 0-otherwise	Dummy	
	ATUS-EHM: 1-if the individual is the main meal preparer, 0-otherwise	Dummy	
	Match: 1-if the individual is the main meal preparer, 0-otherwise Assumption: the primary respondent is the main meal preparer in FoodAPS	Dummy	
D_grocery	FoodAPS: 1-if the individual is the primary respondent, 0-otherwise	Dummy	
	ATUS-EHM: 1-if the individual is the main grocery shopper, 0-otherwise	Dummy	
	Match: 1-if the individual is the main grocery shopper, 0-otherwise Assumption: the primary respondent is the main grocery shopper in FoodAPS	Dummy	
AGE	FoodAPS: age at the time of the survey	Continuous	yrs
	ATUS-EHM: Persons aged 80-84 are assigned a code of 80 and persons age 85+ are assigned a code of 85.	Continuous	yrs
	Match: ATUS-EHM top coding is adapted	Continuous	yrs
SEX	FoodAPS: 1-male, 2-female, .r-refuses	Dummy	
	ATUS-EHM: 1-male, 2-female, 3-Don't know, refuse	Dummy	
	Matches: 1-Female, 0-Male, .-missing	Dummy	
WHITE	FoodAPS: 1- one race: white (67%), 2-black/African American (15%), 3-American Indian or Alaska Native (1%) 4-Asian(3.76%), 5-Native Hawaiian(0.4%), 6-Other Race (10.4%), 7-Multiple Race (2.4%), .r-refuse (0.2%)	Dummy	
	ATUS-EHM: 100-White only; 110-Black only; 131-Asian only; …Others	Dummy	
	Match: 1-White only, 0-others	Dummy	
BLACK	ATUS-EHM: 100-White only; 110-Black only; 131-Asian only; …Others	Dummy	
	FoodAPS: 1- one race: white (67%), 2-black/African American (15%), 3-American Indian or Alaska Native (1%) 4-Asian(3.76%), 5-Native Hawaiian(0.4%), 6-Other Race (10.4%), 7-Multiple Race (2.4%), .r-refuse (0.2%)	Dummy	
	Match: 1-Black only, 0-others	Dummy	

Table 4.6.1(Cont.) Common Variable (variable used for predictions)

Right Hand Side Variables (Individual Level)		Data Type	Unit
ASIAN	FoodAPS: 1- one race: white (67%), 2-black/African American (15%), 3-American Indian or Alaska Native (1%) 4-Asian(3.76%), 5-Native Hawaiian(0.4%), 6-Other Race (10.4%), 7-Multiple Race (2.4%), .r-refuse (0.2%)	Dummy	
	ATUS-EHM: 100-White only; 110-Black only; 131-Asian only; …Others	Dummy	
	Match: 1-Asian only, 0-others	Dummy	
CITIZEN	FoodAPS: Universe of this question is individuals not born in the U.S. (14.5%).	Dummy	
	ATUS-EHM: Variable drawn from CPS (2-5 months lag)	Dummy	
	Match: 1-U.S. citizen, 0- Not U.S. citizen Assumptions: 1: no change of citizenship for ATUS-EHM individual between CPS and ATUS-EHM. 2. For FoodAPS, individual born in the U.S. are considered as U.S. citizen.	Dummy	
EDU1-5	FoodAPS: <=18: less than HS; [19,20]: HS, no college, [21, 22]: some college; 23: College; 24: Master or higher	Dummy	
	ATUS-EHM: <=17: Les than HS; [20-21]: HS, no college, [30-32] some college; 40-Bachelor; [41-43]: Master, Doctoral degree	Dummy	
	Match: 1-below HS, 2-HS, 3- Some college, 4-College, 5-Higher than College Assumption: no change of education level during the 2-5 months gap between ATUS-EHM and CPS	Dummy	
EMPLOY_IND	FoodAPS: Work status of last week: 1- working at a job or business; 2-with a job or business but not at work; 3-looking for work; 4-not working; .v-Valid skip (age<16)	Categorical	
	ATUS-EHM: work status of last 7 days: 1-Employed (at work); 2-Employed (absent); 3-unemployed (layoff); 4-unemployed (looking); 5-Not in labor force	Categorical	
	Match: labor force status of last week (1-Employed, 0-Unemployed, not in labor force, or Valid Skip, age<16)	Dummy	
EARN	FoodAPS: Individual's monthly reported income or average imputed total gross income (over 5 imputations)	Continuous	$
	ATUS-EHM: Respondent's usual weekly earnings in dollars. Weekly earnings are not available for persons who are self-employed. Top coded at $2884.61.	Continuous	$
	Match based on monthly earning: ATUS-EHM monthly = weekly * 52/12 (assuming smooth earning over the year). FoodAPS is top coded accordingly.	Continuous	$

Table 4.6.1(Cont.) Common Variable (variable used for predictions)

Right Hand Side Variables (Individual Level)		Data Type	Unit
D_HEALTH1-5	FoodAPS: Self-reported general health condition (1-Excellent, 2- very good, 3- good, 4- fair, 5-poor)	Dummy	
	ATUS-EHM: Self-reported general health condition (1-Excellent, 2- very good, 3- good, 4- fair, 5-poor)	Dummy	
	Matches	Dummy	
BMI	ATUS-EHM: Calculated based on self-reported height and weight. Universe: EHM respondent. (9998-Blank; 9999-NIU)	Continuous	
	FoodAPS: Calculated based on self-reported weight and height. Universe: age>=2	Continuous	
	Matches	Continuous	
WEEKDAY1-7	FoodAPS: The weekdays of the interview date (1-Mon. 2-Tue., 3-Wed., 4-Thu, 5-Fri. 6-Sat. 7-Sun)	Dummy	
	ATUS-EHM: The weekdays of the interview date (1-Mon. 2-Tue., 3-Wed., 4-Thu, 5-Fri. 6-Sat. 7-Sun)	Dummy	
D_WEEKDAY1-7	Matches	Dummy	
Right Hand Side Variables (Household Level)		**Data Type**	**Unit**
HHSIZE	FoodAPS: Household Size	Continuous	
	ATUS-EHM: Household Size	Continuous	
	Matches	Continuous	
REGION1-4	FoodAPS: 4-regions defined by the census (1-Northeast, 2-Midwest, 3-South, 4-West) based on the state of residency	Dummy	
	ATUS-EHM: 4-regions defined by the census (1-Northeast, 2-Midwest, 3-South, 4-West) based on the state of residency	Dummy	
	Matches	Dummy	
METRO	FoodAPS (NONMETRO): (1- if household does not reside in a Census core based statistical area (CBSA) of more than 10,000 population, 0-otherwise)	Dummy	
	ATUS-EHM: Dummy variable (1- if the household live in metropolitan area with population >=100,000; 0- otherwise)	Dummy	
	Matches: FoodAPS coding is reversed to match with ATUS-EHM coding. Assumption: FoodAPS household resides in area with population (10,000 to 100,000) shares the same pattern to household resides on the area with population larger than 100,000.	Dummy	

Table 4.6.1(Cont.) Common Variable (variable used for predictions)

Right Hand Side Variables (Household Level)		Data Type	Unit
HHINCOME1-16	FoodAPS: Household average (monthly income)	Continuous	$
	ATUS-EHM (FAMINCOME): Family's total annual income: 1-<5000, 2-(5,000 to 7,499), 3-(7,500 to 9,999); 4-(10,000-12,499); 5-(12,500 to 14,999); 6- (15,000 to 19,999); 7-(20,000 to 24,999); 8-(25,000 to 29,999); 9-(30,000 to 34,999); 10-(35,000 to 39,999); 11-(40,000 to 49,999); 12-(50,000 to 59,999); 13-(60,000 to 74,999); 14-(75,000 to 99,999); 15-(100,000 to 149,999); 16-(150,000 and over)	Dummy	
	Matches: ATUS-EHM coding is used. Assumptions for ATUS-EHM: 1. The family income equals to household income. This is true under two cases: 1) there is no non-family household member in the ATUS-EHM surveyed house. 2) if the non-family household member exists, their income is zero. 2. There is no change of family income during the 3-5 months lag between ATUS-EHM and EHM survey time.	Dummy	
HHTYPE	FoodAPS (HOUSINGOWN): 1-rent; 2-own; 3-others, free	Categorical	
	ATUS-EHM: Type of housing unit (1- Own the house; 2- Rented for cash; 3-Occupied w/o payment of cash rent); Drawn from CPS	Categorical	
	Matches: Type of housing unit (1- Own the house, 0- Does not own the house) Assumption: no change of housing type in 2-5 months gap between CPS and ATUS-EHM.	Dummy	
KID0_1	FoodAPS: # of non-guest in the household with age<1. (constructed)	ordinal	
	ATUS-EHM (HH_NUMKIDS): # of children within the age range living in the household.	ordinal	
	Matches	ordinal	
KID1_2	FoodAPS: # of non-guest in the household with age between [1,2]. (constructed)	ordinal	
	ATUS-EHM (HH_NUMKIDS): # of children within age between [1,2] living in the household.	ordinal	
	Matches	ordinal	
KID3_5	FoodAPS: # of non-guest in the household with age between [3,5] . (constructed)	ordinal	
	ATUS-EHM (HH_NUMKIDS): # of children within age between [3,5] living in the household.	ordinal	
	Matches	ordinal	

Table 4.6.1(Cont.) Common Variable (variable used for predictions)

Right Hand Side Variables (Household Level)		Data Type	Unit
KID6_12	FoodAPS: # of non-guest in the household with age between [6,12]. (constructed)	ordinal	
	ATUS-EHM (HH_NUMKIDS): # of children within age between [6,12] living in the household.	ordinal	
	Matches	ordinal	
KID13_17	FoodAPS: # of non-guest in the household with age between [13,17] . (constructed)	ordinal	
	ATUS-EHM (HH_NUMKIDS): # of children within age between [13,17] living in the household.	ordinal	
	Matches	ordinal	
SNAP	FoodAPS: Current SNAP receipt confirmed by administrative match. It also count households with real SNAP card payment as SNAP participants. (1-Yes, 0-No)	Dummy	
	ATUS-EHM: Household receiving SNAP benefit in the past 30 days. (1-Yes, 0-No)	Dummy	
	Matches: 1-SNAP; 0-Non SNAP Assumption: No change of SNAP participation during the 3-5 months gap between CPS and ATUS-EHM survey.	Dummy	

Table 4.6.2: Summary statistics of variables for time prediction model

Variable	ATUS-EHM (N= 15,527)				FoodAPS (N=1,150)*			
	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Individual level								
age	51.25	0.20	50.85	51.65	52.67	0.94	50.75	54.58
age2	3011.48	20.30	2971.68	3051.28	3110.63	99.77	2907.39	3313.86
sex	0.57	0.00	0.56	0.58	0.57	0.02	0.52	0.62
white	0.77	0.00	0.76	0.78	0.76	0.02	0.71	0.81
black	0.19	0.00	0.18	0.19	0.16	0.02	0.11	0.20
asian	0.02	0.00	0.02	0.02	0.02	0.01	0.00	0.04
Less the High School	0.16	0.00	0.15	0.17	0.11	0.01	0.08	0.14
HS Diploma	0.29	0.00	0.28	0.30	0.26	0.02	0.23	0.30
Some College	0.26	0.00	0.25	0.27	0.31	0.02	0.27	0.36
Bachelor	0.19	0.00	0.18	0.19	0.20	0.02	0.15	0.24
Master or above	0.11	0.00	0.10	0.11	0.11	0.02	0.07	0.14
employ_ind	0.56	0.00	0.56	0.57	0.52	0.02	0.47	0.57
earning	1873.18	25.13	1823.94	1922.43	1958.16	155.74	1640.92	2275.40
Excellent	0.17	0.00	0.16	0.18	0.13	0.02	0.10	0.17
Very good	0.32	0.00	0.31	0.32	0.32	0.02	0.28	0.37
Good	0.31	0.00	0.30	0.32	0.33	0.02	0.28	0.38
Fair	0.15	0.00	0.14	0.15	0.19	0.02	0.16	0.22
Poor	0.06	0.00	0.06	0.07	0.03	0.01	0.02	0.04
bmi	27.56	0.06	27.44	27.68	28.09	0.36	27.35	28.83

Note: * for FoodAPS, the summary is based on individual with age 15+ to be consistent with ATUS-EHM

Table 4.6.2 (Cont.): Summary statistics of variables for time prediction model

	ATUS-EHM (N= 15,527)				FoodAPS (N=1,150)*			
Variable	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Household level								
metro1	0.832	0.004	0.825	0.839	0.846	0.048	0.748	0.944
hhtype1	0.539	0.005	0.530	0.549	0.509	0.030	0.449	0.570
Northeast	0.167	0.004	0.160	0.174	0.129	0.026	0.077	0.182
Midwest	0.258	0.004	0.250	0.267	0.320	0.032	0.255	0.386
South	0.384	0.005	0.374	0.393	0.388	0.039	0.308	0.469
West	0.191	0.004	0.184	0.198	0.162	0.032	0.097	0.226
d_hhincome1	0.051	0.002	0.047	0.056	0.040	0.007	0.026	0.054
d_hhincome2	0.038	0.002	0.034	0.042	0.015	0.003	0.008	0.022
d_hhincome3	0.055	0.002	0.050	0.059	0.057	0.007	0.043	0.071
d_hhincome4	0.067	0.002	0.062	0.072	0.064	0.010	0.043	0.084
d_hhincome5	0.053	0.002	0.049	0.057	0.059	0.007	0.046	0.073
d_hhincome6	0.078	0.003	0.073	0.083	0.112	0.013	0.086	0.138
d_hhincome7	0.092	0.003	0.087	0.098	0.095	0.013	0.069	0.122
d_hhincome8	0.081	0.003	0.076	0.087	0.087	0.014	0.059	0.115
d_hhincome9	0.080	0.003	0.075	0.085	0.080	0.012	0.055	0.104
d_hhincome10	0.062	0.002	0.058	0.066	0.045	0.009	0.026	0.064
d_hhincome11	0.092	0.003	0.087	0.098	0.075	0.012	0.050	0.100
d_hhincome12	0.074	0.003	0.069	0.079	0.079	0.021	0.036	0.122
d_hhincome13	0.065	0.002	0.060	0.070	0.096	0.022	0.052	0.141
d_hhincome14	0.055	0.002	0.050	0.059	0.049	0.013	0.022	0.076
d_hhincome15	0.036	0.002	0.032	0.039	0.033	0.012	0.008	0.058
d_hhincome16	0.021	0.001	0.018	0.024	0.013	0.007	0.000	0.027
kid1	0.005	0.001	0.004	0.007	0.004	0.002	0.001	0.007
kid1_2	0.021	0.002	0.018	0.024	0.008	0.002	0.004	0.012
kid3_5	0.037	0.002	0.033	0.041	0.036	0.006	0.024	0.048
kid6_12	0.114	0.004	0.107	0.122	0.110	0.022	0.065	0.155
kid13_17	0.142	0.005	0.132	0.153	0.087	0.011	0.063	0.110
snap	0.126	0.003	0.120	0.133	0.141	0.014	0.112	0.170
d_employ	2.478	0.014	2.451	2.504	2.568	0.063	2.440	2.697

Note: * for FoodAPS, the summary is based on individual with age 15+ to be consistent with ATUS-EHM

Table 4.7.1: Summary Statistics of Time on FAFH: ATUS-EHM Dataset

	Original ATUS individual in the model					Prediction (all individual over the week)				
	Unconditional		Conditional			Unconditional		Conditional		
	N	Mean (mins)	N of non-zeros	Proportion of non-zeros (%)	Mean (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	1577	3.41	181	11.48%	29.71	15527	3.33	15527	100.00%	29.85
Tuesday	1508	3.02	194	12.86%	23.44	15527	3.12	15527	100.00%	25.33
Wednesday	1550	2.80	167	10.77%	25.95	15527	2.90	15527	100.00%	27.04
Thursday	1530	3.51	198	12.94%	27.13	15527	3.70	15527	100.00%	28.94
Friday	1542	3.77	199	12.91%	29.21	15527	3.82	15527	100.00%	29.79
Saturday	3948	3.99	502	12.72%	31.36	15527	4.00	15527	100.00%	31.94
Sunday	3871	3.18	401	10.36%	30.67	15527	3.05	15527	100.00%	29.89
Total*	15526	3.44	1842	11.86%	29.04	108689	3.42	108689	100.00%	28.97

*One outlier with extreme FAFH time is excluded in the prediction model, such leads to 15,526 observations.

Table 4.7.2: Summary Statistics of Time on Grocery Shopping: ATUS-EHM Dataset

	Main Shopper									
	Original ATUS individual in the model					Prediction (all main shoppers over the week)				
	Unconditional		Conditional			Unconditional		Conditional		
	N	Mean (mins)	N of non-zeros	Proportion of non-zeros (%)	Mean (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	1476	6.06	226	15.31%	39.57	14381	6.02	14381	100.00%	38.68
Tuesday	1405	5.64	212	15.09%	37.40	14381	5.72	14381	100.00%	36.74
Wednesday	1428	5.91	233	16.32%	36.23	14381	5.90	14381	100.00%	35.85
Thursday	1424	6.72	234	16.43%	40.89	14381	6.72	14381	100.00%	40.68
Friday	1433	7.70	269	18.77%	41.04	14381	7.09	14381	100.00%	38.00
Saturday	3646	9.52	818	22.44%	42.45	14381	9.90	14381	100.00%	43.90
Sunday	3569	6.79	611	17.12%	39.66	14381	6.85	14381	100.00%	39.55
Total	14381	7.29	2603	18.10%	40.29	100667	6.89	100667	100.00%	39.06

Table 4.7.2 (Cont.): Summary Statistics of Time on Grocery Shopping: ATUS-EHM Dataset

	Nonmain Shopper									
	Original ATUS individual in the model					Prediction (all non-main shoppers over the week)				
	Unconditional		Conditional			Unconditional		Conditional		
	N	Mean (mins)	N of non-zeros	Proportion of non-zeros (%)	Mean (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	95	2.95	6	6.32%	46.67	1146	99.10	1146	100.00%	968.34
Tuesday	98	0.87	5	5.10%	17.00	1146	40.33	1146	100.00%	437.54
Wednesday	111	0.77	1	0.90%	85.00	1146	220.78	1146	100.00%	11553.56
Thursday	96	1.55	5	5.21%	29.80	1146	11.57	1146	100.00%	130.31
Friday	105	1.62	7	6.67%	24.29	1146	10.89	1146	100.00%	99.82
Saturday	288	3.44	19	6.60%	52.11	1146	86.64	1146	100.00%	760.12
Sunday	283	2.32	16	5.65%	41.06	1146	51.01	1146	100.00%	522.69
Total	1076	2.25	59	5.48%	40.95	8022	74.33	8022	100.00%	2067.48

Table 4.7.3: Summary Statistics of Time on Meal Preparation: ATUS-EHM Dataset

	Main Preparer									
	Original ATUS individual in the model					Prediction (all main preparers over the week)				
	Unconditional		Conditional			Unconditional		Conditional		
	N	Mean (mins)	N of non-zeros	Proportion of non-zeros (%)	Mean (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	1476	31.09	947	64.16%	48.45	14446	31.30	14446	100.00%	47.38
Tuesday	1394	29.08	861	61.76%	47.09	14446	29.39	14446	100.00%	46.06
Wednesday	1432	30.47	943	65.85%	46.27	14446	31.41	14446	100.00%	46.36
Thursday	1439	29.32	940	65.32%	44.89	14446	30.10	14446	100.00%	44.87
Friday	1425	27.90	875	61.40%	45.44	14446	27.39	14446	100.00%	43.63
Saturday	3683	29.45	2163	58.73%	50.15	14446	30.25	14446	100.00%	49.94
Sunday	3597	34.32	2182	60.66%	56.58	14446	34.36	14446	100.00%	55.12
Total	14446	30.73	8911	61.68%	49.82	101122	30.60	101122	100.00%	47.62

Table 4.7.3 (Cont.): Summary Statistics of Time on Meal Preparation: ATUS-EHM Dataset

	Nonmain Preparer									
	Original ATUS individual in the model					Prediction (all non-main preparers over the week)				
	Unconditional		Conditional			Unconditional		Conditional		
	N	Mean (mins)	N of non-zeros	Proportion of non-zeros (%)	Mean (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	101	13.50	29	64.16%	47.00	1081	14.57	1081	100.00%	51.22
Tuesday	114	10.48	33	61.76%	36.21	1081	9.45	1081	100.00%	33.39
Wednesday	118	9.97	31	65.85%	37.97	1081	10.88	1081	100.00%	39.93
Thursday	91	12.07	33	65.32%	33.27	1081	11.97	1081	100.00%	34.02
Friday	117	9.74	36	61.40%	31.64	1081	10.91	1081	100.00%	34.13
Saturday	266	9.12	59	58.73%	41.12	1081	8.36	1081	100.00%	38.74
Sunday	274	14.11	81	60.66%	47.73	1081	16.15	1081	100.00%	55.72
Total	1081	11.35	302	61.68%	40.61	7567	11.76	7567	100.00%	41.02

Table 4.7.4: Summary Statistics of Time on FAFH for Individual with FAFH events: FoodAPS

	Unconditional		Conditional		
	N*	Prediction (mins)	N of non-zeros	Proportion of non-zeros over all individuals (%)	Prediction (mins)
Monday	1125	1.38	431	38.31%	29.44
Tuesday	1125	1.28	440	39.11%	24.66
Wednesday	1125	1.32	478	42.49%	26.60
Thursday	1125	1.72	488	43.38%	28.22
Friday	1125	1.82	507	45.07%	29.43
Saturday	1125	1.73	453	40.27%	31.47
Sunday	1125	1.26	441	39.20%	29.54
Total**	7875	1.50	3238	41.12%	28.48

*The prediction of the 2PM is based on individuals with FAFH events only. So we report the unconditional time for this sub-set of individuals.

**Because the FoodAPS is a 7-day survey, the total is a summary of person-date.

Table 4.7.5: Summary Statistics of Time on FAH Grocery: FoodAPS Dataset

	Main Shopper				
	Unconditional		Conditional		
	N*	Prediction (mins)	N of non-zeros	Proportion of non-zeros over all individuals (%)	Prediction (mins)
Monday	1072	1.62	287	26.77%	40.18
Tuesday	1072	1.48	272	25.37%	38.44
Wednesday	1072	1.74	304	28.36%	37.40
Thursday	1072	1.74	266	24.81%	42.83
Friday	1072	1.93	284	26.49%	39.31
Saturday	1072	2.64	272	25.37%	45.88
Sunday	1072	1.81	276	25.75%	40.39
Total**	7504	1.85	1961	26.13%	40.63

*The prediction of the 2PM is based on individuals with FAFH events only. So we report the unconditional time for this sub-set of individuals.

**Because the FoodAPS is a 7-day survey, the total is a summary of person-date.

Note: only 3 non-shopper conducted the grocery shopping event, thus the non-main shopper group is not presented here.

Table 4.7.6: Summary Statistics of Time on FAH Meal: FoodAPS Dataset

	Main Preparer					Nonmain Preparer				
	Unconditional		Conditional*			Unconditional		Conditional		
	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)	N	Prediction (mins)	N of non-zeros	Proportion of non-zeros (%)	Prediction (mins)
Monday	1072	34.12	1072	100.00%	50.34	53	11.79	53	100.00%	47.21
Tuesday	1072	32.10	1072	100.00%	48.94	53	7.64	53	100.00%	30.77
Wednesday	1072	34.19	1072	100.00%	49.26	53	8.78	53	100.00%	36.80
Thursday	1072	32.79	1072	100.00%	47.68	53	9.83	53	100.00%	31.36
Friday	1072	29.94	1072	100.00%	46.36	53	8.90	53	100.00%	31.46
Saturday	1072	33.13	1072	100.00%	53.07	53	6.64	53	100.00%	35.71
Sunday	1072	37.58	1072	100.00%	58.57	53	13.08	53	100.00%	51.35
Total**	7504	33.41	7504	100.00%	50.60	371	9.52	371	100.00%	37.81

* Conditional is for better comparison to the prediction of the ATUS-EHM dataset, which is not used in the HEI model.

**Because the FoodAPS is a 7-day survey, the total is a summary of person-date.

Table 4.8.1: Summary of HEI, money input and predicted time input by HEI groups for FoodAPS dataset

Groups			Total Event			FAH Event			FAFH Event		
			HEI_total	M_Total ($)	T_Total (Mins)	HEI_FAH	M_FAH ($)	T_FAH (Mins)	HEI_FAFH	M_FAFH ($)	T_FAFH (Mins)
SNAP	Non-SNAP	Mean	51.48	78.1	390.97	51.94	46.99	282.89	43.91	31.11	108.08
		SD	(14.54)	(69.1)	(161.1)	(15.71)	(52.25)	(111.31)	(11.48)	(43.51)	(96.79)
		N	717	717	717	610	717	717	637	717	717
	SNAP	Mean	47.33	76.69	494.87	45.96	56.09	390.92	43.31	20.6	103.94
		SD	(13.11)	(72.38)	(221.09)	(15.01)	(57.26)	(147.89)	(10.59)	(42.47)	(118.65)
		N	310	310	310	288	310	310	254	310	310
Household Head Employment	Unemployed	Mean	50.25	69.42	440.12	50.07	49.9	350.38	42.96	19.52	89.74
		SD	(14.51)	(65.54)	(179)	(15.84)	(53.84)	(129.25)	(11.31)	(36.03)	(94.79)
		N	548	548	548	491	548	548	449	548	548
	Employed	Mean	50.2	87.12	401.98	49.97	49.55	275.59	44.54	37.57	126.39
		SD	(13.95)	(73.87)	(194.73)	(15.63)	(54.11)	(125.99)	(11.11)	(48.91)	(110.19)
		N	479	479	479	407	479	479	442	479	479
Total		Mean	50.23	77.67	422.33	50.02	49.74	315.5	43.74	27.94	106.83
		SD	(14.25)	(70.07)	(187.38) ()	(15.73)	(53.94)	(133.02) ()	(11.23)	(43.45)	(103.83)
		N	1027	1027	1027	898	1027	1027	891	1027	1027

Note: standard deviation in the parenthesis.

Table 4.8.2: Summary of HEI, Money input and predicted time input by HEI_SNAP groups for FoodAPS dataset

Groups			Total Event			FAH Event			FAFH Event		
HEI	SNAP Participation		HEI_total	M_Total ($)	T_Total (Mins)	HEI_total	M_Total ($)	T_Total (Mins)	HEI_total	M_Total ($)	T_Total (Mins)
[0,20)	Non-SNAP	Mean	12.03	29.33	236.36	13.08	7.42	197.99	26.81	21.91	38.37
		SD	(7.86)	(32.3)	(49.25)	(6.71)	(8.88)	(58.57)	(12.05)	(33.2)	(50.62)
		N	8	8	8	5	8	8	4	8	8
	SNAP	Mean	9.01	44.06	416.38	13.93	18.27	382.91	41.55	25.79	33.46
		SD	(12.74)	(58.85)	(172.34)	(.)	(25.84)	(166)	(.)	(33.01)	(6.33)
		N	2	2	2	1	2	2	1	2	2
[20,40)	Non-SNAP	Mean	33.24	57.63	347.05	31.9	30.05	247.29	37.27	27.58	99.76
		SD	(4.82)	(56.2)	(125.36)	(8.23)	(37.11)	(96.84)	(10.24)	(40.45)	(81.32)
		N	142	142	142	105	142	142	118	142	142
	SNAP	Mean	33.89	74.05	496.2	31.04	54.78	385.82	38.21	19.27	110.38
		SD	(4.17)	(65.17)	(248.94)	(8.57)	(58.68)	(138.49)	(8.92)	(27.93)	(147.29)
		N	95	95	95	85	95	95	77	95	95
[40,60)	Non-SNAP	Mean	49.63	79.47	401.67	48.85	47.02	286	43.83	32.45	115.67
		SD	(5.57)	(64.8)	(177.81)	(7.1)	(49.83)	(116.77)	(10.33)	(43.32)	(109.25)
		N	368	368	368	308	368	368	333	368	368
	SNAP	Mean	49.11	72.38	502.44	47.3	53.55	392.68	44.71	18.83	109.76
		SD	(5.71)	(64.12)	(215.87)	(8.4)	(56.01)	(156.89)	(10.35)	(27.68)	(111.06)
		N	162	162	162	151	162	162	136	162	162

Table 4.8.2 (Cont.) : Summary of HEI, Money input and predicted time input by HEI_SNAP groups for FoodAPS dataset

Groups			Total Event			FAH Event			FAFH Event		
			HEI_total	M_Total	T_Total	HEI_FAH	M_FAH	T_FAH	HEI_FAFH	M_FAFH	T_FAFH
HEI	SNAP Participation			($)	(Mins)		($)	(Mins)		($)	(Mins)
[60,80)	Non-SNAP	Mean	67.84	92.87	409.03	67.17	59.9	304.03	48.56	32.97	105
		SD	(5.37)	(83.21)	(145.56)	(9.04)	(62.83)	(102.35)	(11.31)	(47.77)	(82.61)
		N	180	180	180	173	180	180	166	180	180
	SNAP	Mean	66.25	99.68	473.54	65.46	68.42	394.02	48.59	31.26	79.52
		SD	(5.15)	(107.58)	(192.33)	(6.59)	(61.43)	(145.4)	(10.86)	(88.43)	(74.79)
		N	46	46	46	46	46	46	37	46	46
[80, 100]	Non-SNAP	Mean	85.2	85.31	405.88	84.43	67.5	324.08	50.91	17.82	81.8
		SD	(4.97)	(60.86)	(146.21)	(6.6)	(52.99)	(127.15)	(15.96)	(26.76)	(62.4)
		N	19	19	19	19	19	19	16	19	19
	SNAP	Mean	86.34	68.14	451.69	86.24	64.83	405.87	46.35	3.31	45.82
		SD	(6.21)	(26.26)	(86.86)	(6.41)	(24.37)	(47.08)	(11.44)	(5.37)	(49.18)
		N	5	5	5	5	5	5	3	5	5
Total		Mean	50.23	77.67	422.33	50.02	49.74	315.5	43.74	27.94	106.83
		SD	(14.25)	(70.07)	(187.38)	(15.73)	(53.94)	(133.02)	(11.23)	(43.45)	(103.83)
		N	1027	1027	1027	898	1027	1027	891	1027	1027

Note: standard deviation in the parenthesis.

Table 4.8.3: Results of OLS, 2SLS, SUR and 3SLS on HEI for single headed households

	OLS		2SLS		SUR		3SLS	
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
main								
E_FAH	0.042***	0.035**	0.05	0.07	0.043**	0.036**	0.05	0.07
	(0.01)	(0.01)	(0.05)	(0.05)	(0.01)	(0.01)	(0.05)	(0.06)
E_FAFH	0.01	0.01	0.09	0.09	0.02	0.02	0.09	0.09
	(0.02)	(0.02)	(0.05)	(0.06)	(0.01)	(0.02)	(0.05)	(0.06)
T_FAH		0.02		0.01		0.02		0.02
		(0.01)		(0.02)		(0.01)		(0.01)
T_FAFH		0.00		-0.01		0.00		0.00
		(0.01)		(0.01)		(0.01)		(0.01)
Some College	3.01	3.41	2.67	3.05	2.98	3.38	2.67	2.99
	(1.85)	(1.90)	(1.92)	(2.14)	(1.84)	(2.03)	(1.94)	(2.06)
Bachelor	7.115**	7.346**	5.944*	6.213*	7.008**	7.233**	5.944*	5.945*
	(2.50)	(2.48)	(2.68)	(2.73)	(2.30)	(2.52)	(2.70)	(2.78)
Master or above	6.803*	7.091*	6.216*	6.34	6.740*	7.035*	6.22	6.11
	(3.04)	(3.02)	(3.15)	(3.33)	(3.13)	(3.27)	(3.23)	(3.41)
…								
E_FAH								
Some College					5.27	5.27	5.12	4.81
					(8.50)	(7.58)	(7.23)	(7.20)
Bachelor					12.63	12.63	12.53	12.22
					(10.03)	(10.08)	(9.11)	(9.08)
Master or above					15.23	15.22	15.17	14.96
					(14.32)	(13.84)	(12.72)	(12.75)
…								
E_EAFH								
Some College					0.82	0.80	0.17	0.16
					(7.84)	(7.54)	(7.14)	(7.17)
Bachelor					6.13	6.10	5.66	5.54
					(10.47)	(9.54)	(9.71)	(9.77)
Master or above					-5.93	-5.97	-6.20	-6.31
					(11.94)	(11.85)	(10.19)	(10.22)
…								
R-sqr	0.13	0.13	0.07	0.07	0.13	0.13	0.07	0.07
dfres								
BIC	8399.5	8406.8	.	.	30434.9	30442	30429	30434.5

* p<0.05, ** p<0.01, ***p<0.001

Table 4.8.4: Results of OLS, 2SLS, SUR and 3SLS on HEI for single headed SNAP households

	OLS		2SLS		SUR		3SLS	
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
main								
E_FAH	0.01	0.01	0.08	0.07	0.02	0.01	0.08	0.07
	(0.02)	(0.02)	(0.05)	(0.06)	(0.02)	(0.02)	(0.06)	(0.06)
E_FAFH	0.01	0.02	0.05	0.06	0.02	0.03	0.05	0.05
	(0.04)	(0.04)	(0.07)	(0.08)	(0.04)	(0.04)	(0.06)	(0.07)
T_FAH		0.02		0.01		0.02		0.01
		(0.01)		(0.02)		(0.01)		(0.01)
T_FAFH		-0.01		-0.01		-0.01		-0.01
		(0.01)		(0.01)		(0.01)		(0.01)
Some College	-0.80	-0.14	-0.75	-0.28	-0.74	-0.10	-0.75	-0.16
	(2.46)	(2.56)	(2.69)	(2.85)	(2.82)	(2.95)	(2.93)	(3.13)
Bachelor	7.59	7.85	7.23	7.66	7.52	7.79	7.23	7.56
	(4.79)	(4.92)	(4.88)	(4.5.07)	(4.70)	(5.16)	(5.78)	(5.98)
Master or	19.28	19.03	18.01	17.90	18.89	18.70	18.01	17.93
above	(13.22)	(13.24)	(13.94)	(13.82)	(13.51)	(13.25)	(12.69)	(12.67)
…								
E_FAH								
Some College					2.23	2.22	1.68	1.60
					(14.50)	(12.67)	(11.20)	(11.18)
Bachelor					3.14	3.11	3.37	3.07
					(22.69)	(18.51)	(15.41)	(15.49)
Master or above					-10.97	-10.93	-10.65	-10.15
					(19.71)	(18.24)	(21.22)	(21.23)
…								
E_EAFH								
Some College					-4.13	-4.13	-4.30	-4.31
					(7.43)	(9.69)	(10.20)	(10.21)
Bachelor					10.67	10.63	10.72	10.60
					(11.12)	(12.90)	(12.69)	(12.61)
Master or above					39.05	39.11	39.20	39.38
					(40.73)	(39.43)	(35.07)	(35.13)
…								
R-sqr	0.20	0.21	0.13	0.13	0.20	0.21	0.13	0.14
dfres								
BIC	2542.2	2548.8	.	.	9000.2	8972.8	9013.4	9009.2

* p<0.05, ** p<0.01, ***p<0.001

Table 4.8.5: Results of OLS, 2SLS, SUR and 3SLS on HEI for single headed Non-SNAP households

	OLS		2SLS		SUR		3SLS	
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
main								
E_FAH	0.057***	0.046***	0.09	0.08	0.061***	0.050***	0.094*	0.08
	(0.01)	(0.01)	(0.05)	(0.05)	(0.01)	(0.01)	(0.05)	(0.05)
E_FAFH	0.01	0.00	0.04	0.03	0.01	0.01	0.04	0.03
	(0.01)	(0.01)	(0.05)	(0.06)	(0.01)	(0.02)	(0.05)	(0.06)
T_FAH		0.032**		0.02		0.032**		0.031*
		(0.01)		(0.02)		(0.01)		(0.01)
T_FAFH		0.01		0.01		0.01		0.01
		(0.01)		(0.02)		(0.01)		(0.01)
Some College	5.02	5.489*	4.29	4.74	4.942*	5.400*	4.29	4.74
	(2.57)	(2.54)	(2.71)	(2.69)	(2.24)	(2.52)	(2.72)	(2.67)
Bachelor	7.663**	7.816**	6.270*	6.606*	7.518**	7.638**	6.270*	6.403*
	(2.91)	(2.88)	(3.18)	(3.12)	(2.43)	(2.72)	(3.19)	(3.16)
Master or	6.755*	6.728*	5.47	5.69	6.635*	6.58	5.47	5.41
above	(3.33)	(3.30)	(3.61)	(3.60)	(2.70)	(3.40)	(3.64)	(3.65)
…								
E_FAH								
Some College					9.98	9.97	9.78	9.83
					(7.19)	(7.32)	(6.64)	(6.63)
Bachelor					19.913*	19.901*	19.789*	19.783*
					(8.58)	(8.87)	(8.30)	(8.31)
Master or above					22.934*	22.92	23.02	22.98
					(10.56)	(12.06)	(11.89)	(11.85)
…								
E_EAFH								
Some College					4.31	4.30	4.19	4.21
					(6.81)	(6.51)	(6.73)	(6.75)
Bachelor					8.15	8.14	8.08	8.07
					(7.36)	(7.10)	(7.98)	(7.97)
Master or above					-7.40	-7.40	-7.34	-7.36
					(7.94)	(8.60)	(8.17)	(8.20)
…								
R-sqr	0.13	0.15	0.11	0.13	0.13	0.14	0.11	0.12
dfres								
BIC	5931.8	5931.9	.	.	21370	21370	21368	21367

* p<0.05, ** p<0.01, ***p<0.001