The Pennsylvania State University
The Graduate School

ESSAYS ON FOOD WASTE AND CONSUMER DEMAND ANALYSIS

A Dissertation in

Agricultural, Environmental and Regional Economics

by

Yang Yu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2020

The dissertation of Yang Yu was reviewed and approved by the following:

Edward C. Jaenicke
Professor of Agricultural Economics
Dissertation Advisor
Chair of Committee
Graduate Program Director, Agricultural, Environmental and Regional Economics

Yizao Liu
Assistant Professor of Agricultural Economics

Douglas H. Wrenn
Associate Professor of Environmental and Resource Economics

Peter Newberry
Assistant Professor of Economics

Abstract

Unnecessary food waste is a global economic and environmental problem. In the United States alone, consumer welfare loss from food waste amounts to a massive $160 billion annually, which is about 30% of the total food supply. Moreover, discarded food is a major source of greenhouse gas emission globally, generating about 3.3 gigatons of carbon dioxide and methane each year. If regarded as a country, food waste is the third-largest carbon-emitting country after the U.S. and China. Despite the importance of the food-waste problem, researchers have had only limited success in studying the underlying issues behind food waste, partly because no public or private organization is measuring actual food waste on a wide scale. At best, researchers have been able to investigate food-waste issues either at the national level by comparing separate datasets on food consumption and food acquisition or at the small scale by conducting experiments or surveys.

The three essays in this dissertation study attempts to fill this gap by (i) employing an indirect but creative method to examine household-level food waste in a national survey of food acquisition, thus allowing us to investigate how household characteristics are linked to the estimated levels of food waste, (ii) incorporating food waste into a theoretical model of household behavior, thereby showing that waste is a rational outcome of utility maximization and an important factor to account for in other models of household-level food behavior, and (iii) finding empirical evidence in consumer and market data that policy changes (i.e., extending the sell-by date on milk cartons) can and do reduce food waste.

To overcome the lack of observed data on food waste, the first essay begins by formulat-

ing household food consumption as a production process that transforms food inputs into chemical energy required for the human body's metabolic process and physical activities. Household-level food waste is estimated as input inefficiency via a stochastic frontier production model. Applying the method to a nationally representative sample of households, the essay shows that on average, U.S. households waste about 31% of their food, and that this level of annual waste corresponds to $240 billion. In addition, by accommodating heterogeneous wasting behavior, the results indicate that healthier diets and higher income lead to more household food waste, whereas lower household food security, food-assistance program participation, and larger household sizes are associated with less food waste.

The second essay shows that without modeling or at least partially accounting for wasting behavior, demand estimates in traditional models are potentially biased. The reason for the bias is that the omitted food waste is often a rational and heterogeneous choice made by households and linked to other consumer choices. This point is illustrated by both theoretical and empirical examples. Two structural approaches to identifying and estimating rational food waste are introduced. The first approach partially identifies the waste function through economic constraints. The second approach considers behavioral assumptions on household utility maximization. Taken together, these efforts represent one of the first attempts to incorporate food waste into utility-maximizing models of consumer behavior and provide useful estimates to study the rationales of wasting food. Policymakers could apply the models and utilize the results to calibrate the amounts of actual consumption and to find more effective mechanisms to incentivize food waste reduction.

The third essay examines a real-world policy change that was intended to reduce food waste. Consumers often find sell-by labels confusing and misinterpret their meanings as "safe-until" dates. Consequently, a significant portion of perishable food is mismanaged and disposed of earlier than necessary. As an effort to reduce food waste, in September 2010, New York City's Board of Health repealed its regulation on sell-by dates of pasteurized milk products. This policy change, in effect, increased the shelf-life of milk from 9 days to about 15

days. Based on a theoretical model of rational food waste and various empirical verifications using micro-level scanner data, the essay finds that the city's new policy effectively reduced food waste by more than 10%. This result translates to a reduction in wasted milk of more than 5.2 million pounds annually in New York City, an approximately $3.4 million value. This study is the first to find empirical evidence that policy changes can reduce food waste.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to gratefully acknowledge my mentor and committee chair, Dr. Ted Jaenicke. It would be difficult to envision the completion of this dissertation without his continuous support and encouragement throughout the past five years. He has not only provided intellectual guidance in conducting research through our countless conversations, brainstorming, and collaboration, but also generously offered his wisdom and thoughtful advice when I face tough times in life. I would also like to thank my committee members, Dr. Yizao Liu, Dr. Doug Wrenn, and Dr. Peter Newberry, for offering their expertise and valuable suggestions in research and teaching.

I am grateful for the financial support I received from the U.S. Department of Agriculture. My dissertation research is funded, in part, by the USDA-NIFA-AFRI Exploratory Program grant 2017-67030-26611. However, the views expressed here are those of the author and do not necessarily reflect the views of the U.S. Department of Agriculture. The results are the researcher's own analyses calculated and derived based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Dedication

To my wife, Jiacheng, and my parents, Yaxian and Renli.

Chapter 1

Estimating Food Waste as Household Production Inefficiency

1.1 Introduction

Several important studies show that, at the aggregate level, 30% to 40% of the total food supply in the United States goes uneaten, representing more than $160 billion in economic losses (Muth et al., 2011; Leib et al., 2013; Buzby et al., 2014). Moreover, these waste figures mean that resources used to produce the uneaten food, including land, water, and labor, are wasted as well. Throughout its life cycle, discarded food is also a major source of greenhouse gas emissions (Chapagain and James, 2011; Quested and Parry, 2011; Venkat, 2011; Beretta et al., 2013; FAO, 2013). According to the U.N.'s Food and Agriculture Organization (FAO), food waste is responsible for about 3.3 gigatons of greenhouse gas annually, which would be, if regarded as a country, the third-largest carbon-emitting country after the U.S. and China (FAO, 2013).

This chapter focuses on food waste at the consumer level and proposes a novel approach to empirically identify household food waste as input inefficiency in a production context. Along the food supply chain, the final consumption stage constitutes the largest share of food waste in the U.S. (Griffin et al., 2009; Buzby et al., 2014; Bellemare et al., 2017). In addition, important food-related policies and assistance programs may face inaccurate cost-benefit calculations because of the hidden costs of food waste. For instance, our results show that healthy dietary practices are associated with significantly more waste, suggesting programs aimed at promoting healthy eating should be evaluated for their implications on food waste. Methodologically, an accurate measurement of consumer food waste provides a means of calibrating actual consumption in both the traditional consumer demand models and newer models that treat food waste as a rational choice (Bellemare et al., 2017; Lusk and Ellison, 2017; Hamilton and Richards, 2019).

Specifically, we formulate household food consumption as a production process that converts food inputs, categorized by types of food and measured by gram weights of food acquired, into chemical energy required for the human body's metabolic process and physical activities. We then identify food waste as input inefficiency in this production process using a stochastic production frontier model (Aigner et al., 1977; Jondrow et al., 1982). Importantly, this strategy for identifying food waste can be reasonably employed at other stages of the food supply chain as well. For instance, at the farm and retail levels, the output and input measures are generally well-defined and observed; therefore, input inefficiencies can be directly estimated using various methods from the productivity analysis literature.

Current empirical studies that focus on measuring consumer food waste largely fall into two strands of literature. One strand considers the difference between reported purchases and actual food intakes, either across different data sources (Muth et al., 2011; Buzby et al., 2014) or within a single dataset (Landry and Smith, 2019). While these studies provide a straightforward calculation of food waste that is easy for interpretation, they are often dependent on the choice of datasets and the availability of food intake data.[1]

Another strand utilizes innovative survey and experimental methods to obtain a set of

[1] For instance, although Landry and Smith (2019) successfully obtain household-level food waste estimates and find variables that influence the variation in waste, their particular dataset, the 1977-78 Nationwide Food Consumption Survey, may not reflect the most recent patterns in food consumption behaviors.

"first-hand" observations (for examples, see Stefan et al. (2013); Reynolds et al. (2014); Neff et al. (2015); Secondi et al. (2015); Qi and Roe (2016); Ellison and Lusk (2018); Roe et al. (2018a)). These methods are extremely useful in assessing attitudes toward food waste and the effects of household- or product-specific characteristics, e.g., date labels (Wilson et al., 2017; Roe et al., 2018a). On the other hand, the accuracy of waste measures from survey data is influenced by participants' ability to effectively track and recall various wasting occasions, while experiments such as visual estimation, manual weighing, and digital photography are usually constrained to small-scale settings with limited application for households more generally. Overall, in the emerging body of food waste studies, there is a need for comprehensive food-waste estimates at the individual household level that can be generalized to a wide range of household groups. Consequently, the direct link between household-specific characteristics and food waste has not been completely documented.

In this chapter, we overcome the data obstacle by conducting a productivity analysis of household production to obtain an input inefficiency measure that is interpreted as excess food inputs used to produce the current level of output in the form of energy expenditure. By construction, our model considers food diverted or recovered for nonfood purposes as food waste, which is consistent with the definitions used by the U.S. Department of Agriculture's Economic Research Service (ERS), the FAO, and the EU FUSIONS program (FAO, 2013; Buzby et al., 2014; FUSIONS, 2016), but different from the latest categorization proposed by Bellemare et al. (2017). In addition, because we use the edible parts of food as inputs in the estimation, the waste estimates in this chapter point to avoidable food waste according to the definitions by the Waste and Resources Action Programme (WRAP) (Quested and Parry, 2011).[2]

A useful advantage of our approach is that it only requires food acquisition data plus some biological measures (age, height, weight, and gender) of household members. Therefore, it can be replicated and tested by various consumer datasets that are commonly used. The

[2] Detailed comparisons between our food waste measure and the existing definitions and studies are provided in a later section and in the Appendix.

particular data source used in this chapter is the 2012 National Household Food Acquisition and Purchase Survey (FoodAPS). Because the FoodAPS data lack sufficient information on physical activities, our baseline model treats only the sum of household members' basal metabolic rates as the output measure. We address this issue by examining two additional models. The second model treats employment status as an indicative proxy for physical activities and weekend shopping frequency as its instrument. The third model employs a two-step procedure to impute individual physical activity levels by a person's biological measures and a set of demographic variables, using the 2011-2012 National Health and Nutrition Examination Survey (NHANES) and its reported Metabolic Equivalents for different types of activities (Institute of Medicine, 2005). All three models yield similar results.

Our estimates show that the average amount of food wasted at the household level is 31.9% in the baseline model, which is in line with the existing findings at the aggregate level. By using the sample weights assigned to each household and their food expenditures, this estimated percentage translates to annual U.S. consumer-level food waste valued at $240 billion. In addition, by allowing for heterogeneous wasting behavior across households, we examine how household-specific attributes explain the variation of our food waste estimates. We consider three variables directly related to food management and eating behavior, and we find that better household food security, healthy dietary practices, and higher income lead to more household-level food waste.

Because our stochastic frontier models are built upon a household production process that takes food inputs as given, we also conduct a post-estimation analysis that focuses on factors related to shopping behavior and purchase decisions (Stefan et al., 2013; Porpino et al., 2015). Our results show that shopping with a grocery list, participation in food-assistance programs, longer distance to primary stores, and larger household sizes are all associated with lower waste estimates. These results, therefore, provide useful reference points for studies investigating the feasibility and effectiveness of possible food waste prevention policies that are aimed at particular food types, the retail environment, and, more importantly, particular household types.

Finally, we present a series of robustness and validity checks. For example, we include alternative input and output measures and examine the presence of households currently receiving food-assistance benefits or undergoing dieting practices. We also apply copula estimation, an instrument-free approach, to address possible endogeneity of the food-input variables. In addition, we discuss potential obstacles in incorporating more contextual variables into the stochastic frontier estimation. To test the general validity of the method, we show that when applied to a pure food-intake dataset, the NHANES data, our model predicts waste estimates close to zero (as it should). Perhaps more importantly, though FoodAPS includes both food-at-home and food-away-from-home consumption, we find that our method is also valid when applied to only food-at-home acquisition data with minor revisions. This last finding suggests that our approach is replicable to other widely used scanner datasets where food-away-from-home information is absent.

The rest of the chapter is organized as follows: the next section presents the model specification and econometric approach, followed by discussions of the data and main results, including the variation in food waste across demographic groups and a post-estimation of the effects of shopping behavior. Finally, a set of robustness checks and validations on replicating the model are provided.

1.2 Empirical Models and Estimation

In most cases, directly measuring food waste for a large sample of households is not feasible due to the difficulty of tracking and recording. We propose modeling household food consumption as a production process that converts food inputs into chemical energy required to meet household members' metabolic processes and additional energy demand from physical activities. This production function shows, from a nutritional perspective, how various food-group contents are transformed into energy expenditure. We then treat the input inefficiency as a consequence of uneaten food, taking heterogeneous wasting behaviors into consideration. Intuitively, input inefficiency is interpreted as the excess input usage that can be reduced to yield the current level of output if efficiency were maximized. Thus, uneaten food, indirectly measured, becomes our operational definition of household food waste.

1.2.1 The Baseline Model

Household h's production process takes the form in equation (1.1). The output, $Y(b_h, PA_h)$, is the sum of total energy expenditures across all household members. The vector b_h contains the members' biological measures, e.g., weights, heights, ages, and genders, that capture the basal metabolic rate; and PA_h is a vector representing additional physical activities. The production technology $F(x_h, d_h)$ is a function of acquired food inputs given by the vector x_h, measured either in weights or calorie contents, and a set of household demographic variables d_h that determine inefficiency. Similar to the base case in Hall et al. (2009), we assume that each individual maintained a state of energy balance during the 7-day survey period.

$$Y(b_h, PA_h) = F(x_h, d_h) \tag{1.1}$$

Following the common practice in nutritional and medical research, we calculate total energy expenditures using Basal Metabolic Rate (BMR) and Physical Activity Level (FAO/WHO/UNU, 1985; Scrimshaw et al., 1996; Institute of Medicine, 2005). BMR reflects the at-rest energy required to maintain basic body functioning and is computed by the revised Harris-Benedict equations using weight, height, age, and gender (Roza and Shizgal, 1984).[3] Typically, BMR accounts for 65 to 75% of an individual's total energy expenditure (Institute of Medicine, 2005). Physical Activity Level is a multiplier, generally ranging from 1 to 2.5, which represents the ratio of total energy expenditure to BMR. It includes, in addition to BMR, the thermal effect of food and additional energy needed to perform daily

[3]The equations are provided in the Appendix.

household tasks and exercise. For example, if a person's BMR is 1000 Kcal per day and his/her physical activity level is 1.7, then the total energy required to maintain a balanced state is $1000 \times 1.7 = 1700$ (Kcal) per day. Letting S_h be the household size, m the index of its members, and $BMR(\cdot)$ as the revised Harris-Benedict equation, then the output measure of household h is calculated as follows:

$$Y(b_h, PA_h) = \sum_{m=1}^{S_h} BMR(b_{m,h}) \cdot PA_{m,h} \tag{1.2}$$

In our main data source, FoodAPS, individual biological measures in $b_{m,h}$ are recorded. However, this dataset lacks sufficient information on physical activities. In the baseline model, we tackle this issue by first rewriting the output measure: $Y(b_h, PA_h) = y(b_h) \cdot PA'_h$, where $y(b_h)$ is the household total BMR, and PA'_h is the aggregated physical activity level that is the ratio of the household total energy expenditure to its total BMR. This specification allows us to separate $y(b_h)$ and PA'_h by taking logarithm: $\log Y(b_h, PA_h) = \log y(b_h) + \log PA'_h$. Furthermore, distributional assumptions on $\log PA'_h$ are imposed to enable a maximum likelihood estimation, as discussed below, where $\log PA'_h$ is subsumed by a white-noise term. Thus, in our baseline model, household total BMR, $y(b_h)$, is the operational output measure, which is given by $y(b_h) = \sum_{m=1}^{S_h} BMR(b_{m,h})$.

The full specification of the baseline model is an extension of the stochastic production frontier models (Aigner et al., 1977; Fried et al., 2008). Denote $x_h = (x_{1,h}, x_{2,h}, ...x_{I,h})^T$ as the input quantities from I groups of food, including both at-home and away-from-home purchases. Our main analysis is based on input weights (grams) of the edible parts of food, whereas the calorie-content-based estimation is presented as a robustness check. We formulate the production technology $F(x_h, d_h)$ in the translog form where v_h is a white noise and u_h is the output inefficiency due to food waste:

$$\log y_h = \alpha_0 + \sum_{i=1}^{I} \alpha_i \log x_{i,h} + \sum_{i=1}^{I} \sum_{j \leq i} \beta_{i,j} \log x_{i,h} \log x_{j,h} + v_h - u_h \tag{1.3}$$

Here the term $-\log PA'_h$ does not appear on the right-hand side of the above equation. In the baseline model, we assume that it is independent of the explanatory variables, and its population distribution is completely captured by the distributions of α_0 and v_h. This assumption is dropped in the second and third models where we tackle the issue of missing physical activities by either adding proxy variables or imputing activity levels. As is usually assumed in normal–half-normal stochastic frontier models, the white noise term v_h follows a normal distribution $N(0, \sigma_v^2)$, and the inefficiency term u_h follows a half-normal distribution $N^+(0, \sigma_{u_h}^2)$ and is heteroskedastic:

$$\sigma_{u_h}^2 = \exp(\gamma_0 + \gamma' d_h) \tag{1.4}$$

It is noteworthy that when a demographic variable in d_h generates larger $\sigma_{u_h}^2$, it also induces higher inefficiency, on average. The econometric method we use for estimating the stochastic frontier model is based on maximum likelihood. The likelihood function is derived on $\varepsilon_h = v_h - u_h$ whose density can be solved analytically:

$$f_{\epsilon_h}(\varepsilon_h) = \frac{2}{\sigma_h}\phi(\frac{\varepsilon_h}{\sigma_h})\Phi(-\frac{\lambda_h \varepsilon_h}{\sigma_h}) \tag{1.5}$$

where $\sigma_h^2 = \sigma_v^2 + \sigma_{u_h}^2(d_h)$ and $\lambda_h = \sigma_{u_h}/\sigma_v$. $\phi(\cdot)$ and $\Phi(\cdot)$ are density and cumulative distribution functions of the standard normal distribution, respectively. Maximization is performed on $\sum_h \log f_{\varepsilon_h}(\varepsilon_h)$ to obtain parameter estimates $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\sigma}_v^2)$. Intermediate household-specific parameters $\hat{\sigma}_{u_h}^2$, $\hat{\sigma}_h^2$, and $\hat{\lambda}_h$ are then calculated for each observation.

The translog specification in equation (1.3) is a flexible functional form that is adequate in many cases. Nonetheless, as we use household total energy expenditure as the output, one might wonder if we can simply treat total calorie contents from all food groups as the input, that is, $\log y_h = \alpha_0 + \alpha_1 \log(\text{total calories}) + v_h - u_h$.[4] This single-input production process faces several weaknesses due to its simplification. First, aggregating calorie values of food

[4]This specification yields a slightly higher estimate of average food waste at about 40%.

products based on their nutrition labels have been criticized for ignoring other substantial factors such as food composition (Trivedi, 2009). Second, food digestion itself requires energy (the thermal effect of food), which typically accounts for about 10% of total energy expenditure (McArdle et al., 1986). Different types of food need different amounts of energy to digest, even when they contain the same calorie content on the nutrition labels. For instance, protein-intense food generates more heat in postprandial thermogenesis than carbohydrate and lipids-intense food, thereby provides less "effective" energy that is eventually absorbed by the body (Johnston et al., 2002). In modeling the production function, calorie contents from different types of food are not perfect substitutes, hence not linearly additive.

1.2.2 Food Waste Measure

Our primary goal is to estimate the percentage of food waste at the individual household level. This task can be accomplished once we have an estimate of the output inefficiency term $\hat{u_h}$ for each household and transform it into an input inefficiency measure. The closed-form prediction of u_h post-estimation is well established in the stochastic frontier literature (for example, Jondrow et al. (1982)). The solution is given as follows, where $\hat{b}_h = \hat{\varepsilon}_h \hat{\lambda}_h / \hat{\sigma}_h$:

$$\hat{u_h} = E(u_h|\hat{\varepsilon}_h) = \frac{\hat{\sigma}_{u_h}\hat{\sigma_v}}{\hat{\sigma_h}}[\frac{\phi(\hat{b}_h)}{1-\Phi(\hat{b}_h)} - \hat{b}_h] \tag{1.6}$$

For exploratory purpose, let us assume that, for household h, food from all I groups are wasted in the same proportion, δ_h. Then we have the following identity:

$$\begin{aligned} &\sum_{i=1}^{I} \hat{\alpha}_i \log x_{i,h} + \sum_{i=1}^{I}\sum_{j\leq i} \hat{\beta}_{i,j} \log x_{i,h} \log x_{j,h} - \hat{u}_h \\ =&\sum_{i=1}^{I} \hat{\alpha}_i \log(1-\hat{\delta}_h)x_{i,h} + \sum_{i=1}^{I}\sum_{j\leq i} \hat{\beta}_{i,j} \log(1-\hat{\delta}_h)x_{i,h} \log(1-\hat{\delta}_h)x_{j,h} \end{aligned} \tag{1.7}$$

Here, the $-\hat{u}_h$ term is transformed into a multiplication factor $(1-\hat{\delta}_h)$ on each $x_{i,h}$.[5]

[5] Similar transformation is used in Reinhard et al. (1999) for a single-input case, whereas Kurkalova and

Solving for $\hat{\delta}_h$ using the quadratic formula gives us two sets of solutions. Even though we did not impose any theoretical restrictions on the parameters, only one of them makes economic sense. The rationale is that we expect a positive correlation between $\hat{u}_h$ and $\hat{\delta}_h$ so that more output inefficiency implies more input waste. Verifying the correlation through partial derivatives gives the following solution for $\hat{\delta}_h$, where $\hat{A} = \sum_{i=1}^{I} \sum_{j \leq i} \hat{\beta}_{i,j}$, $\hat{B}_h = \sum_{i=1}^{I} \hat{\alpha}_i + \sum_{i=1}^{I} \sum_{j \leq i} \hat{\beta}_{i,j}(\log x_{i,h} + \log x_{j,h})$, and $C_h = \hat{u}_h$:

$$\text{Food waste: } \hat{\delta}_h = 1 - \exp\left(\frac{-\hat{B}_h + \sqrt{\hat{B}_h^2 - 4\hat{A}\hat{C}_h}}{2\hat{A}} \right) \tag{1.8}$$

This key estimate is essential to our analysis of household food waste. To the best of our knowledge, this is the first study that provides econometrically estimated food waste for individual households using food acquisition data. Moreover, it opens a channel for conducting post-estimation analysis on sub-group comparisons based on various demographic measures, as well as implications for food policies aimed at particular household types.

It is important to compare our concept of food waste, as input inefficiency in a household production setting, to the existing definitions of food waste that are widely used. Bellemare et al. (2017) note that the definitions provided by the FAO, the ERS, and the EU FUSIONS project all consider food diverted for nonfood purposes as being wasted, which ignores the economic value of nonfood uses and imposes potential difficulty in tracking such waste. As a response, they propose a refined definition that only counts the amount of food that ends up in landfill without being used for any purpose. Because our output measure is the total energy provided by food consumption, food diverted for other purposes is considered as waste. Hence our definition resembles those of the FAO's and the ERS's, which can be considered as an upper bound for the one defined in Bellemare et al. (2017).

Another important aspect of food waste measurement concerns the degree to which it

Carriquiry (2003) considers transformation for multiple inputs in a Cobb-Douglas model. Alternatively, Kumbhakar and Tsionas (2006) provide an approach that directly formulates input-inefficiency therm δ_h as a random variable and uses a simulated ML estimation.

is avoidable. The WRAP program provides a three-level categorization on this subject.[6] According to its definitions, "avoidable" food waste is food thrown away but perfectly edible otherwise; "possibly avoidable" waste is food that only some people find edible or only edible when prepared in certain ways (e.g., potato skins); and "unavoidable" waste is part of food that is not edible for all consumers (e.g., eggshells) (Quested and Parry, 2011). The food input quantities in our analysis are based on edible portions of food.[7] Therefore, the waste estimates presented in this chapter correspond to avoidable food waste, the first type in the WRAP definition. In the Appendix, we show that when using total food acquisitions as input quantities, the waste estimates are identical to the previous model where inputs are the edible portions. This result could be useful when applying the method to datasets that do not contain information on edibility.[8]

1.2.3 Proxy and Instrumental Variables

Our primary concern about the baseline model is the missing physical activities in the FoodAPS data. Archer et al. (2016) show that physical activities play an important role in assessing food waste as they form significant shares of total energy expenditure. From a technical point of view, large shares do not necessarily imply inaccuracy of the baseline model. The distributional properties of the missing variable $\log PA'_n$, however, have a direct influence in determining the consistency of parameter estimates. If they fail to meet the conditions specified, inconsistent parameters would possibly generate biased food waste estimates.[9]

[6] The Food Loss and Waste Protocol standard also encourages agencies to distinguish edible/inedible parts of food. And it acknowledges that the edibility of food is influenced by a number of variables, including the socio-economic and cultural factors (FLW Protocol, 2016).

[7] FoodAPS calculates the edible portion of food by matching the product information with either the National Nutrient Database for Standard Reference (SR) or the Food and Nutrient Database for Dietary Studies (FNDDS).

[8] The reason behind this result is that the portion of unavoidable food waste is systemic; e.g., it is the same for all consumers, and that the inedible part has zero marginal productivity. However, we note that our method cannot account for the variation in the possibly avoidable waste–the second type in WRAP's definition.

[9] We provide an intuitive explanation as to why the baseline model overestimates waste by about 3% in the next section.

Therefore, we propose a proxy variable for the missing physical activities. Though FoodAPS does not contain direct measures of physical activities, it provides some highly indicative variables. One example is the employment status of working-age individuals, a discrete variable of four levels, with 1 meaning unemployed while not searching for a job, and 4 representing employed and working regularly. For each household, we obtain an age-weighted average employment status and normalize it to a value between 0 and 1. The rationale of using employment as a proxy is that employed people generally have a higher level of mandatory physical activities, all else equal. In addition, among the unemployed individuals in FoodAPS, about 44% are due to retirement, health issues, or disabilities, who are likely to have less physical activities.[10] The validity of the proxy is further supported by the NHANES 2011-2012 data which follow the same coding rule of employment as FoodAPS. The Appendix Table A.6 shows that, at the individual level, a higher value of employment implies a higher level of physical activities.

Despite these features of the employment status variable, the proxy itself is not entirely free of endogeneity concerns. Indeed, employment does not represent all types of physical activities. Recreational activities, for instance, may not be fully explained by employment status. To mitigate the endogeneity issue of the proxy variable, we adopt an instrumental variable approach and apply a version of the Limited Information Maximum Likelihood (LIML) specifically derived for stochastic frontier analysis.

Our choice of instrument is the frequency of weekend shopping. It is calculated as the percentage of a household's shopping trips to grocery stores and supermarkets that occurred during weekends. On the one hand, whether a household shops on weekends or weekdays is highly correlated with its members' employment status. In FoodAPS, households of the highest 25% employment status spend 34% of their trips on weekends, while the percentage of weekend trips is 26% for those of the lowest 25% employment status. On the other hand, the

[10] The implications of health issues and disabilities are straightforward. As for retirement, it can be regarded as an indicator of age. Our results in the next section on NHANES show that age is negatively correlated with physical activity level (Appendix Table A.6).

instrument is exogenous in the sense that it merely represents a choice of shopping schedule, not purchase decisions, e.g., it is unlikely to affect total food purchases over a whole week. In addition, it is reasonable to assume that such shopping schedule is uncorrelated with physical activities not represented by employment status such as recreational activities. Hence the instrument is correlated with the output only through the proxy variable.

There are several recent studies that tackle the issue of endogeneity in stochastic frontier models. Maximum likelihood methods are studied in Kutlu (2010) and Amsler et al. (2016). Shee and Stefanou (2014) consider using a proxy variable in a panel data estimation, and Tran and Tsionas (2015) develop a copula approach without requiring external instruments. We follow the LIML method in Amsler et al. (2016) by first adding household average employment status $\widetilde{PA}'_h$ as a proxy variable for physical activities to the right-hand side of equation (1.3). We then add the following reduced-form equation for the endogenous variable in which weekend shopping frequency z_h is the instrument, and estimate it jointly with the original equation:

$$\log \widetilde{PA}'_h = \pi_0 + \pi_{IV} \log z_h + \sum_{i=1}^{I} \pi_i \log x_{i,h} + \sum_{i=1}^{I} \sum_{j \leq i} \pi_{i,j} \log x_{i,h} \log x_{j,h} + \eta_h \tag{1.9}$$

Following Kutlu (2010) and the LIML case in Amsler et al. (2016), we assume that η_h is correlated with v_h, but not with u_h. More details on the derivation of the likelihood function and calculation of the predicted food waste are provided in the Appendix.

1.2.4 Fitted Physical Activities

We consider a third model in which physical activity levels $PA_{m,h}$ of individual household members are fitted using information from another nationally representative dataset–the 2011-2012 NHANES data. Physical activities are categorized into three types in NHANES: sedentary, moderate, and vigorous. Each survey participant reports how much time he or she spends on each type of activities in a typical day, including recreational activities. The

Table 1.1: METs and Increase in Physical Activity Level

Activity Type	METs	$\Delta PAL/10min$	$\Delta PAL/1h$
Sedentary ($Type = 1$)	1.5	0.005	0.03
Moderate ($Type = 2$)	4.0	0.029	0.17
Vigorous ($Type = 3$)	8.0	0.067	0.4

Note: NHANES contains suggested METs values of moderate and vigorous activities, whereas the METs of sedentary activities is taken from Table 12-3 of Institute of Medicine (2005). The last two columns are taken from Tables 12-1, 12-2, and 12-3 of Institute of Medicine (2005). The per 1-hour values in the last column are not exactly six times the per 10-min values due to nonlinear relationship. When calculating the $Time_{Type}$ variable, we use the per 10-min values. Time spent on activities is aggregated to a daily value.

physical activity levels in NHANES are calculated using Metabolic Equivalents (METs) of each type of activity (Institute of Medicine, 2005). METs represent the multiples of an individual's resting oxygen uptake, which is used to translate the amount of time spent on activities to an increase in the standard physical activity level, as shown in Table 1.1. Individual t's physical activity level is calculated by the following equation:

$$PA_t^{NHANES} = 1.1 + \sum_{Type=1}^{3} \Delta PAL_{Type} \cdot Time_{Type} \tag{1.10}$$

The constant number 1.1 reflects the base energy requirement plus 10% thermal effect of food. For example, let us consider a person that only performs 3 hours of daily moderate activities. Since the METs of moderate activities is 4.0, corresponding to a 0.17 increase in activity level for each hour's exercise, the person's physical activity level is calculated as $1.1 + 0.17 \times 3 = 1.61$.

We fit the missing values of $PA_{m,h}$'s in FoodAPS in two steps. First, in NHANES, we regress the physical activity level PA_t^{NHANES} on a set of individual characteristics g_t^{NHANES} containing biological measures and a set of demographic variables such as race/ethnic groups and overall health status. For individuals older than 20 years, additional variables include marital status, employment, presence of children in the household, and smoking habit. The list of variables and estimation results of the first-stage regression are contained in Appendix Table A.6. Males consistently have higher physical activity levels than females in both age groups. Weight and age are negatively correlated with physical activities, while height has a positive correlation. Self-reported healthier individuals and non-Hispanic whites are found to have higher physical activity levels. We also include a variable reported by NHANES that shows whether the survey was taken in colder weather (from November to April). Though estimated to have a negative influence on physical activities, the variable is not statistically significant. For persons of age 20 and above, employment status, higher education, and income are associated with higher activity levels, while being married or a smoker indicate lower activity levels.

The second step involves applying the coefficient estimates, $\widehat{\theta}_0$ and $\widehat{\theta}'$, to FoodAPS to impute the physical activity levels for each household member. Once we obtained the fitted values of individual physical activity levels $\widehat{PA}_{m,h}$, the fitted household total energy expenditure is applied as the output measure in the analysis, that is: $\hat{Y}_h = \sum_{m=1}^{S_h} BMR(b_{m,h}) \cdot \widehat{PA}_{m,h}$.

$$PA_t^{NHANES} = \theta_0 + \theta' g_t^{NHANES} \qquad \text{(First Step)} \tag{1.11}$$

$$\widehat{PA}_{m,h} = \widehat{\theta}_0 + \widehat{\theta}' g_{m,h}^{FoodAPS} \qquad \text{(Second Step)} \tag{1.12}$$

In general, whether the percentage food waste is overestimated or underestimated in the baseline model is a complicated matter that involves many factors including the signs of various correlations and shape of the distribution of the missing $\log PA_h'$. In this chapter, we do not explicitly explore the mechanisms that determine the bias. Nonetheless, the results of the third model suggest that the baseline model overestimates by 3%, on average.

As an intuitive yet not fully rigorous analysis of this bias, we calculate the implied household aggregate physical activity level $\log PA_h'$ by dividing the fitted household total energy expenditure $\hat{Y}_h$ by its total BMR, y_h. Its distribution approximates a nor-

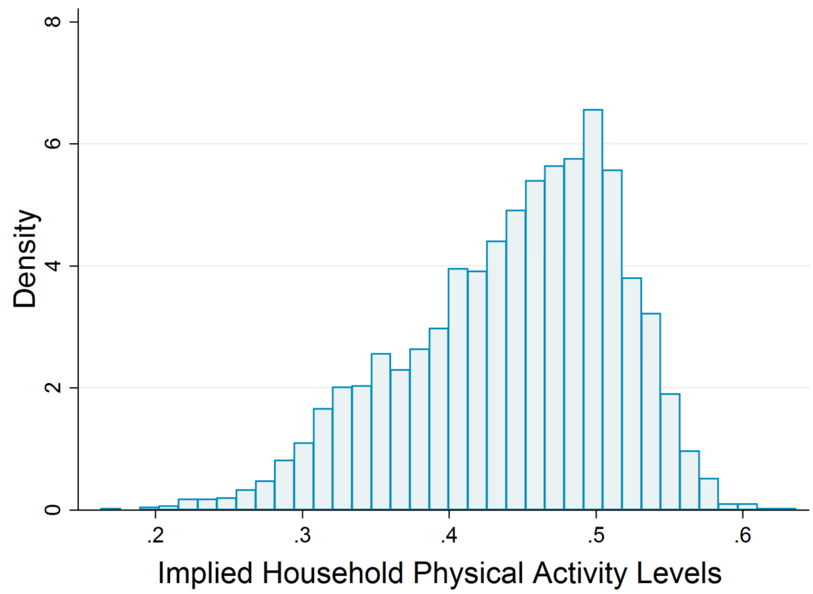


Figure 1.1: Distribution of logarithm of implied household physical activity level, $\log PA'_h$

mal distribution with slight negative skewness as shown in Figure 1.1. Suppose we can write $\log PA'_h$ in terms of a normally distributed variable minus a half-normal variable, the latter denoted as w_h. Since $-\log PA'_h$ is the missing variable, what is estimated as output inefficiency in the baseline model can be regarded as $u'_h = u_h - w_h$. Note that $Var(u'_h) = Var(u_h) + Var(w_h) - Cov(u_h, w_h) > Var(u_h)$ when $Var(w_h) - Cov(u_h, w_h) > 0$. Based on our distributional assumptions on $\log PA'_h$, we apply a maximum likelihood estimation to obtain the predicted values of w_h, which is similar to estimating a stochastic frontier model without inputs. And we find that the variance of $\hat{w}_h$ is 0.0035 and its covariance with $\hat{u}_h$ is 0.000498, meaning the inequality indeed holds.

1.3 Data and Variables

Our choice of directly measurable quantities, e.g., energy requirement and food purchases are made feasible by utilizing the USDA's FoodAPS data. For a nationally representative sample of 4826 households, FoodAPS provides reasonably complete information on (i) household demographic variables, including income, education, and health outcomes, (ii) biological measures of each household member, and (iii) detailed data, including food categories, food quantities, and nutrition information, on food purchased or acquired for both at-home and away-from-home consumption, for a period of seven days.

The output in the baseline and proxy-instrument models, y_h, is the sum of household members' BMR, while the third model uses the fitted total household energy expenditure, $\hat{Y}_h$, as the output. There are 412 households in the data, out of 4826 total, that have missing information on their members' biological measures, and are dropped from the analysis. Table 1.2 shows the summary statistics of the input and output measures. The average household total BMR is 4293 Kcal, and average total energy expenditure is 6861 Kcal, which translate to per-member BMR at 1550 Kcal and energy expenditure at 2441 Kcal, implying an average physical activity level of about 1.6.

The input variables, $x_{1,h}, x_{2,h}, ...x_{I,h}$, are gram weights of nine food-group acquisitions, including both at-home and away-from-home occasions.[11] They are classified by the USDA's What We Eat in America (WWEIA) category codes: milk and dairy, protein foods, mixed dishes, grains, snacks, fruit and vegetables, beverages, condiments, and infant formula. The protein foods category includes all uncooked meat products such as pork, beef, fish, and chicken. It also contains protein-intense food like beans, nuts, and seeds. Mixed dishes generally refer to processed or cooked meals. The three types of mixed dishes with the largest shares in this category are pasta dishes, pizzas, and canned soups. The other types include burritos, tacos, sandwiches, and so on. Note that FoodAPS has a tenth, catch-all group for food items without an assigned group code. We merge this last group with the ninth group to generate a combined group called "infant formula and all other food without a code". In addition, we exclude households with zero total purchase (21 observations) and those with extremely large purchase quantities that exceed 100 kilograms (162 observations).

[11] Since for some households, there are food groups with zero values, we used $\log(x_{i,h} + 1)$ in estimation. The mean amounts of food in the data are typically in thousands. Hence we believe the bias, if any, is negligible. In fact, using $\log(x_{i,h} + 0.001)$ would produce the same amount of waste estimates.

Table 1.2: Summary Statistics–Output and Input Measures

Variable	Mean	Std. Dev.	5% percentile	95% percentile
y_h, Total Household BMR	4293.2	2401.8	1312.9	8699.6
$\hat{Y}_h$, Total Energy Expenditure	6861.4	4019.8	1871.1	14220.2
$x_{1,h}$, Milk and Dairy	3388.0	4276.6	0.0	11712.0
$x_{2,h}$, Protein Foods	1746.4	2264.6	0.0	5692.6
$x_{3,h}$, Mixed Dishes	2661.4	2629.0	0.0	7777.6
$x_{4,h}$, Grains	1548.0	2136.8	0.0	5264.3
$x_{5,h}$, Snacks	1428.4	1861.7	0.0	4951.2
$x_{6,h}$, Fruit and Vegetables	2594.7	2950.9	0.0	8361.0
$x_{7,h}$, Beverages	11099.4	11602.0	0.0	34852.2
$x_{8,h}$, Condiments	1509.3	2286.6	0.0	5972.0
$x_{9,h}$, Infant formula & Uncoded	90.1	610.7	0.0	340.2

Note: BMR and total energy expenditure are in calories (Kcal). The amounts of food acquisitions are measured in grams.

In the Appendix, we provide further details on how random-weight products and food-away-from-home items are measured, whether free or donated food is under-reported, and the effect of inventory accumulation of storable food.

Three household-level demographic variables are used as determinants of the variance of the output inefficiency term u_h: monthly income per adult equivalent, overall self-evaluated dietary healthfulness, and food security. These variables are selected as they directly influence food management, eating behavior, and people's attitudes toward wasting food. Family monthly income in thousand dollars is divided by adult equivalent household size, which is a frequently used method to analyze income effect. In calculating adult equivalence, we assign children under the age of 6 years a weight of 0.2, between 7 to 12 years a weight of 0.3, and

13 to 17 years a weight of 0.5 (World Bank, 2005).[12] There are 110 households that reported monthly income per adult equivalence less than \$100 among which 89 reported zero income, and 47 households that reported more than \$10,000 of which the largest number is \$66,000 for a single-member household. These observations are excluded from the analysis to avoid either misreporting or recording errors. The other two variables take discrete values and are normalized to a range between 0 and 1. Dietary healthfulness has values 1 to 4, with 1 as the least healthy diet and 4 as the healthiest. Food security is defined by three levels–low, medium, and high. Two observations with missing dietary healthfulness are dropped.[13]

Our second model uses household employment status as a proxy for physical activities. For each household, we first normalize the individual employment status to a value between 0 and 1, and then take the sum across all working-age members, divided by age-weighted household size. In addition, the instrumental variable for the proxy is frequency of household weekend shopping trips, measured by its percentage share of all shopping trips during the week. Because about half of the sample reported zero occurrences of weekend shopping, we use $\log(z_h + 0.1)$ as the operational instrument instead. In the third model, an additional non-binary demographic variable, education level, is used, which has five levels–from 9th-grade graduates to college graduates. The original education variable in FoodAPS has more levels and was re-categorized to match the one in NHANES. The summary statistics for all variables are presented in the Appendix Tables A.1-A.5.

1.4 Main Results

This section presents the major results from the three stochastic frontier models.

[12]While there is no definitive rule in choosing adult equivalent scales, we prefer a scale that includes multiple age groups of children instead of a simple weight for all children. An alternative method that provides similar results is to divide ages by 18 as weights for children. Note that although food expenditure is an essential component of household spending, the adult equivalent income is not limited to food but also reflects other types of spending, and, therefore, works as a general measure of income on household behavior.

[13]The ranking orders of dietary healthfulness and food security are reversed from the original FoodAPS data. For instance, in FoodAPS, 1 represents the healthiest or most secure. We change the orders to avoid confusion in relating the values to their meanings. Moreover, the number of levels is reduced by combining small marginal groups.

Table 1.3: Mean Elasticities of Food Groups

Food Groups	Baseline	Proxy-instrument	Fitted $PA_{m,h}$
1. Milk and Dairy	0.0634	0.0732	0.0669
2. Protein Foods	0.0243	0.0280	0.0219
3. Mixed Dishes	0.1434	0.1295	0.1621
4. Grains	0.0395	0.0462	0.0414
5. Snacks	0.0045	0.0159	0.0045
6. Fruit and Vegetables	0.0341	0.0409	0.0285
7. Beverages	0.0818	0.0748	0.0895
8. Condiments	-0.0109	-0.0074	-0.0128
9. Infant formula & Uncoded	0.0103	0.0027	0.0085
Number of Observations	3304	3261	3320

1.4.1 Elasticities of Food Groups

The number of parameters in our model is more than 50, and the coefficient estimates do not have direct interpretations. Instead, we show the estimated output elasticities of each food group, as a means to display the direction and magnitude of the marginal effects. For each household, we calculate the output elasticity of group k food as follows:

$$e_{k,h} = \frac{\partial \log y_h}{\partial \log x_{k,h}} = \alpha_k + \sum_{j<k} \beta_{k,j} \log x_{j,h} + 2\beta_{k,k} \log x_{k,h} + \sum_{i>k} \beta_{i,k} \log x_{i,h} \tag{1.13}$$

Because the above elasticity is observation-dependent, we take the sample average for each food group. The results are shown in Table 1.3. An input's measured output elasticity represents the percentage increase in output, in response to a 1% increase of this input. For instance, the elasticity on mixed dishes is about 0.14, meaning a 10% increase in mixed dishes consumption would, on average, lead to a 1.4% increase in household energy output.

It is clear that the average elasticities do not sum to one. Therefore, the estimated household production function does not exhibit the constant returns to scale property that is often assumed for a firm's production of material goods. This is a reasonable result considering that the human body, unlike a firm's manufacturing equipment, only has limited capacity to digest food. Most of the elasticities in these models are positive while group 8 (condiments) has negative values in all three models. The negativity on condiments does not undermine the overall validity of the results. Based on the estimates, condiments are not major sources in producing output in the sense that its first-order and second-order coefficients α_8 and $\beta_{8,8}$ are not statistically significant (Appendix Table A.10). Among other groups, the mixed dishes category persistently has the highest elasticity, followed by beverages. This result is consistent with common sense as they are major sources of gaining energy: mixed dishes are typical meals such as pizzas and sandwiches, and the majority of beverage items consist of sweetened products such as soda and tea.

1.4.2 Percentage Food Waste

The average percentage food waste across all households in our baseline model is 31.9% with a standard deviation of 15.8%. By using the sample weights assigned to each household in the data, this result translates to annual U.S. consumer-level food waste valued at $240 billion. After accounting for proxied or imputed physical activities, the average waste decreases to 31.1% (standard deviation 17.1%) in the second model, and to 28.4% (standard deviation 15.6%) in the third model. As discussed earlier, the slightly negative skewness of the distribution of physical activity level and its correlation with output inefficiency may help explain the small bias in the baseline. It is noteworthy that while the total number of available observations in our baseline model is 4072, the number of observations that yield estimates for food waste $\hat{\delta_h}$ is 3304. This discrepancy is a result of transforming the output inefficiency $\hat{u_h}$ into food waste estimate $\hat{\delta_h}$. As some of the transformations do not yield a quadratic solution, the number of observations with waste estimates decreases accordingly, and similar decreases occur with the other two models.

The results suggest that about 30% of the total available food goes uneaten at the consumer level. Since our input measure represents the edible part of food, the estimated amount of food waste should be regarded as avoidable waste. As discussed earlier, because we track household total energy expenditure as the production output, any food recovered or recycled for nonfood or nonhuman uses is considered waste. Therefore, our results do not reveal which households are more efficient in recovering wasted food. In addition, there might be specific parts of food that only some households find edible, such as potato skins. This type of heterogeneous taste preference is not modeled and assumed to be contained in the randomness of the error terms.

Figure 1.2 shows the distribution of the predicted output inefficiency and food waste for the baseline model; the two additional models generate similar results. As the graphs suggest, output inefficiency $\hat{u_h}$ follows a half-normal shape as we assumed. We observe that most households (about 70%) have waste estimates between 20% and 50%. In addition, the results are consistent with increasing marginal costs of reducing food waste (Ellison and Lusk, 2018; Hamilton and Richards, 2019), and we find that even the most efficient household still wastes about 8.7%. Our view is that the baseline model gives a good approximation to the problem despite the issue of physical activities. This point is further supported by the three models' similar results regarding the effects of food waste determinants, as shown below.

1.4.3 Food Waste Determinants

We now turn to the variations in estimated food waste across various households. Table 1.4 lists the coefficient estimates of $\hat{\gamma}$ and their standard errors. Since larger output inefficiency $\hat{u_h}$ corresponds to more food waste $\hat{\delta}_h$, these parameters indicate the effects of demographic variables on the estimated percentage food waste. In all models, the three variables–income per adult equivalent, dietary healthfulness, and food security–are statistically significant

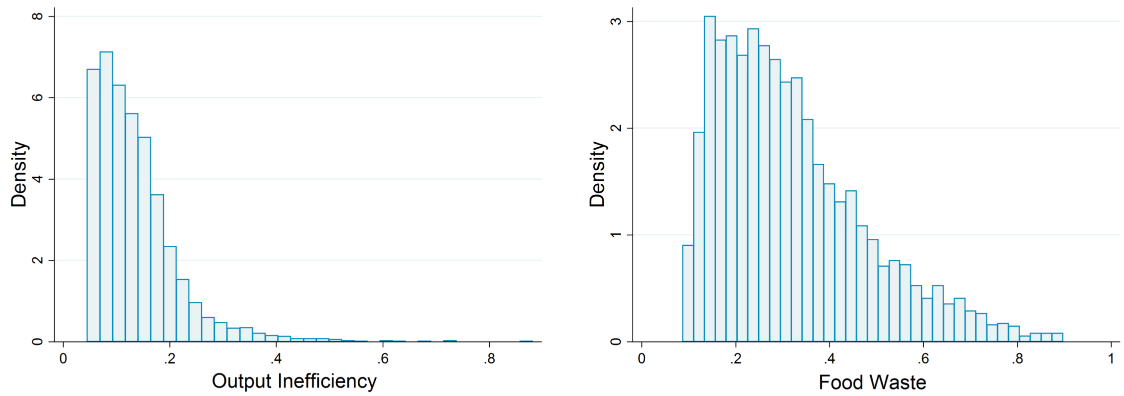


Figure 1.2: Distributions of output inefficiency $\hat{u}_h$ (left panel) and food waste $\hat{\delta}_h$ (right panel) in the baseline model

except food security in the third model. In that one case, the p-value of 0.111 is not far from a 10% significance-level threshold.

The signs of these demographic variables make good economic sense. Income has a positive impact on food waste. Most obviously, households facing less constrained budgets are less efficient managers of food purchases and allocations. Additionally, Figure 1.3 shows that the marginal effect of income on food waste decreases at higher ranges of income. Thus, we discover that "affordable" waste as a share of total food increases at a slower pace than income. This result may relate to Engel's law, which states that the expenditure share of food falls when income rises.

Dietary healthfulness has a positive influence on food waste, as well (Figure 1.4). While this result is unfortunate from a public health viewpoint, perishable produce such as fruit and vegetables are necessary components of healthy eating and a major source of food waste (Leib et al., 2013; Buzby et al., 2014). We examine actual purchase amounts and find that households with the highest self-reported diet quality consume 60% more fruit and vegetables per person than those with the lowest diet quality. This significant finding has important policy implications: policies aimed at promoting healthier diets and increased consumption of perishables have substantial hidden costs from food waste.

Our results for the last determinant, food security, also have a sound and intuitive inter-

Table 1.4: **Food Waste Determinants**

Variable	Baseline	Proxy-instrument	Fitted $PA_{m,h}$
Income	0.3458***	0.3980***	0.3102***
	(0.0601)	(0.0572)	(0.0617)
Healthy Diet	1.5147***	1.0986*	1.7552***
	(0.5871)	(0.5050)	(0.7000)
Food Security	1.9550*	2.6549**	2.6025
	(1.1395)	(1.2281)	(1.6317)
Constant	-6.5729***	-7.0176***	-7.372***
	(1.7434)	(1.5743)	(2.3028)
Number of Obs.	4072	3465	4049

Note: The estimated coefficients are γ and γ_0 in the heteroskedastic specification $\sigma^2_{u_h} = \exp(\gamma_0 + \gamma' d_h)$. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

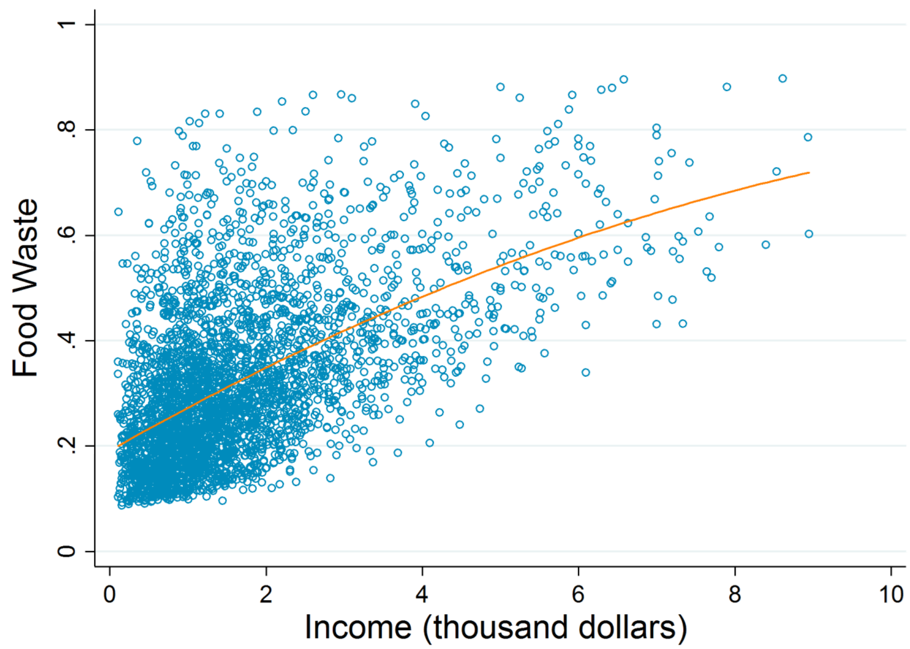


Figure 1.3: Higher income leads to more waste

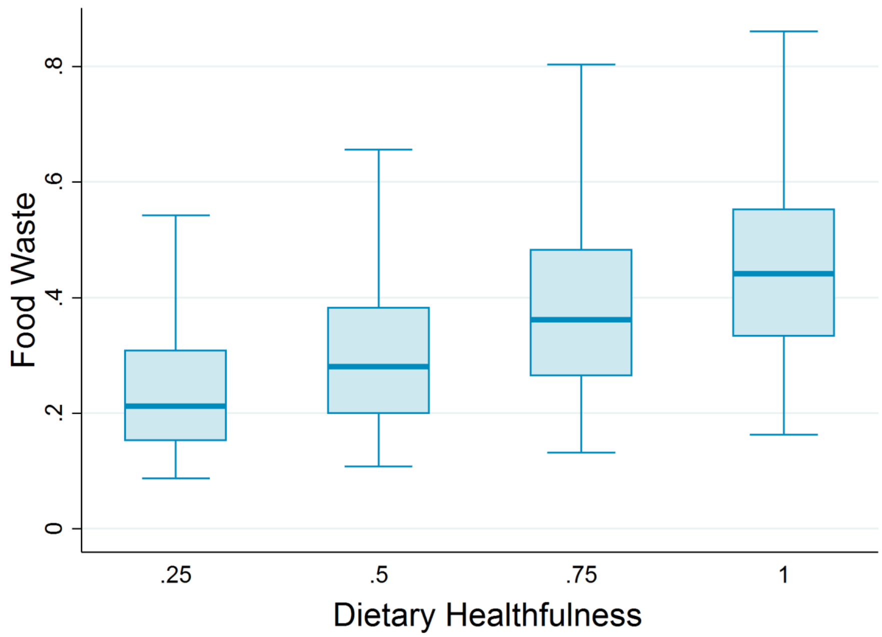


Figure 1.4: Healthy diet leads to more waste (a higher value of dietary healthfulness means healthier diet)

pretation. Table 1.5 lists the average percentage waste along three levels of food security. Food-insecure households waste significantly less than secure ones. On average, the low-food-security households waste only about half the amount of what the high-food-security households waste, e.g., 20.5% vs. 39.9%, 17.7% vs. 40.7%, and 15.7% vs. 37.2% in the three models, respectively. This persistent pattern serves as strong evidence that food-insecure households are less wasteful food managers.

1.4.4 Food Waste and Household Characteristics

The food waste estimates, calculated by equation (1.8), depend on both the output inefficiency $\hat{u_h}$ and food inputs $x_{1,h}, x_{2,h}...x_{I,h}$. The three demographic variables in d_h from the previous section are modeled as affecting output inefficiency $\hat{u_h}$ through food management, taking food inputs as given. Since we have waste estimates for individual households in the sample, it is also feasible to conduct post-estimation analyses on other household-specific

Table 1.5: Food Waste and Food Security

	Baseline		Proxy-instrument		Fitted $PA_{m,h}$	
Food Security	Avg. Waste	S.D.	Avg. Waste	S.D.	Avg. Waste	S.D.
Low	20.5%	(10.4%)	17.7%	(10.4%)	15.7%	(8.8%)
Medium	26.9%	(11.7%)	25.2%	(11.9%)	23.3%	(11.2%)
High	39.9%	(15.0%)	40.7%	(15.9%)	37.2%	(14.4%)

characteristics, including those relating to shopping behavior and purchase decisions that affect food waste through food inputs $x_{1,h}, x_{2,h}...x_{I,h}$.

Here we consider six variables that are of potential interest to policymakers: shopping with a grocery list, distance from primary store (driving time), rural/urban residential status, participation in the Supplemental Nutrition Assistance Program (SNAP), employment status, and education. It is important to clarify that the distinction between these variables and those in d_h is that the former are assumed to reflect shopping decisions before the household production takes place, while d_h takes effect in production. In fact, including these variables in d_h would find them statistically insignificant even though they are strongly correlated with the food waste estimates. In the robustness check section, we discuss the technical challenges in incorporating more variables into d_h and their implications to our results. Lastly, we add household size as another important policy-relevant variable to the post-estimation analysis. Previously, household size was implicitly accounted for in calculating household per adult equivalent income and was not explicitly included to avoid correlation with the food-input quantities. Here we decompose the previous income variable into household monthly total income, household size, and their interactions.

Table 1.6 provides the results of the post-estimation analysis. The dependent variable is estimated food waste measured in percentages. The three demographic variables in d_h, with the modification on income mentioned above, have effects in line with the previous results.

Household size is shown to have a negative impact on food waste. For each household size, we show the average percentage food waste and standard deviation in Table 1.7. We only list households up to six members, which account for more than 95% of the sample. There is a uniform decline in food waste as we move from small households to larger ones. In all models, single-member households are associated with the highest rate of food waste–more than 40% on average; and the percentage is reduced to about half the value for six-member households. These findings suggest that larger households are flexible in managing food purchases and more efficient in allocating the purchases among their members. A single-member household, on the other hand, is less flexible in remedying over-purchased or near-expiring food.

We observe that households who shop with a grocery list generate about 1.5% less food waste. The relationship between households' food waste and the distance from their primary stores is negative, and 30 minutes more driving is associated with about 2% less food waste. Although households farther away from their stores may be less flexible in arranging shopping trips, this result suggests that they might organize a better shopping plan because of the long distance. Moreover, if the store choices are endogenously made, then households who are willing to travel a long distance are likely to be those with relatively lower time cost, implying they spend more time in food management as well.[14] We do not see a significant difference between rural and urban households, and the role of education is not supported in two of the three models. SNAP participation has a negative coefficient yet small in magnitude. We suspect that its explanatory power is mainly reflected by the income and household size variables that are used by the program to determine eligibility.

Here we give a closer examination of food-assistance programs by analyzing two national

[14] We also checked whether households farther away from their primary stores under-reported their purchases at other retail outlets that are possibly closer to their residence. The sample average travel time to primary stores is 9.2 minutes (one-way) and the maximum is 102 minutes. We look at the total food purchases of households who travel more than 10 minutes (1135 households). The average total calories of food purchased by these households is 39877 Kcal during the survey period, higher than the sample average of 37861 Kcal. Moreover, these households reported an average purchase of 690 Kcal from convenience stores, which is also higher than the sample average of 596 Kcal. Although these households report a lower-than-average purchase at farmers' market, the shares of such purchases are extremely small (22 Kcal vs. 34 Kcal). Overall, we do not see clear evidence showing this under-reporting issue.

Table 1.6: **Post-Estimation Analysis of Food Waste**

Variable	Baseline	Proxy-instrument	Fitted $PA_{m,h}$
Total income	1.7226***	2.0549***	1.6703***
	(0.1924)	(0.1981)	(0.1866)
Household size	-3.8184***	-3.8837***	-3.1907***
	(0.2199)	(0.2480)	(0.2156)
Income*household size	-0.1796***	-0.2347***	-0.1944***
	(0.0516)	(0.0558)	(0.0511)
Healthy diet	18.4114***	13.4092***	19.1993***
	(0.9342)	(1.0342)	(0.9000)
Food security	19.0437***	24.3318***	23.0748***
	(0.6971)	(0.7415)	(0.6570)
Shopping list	-1.5037**	-1.1684*	-1.1260*
	(0.6694)	(0.7091)	(0.6340)
Distance primary store	-0.0712***	-0.0652**	-0.0611**
	(0.0253)	(0.0288)	(0.0244)
Rural	-0.6129	-0.4164	-0.5207
	(0.4455)	(0.4793)	(0.4378)
SNAP	-0.8718**	-1.0623**	-0.5478
	(0.4272)	(0.4467)	(0.4144)
Employment	-1.1006	2.6886***	-2.4221***
	(0.7371)	(0.7838)	(0.7216)
Education	2.7483**	1.9281	1.4154
	(1.2646)	(1.3372)	(1.1669)
Constant	16.2976***	11.6882***	9.2042***
	(1.3119)	(1.3819)	(1.2373)
Adjusted R-squared	0.5733	0.5924	0.5947
Number of Obs.	3101	3060	3117

Note: The dependent variable is estimated food waste in percentage terms: $\hat{\delta}_h \cdot 100$. The estimation is ordinary least squares. The robust standard errors are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.7: Food Waste and Household Size

	Baseline		Proxy-instrument		Fitted $PA_{m,h}$	
Household Size	Avg. Waste	S.D.	Avg. Waste	S.D.	Avg. Waste	S.D.
1	44.7%	(18.2%)	44.6%	(20.8%)	39.2%	(18.6%)
2	36.3%	(15.1%)	35.7%	(16.5%)	32.6%	(15.4%)
3	29.7%	(12.9%)	28.9%	(13.7%)	26.4%	(13.0%)
4	25.2%	(10.0%)	24.4%	(11.3%)	22.3%	(10.3%)
5	23.0%	(9.8%)	22.0%	(11.0%)	20.2%	(9.8%)
6	20.3%	(9.7%)	18.9%	(10.6%)	19.0%	(10.7%)

Table 1.8: Food Waste, SNAP and WIC

	Baseline model		Proxy-instrument		Fitted $PA_{m,h}$	
Programs	Avg. Waste	S.D.	Avg. Waste	S.D.	Avg. Waste	S.D.
SNAP						
Not eligible	36.5%	(16.0%)	36.8 %	(17.5%)	33.0%	(16.0%)
Eligible, unenrolled	31.0%	(15.4%)	29.9%	(16.6%)	27.3%	(15.0%)
Eligible, enrolled	24.4%	(12.3%)	22.1%	(12.1%)	21.1%	(12.3%)
WIC						
Not eligible	39.4%	(16.6%)	38.7%	(18.1%)	35.6%	(16.7%)
Eligible, unenrolled	28.0%	(12.7%)	27.1%	(13.8%)	24.7%	(12.5%)
Eligible, enrolled	23.5%	(10.7%)	22.0%	(11.3%)	20.7%	(11.2%)

programs: the SNAP program and the Women, Infants, and Children (WIC) program. In Table 1.8, we list average waste estimates based on the households' eligibility and program participation. Ineligible households (with generally higher income) waste up to 50% more food. Among the eligible households, those actually enrolled in the programs waste less food. We identify two possible reasons: First, the total monthly income of the participating households in the two programs is about \$850 and \$1430 lower, respectively, than non-participating households. In addition, the distributions of household sizes show that the benefit-receiving group consists of a higher number of large households.

1.5 Model Robustness and Validity

To demonstrate the robustness of our approach, we first investigate other choices of input and output measures, as well as demographic variables that determine inefficiency. Next, we examine the possible endogeneity issue of food inputs and apply an instrument-free method to re-estimate the models. In addition, we discuss both the importance of and challenges in incorporating more contextual variables into the analysis. Finally, we discuss the general validity of the approach and the feasibility of replicating the method.

1.5.1 Choice of Variables

In our main specification, food inputs are measured by their weights in grams. Here we re-estimate the models based on their calorie contents. That is, $x_h = (x_{1,h}, x_{2,h}, ...x_{I,h})^T$ now represents the amounts of calories of each food group. In the second and third tests, we use two alternatives for the output measure. The second test re-estimates the models using the Mifflin-St Joer equations, a more recently developed method in medical and nutritional research, to calculate a person's BMR (Frankenfield et al., 2005) (equations are provided in the Appendix). For the third test, based on the rationale in Hall et al. (2009), we take the output as simply the sum of household members' body weights. In all three tests, the average waste estimates are very close to the previous results, ranging from 29% to 37%. The food waste determinants are also in line with our previous numbers (Appendix Tables A.7-A.9). As before, adding physical activities produces lower food waste estimates than the baseline model. Overall, different choices of input and output measures do not result in significant changes in major estimates.

We included three demographic variables in d_h that determine the variance of output inefficiency u_h: household monthly income per adult equivalent, overall self-evaluated dietary healthfulness, and household food security. They are chosen for their direct influence on food management and eating behavior, as well as people's attitudes toward wasting food. As discussed earlier, there are other household-specific variables that are potentially correlated with food waste, such as the use of shopping lists and SNAP benefits. Their effects on food waste are considered indirect, through shopping decisions that are made before the production takes place. Additionally, testing on different demographic variables may shift our estimates of average food waste while patterns across different household groups maintain. For instance, when SNAP benefit status is included in d_h, the average food waste estimates are around 27% in the three models while other waste determinants' coefficients remain similar. These modest robustness checks lead to two main conclusions: (i) Average food waste estimates may vary by several percentage points when using different combinations of demographic variables; nonetheless, ii) the same post-estimation patterns persist across household groups, e.g., food-insecure households waste less.

We also test whether the presence of households who recently received SNAP payments and who practice dieting would affect our results, and we find they do not have significant impacts. The discussions and results are provided in the Appendix.

1.5.2 Input Endogeneity

Like many other applications of stochastic frontier analysis, our model is based on a straightforward estimation of the production technology and its corresponding technical inefficiencies. This approach focuses on the household production stage without modeling the first-stage food purchase decisions. By taking food purchases as given, we analyze how efficient the households are in food management and utilization. Admittedly, there are perhaps some unobserved factors that affect both the output (total energy) and the choice of input quantities. The existence of these factors could make food inputs endogenous variables and consequently undermine the food waste estimates.

In the context of household production, there are two types of sources leading to endogeneity. The first type is related to the estimated household production frontier. Some studies have shown that an individual's food choices are influenced by his/her body measures and physical activities (Simoes et al., 1995; Drewnowski, 1997). For instance, people with higher physical activity levels are found to eat less fat-intense food than those with lower physical activity levels (Simoes et al., 1995). Such correlation itself does not affect the appropriateness of the productivity analysis as long as physical activity is correctly measured. However, if the underlying mechanism is partly determined by some omitted variables, e.g., income and education, then the estimated production frontier is biased. That is, lower income could lead to less physical activities while also leading to unhealthy food choices, which constitutes a traditional "omitted variable" issue. Empirically, these unmeasured factors reside in the symmetric error term v_h that shifts the production frontier.

The second type of endogeneity comes from the correlation between food inputs and the output inefficiency term u_h that is related to food waste. When households make purchase decisions, they have expectations about the likelihood and approximate amount of food waste. In this sense, the first-stage purchase quantities depend on the second-stage food waste determinants. Clearly, there might be some household-specific variables affecting u_h that are not modeled in the analysis. What is less obvious is the role of market or product-specific characteristics. For example, price-sensitive households may lower their purchase quantities when the price of a food product increases. At the same time, they may also improve the utilization of this product because it is now more "expensive" to waste it.

In this section, we use copula estimation, an instrument-free method, to address the above-mentioned endogeneity issues (Nelsen, 2006; Tran and Tsionas, 2015). The idea of this method is very straightforward–since endogeneity takes places when food inputs x_i's are correlated with the error terms, we allow such correlation and directly estimate it. Specifically, we model the joint distribution of the endogenous variables x_i's and the composite error term $\varepsilon = v - u$ through a Gaussian copula:

$$C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = \Phi_{R,I+1}\left(\Phi^{-1}(\xi_1), \Phi^{-1}(\xi_2), ...\Phi^{-1}(\xi_I), \Phi^{-1}(\xi_\varepsilon)\right) \tag{1.14}$$

Here $C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = F(x_1, x_2, ...x_I, \varepsilon)$ for $\xi_i = F_i(x_i)$ and $\xi_\varepsilon = F_\varepsilon(\varepsilon)$ (Sklar, 1959), Φ is the standard normal distribution, and $\Phi_{R,I+1}$ is a $(I+1)$-dimension normal distribution with correlation matrix $R \in [-1, 1]^{(I+1)\times(I+1)}$ as its covariance matrix. A copula can be considered as a flexible parametrization of the original joint distribution $F(x_1, x_2, ...x_I, \varepsilon)$ by exploiting the relations among its marginal distributions. We then derive the density function of the copula and conduct a maximum likelihood estimation.

The Appendix provides full details of the method, including the equivalence between a copula and its corresponding original joint distribution, the derivation and estimation of the likelihood function, as well as a comparison between copula and traditional instrumental-variable and structural approaches.

Because of its flexibility and instrument-free technique, copula estimation has become an emerging method in economic studies. Danaher and Smith (2011) use copula to model correlations in a multivariate analysis of marketing data. Park and Gupta (2012) utilize the method to tackle endogeneity issues in linear regressions and logit models. Goodwin and Hungerford (2014) apply a set of different copulas to evaluate U.S. agricultural insurance programs. Our specification of copula estimation is similar to Tran and Tsionas (2015), which is the first study that applies the method to stochastic frontier analysis. They find that the copula method works as effectively as the generalized method of moments, while the latter requires using instrumental variables.

The results of the copula estimation largely confirm our previous findings. The average wastes across all households are 33.2% (standard deviation 16.1%) and 29.7% (standard deviation 15.8%) when the output measures are total BMR and total energy expenditure, respectively. An interesting observation is that now we have slightly larger mean elasticities for some food categories. For instance, the elasticity on beverages is now about 0.16, comparing to 0.08 in the baseline model. We suspect that this is a result of the fact that beverages are storable and that households maintain an inventory for this category.[15]

Table 1.9 shows the estimated 10×10-dimension correlation matrix R when the output measure is total energy expenditure. The parameters in R measure the dependence between the endogenous variables and the error term, as well as the dependence among endogenous variables themselves. The first column contains the correlations between the transformed error term $\Phi^{-1}(\xi_\varepsilon)$ and food inputs $\Phi^{-1}(\xi_1), \Phi^{-1}(\xi_2), ...\Phi^{-1}(\xi_9)$, which roughly indicate how "endogenous" the food inputs are. Other than the ninth food group, the input variables are positively correlated with the error term. Because larger ε corresponds to smaller output inefficiency, this result supports our previous findings regarding household size–larger households who purchase more food actually waste less. All the estimated correlations are statistically significant at 10% level, except for the correlation between $\Phi^{-1}(\xi_\varepsilon)$ and $\Phi^{-1}(\xi_9)$. We also see that the correlations among food inputs themselves are generally much larger than the "endogeneity" correlations represented by the first column, with the exception of the ninth food group which is a small "catch-all" category for the uncoded food items.

1.5.3 Contextual Variables

In studies of consumer economics, researchers often emphasize the role of contextual variables. We are interested in knowing what particular household characteristics or environmental factors are associated with the behavior under study, such as wasting food. In this

[15] As discussed in the Appendix, when the frequency of replenishing the inventory is correlated with some household characteristics, the coefficient estimates could be biased if endogeneity is not dealt with.

Table 1.9: **Estimated Correlation Matrix, Copula Estimation**

	$\Phi^{-1}(\varepsilon)$	$\Phi^{-1}(\xi_1)$	$\Phi^{-1}(\xi_2)$	$\Phi^{-1}(\xi_3)$	$\Phi^{-1}(\xi_4)$	$\Phi^{-1}(\xi_5)$	$\Phi^{-1}(\xi_6)$	$\Phi^{-1}(\xi_7)$	$\Phi^{-1}(\xi_8)$	$\Phi^{-1}(\xi_9)$
$\Phi^{-1}(\varepsilon)$	1.000									
$\Phi^{-1}(\xi_1)$	0.076	1.000								
$\Phi^{-1}(\xi_2)$	0.076	0.369	1.000							
$\Phi^{-1}(\xi_3)$	0.097	0.282	0.377	1.000						
$\Phi^{-1}(\xi_4)$	0.067	0.472	0.516	0.323	1.000					
$\Phi^{-1}(\xi_5)$	0.082	0.396	0.394	0.387	0.446	1.000				
$\Phi^{-1}(\xi_6)$	0.070	0.365	0.472	0.317	0.450	0.383	1.000			
$\Phi^{-1}(\xi_7)$	0.089	0.300	0.383	0.445	0.358	0.460	0.315	1.000		
$\Phi^{-1}(\xi_8)$	0.078	0.374	0.455	0.241	0.488	0.437	0.422	0.309	1.000	
$\Phi^{-1}(\xi_9)$	-0.023	0.039	0.064	0.351	0.058	0.070	0.081	0.035	0.038	1.000

Note: $\xi_i = F_i(x_i)$ and $\xi_\varepsilon = F_\varepsilon(\varepsilon)$. All estimates are statistically significant at 10% level, except for the correlation between $\Phi^{-1}(\xi_\varepsilon)$ and $\Phi^{-1}(\xi_9)$.

chapter, we consider three variables in d_h that directly affect the variance of the output inefficiency term. Our post-estimation analysis partly compensates for this relatively simple specification by showing how household characteristics could indirectly affect food waste through the shopping planning and purchase decisions. Below, we will discuss in further detail two major obstacles in formulating more contextual variables within the framework of stochastic frontier analysis and their implications on the accuracy of our waste estimates.

The first issue is closely related to the input endogeneity concern discussed in the preceding section. In stochastic frontier analysis, the variables in d_h are called external or environmental factors (recall that $\sigma^2_{u_h} = \exp(\gamma_0 + \gamma' d_h)$). When directly adding more contextual variables into d_h, we are facing the possibility that the new contextual variables are correlated with food inputs $x_{i,h}$'s, causing potential correlation between u_h and food inputs as well. Because of this concern, incorporating a comprehensive set of contextual variables has been a challenging task in productivity analysis research (Wang and Schmidt, 2002; Alvarez et al., 2006; Simar and Wilson, 2007; Banker and Natarajan, 2008). The Appendix provides a detailed review of the literature and recent development of this subject. Using the methods in Alvarez et al. (2006), we explicitly test if more contextual variables can be added as food waste determinants while still maintaining consistency of the parameter estimates. The results show that even adding one or two variables would likely lead to endogeneity issues.

Second, there is an identification problem involved in distinguishing two types of contextual variables: those that shift the production frontiers and those that shift the output inefficiency. In other words, households that produce less output may exhibit a different production technology rather than being inefficient. Let us write the production function as $\log y_h = \text{translog}[x_h] + s(c_h) + v_h - u_h(d_h)$, where c_h is a set of contextual variables that shift the production frontier. Few studies have considered the frontier-shifting contextual variables (Linna, 1998; Greene, 2004). The obvious challenge, in our case, is that c_h and d_h are likely to be correlated or even overlap in some of their components. The more difficult problem is that by saying $s(c_h)$ controls the production frontier and $u_h(d_h)$ controls the inefficiency, we are imposing very strong assumptions on the model structure. Econometrically, this model is not distinguishable from the one that considers $s(c_h) - u_h(d_h)$ as the overall output inefficiency as in Kumbhakar et al. (1991).[16] Therefore, unless we have absolute knowledge about which contextual variables affect production frontier/output inefficiency and how this mechanism works, the way of interpretation is up to the researchers.

In sum, contextual variables are important in studying household consumption behaviors as they provide useful guidance for policy implementation. The endogeneity issue and the difficulty in correctly specifying the parameter structure constrain us from including more contextual variables into our stochastic frontier model. One possible solution for future research is to build structural models that explicitly formulate how contextual variables

[16]In this case, c_h is not a frontier shifter but an inefficiency shifter. This identification challenge is called by William Greene as the "where do we put the z's" problem in Fried et al. (2008).

enter the production function and the output inefficiency term. The bottom line is that (a) the copula method provides consistent estimates for the model parameters and food waste, even when the omitted contextual variables correlate with food inputs, and (b) these copula-method estimates are similar to our original estimates of the baseline model.

1.5.4 Validity and Replication

Finally, we check the general validity of our approach and the feasibility of replicating the method. First, we test our method using NHANES data, which reflect pure food intakes. Since there is no household production or food management involved, we re-estimate the model without the three demographic variables that determine inefficiency. Theoretically, because NHANES only records actual food intakes, our method should predict zero estimates of food waste if it is valid. In fact, using the NHANES data, we estimate the average waste to be 1.05% (standard deviation 0.8%) when output is BMR; and 0.66% (standard deviation 0.56%) when output is total energy expenditure. Thus, we feel confident that the food waste estimates from our baseline model using the FoodAPS data are not some artifact of the stochastic frontier estimation.

Recent studies have shown that there is systemic under-reporting of dietary intakes in the NHANES data (Archer et al., 2013; Briefel et al., 1997; Subar et al., 2015). It may seem that this pattern of under-reporting would generate overly efficient productivity and, thus, underestimation of food waste. However, the production frontier model implies quite the opposite. On the one hand, if the under-reporting is truly systemic, then food waste estimates will remain the same since all the individuals face the same shift in the estimated production frontier. On the other hand, if the under-reporting only applies to a certain group of people, then it will decrease the estimated food waste for this group and, at the same time, increase the waste estimates for the rest of the population. Because food waste is already constrained at zero in the pure food-intake NHANES data (cannot decrease further below zero), the dominant effect is an increase in estimated food waste for the individuals who do not under-report. Thus, the overall effect is, in fact, an overestimation of food waste. The intuition is that food waste is a deviation from the "best practice" and that the presence of the under-reporting group would lift the standard of "best practice", making all others less efficient.

Second, as many widely used consumer datasets contain only food-at-home purchases, we investigate whether the approach proposed by this chapter remains useful for these datasets. More specifically, we re-estimate our original models for the FoodAPS data but omit food acquired for away-from-home consumption. Our analysis shows that the baseline and the proxy-instrument models give average waste estimates of 28.2% and 32%, respectively. Moreover, the effects and significance of demographic variables are preserved as well. As for the third model where physical activities are imputed, it gives fairly high estimates that are above 50%, on average. This issue can be possibly attributed to the correlation between the imputed physical activities and the missing away-from-home consumption, which can be tackled by adding a proxy variable. After adding employment status as a proxy to account for food-away-from-home, the third model gives an average waste at 30.2% and consistent waste-determinant estimates.

Alternatively, in practice, one can also impute the shares of food-away-from-home consumption based on household characteristics when replicating the method. One of the key reasons for the model to work in absence of away-from-home data is that most of the food groups consist of mainly at-home consumption. With the exception of mixed dishes and uncoded food, the shares of food-at-home for other groups are close to or above 70% of total food acquisitions. These results suggest that the stochastic frontier approach is highly replicable to other widely used datasets where food-away-from-home information is not available.

1.6 Conclusions

This study overcomes data insufficiencies by implementing an inefficiency analysis to indirectly estimate household food waste. Our estimates of average food waste, which range from 28.4% to 31.9%, are in line with the existing findings at the aggregate level. More importantly, by obtaining a waste estimate for each household, we are able to conduct a series of analyses on the relationship between food waste and various household characteristics. For example, we see a clear link between food waste and levels of dietary healthfulness, and this link may be crucially important for policies aimed at either food waste reduction or dietary improvement. Our results are generally consistent across a wide set of robustness checks, including different choices of input and output measures, the presence of households participating in food-assistance programs, and the issue of endogenous food-input variables. It is also important to emphasize that although our output measure is built upon an explicit physiological foundation, the inefficiency-analysis approach can be potentially applied to other output choices given that there exist well-defined functional relationships. For instance, as we show in the robustness checks, simply using the sum of household members' body weights as an output measure for our sample would generate similar waste estimates.

The highly replicable method employed in this study can advance further food waste research and extensions such as calibrating differences between widely available purchase data and unobserved actual consumption. Policymakers who are interested in knowing the amounts of actual consumption may find our method useful and convenient. This usefulness is particularly important when the policy concerns nutrition intakes or food waste reduction. Because we use the edible portions of food as inputs in the production function, the waste estimates in this chapter directly correspond to avoidable food waste. We also show that our percentage waste estimates are identical to those resulting from a model where inputs are total purchase quantities. Potentially, our research could be extended by estimating different waste rates for separate food groups. In addition, developing a structural or behavioral model would help investigate the underlying food-wasting mechanisms by rational households.

Collectively, our results help illustrate our contribution in the context of previous research on food waste. While the precise measurement of food waste is important, it may be equally important to investigate how household-specific factors influence food waste. Our indirect estimation strategy allows us to accomplish this task. Thus, we hope that this approach provides other researchers working on the topic a new lens through which estimating individual household food waste becomes feasible, and that it encourages them to extend the idea of indirect measurement to applications on other data and interesting cases.

Chapter 2

Rational Food Waste and Consistent Estimation of Consumer Demand

2.1 Introduction

On average, more than 30% of all perishable foods are wasted at the consumer level (Yu and Jaenicke, 2020a; Buzby et al., 2014). Food waste is widely recognized as both an economic and environmental issue. Agencies such as the Economic Research Service of the U.S. Department of Agriculture (USDA-ERS), the U.N.'s Food and Agriculture Organization (FAO), and the Waste & Resources Action Programme of U.K. (WRAP) provide aggregate-level estimates of food waste that indicate the significant welfare and environmental impacts of food waste (Buzby et al., 2014; FAO, 2013; Quested and Parry, 2011).

More recently, researchers have made efforts to reach consistent definitions of food waste and more comprehensive waste measurements at the micro household level (Bellemare et al., 2017; Landry and Smith, 2019; Yu and Jaenicke, 2020a). These definitions and estimates offer new insights into the underlying mechanism that generates household food waste, as well as useful policy guidance for programs that aim to reduce consumer food waste. In addition, there is an increasing number of studies that utilize various survey and experimental methods to uncover the behavioral and economic drives of food waste (see *inter alia* Stefan et al. 2013; Reynolds et al. 2014; Neff et al. 2015; Secondi et al. 2015; Qi and Roe 2016; Ellison and Lusk 2018; Roe et al. 2018a; Wilson et al. 2017).

However, in most scholarly discussions, the major attention is paid to the importance of many policy-and welfare-related issues. This chapter studies food waste from another important, yet under-studied, perspective–its implications on the methodological approaches used in traditional demand analysis. In short, it demonstrates, both theoretically and empirically, that omitting food waste in consumer demand models would potentially generate inconsistent parameter estimates. More specifically, I first illustrate this point through a simple, yet a general example of utility-maximizing households. The intuition is that wasting food is often a rational choice and, hence, is linked to many household-or product-specific characteristics (Lusk and Ellison, 2017; Yu and Jaenicke, 2020b; Hamilton and Richards, 2019). The most obvious case that leads to biased demand parameters is where those omitted characteristics are correlated with the observed demand shifters, which is a typical omitted-variable problem. For instance, a product's attributes such as its shelf-life directly affect household food management and the resulting food waste, while, at the same time, also influence the purchase amount of the product through their correlations with its price.

Furthermore, the chapter points out that the consequence of not modeling food waste is often more than an endogeneity issue as a result of omitted variables. More importantly, it is also a misspecification problem in which the underlying decision-making structure is not modeled. For this reason, traditional instrumental-variable approaches could not sufficiently address the issue of biased demand estimates.[1]

Empirically, the presence of food waste adds an extra dimension to the space of households' decision variables, e.g., they need to choose both how much to purchase and how much to waste. This causes analyses that only consider the observed purchase quantity to likely underestimate consumer responses to price changes, as found in this chapter's empirical anal-

[1] A detailed discussion on the (in)validity of the traditional instrumental-variable approach in addressing the issue of food waste is provided in the next section.

ysis. As Hamilton and Richards (2019) note, food waste makes the demand of perishable products more price-elastic. The intuition is that perishability constrains households from fully utilizing their inventories when facing price discounts. On the other hand, they have additional flexibility by reducing amounts of food waste when price increases. This observation is a direct result of perishability of food products. Indeed, if the products are non-perishable, then different conclusions could be drawn. As found in the studies of household inventory of storable goods, own-price elasticities are often overestimated in models that overlook the households' stockpiling behavior (Hendel and Nevo, 2006; Wang, 2015).

The lack of existing studies that take into account food waste is mainly due to the fact that food waste is almost always unobservable in empirical datasets.[2] To address the adverse consequences of omitting food waste in analyzing consumer demand, this chapter examines two approaches that structurally identify and estimate food waste. The first approach partially identifies food-waste cost through economic constraints. Specifically, I impose conditions to ensure that the indirect utility function is non-decreasing in food prices and in the cost of reducing food waste. Additionally, food waste is bounded between zero and the purchase amount. I use a Bayesian estimation method that incorporates these constraints into a prior distribution (Tsionas and Izzeldin, 2018). This approach is built upon a general specification of consumer demand, and, hence, can be applied to a wide range of consumer datasets.

The second approach imposes behavioral assumptions to the household utility function, in which households must sustain a certain "minimum" or "target" level of utility drawn from actual consumption, regardless of the choice of food waste. Similar formulations can be found in the literature of productivity analysis, which analyze the inefficiency associated with meeting a target output (Bogetoft and Hougaard, 2003; Asmild et al., 2013; Asmild and Matthews, 2012). An example is that, despite the differences in taste preference and ability to manage food waste, individuals must obtain enough calories from food consumption that meets basic energy requirements for body functioning and physical activities.[3]

[2]An exception is the 1977-78 Nationwide Food Consumption Survey used in Landry and Smith (2019).

[3]This approach can be generalized to a case where taste preference is also modeled, which is briefly

This minimum-level requirement, whose value is determined by household-specific characteristics, provides a limiting bound for actual consumption and helps identify food waste. The approach's identification power generally depends on how informative this bound is in explaining the variation among household food purchases.

The rest of the chapter is organized as follows: First, I provide a general theoretical example to illustrate the importance of considering food waste in modeling consumer demand. I also discuss the reasons why the instrumental-variable methods do not address the issue of inconsistent estimates. Next, I present two structural approaches and apply them to a nationally representative food acquisition dataset. I also briefly discuss another possible structural model based on directional distance function that may be useful for identifying food-group-specific waste rates. And the final section compares the proposed approaches and concludes the chapter.

2.2 A Theoretical Example

To provide an intuitive example as to why omitting food waste could be associated with biased demand estimates, consider the following specification of household utility function:

$$U = U(\boldsymbol{q}, \boldsymbol{\delta})$$

where $\boldsymbol{q} \in R_{+}^{J}$ is the vector of purchased amounts of food from J categories, and $\boldsymbol{\delta} \in R_{+}^{J}$ is the amounts of food waste (unobserved). Households maximize their utility by optimizing over these two variables. Equivalently, one of the choice variables can be replaced by the amount of actual consumption $\boldsymbol{x} \in R_{+}^{J}$ since $\boldsymbol{q} = \boldsymbol{x} + \boldsymbol{\delta}$. In addition, the budget is given by m and price vector by $\boldsymbol{p}$ so that the budget constraint is $\boldsymbol{p}'\boldsymbol{q} \leq m$.

More importantly, households' decisions on how much to waste are typically conditioned on several factors, including food management ability and time cost, which is denoted as a

discussed later.

vector $\boldsymbol{w}$. In particular, $\boldsymbol{w}$ may contain exogenous product-specific variables such as sell-by dates (shelf-lives). I assume that $\boldsymbol{w}$ enters the model through the utility function $U(\cdot)$ such that it is costly to reduce waste, e.g., efforts in reducing waste generates disutility. Next, consider the indirect utility function $V(\boldsymbol{p}, \boldsymbol{w}, m)$ and expenditure function $e(\boldsymbol{p}, \boldsymbol{w}, u)$ generated from the household utility-maximization problem. By Roy's identity and Shephard's lemma (duality theorems), the Marshallian demand and Hicksian demand are found as:

$$\boldsymbol{q} = -\frac{1}{\partial V(\boldsymbol{p}, \boldsymbol{w}, m)/\partial m} \nabla_{\boldsymbol{p}} V(\boldsymbol{p}, \boldsymbol{w}, m) \quad \text{(Marshallian)} \tag{2.1}$$

$$\boldsymbol{h} = \nabla_{\boldsymbol{p}} e(\boldsymbol{p}, \boldsymbol{w}, u) \quad \text{(Hicksian)} \tag{2.2}$$

These two demand functions are foundations for many traditional specifications based on Gorman form and PIGLOG-class demand systems, such as AIDS and Translog models. As we can see, unless $V(\boldsymbol{p}, \boldsymbol{w}, m)$ and $e(\boldsymbol{p}, \boldsymbol{w}, u)$ are linear in food-waste-determinant factors $\boldsymbol{w}$, it constitutes an omitted variable or misspecification issue if a demand estimation only includes prices and income. However, in most flexible functional specifications such as the quadratic form, the indirect utility and expenditure functions are nonlinear in $\boldsymbol{w}$. It is noteworthy that there shall be a demand function for food waste as well. In this case, in addition to the above-mentioned problems, there is the so-called seemingly unrelated regressions issue if we only estimate the observed purchase quantities $\boldsymbol{q}$.

When the only issue of concern is the endogeneity of price $\boldsymbol{p}$ and income m resulting from omitting $\boldsymbol{w}$, then one would naturally consider an instrument-variable approach such as the Generalized Method of Moments (GMM). However, there are several cases where the instrument-based methods do not offer a satisfactory solution to this problem.

First, since $\boldsymbol{w}$ may include product-or food-group-specific characteristics, e.g., perishability, it is relatively hard to find instruments that are correlated with prices but not with these characteristics. Similarly, many household-specific variables such as income and household size may affect both purchase quantity $\boldsymbol{q}$ and food waste $\boldsymbol{\delta}$, and such complex features render it a difficult task to construct good instruments that satisfy the exclusion criteria.

Second, food waste represents a large portion of the total available food at the aggregate level (Buzby et al., 2014; Leib et al., 2013). At the individual-household level, both Yu and Jaenicke (2020a) and Landry and Smith (2019) observe significant variation in food waste across household groups. While the most efficient households manage to utilize more than 90% of their purchases, the less efficient ones could waste half of their food. As noted by Yu and Jaenicke (2020a) in their discussion of input endogeneity, this fact implies that when an instrument is statistically valid, it is possibly a "weak instrument" because it does not have sufficient capacity to account for a large portion of the variation in observed purchases (Angrist and Krueger, 1991; Bound et al., 1995).

Third, and most importantly, misspecification is a particular issue for the problem of food waste. The instrumental-variable approaches work well only when the omitted variables are separated from the observed variables, i.e., they are linearly additive in equation (2.1). On the other hand, they could interact with each other, which constitutes a misspecification problem. Consider the following simple one-good example, where quantity measures q and δ, as well as waste determinant w are all single-dimensional variables:

$$\max_{q,\delta} u(q-\delta) + v(\delta) - pq = (q-\delta)^{\alpha} + w\delta^{\beta} - pq$$

Here the cost of reducing waste is a function of the amount of δ only and independent of q, with $v'(\cdot) \geq 0$. This specification implies that a household generates utility $v(\delta)$ from wasting food (Bogetoft and Hougaard, 2003; Asmild and Matthews, 2012; Asmild et al., 2013). It is equivalent to the negative of a cost function: $v(\delta) = -C(\delta)$, where $C'(\cdot) \leq 0$, i.e., more waste results in less cost.[4] It is easy to verify that the optimal purchase quantity

[4]Some studies use the percentage-based waste cost function: $u(\theta q) - C(\theta)$, where θ is the utilization rate of food $\theta = (q-\delta)/q$ and $C'(\cdot) \geq 0$ (Hamilton and Richards, 2019; Yu and Jaenicke, 2020b; Landry and Smith, 2019). The amount-based formulation is appropriate as long as important household-specific factors such as household size are controlled.

q^* and food waste δ^* are given in the following form:

$$q^* = \left(\frac{\alpha}{p}\right)^{1/(1-\alpha)} + \left(\frac{w\beta}{p}\right)^{1/(1-\beta)}$$
$$\delta^* = \left(\frac{w\beta}{p}\right)^{1/(1-\beta)}$$

Here the omitted variable w interacts with price p in the first equation of the demand system. In this case, it is practically difficult for the researcher to find a transformation that would make a regression of q^* on p still consistent, even with the help with instrumental variables. Mathematically, such interaction implies that the estimated coefficient on p is also a random variable–the so-called "slope endogeneity" problem (Heckman, 1976), which happens when there is an underlying decision-making structure not being modeled.

In addition, there is also a risk of misinterpretation that researchers may face. Note that the first-order condition for the purchase quantity in the above problem is:

$$q = \left(\frac{\alpha}{p}\right)^{1/(1-\alpha)} + \delta$$

This is a structural equation, not a reduced-form equation. If the researcher ignores the role of food waste, a separate decision variable, and only considers the utility function as $u(q) = q^\alpha - pq$. Then he/she may estimate $q = \left(\frac{\alpha}{p}\right)^{1/(1-\alpha)}$ instead and interpret the coefficients as reduced-form parameters, which may lead to incorrect conclusions in counterfactual analysis.

In sum, structural approaches that are derived from the economic theory and incorporated with mechanisms describing rational food waste are better candidates to produce consistent demand estimates. In addition, the instrumental-variable approach does not provide waste estimates, while structural analysis can provide useful information to infer the amount of food waste. Many policymakers are interested in knowing the actual amount of consumption, not just purchase quantities. This is especially true when the policy concerns nutrition intakes and healthy eating behavior. Next, I will discuss the identification challenges in estimating demand models involving food waste and present two structural approaches in detail.

2.3 Structural Identification of Food Waste

Theoretically, once we have a specification of $U = U(\boldsymbol{q}, \boldsymbol{\delta})$ and the corresponding constraints, solving for optimal level of purchase and waste is straightforward. The maximization problem can be carried out in either a two-stage decision making, e.g. deciding $\boldsymbol{q}$ first and $\boldsymbol{\delta}$ second, or a simultaneous decision involving the two. Because I do not explicitly incorporate uncertainty, the optimal solutions from the two scenarios are identical, given that both solutions exist.[5] Various theoretical properties regarding optimal food waste can be explored (as seen in Hamilton and Richards 2019). For instance, if the cost of reducing food waste is only a function of $\boldsymbol{\delta}$, then given the purchase quantities, households solve the problem of allocating the purchased food between actual consumption and food waste, which is similar to the optimal distribution within an Edgeworth box.

The main challenges reside in the empirical estimation, which requires appropriate and effective strategies to identify food waste. Since waste is not observed, it is generally not feasible to directly estimate it or its associated cost function. For instance, in the earlier example where $U = u(\boldsymbol{q} - \boldsymbol{\delta}) + v(\boldsymbol{\delta})$, distinguishing $v(\cdot)$ from $u(\cdot)$ would be an empirically difficult task without further structural assumptions. The problem can be seen more clearly in the following relation, where the vector $\boldsymbol{b}_h$ represents variables that affect the demand of actual consumption, and $\boldsymbol{w}_h$ is food-waste-determinant factors.

$$\begin{array}{cccc} \text{Purchase Quantities} = & \text{Actual Consumption} & + & \text{Food Waste} \\ \boldsymbol{q}(\boldsymbol{b}_h, \boldsymbol{w}_h; \boldsymbol{p}_h, m) = & \boldsymbol{x}(\boldsymbol{b}_h; \boldsymbol{p}_h, m) & + & \boldsymbol{\delta}(\boldsymbol{w}_h; \boldsymbol{p}_h, m) \end{array}$$

In order to disentangle the observed purchase quantities, we need to identify either actual

[5]This can be easily verified by the first-order conditions.

consumption $\boldsymbol{x}(\boldsymbol{b}_h; \boldsymbol{p}_h, m)$ or food waste $\boldsymbol{\delta}(\boldsymbol{w}_h; \boldsymbol{p}_h, m)$, neither of which are generally observed in market data. Based on this idea, the two approaches proposed in this chapter take two different routes to structurally estimate food waste.

The first approach attempts to (partially) identify the utility drawn from food waste $\boldsymbol{\delta}(\boldsymbol{w}_h; \boldsymbol{p}_h, m)$. It applies duality theorems on utility maximization or expenditure minimization and parametrizes the cost structure based on $\boldsymbol{w}$. In this model, the marginal cost of reducing food waste can be regarded as a "shadow price" for the utility generated from wasting food. By taking the partial derivative of the indirect utility function with respect to this price, one can obtain a demand function for food waste. Hence the identification comes from the variation in observed variables in $\boldsymbol{w}$ that determine the shadow price. However, since this demand is not directly observed, I exploit bounds for food-waste cost using a Bayesian estimation in which economic restrictions, e.g., the indirect utility is non-increasing in prices and food waste is non-negative, are used to reach partial identification.

The second approach, on the other hand, explores a behavioral perspective related to actual consumption $\boldsymbol{x}(\boldsymbol{b}_h; \boldsymbol{p}_h, m)$. This model assumes that regardless of the choice of food waste, a household must sustain a certain "minimum" utility level drawn from the actual consumption. Hence, households with certain characteristics, e.g., higher wage, busy schedules, may waste more food while having the same amount of actual food intakes. This theoretical implication enables the empirical identification by first estimating this "minimum" output and attributing the rest of the variation to food waste.

The two approaches are largely built upon recent studies in productivity economics and firm-level efficiency analysis (Asmild and Matthews, 2012; Asmild et al., 2013; Atkinson and Tsionas, 2016; Bogetoft and Hougaard, 2003; Feng and Serletis, 2014; Tsionas and Izzeldin, 2018; Malikov et al., 2016), which are adopted and refined for the purpose of studying household production. For simplicity, in both approaches, I assume that $U(\cdot)$ is quasilinear and the price of the outside good is one so that there is no income effect. Below, I will discuss the detailed model assumptions, the estimation strategies and results of the proposed approaches to uncovering food waste.

2.4 Approach 1: Partial Identification by Duality

This approach explores the duality theorems applied to the household utility maximization or expenditure minimization. I consider the following utility maximization problem for household h as an example, whereas the case of expenditure minimization is similar:

$$\max_{\boldsymbol{q_h},\boldsymbol{\delta_h}} U(\boldsymbol{q}_h, \boldsymbol{\delta}_h) = u(\boldsymbol{q}_h - \boldsymbol{\delta}_h) + c(\boldsymbol{w}_h) \cdot v(\boldsymbol{\delta}_h) - \boldsymbol{p}_h' \boldsymbol{q}_h \tag{2.3}$$

Here the marginal cost of reducing food waste $c(\cdot)$ is determined by a set of exogenous variables $\boldsymbol{w}_h$ that is related to household food management and production. Denote the optimal amounts of purchase and waste as $\boldsymbol{q}_h^*$ and $\boldsymbol{\delta}_h^*$. Then by duality (envelope theorem), the following relations hold:

$$-\frac{\partial U_h^*}{\partial p_{h,j}} = q_{h,j}^* \quad \text{for } j = 1, ...J$$
$$\frac{\partial U_h^*}{\partial c_h} = v(\delta_h^*)$$

Consider the indirect utility function $V(\boldsymbol{p}_h, c_h)$ that is parametrized to a quadratic form:

$$V(\boldsymbol{p}_h, c_h) = \beta_0 + \beta_p' \boldsymbol{p}_h + \beta_c c_h + \frac{1}{2} \boldsymbol{p}_h' \Gamma_{pp} \boldsymbol{p}_h + \boldsymbol{p}_h' \Gamma_{pc} c_h + \frac{1}{2} \gamma_{cc} c_h^2 \tag{2.4}$$

Then the duality results, with error terms added, translate to:[6]

$$-\boldsymbol{q}_h^* = \beta'{}_p + \Gamma_{pp} \boldsymbol{p}_h + \Gamma_{pc} c_h + \varepsilon_{q,h} \tag{2.5}$$

$$v_h^* = \beta_c + \Gamma_{pc}' \boldsymbol{p}_h + \gamma_{cc} c_h + \varepsilon_{v,h} \tag{2.6}$$

[6] Alternatively, one can use logarithms of quantities and prices, in which case, the demand equations are expressed in terms of expenditure shares.

The system is complete with a specification for $c_h(\boldsymbol{w}_h)$, where μ_h and η_t are household and time fixed effects, respectively:

$$c_h = \mu_h + \eta_t + \alpha' \boldsymbol{w}_h + \varepsilon_{c,h} \tag{2.7}$$

If food waste is observed and the functional form of v_h^* that describes the utility drawn from food waste is specified, the system can be easily estimated simultaneously using maximum likelihood. However, as data on individual-household level food waste is typically unavailable, I adopt a Bayesian estimation in which prior distributions of β_c and γ_{cc} in equation (2.6) are updated through a set of constraints.[7] The estimation procedure is fully developed in Tsionas and Izzeldin (2018). Specifically, I use restrictions that are consistent with economic theory: $\partial V/\partial \boldsymbol{p}_h \leq 0$, $\partial V/\partial c_h \leq 0, c_h > 0$. Also note that it is assumed that $v(\boldsymbol{\delta}_h) < 0$ and $v' > 0$ so that $c(\boldsymbol{w}_h) \cdot v(\boldsymbol{\delta}_h)$ can be regarded as a utility function of wasting food that has negative values. These constraints are evaluated at the expectations conditional on $\boldsymbol{p}_h$ and $\boldsymbol{w}_h$, i.e., $E(\partial V/\partial \boldsymbol{p}_h | \boldsymbol{p}_h, \boldsymbol{w}_h) > 0$. They must be satisfied for every observation and, hence, provide a channel for partial identification. Meanwhile, the power of identification depends on the observed variations in prices and cost. Note that the concept of partial identification here is not identical to the usual concept of set identification used in the frequentist view. However, it can be related to the estimation procedure of conditional moment inequalities which can be applied to estimate this model and generate set estimates.

Denote all the model parameters as a vector $\boldsymbol{\theta}$, and the observation-specific constraints as $\mathbb{M}(\boldsymbol{\theta}; \text{data}) \leq 0$. The prior is then defined as a flat distribution over the region that satisfy the restrictions:

$$p(\boldsymbol{\theta}) \propto \mathbb{I}(\mathbb{M}(\boldsymbol{\theta}, \text{data}) \leq 0) \tag{2.8}$$

The likelihood function is based on equations (2.5) and (2.7) where the outcome variables are observed. The joint distribution of $\varepsilon_{q,h}$ and $\varepsilon_{c,h}$ is assumed to follow a multi-variate

[7] Another possible technique is to use moment inequalities to estimate the model with the inequality constraints.

normal $N(0,\Sigma)$. Denote the likelihood function of the system as $\mathbb{L}(\text{data}|\boldsymbol{\theta})$. Then the posterior distribution is given as follows:

$$p(\boldsymbol{\theta}|\text{data}) \propto \mathbb{L}(\text{data}|\boldsymbol{\theta})\mathbb{I}(\mathbb{M}(\boldsymbol{\theta},\text{data})\leq 0) \tag{2.9}$$

Note that I do not fully estimate the second equation $v_h^* = \beta_c + \Gamma_{pc}\boldsymbol{p}_h + \gamma_{cc}c_h + \varepsilon_{v,h}$. Rather, once the parameters are estimated (partially), I fit the value of $\hat{v}_h^*$ for each observation and each draw from the posterior distribution.

An interesting extension of the model is that we can further explore how purchase quantities from different food categories affect the level of waste. This can be done by modeling the optimal waste as a second-stage decision that is a function of purchase quantities: $\boldsymbol{\delta}_h^* = \boldsymbol{\delta}_h^*(\boldsymbol{q}_h)$, which implies the utility of wasting food is also a function of purchases: $v_h^* = v_h^*(\boldsymbol{q}_h)$. Assuming Γ_{pp} is invertible then equation (2.6) can be written as:

$$v_h^*(\boldsymbol{q}_h^*, c_h) = \widetilde{\beta}_v + \widetilde{\Gamma}_{vq}\boldsymbol{q}_h + \widetilde{\gamma}_v c_h + \widetilde{\varepsilon}_{v,h} \tag{2.10}$$

where $\widetilde{\beta}_v = \beta_c - \Gamma'_{pc}\Gamma_{pp}^{-1}\beta'_p$, $\widetilde{\Gamma}_{vq} = -\Gamma'_{pc}\Gamma_{pp}^{-1}$, $\widetilde{\gamma}_v = \gamma_{cc} - \Gamma'_{pc}\Gamma_{pp}^{-1}\Gamma_{pc}$, and $\widetilde{\varepsilon}_{v,h} = \varepsilon_{v,h} - \Gamma'_{pc}\Gamma_{pp}^{-1}\varepsilon_{q,h}$. Estimating this equation will inform us, how food purchases $\boldsymbol{q}_h$ and household management cost c_h affect wasting behavior. Through normalization, we could also reach an estimate for group-specific waste estimates, e.g., a percentage waste estimate for each food category. Note that equation (2.10) only provides an analysis of relationship between purchases quantities and the food-waste utility, which does not always directly link to waste of individual food groups (as seen in Tsionas and Izzeldin 2018). For instance, if the coefficient on group i food is positive, it may simply reflect the large volume of purchase (and waste) of this group, not the perishability of it. For this reason, to directly estimate the effects of food composition, one needs to rely on expenditure shares instead of purchase volumes.

Table 2.1: **Summary Statistics**

	Mean	Std. dev.	5% percentile	95% percentile
q_1, Milk & dairy	5.811	4.801	0.454	14.356
q_2, Protein foods	2.829	2.685	0.405	7.450
q_3, Mixed dishes	3.888	2.990	0.568	10.011
q_4, Grains	2.683	2.202	0.454	6.974
q_5, Snacks	2.476	2.291	0.286	6.721
q_6, Fruit & Vegetables	4.099	3.373	0.582	10.135
q_7, Beverages	16.734	12.641	2.289	42.478
q_8, Condiments	2.877	2.803	0.227	8.024
w_h, Income	1.756	1.330	0.351	4.508

Note: Purchase quantities are measured in kilograms, and household monthly per adult equivalent income is measured in thousand dollars.

2.4.1 Empirical Example

I apply the above method to a nationally representative sample of households and their food acquisition data. The dataset is the USDA's National Household Food Acquisition and Purchase Survey (FoodAPS). The FoodAPS data contains a detailed record of household food purchases, for both at-home and away-from-home events, for a period of seven days within the timeframe of April 19, 2012 to January 22, 2013. It also contains useful information, e.g., age, body measure, employment status, and education, of each individual household member, in addition to a rich set of household-level variables. I look at purchases of eight groups of food: milk and dairy, protein foods, mixed dishes, grains, snacks, fruit and vegetables, beverages, and condiments. The acquisition amounts are measured in kilograms, and prices in dollars per kilogram. This dataset is extensively explored in Yu and Jaenicke (2020a) which contains a detailed description of the variables used in this chapter.

I control for a set of demographic variables, including income and household size, in es-

Table 2.2: Bias in Estimated Own-price Effects

	Approach 1		Simple analysis		Bias
	Posterior mean	s.d.	Posterior mean	s.d.	
1. Milk & dairy	-0.4142	(0.0273)	-0.3873	(0.0327)	-6.49%
2. Protein foods	-0.0441	(0.0038)	-0.0384	(0.0035)	-12.93%
3. Mixed dishes	-0.1166	(0.0109)	-0.0900	(0.0042)	-22.81%
4. Grains	-0.1248	(0.0149)	-0.0893	(0.0159)	-28.45%
5. Snacks	-0.0725	(0.0068)	-0.0572	(0.0072)	-21.10%
6. Fruit & Veg.	-0.2393	(0.0265)	-0.2145	(0.0226)	-10.36%
7. Beverages	-2.4217	(0.1354)	-2.3649	(0.1651)	-2.35%
8. Condiments	-0.106	(0.0121)	-0.1013	(0.0117)	-4.43%

Note: Own-price effects are the diagonal elements of matrix parameter Γ_{pp}, which measures $\partial q_i/\partial p_i$ for food group i. The last column list the estimated bias in terms of the absolute magnitude of own-price effects.

timating the observed demand equation (2.5). To capture the cost shifters associated with reducing food $\boldsymbol{w}_h$, I use monthly income per adult equivalent household size, self-reported diet healthfulness and food security proxies for the cost of food management. Households that reported zero purchase amounts or missing price information are excluded from the analysis. Table 2.1 provides the summary statistics of the key variables. The posterior distribution in equation (2.9) can be analyzed by Markov Chain Monte Carlo (MCMC) methods. Specifically, I use an adaptive Metropolis-Hastings algorithm to draw 25,000 simulations from the distribution, for which the first 5,000 are discarded to avoid the effect of initial values.

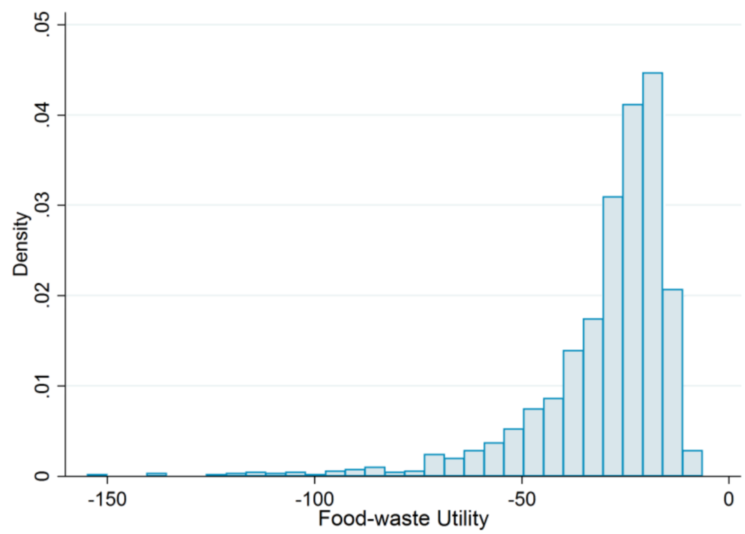
The first two columns in Table 2.2 show the posterior mean and standard deviations of the estimated parameters that measure the own-price effect, $\partial q_i/\partial p_i$. To demonstrate the potential bias in a demand model without incorporating food waste, I run the same Bayesian estimation procedure without introducing the cost function $c(\boldsymbol{w}_h)v(\boldsymbol{\delta}_h)$ and its corresponding constraints. The results of this unconstrained model are reported in the third and fourth columns in Table 2.2, whereas the last column presents the estimated percentage bias.

Except for the case of condiments, own-price effects of all other categories are underestimated in the unconstrained model. The largest bias is on the fourth food group, Grains, followed by Mixed Dishes and Snacks. This is an important observation especially for counter-factual analysis. If we are interested in knowing how demand would respond to an increase in tax, the simple demand analysis would likely under-estimate the drop in purchase volume and, hence, propose an excessive taxation recommendation. Similarly, when the policymakers consider price subsidies to healthy food, the simple model omitting food waste will undermine the effectiveness of such policies. Note that these estimates measure the absolute change in purchase volumes in response to own-price changes, which do not directly translate to price-elasticities, $\partial \log q_i / \partial \log p_i$, that are often used in counter-factual analysis. For instance, the prices of beverage products are relatively low in the sample while its purchase volumes are high, implying that its price elasticity is not the highest, even it has the largest own-price effect.

Figure 2.1 plots the distribution of the estimated $v(\boldsymbol{\delta}_h)$ evaluated at the mean value of the parameter estimates. Table 2.3 presents the estimation result of equation (2.6). Food waste responds negatively to most food prices, indicating it is a normal good (Landry and Smith, 2019). Mixed dishes have the highest price impact on food waste, suggesting households are able to effectively reduce waste in this category when facing a price increase. On the other hand, food waste resulting from the consumption of fruit and vegetables has limited potential to be reduced because of their high degrees of perishability. Management cost is shown to have a positive impact on food (also see Figure 2.2). In particular, households with higher incomes may face higher costs when they devote time to food management and meal preparation.

Finally, it is important to note that the utility of food waste presented here, $v(\boldsymbol{\delta}_h)$, is an

Figure 2.1: Distribution of predicted food-waste utility $v(\boldsymbol{\delta}_h)$



Note: $v(\boldsymbol{\delta}_h)$ is predicted at the posterior-mean values of the parameters.

Figure 2.2: Relationship between food-waste utility and management cost

Note: Scatter plot of the 20,000 MCMC samples, evaluated at the median of the sample data.

Table 2.3: Food Waste, Food Prices, and Management Cost

Variables	Posterior mean	Posterior s.d.
p_1, Milk & dairy	-1.2116	(0.2024)
p_2, Protein foods	-0.0639	(0.0724)
p_3, Mixed dishes	-0.2119	(0.1337)
p_4, Grains	-0.1633	(0.1640)
p_5, Snacks	-0.9770	(0.1170)
p_6, Fruit & Vegetables	-0.8503	(0.3085)
p_7, Beverages	-0.7103	(0.2107)
p_8, Condiments	-1.3766	(0.2002)
c_h, Management cost	1.9638	(0.1458)
$w_{1,h}$, Income	0.0400	(0.0252)
$w_{2,h}$, Healthy diet	1.0967	(0.1904)
$w_{3,h}$, Food security	0.3630	(0.1309)

Note: This table presents the estimation result of equation (2.6). The coefficients on the last three variables, are α's in equation (2.7).

aggregate measure that indicates the overall changes in food waste. This general analysis can be modified to accommodate group-specific waste rates after some normalization techniques. For instance, we can specify $v(\boldsymbol{\delta}_h)$ as Cobb-Douglas or quadratic and estimate the corresponding parameters, which will give us the optimal levels of $\boldsymbol{\delta}_h$ under utility maximization.

2.5 Approach 2: Minimum-Output Household Production

The second approach incorporates household behavioral assumptions to help identify food waste. In order to distinguish the sub-utility generated by actual consumption $u(\boldsymbol{q}_h - \boldsymbol{\delta}_h)$

from the one generated by food waste $v(\boldsymbol{\delta}_h)$, I consider that households produce meals using food inputs to meet a "minimum" level requirement. This requirement, in general, could be a minimum utility level: $u(\boldsymbol{q}_h - \boldsymbol{\delta}_h) \geq \overline{u}_h$; or in the case of Yu and Jaenicke (2020a), it is the minimum amount of calories required to maintain an energy-balanced state. In this sense, the sub-utility function $u(\cdot)$ could also be interpreted as a production function. For exploratory purposes, let us consider the following specification of household utility:

$$U_h = \log \left\{ \mathbb{I} \left[u(\boldsymbol{q}_h - \boldsymbol{\delta}_h) \geq \overline{u}_h \right] \right\} + v(\boldsymbol{\delta}_h; \boldsymbol{w}_h) - \boldsymbol{p}_h' \boldsymbol{q}_h \tag{2.11}$$

Here $\mathbb{I}(\cdot)$ is the indicator function, and $\overline{u}_h$ is the household-specific target utility level. This utility function states that households will meet an exact target output $\overline{u}_h$–anything less than that results in $-\infty$ utility and anything more than that adds zero marginal contribution. A more general formulation that allows for variations above the minimum threshold is presented at the end of this section. Since $\overline{u}_h$ is household-specific, it could capture a wide range of household demographic variables, e.g., household size, geographic location, member composition etc., by writing it as a function of such variables: $\overline{u}_h = \overline{u}_h(b_h)$. This formulation is related to the analysis in Bogetoft and Hougaard (2003) and Asmild et al. (2013), who describe productivity inefficiency (slacks) as "an indirect, on-the-job compensation to agents in an organization". In the context of our model, such on-the-job utility is presented by $v(\boldsymbol{\delta}_h; \boldsymbol{w}_h)$.

To see how the model can be estimated, it is useful to first look at the optimal decisions made by households. It is straightforward that regardless of the choice of $\boldsymbol{\delta}_h$, the optimal actual consumption bundle $\boldsymbol{x}_h = \boldsymbol{q}_h - \boldsymbol{\delta}_h$ must lie on the indifference curve $u(\boldsymbol{x}_h) = \overline{u}_h$. Furthermore, cost minimization implies that optimal $\boldsymbol{x}_h$ is the point where the indifference curve is tangent to the price hyperplane.

Therefore if we know or have already estimated the production technology $u(\boldsymbol{x}_h)$, then we can analytically solve for $\boldsymbol{x}_h$ from the first-order conditions $\partial u(\boldsymbol{x}_h)/\partial \boldsymbol{x}_h = \boldsymbol{p}_h$ and the production target level $u(\boldsymbol{x}_h) = \overline{u}_h$. The estimates of food waste follow directly from $\boldsymbol{\delta}_h = \boldsymbol{q}_h - \boldsymbol{x}_h$. The preceding method works best if $\overline{u}_h$ is well-defined and directly observed in the data, for instance, the total energy expenditure in Yu and Jaenicke (2020a); and the estimation of the function $u(\boldsymbol{x}_h)$ can be carried out using a stochastic frontier model that takes into account endogeneity of inputs $\boldsymbol{x}_h$.

In cases where $\overline{u}_h(\boldsymbol{b}_h)$ is not directly measurable in the data but has to be estimated, it makes more sense to turn to estimating the demand function $\boldsymbol{q}(\boldsymbol{p}_h, \overline{u}_h, \boldsymbol{w}_h)$. After deriving the explicit formulation of the demand functions $\boldsymbol{x}(\boldsymbol{p}_h, \overline{u}_h)$ and $\boldsymbol{\delta}(\boldsymbol{p}_h, \boldsymbol{w}_h)$, we can estimate the following:

$$\boldsymbol{q}_h^* = \boldsymbol{x}^*(\boldsymbol{p}_h, \overline{u}_h) + \boldsymbol{\delta}^*(\boldsymbol{p}_h, \boldsymbol{w}_h)$$

This is a structural estimation of the observed purchase quantities on prices $\boldsymbol{p}_h$, target utility level $\overline{u}_h$, and cost variables $\boldsymbol{w}_h$, where target utility can be substituted by a parametrized form $\overline{u}_h(\boldsymbol{b}_h)$. This system of equations differ from those in typical demand analysis in that it includes two additional sets of variables whose parameter structures are derived within the rational-food-waste framework. The identification power is determined by two factors. First, there must be a sufficient number of distinct variables between $\boldsymbol{b}_h$ and $\boldsymbol{w}_h$. Particularly, when $\boldsymbol{w}_h$ contains product-specific variables such as sell-by dates, identification can be greatly improved. Second, the interactions between these two variables and the price vector improve identification. For instance, if $\boldsymbol{\delta}_h^* = f(\boldsymbol{w}_h)\boldsymbol{g}(\boldsymbol{p}_h)$, then it is possible to identify most of the parameters for $\boldsymbol{\delta}_h^*$. Of course, this type of estimation procedure where parameters of $\boldsymbol{x}^*(\boldsymbol{p}_h, \overline{u}_h)$ and $\boldsymbol{\delta}^*(\boldsymbol{p}_h, \boldsymbol{w}_h)$ are simultaneously estimated also works for cases where $\overline{u}_h(\boldsymbol{b}_h)$ is observed, as we will see in an empirical example later.

A more general specification that allows for positive marginal utility when consuming more than the minimum level $\overline{u}_h$ is given by the following:

$$U_h = \mathbb{I}\left[u_1(\boldsymbol{q}_h - \boldsymbol{\delta}_h) \geq \overline{u}_h\right] u_2(\boldsymbol{q}_h - \boldsymbol{\delta}_h) + v(\boldsymbol{\delta}_h) - \boldsymbol{p}_h'\boldsymbol{q}_h$$

Here the production function $u_1(\cdot)$ and marginal utility function $u_2(\cdot)$ could be different. We could still impose $\log(\cdot)$ function on the first term to ensure the minimum level is met, but this is not necessary when $\partial u_2/\partial \boldsymbol{x}_h|_{\boldsymbol{x}_h=0} = +\infty$ and the cost is minimized. Moreover, there may be some households that meet exactly the minimum requirement: $U_h = u_2(u_1^{-1}(\overline{u}_h)) + v(\boldsymbol{\delta}_h) - \boldsymbol{p}_h'\boldsymbol{q}_h$, and some surpass the minimum level: $U_h = u_2(\boldsymbol{q}_h - \boldsymbol{\delta}_h) + v(\boldsymbol{\delta}_h) - \boldsymbol{p}_h'\boldsymbol{q}_h$. This general approach then requires more assumptions on how such thresholds are determined and faces more identification challenges when estimating those households exceeding minimum output.

2.5.1 Empirical Example

I provide a simple example where $\overline{u}_h$ is observed in data. Following Yu and Jaenicke (2020a), I focus on the household members' expected total energy requirement as the minimum output. Specifically, for member m at household h, it is measured as:

$$\overline{u}_{m,h} = BMR_{m,h} \cdot PA_{m,h} \tag{2.12}$$

where $BMR_{m,h}$ is the predicted basal metabolic rate (BMR), calculated by the revised Harris-Benedict equation as a function of age, weight, height, and gender (Roza and Shizgal, 1984). $PA_{m,h}$ is a positive multiplier on BMR that represents the level of physical activities. For household h, the target level of output $\overline{u}_h$ is the sum of the expected total energy requirement across its members: $\overline{u}_h = \sum \overline{u}_{m,h}$.

In Yu and Jaenicke (2020a), the production function $u(\cdot)$ is estimated by a stochastic frontier model where food waste is modeled as an one-sided inefficiency term. In the context of this chapter, it is possible to conduct a separate stochastic frontier analysis to recover $u(\cdot)$, as long as the endogeneity issue resulting from the rationally-chosen inefficiency is taken care of. Once $u(\cdot)$ is estimated, amounts of actual consumption can be calculated by closed-form cost minimization solutions. However, a more efficient way is to simultaneously estimate $u(\cdot)$

and $v(\cdot)$, which can be done by explicitly deriving food waste as a function of food-waste determinants and food prices.

To be consistent with Yu and Jaenicke (2020a) and for analytical traceability, I formulate food waste in its percentage terms:

$$\delta_{i,h}^P = \frac{q_{i,h} - x_{i,h}}{q_{i,h}} \tag{2.13}$$

The production function is assumed to be translog, as in Yu and Jaenicke (2020a), while the waste function takes the form of $v(\boldsymbol{\delta}_h^P) = \sum_i c(\boldsymbol{w}_h, \gamma_i) \log \delta_{i,h}^P$. Here $\boldsymbol{w}_h$ indicates factors that affect overall household management ability whereas γ_i is a food category-specific variable. For simplicity, I only consider $\boldsymbol{w}_h$ in this chapter so that $v(\boldsymbol{\delta}_h^P) = c(\boldsymbol{w}_h) \sum_i \log \delta_{i,h}^P$. The household utility maximization takes the following form:

$$\begin{aligned}
\max_{\boldsymbol{q},\boldsymbol{\delta}} \quad & c(\boldsymbol{w}_h) \textstyle\sum_i \log \delta_{i,h}^P - \sum_i p_i q_{i,h} \\
\text{subject to} \quad & \overline{u}_h = \alpha_0 + \textstyle\sum_{i=1}^I \alpha_i \log \left[(1 - \delta_{i,h}^P) q_{i,h}\right] \\
& + \textstyle\sum_{i=1}^I \sum_{j \le i} \beta_{i,j} \log \left[(1 - \delta_{i,h}^P) q_{i,h}\right] \log \left[(1 - \delta_{i,h}^P) q_{j,h}\right]
\end{aligned}$$

A very useful result from the first-order conditions is:

$$\frac{\partial v(\boldsymbol{\delta}_h^P)}{\partial \delta_{i,h}^P}(1 - \delta_{i,h}^P) = p_{i,h} q_{i,h} \tag{2.14}$$

Note that this result is independent of the functional forms of the waste function $v(\cdot)$ and the production function $u(\cdot)$. Moreover, it also holds true when we introduce a taste-preference component, i.e., the utility is $f((1 - \boldsymbol{\delta}_h^P) \odot \boldsymbol{q}_h) + v(\boldsymbol{\delta}_h^P) - \boldsymbol{p}_h' \boldsymbol{q}_h$, where $f(\cdot)$ represents subjective utility draw from consumption. Equation (2.14) provides a solution of the optimal percentage waste in terms of the management cost $c(\boldsymbol{w}_h)$ and expenditure of food category i:

$$\delta_{i,h}^P = \frac{c(\boldsymbol{w}_h)}{c(\boldsymbol{w}_h) + p_{i,h} q_{i,h}} \tag{2.15}$$

With this expression on hand, we can substitute $\log(1-\delta^P_{i,h}) = \log p_{i,h}q_{i,h} - \log(c(\boldsymbol{w}_h) + p_{i,h}q_{i,h})$ into the production function and conduct a single-equation estimation based on the energy production only. It is important to note that when we specify a production function that yields closed-form solutions of optimal purchase quantities $q_{i,h}$, such as Cobb-Douglas, it is also feasible to estimate a complete system of equations, instead of a single equation. To enable identification and to restrict management cost to be positive, I specify $c(\boldsymbol{w}_h) = \exp(\rho_0 + \boldsymbol{\rho}'\boldsymbol{w}_h)$.

The basic idea of identifying food waste and actual consumption in this model is that we have two types of distinct variations that affect only one of them. The energy requirement $\overline{u}_h$ only has its impact on actual consumption and food management factors $\boldsymbol{w}_h$ affect how much a household wastes. In a more general setup, we could allow for taste preference so that actual consumption not only serves to meet energy requirements but also provides "utility," as long as there is different sources of variations that help identification. As previously mentioned, the result from equation (2.14) still holds when adding taste preference.

Again, I apply the method to the FoodAPS data. The estimated food waste function $c(\boldsymbol{w}_h) = \exp(\rho_0 + \boldsymbol{\rho}'\boldsymbol{w}_h)$ is presented in Table 2.4. All three variables have a positive effect on food waste. The most significant factor in determining household's effort in managing food is income while diet healthfulness and food security are marginally insignificant with p-values at 0.143 and 0.114, respectively. A summary of the estimated percentage food waste $\boldsymbol{\delta}^P_h$ are reported in Table 2.5. These numbers are largely consistent with the estimates in other studies that suggest at the aggregate level, about 30% food is wasted. The contribution of this chapter is its set of waste estimates for different food categories. Note that in this analysis, the variation in food waste across households are mainly determined by management cost $c(\boldsymbol{w}_h)$. The variation across food categories within a household, on the other hand, is determined by the relative ratio of $c(\boldsymbol{w}_h)$ and expenditure for a category (see equation (2.15)). As a intuitive explanation, households tend to spend more money on food categories with lower waste rates. In Table 2.5, protein food not only has the lowest waste rate but also the

Table 2.4: **Food Waste Management**

Variables		
Income	0.3488***	(0.0736)
Healthy diet	0.9321	(0.6357)
Food security	1.4927	(0.9436)
Constant	-0.7630	(1.6227)

Table 2.5: Estimated Food Waste

	Average waste	Standard deviation
1. Milk & dairy	36.0%	(20.7%)
2. Protein foods	18.6%	(16.4%)
3. Mixed dishes	21.1%	(18.3%)
4. Grains	36.2%	(21.5%)
5. Snacks	29.5%	(20.6%)
6. Fruit & Veg.	28.9%	(19.5%)
7. Beverages	25.4%	(19.5%)
8. Condiments	38.2%	(22.5%)

highest expenditure share among all food categories.

2.6 Directional Distance Function

In this section, I briefly discuss another possible approach, the directional distance model, which is able to provide group-specific waste. Like the empirical example in Approach 2, it is an analysis of inefficiency in household production. It incorporates first-order conditions derived from cost minimization and an optimal "direction" of input slacks (relative proportions of food waste across food groups). The identification is achieved by estimating the overall productivity inefficiency and decomposing it according to the direction of input slacks.

A directional distance function of a production technology is defined as the following (Chambers et al., 1998; Atkinson and Tsionas, 2016; Malikov et al., 2016):

$$D(\boldsymbol{q}_h, y_h, \boldsymbol{g}) = \max\{\beta : (\boldsymbol{q}_h + \beta \boldsymbol{g}_q, y_h + \beta g_y) \in \mathbb{T}\} \tag{2.16}$$

where $\mathbb{T}$ is the production possibility set, $\boldsymbol{q}_h$ is input vector, y_h is the output level (total energy expenditure), and $\boldsymbol{g} = (\boldsymbol{g}_q, g_y)$ is a vector of directions with $\boldsymbol{g}_q \leq 0$ and $g_y \geq 0$. The value of $D(\boldsymbol{q}_h, y_h, \boldsymbol{g})$ is interpreted as the distance the current production profile $(\boldsymbol{q}_h, y_h)$ must travel to reach the production frontier, at the direction given by $\boldsymbol{g}$. It is an alternative formulation of the typical production functions that is used to describe $\mathbb{T}$. The advantage of using directional distance function in this setting is that we can estimate $\boldsymbol{g}_q$, which indicates the "direction" of wasting food (Färe et al., 2017; Atkinson and Tsionas, 2016). After normalization, we can then reach a separate waste measure for each food group.

The econometric procedure of this approach is built upon the cost minimization problem:

$$\min_{\boldsymbol{q}_h} \boldsymbol{p}_h' \boldsymbol{q}_h : D(\boldsymbol{q}_h, y_h, \boldsymbol{g}) \geq 0 \tag{2.17}$$

Assuming differentiability, the first-order conditions can be expressed as:

$$\frac{p_{h,j}}{p_{h,k}} = \left(\frac{\partial D(\boldsymbol{q}_h, y_h, \boldsymbol{g})}{\partial \boldsymbol{q}_{h,k}} \right)^{-1} \frac{\partial D(\boldsymbol{q}_h, y_h, \boldsymbol{g})}{\partial \boldsymbol{q}_{h,j}} \tag{2.18}$$

In addition, parametrization is imposed for the directional distance function, where $f(\cdot)$ is a flexible functional form such as the quadratic form and $\xi_h(\boldsymbol{w}_h)$ is a strictly positive inefficiency term:

$$D(\boldsymbol{q}_h, y_h, \boldsymbol{g}) = f(\boldsymbol{q}_h, y_h, \boldsymbol{g}) - \xi_h(\boldsymbol{w}_h) \tag{2.19}$$

The above system of equations can be estimated using either Bayesian or maximum likelihood (Atkinson and Tsionas, 2016). As the inefficiency $\xi_h(\boldsymbol{w}_h)$ is rationally chosen, it is correlated with $\boldsymbol{q}_h$ and y_h, making the latter two endogenous. Here, endogeneity is dealt with by adding the first-order conditions to construct a complete structural system. However, certain Jacobian transformations are needed to re-write the equations so that the endogenous variables are on the left-hand side of the equations (Malikov et al., 2016).

Note that although the direction $\boldsymbol{g}$ is not a choice variable for households, it can be estimated as a set of parameters. That is, we search for the direction of food wasting behavior based on the information provided by the data. The idea of a constant direction for all households is to capture the common features of food groups, e.g., various degrees of perishability of different products. In a more complex Bayesian setup, Atkinson and Tsionas (2016) achieved an estimate of $\boldsymbol{g}$ for each individual household. Finally, normalization on $\boldsymbol{g}_q$ can be used to reach an estimate for group-specific food waste, that is, $\boldsymbol{\delta}_h = \boldsymbol{\delta}(\xi_h, \boldsymbol{g}_q)$. However, this approach relies on the availability of the observable outcome variable and typically involves much more burdensome computational tasks, and, therefore, is not explored in this chapter.

2.7 Remarks and Conclusion

In this chapter, I present three structural approaches to estimating demand models involving food waste, and empirically implement two of them. These approaches utilize different identification strategies and require different types of data. The usefulness and effectiveness of each approach depend on the particular question the researcher attempts to answer and the type of dataset on hand. Table 2.6 provides a summary of the three approaches.

Approach 1 only requires the observed purchase quantities and some variables related to food management cost, and therefore, can be applied to commonly available scanner datasets such as the Nielsen Consumer Panel. Though this approach is the most straightforward in application, it only produces partially identified results. Whether the bounds provided by the estimation is meaningfully informative depends on the richness of the data, i.e.,

Table 2.6: Structural Approaches to Estimating Food Waste

	Approach 1	Approach 2	Approach 3
Identification	partial	identified	identified
Optimization objective	utility max./ expenditure min.	utility max.	expenditure min.
# of food groups	any	any	all
Group-specific waste	normalization	yes	normalization
Dataset(s)	Nielsen/FoodAPS	Nielsen/FoodAPS	FoodAPS
Estimation Method	Bayesian	Copula or ML	Bayesian or ML

Note: Approach 1: partial identification using duality theorems; Approach 2: minimum-output household production; Approach 3: directional distance function model.

with sufficiently more variations in prices, we can obtain tighter bounds for the parameter estimates.

The directional distance function approach (Approach 3) is a productivity-based analysis of inefficiency in household production and requires a well-defined measurement of output, i.e., total energy expenditure. Therefore, we must employ a dataset like the FoodAPS data which contains useful information to calculate the output, as well as both food-at-home and away-from-home purchases. The advantage of this approach is that the estimated direction of input slacks gives a very useful indicator of group-specific perishabilities. However, by demanding an operational total output measure and food consumption from all categories, Approach 3 is not suited for many scanner datasets. In addition, it cannot be applied to single-category demand analysis.

Approach 2 exploits the behavioral aspect of household food consumption. It is similar to the third approach in that it attempts to identify actual consumption by tracking the production output. I use household total energy expenditure as an empirical example to test this approach. However, the approach is applicable to cases where the requirement of the reported output measure is relaxed. In fact, it allows the minimum level of output to be a general utility target that can be expressed in terms of household-specific variables. In this sense, the directional distance function approach is a special case of Approach 2 in which this utility level is directly observed. The more general setup that takes arbitrary utility measure, on the other hand, may reduce identification power if the two sets of variables affecting actual consumption and food waste overlap, as discussed earlier.

An important issue that I do not explicitly address in the chapter is the role of uncertainty. In reality, the ability of a household to plan its shopping and consumption plays a central role in determining how much food it wastes. In addition to planning, household food management cost is another source of wasting behavior, in a sense that when a realized consumption occasion is different from the ex-ante expectation, the household needs to make efforts in adjusting its plan to minimize loss. In all three approaches, I have "nested" the two sets of factors–shopping planning and food management–together. In other words, the utility functions under optimization can be regarded as the parameterized expectation of actual utility so that some of the parameters are related to the parameters of the distribution of the random shocks. Since household may adjust consumption and waste according to realized random shocks, what are estimated in the approaches are the "expected" food waste, not the actual amounts. Note that this may sound counter-intuitive for Approach 2, as the household is committed to achieving a minimum level of utility, regardless of the random shock. The explanation is that the second stage decision is independent of prices as they are already paid for. Therefore households may deviate from the cost-minimizing consumption allocation to another point on the utility indifference curve.

Finally, if we prefer to explore the specific mechanism of how uncertainty influences the amount of food waste, we must impose distributional assumptions on the random shocks and specify household risk-aversion parameters. This can be done by extending Approach 2 where the utility maximization problem is directly formulated. In particular, we can model the households' purchase decisions as shopping for future consumption occasions and that they have a rational expectation regarding the number of such occasions. And food waste occurs when the ex-post realized occasions turn out to be fewer than expected. This type of model, though not being applied to the food waste studies, has been estimated by Hendel (1999), as well as by Dubé (2004) who assumes the number of occasions follows a Poisson distribution.

In conclusion, food waste is a result of rational choices made by households, conditional on their management and inventory costs. Households with particular characteristics, for example, those with lower storage costs, lower transaction costs, or higher income, are more flexible in making purchase decisions or more efficient in transforming raw food into meals. Consequently, the share of wasted food is an outcome that simultaneously affects demand. To obtain consistent parameter estimates in consumer demand analysis, food waste and its corresponding waste-determinant factors need to be structurally modeled.

This chapter represents one of the first attempts to incorporate food waste into utility-maximizing models of consumer behavior and provide useful estimates to study the rationales of wasting food. Many interesting extensions and important questions can be further studied. For instance, researchers can utilize the results to calibrate the amount of food waste, based on household demographic variables and their purchase quantities. For policies that concern actual nutritional intakes, this type of calibration provides useful guidelines and reference points. In addition, structural models presented in this chapter could serve as a means to check the robustness of other consumer demand models, especially for the purpose of evaluating counter-factual results.

Chapter 3

The Effect of Sell-by Dates on Purchase Volume and Food Waste

3.1 Introduction

Sell-by dates play a key part in determining the consumption of perishable food (Leib et al., 2013; Ellison and Lusk, 2018; Qi and Roe, 2016; Roe et al., 2018b). Consumers often find sell-by labels confusing, which leads them to misinterpret the date labels as "safe-until" dates (Neff et al., 2015; Wilson et al., 2017; Roe et al., 2018b). Consequently, a significant portion of perishable food is mismanaged and disposed of earlier than necessary, resulting in more food waste. In fact, sell-by dates generally do not refer to food safety concerns but merely serve as an indicator of "best quality" (Leib et al., 2013). To reflect the up-to-date technology in food processing and transportation, and to reduce unnecessary waste of milk products, many states in the U.S. have changed their regulations of the maximum sell-by dates. Most recently, the New York City Board of Health changed its dating regulation of milk products in September 2010 (New York City Board of Health, 2010). Previously, the New York City Code required all milk products to be sold within 9 days after pasteurization.[1]

[1] The City Board of Health, in its notice, addresses the previously required date labels as "expiration dates". We use "sell-by dates" throughout the chapter as they are more frequently used in literature.

The new policy repealed this 9-day rule and left the sell-by dates to be determined by milk manufacturers and processors. This change effectively increased the shelf life of milk to about 14 to 15 days (Hager, 2010).

In this chapter, we first construct a theoretical model of household utility maximization that accommodates food-wasting behavior. We assume that it is costly to reduce food waste and that households make rational choices on how much to waste, conditional on their food management abilities. We explore two different specifications of the food management cost function, with one based on the percentage of food waste and the other based on the amount of food waste. A major finding of the theoretical analysis is that when the demand for milk is price-inelastic, an extended sell-by date will reduce both the observed purchase volume and the unobserved food waste while increasing the amount of actual consumption. The cost structure based on the percentage waste can be considered as a special case of the theoretical model studied by Hamilton and Richards (2019) who focus on two broader categories – perishable food and non-perishable food–and analyze how the changes in food policy and food prices affect household food utilization rate. Although Hamilton and Richards (2019) emphasize that the demand for fresh food is often price-elastic, we find the demand for milk is price-inelastic and, therefore, draw different conclusions for policy effects. In addition, we provide detailed discussions on the policy's impact on the supply side, including its influence on the retailers' loss-leading strategies and the upstream market competition.

Next, we take New York City's regulation change in 2010 as an empirical case and examine whether the City's new policy reduces food waste and improves consumer welfare. We first conduct a series of store fixed-effect difference-in-difference models on the sales volumes reported in the Nielsen Retail Scanner data, for a period of 12 months before and after the policy change. We consider three models that are based on yearly, monthly, and quarterly sales volumes, respectively. We also assess the policy's impact at the household level by performing difference-in-difference estimations on the household monthly purchase volumes using the Nielsen Consumer Panel Data. To test the robustness of the estimates, we examine the issues of price endogeneity and the selection of time frame. We also apply a synthetic control method to address the parallel trends assumption. Finally, we estimate a simple structural model to identify the exact changes in household utilization rate and food waste.

Our results support the assertion made by the City's Board of Health, which stated that the previous 9-day rule led to unnecessary disposal of milk. Based on the retail scanner data, the estimated effect of the new regulation is a reduction in store sales volumes by about 10% in New York City. This percentage is confirmed by the analysis of the consumer panel data, which shows that the average household monthly purchase decreases by about 23 fluid ounces. In addition, we conduct several robustness checks and a synthetic control method to validate the results. The implications of our theoretical model suggest that this 10% change is a lower bound of the reduction in food waste. Our structural model, which does not rely on the theoretical implications on price elasticities, identifies that the actual reduction in food waste is about 10-14%, which is more than the reduction in the purchase volumes. These results suggest that even the observed purchase quantities are now lower, households actually consume more milk. It is also important to note that the elimination of the City's outdated sell-by regulation of milk products does not impose additional health and food-safety risks to consumers who now store milk for a longer period of time. As mentioned in the City's Notice of Adoption, as long as milk is properly processed, it is a safe product, and that milk processors have a strong "business interest" to maintain the product quality. In addition, the notice also points out that the rest of the New York State regions had not reported any "adverse public health effects" in the absence of a sell-by regulation (New York City Board of Health, 2010).

To the best of our knowledge, this is the first study that quantitatively and systemically estimates the effect of sell-by dates on the consumption of a perishable food product. Each year, about 50 billion pounds of pasteurized milk products are sold within the United States, representing a value of approximately $21 billion (USDA-ERS, 2016). In the case of New

York City, we estimate that the elimination of the 9-day rule reduced food waste by at least 5.2 million pounds annually, which translates to about \$3.4 million using our data. Methodologically, this study contributes to the emerging body of food-waste literature by illustrating how to utilize the variation in sell-by dates as an identification strategy for food waste estimates. Lastly, our study directly aids policymakers and food-industry managers who oversee date-labeling programs. Currently, there are more than 20 states that require dating of some food types. Our results suggest that it is important to keep the dating regulations consistent with technological improvement in the food industry and to provide clearer meanings of the date labels.

The rest of the chapter is organized as follows: Section 2 provides a brief overview of the milk sell-by regulations in the U.S. Section 3 provides a theoretical analysis of the effect of an extended sell-by date. Section 4 presents the empirical approach and estimation results, including a set of robustness checks. Section 5 discusses the policy's impact on the supply side, and the final section concludes the chapter.

3.2 Milk Sell-by Regulations in the U.S.

In the United States, there have been no federal regulations of milk products' date labeling since 1934 when the U.S Public Health Service Ordinance and Code eliminated the relevant rules. Currently, 15 states and the District of Columbia require some date labeling on milk products (Leib et al., 2013). While most of these states let the milk manufacturers and processors decide the length of sell-by dates, Pennsylvania, Maryland, and Montana set explicit requirements on the sell-by dates. Historically, the number of states regulating maximum sell-by dates had dramatically decreased. Based on a search of the historical changes in the state legislation or administrative codes, we have compiled in Table 3.1 a list of recent regulation changes regarding the sell-by dates of milk products.

Table 3.1 notes that four states, in addition to New York City, have substantially changed

Table 3.1: Recent Changes in State/City Regulations of Milk Sell-by Dates

State/City	Change Year	Before	After	Regulation Citation
New York City	1987	4 days	9 days	Rules of NYC, Title 24 Health Code, §111, §117
	2010	9 days	n/a	[Repealed]
Connecticut	1982	10 days	12 days	C.T. General Statutes, §22-197b
	2005	12 days	n/a	[Repealed]
Maryland	1996	7 days	14 days	Code of M.D., Health General §21.426
	2003	14 days	n/a	[Repealed]
	2017	n/a	18 days	COMAR, §10.15.06.10
New Mexico	1991	14 days	n/a	N.M. Admin. Code, §21.34.5.9 [Repealed]
Pennsylvania	1996	12 days	14 days	7 P.A. Code §59.22
	2003	14 days	17 days	7 P.A. Code §59.22

Note: This list is by no means comprehensive. Only the sell-by regulations on the regular pasteurized milk are listed. The ultra-pasteurized milk products typically have a shelf-life more than 30 days. "n/a" means there is no regulation on maximum sell-by dates and the manufacturers/processors make the decision on the sell-by dates of their products. 7 P.A. Code §59.22 was replaced/renamed with §59A.15, effective May 21, 2011.

their sell-by regulations of milk products by either extending the length of the sell-by dates or simply eliminating the regulation.[2] For instance, Connecticut extended the sell-by dates from 10 days to 12 days in 1982 and eventually repealed the rule in 2005. Even during the time between 1982 and 2005, milk processors could apply to the Connecticut Department of Agricultural for approvals of dates longer than 12 days. Pennsylvania increased the maximum length from 14 days to 17 days in 2003, seven years after the previous 2-day extension in 1996. In 1991, New Mexico became one of the early states that forwent regulating sell-by dates.

Interestingly, although Maryland doubled its maximum allowable date from 7 to 14 days in 1996 and subsequently abolished the rule in 2003, it reinstated the regulation with a relaxed 18-day cap in 2017. In fact, before 2003, the milk product dating was regulated by the state's statutory law, *Code of Maryland*. In 2010, the regulatory authority was taken over by the state's administrative law, *Code of Maryland Regulations* (COMAR), which re-introduced an 18-day rule. We believe that a sell-by date of 18 days generally reflects the industry's current practice and, hence, does not generate substantial impacts that are comparable to its previous (temporary) elimination of the regulation in 2003.

This chapter pays close attention to the case of New York City's policy change in 2010. Its now-repealed 9-day rule was considered rather stringent, and the resulting effects on consumer behavior were expected to be significant. We investigate the policy's impact through both theoretical analysis and empirical verification.

3.3 The Theoretical Model

In this section, we develop a theoretical model to illustrate how the extension of sell-by dates affects consumer behavior. We assume that households make decisions on how much milk to purchase and how much to waste. In this sense, the amount of food waste is a rational choice

[2] An exception is Montana, where pasteurized milk products have been required to be sold within 12 days since 1980 (M.T Administrative Rule §32.8.101).

made by the households given their food management ability (Hamilton and Richards, 2019; Lusk and Ellison, 2017).

Similar to Hamilton and Richards (2019), we specify that the utility function consists of two parts (see equation 3.1 below). The first component, $u(x)$, represents the utility drawn from consuming x amount of milk and satisfies two properties: $u'(x) > 0$ and $u''(x) < 0$. Letting q denote the observed purchase quantity of milk, then the amount of wasted milk is simply given by $\text{FW} = q - x$. The second component, $c(q, x; L)$, is the cost function of reducing food waste such that lower levels of food waste are associated with higher values of $c(q, x; L)$.[3] Mathematically, this means that $\partial c/\partial q \leq 0$ and $\partial c/\partial x \geq 0$. Note that we have incorporated the length of sell-by dates, L, as a parameter of the cost function. We assume that an extended sell-by date, hence a longer shelf-life, lowers the marginal cost of reducing food waste, which is the type of food policy analyzed in Hamilton and Richards (2019). Finally, we denote the numeraire good as y, the total budget as w, and the price of milk as p. The utility-maximizing household solves the following problem:

$$\begin{aligned} \max_{q,x} \quad & u(x) - c(q, x; L) + y \qquad (3.1) \\ \text{subject to} \quad & pq + y \leq w \\ & 0 \leq x \leq q \end{aligned}$$

Our choice of the quasi-linear specification is based on the fact that the expenditure on milk products represent a very small share (less than 3%) of households' total food expenditure. Note that our model, as well as the existing studies of rational food waste, do not explicitly account for the role of uncertainty and its underlying mechanism but formulates the households' decisions on q and x to be made simultaneously. In the absence of uncertainty, the model is equivalent to a two-stage decision-making process in which food purchase is the

[3] We do not explicitly model how the sell-by date enters the inventory management process. More comprehensive analysis on the subject can be found in the operational research literature of inventory models of perishable food (for examples, see Chazan and Gal (1977); Van Donselaar and Broekmeulen (2012) and Duan and Liao (2013)).

first-stage decision, and the household food waste management takes place at the second-stage. This equivalence can be proven by comparing the first-order conditions resulting from the two specifications. In reality, the ability of a household in planning its shopping and consumption when facing uncertainty plays a central role in determining how much food it wastes. By considering uncertainty, it is naturally a two-stage decision process. When a household makes its shopping plan, the decision variables are the purchase quantity and the "expected" actual consumption (hence food waste) that maximize its expected utility. After the realization of random shocks, the household adjusts consumption and waste according to the ex-post utility. In this sense, what we model in equation 3.1 is the parametrized expectation of the utility function, and the estimated value of $q - x$ can be interpreted as the "expected" food waste.[4]

To investigate the effect of a longer sell-by date on the optimal purchase quantities and food waste, we need to specify the functional properties of $c(q, x; L)$. Here we consider two cases that have different measures of food waste as the argument of the cost function:

$$\begin{array}{lll} \text{Proportional Waste Measure:} & c(q,x;L) = c(\delta^P; L), & \delta^P = (q-x)/q, 0 \leq \delta^P \leq 1 \\ \text{Quantity Waste Measure:} & c(q,x;L) = c(\delta^Q; L), & \delta^Q = q - x, 0 \leq \delta^Q \leq q \end{array}$$

The case of proportional waste is very common in studies that measure household food waste. It provides a percentage waste measure with an intuitive interpretation (Buzby et al., 2014; Muth et al., 2011; Buzby et al., 2009; Bellemare et al., 2017; Yu and Jaenicke, 2020a; Hamilton and Richards, 2019). By regarding δ^P as the determinant of the household food management cost, we are able to account for the variation in household demand for milk that is due to the household structure instead of preference. For instance, a three-member household might naturally purchase more milk than a single-member household. In terms of volumes, larger households have more food waste than smaller households, ceteris paribus. But the larger households are not necessarily more "wasteful" if we consider per-household-

[4]If we need to infer the actual amount of food waste, we must impose distributional assumptions on the random shocks and specify household risk-aversion parameters.

member waste. Therefore, we generally need a percentage measure such as δ^P that captures the relative size of food waste in relation to total purchases and takes into account the differences in household sizes.

However, milk products are often not offered in continuous volumes. Instead, consumers must choose from a set of packages with fixed sizes. This implies that smaller households who prefer buying less milk may be constrained at buying large package sizes. As a consequence, the proportional waste measure may no longer provide an accurate comparison between large and small households. Therefore, we introduce a second type of waste measure δ^Q that is simply based on the quantity of waste. We believe that this specification gives additional insights into the case of sell-by date changes. In a structural model with micro-level household data, researchers could control for household demographic variables such as household size that lead to heterogeneity in the cost function, making δ^Q an appropriate waste measure across different household types.

Given the specifications of the waste measure, we rewrite the utility maximization problem in a simplified form by substituting the budget constraint and assuming interior solutions:

$$\max_{q,\delta} \qquad u[x(q,\delta)] - c(\delta; L) + w - pq \tag{3.2}$$

Here $x(q,\delta) = (1-\delta^P)q$ for the proportional waste measure, and $x(q,\delta) = q - \delta^Q$ for the quantity waste measure. The first-order conditions for the household's optimal decisions of q^* and δ^* are straightforward:

$$\begin{aligned} u'[x(q^*,\delta^*)] \cdot \partial x(q^*,\delta^*)/\partial q - p &= 0 \\ u'[x(q^*,\delta^*)] \cdot \partial x(q^*,\delta^*)/\partial \delta - \partial c(\delta; L)/\partial \delta &= 0 \end{aligned} \tag{3.3}$$

3.3.1 Comparative Statics

Because we are interested in evaluating the effect of an extended sell-by date on the purchase volume q^* and food waste measure δ^*, we conduct an analysis of comparative statics to uncover their relations to the sell-by date L. By the implicit function theorem applied to the first-order conditions and assuming all the derivatives exist, we derive the following results:

$$\text{Proportional Waste:} \quad \frac{\partial(q^*, \delta^{P*})}{\partial L} = \frac{1}{|J|}\frac{\partial^2 c}{\partial \delta^P \partial L}\begin{bmatrix} x^*u'' + u', & (1-\delta^{P*})^2 u'' \end{bmatrix}^T \tag{3.4}$$

$$\text{Quantity Waste:} \quad \frac{\partial(q^*, \delta^{Q*})}{\partial L} = \frac{1}{|J|}\frac{\partial^2 c}{\partial \delta^Q \partial L}\begin{bmatrix} u'', & u'' \end{bmatrix}^T \tag{3.5}$$

Here $|J| > 0$ is the determinant of the Hessian, and it is positive to ensure the concavity of the utility function. Since we assume that an extended sell-by date leads to a lower marginal cost of reducing food waste, we have $\frac{\partial^2 c}{\partial \delta \partial L} > 0$. Hence, whether the policy increases/decreases the optimal purchase quantity q^* and food waste δ^* will depend on the elements inside the Jacobian matrix.

In the case of the proportional waste measure, it is obvious that $\partial \delta^{P*}/\partial L < 0$. Therefore the household food waste, in percentage terms, decreases when sell-by date is extended. What is not clear is how the total purchase q^*, actual consumption $x^* = (1-\delta^{P*})q^*$, and the absolute amount of food waste $\delta^{P*}q^*$ change as a result of the new policy. Note that $\partial q^*/\partial L \gtreqless 0$ whenever $x^*u'' + u' \gtreqless 0$. Let us denote $\eta_D^* = \frac{\partial q^*}{\partial p}\frac{p}{q^*}$ as the price elasticity of the demand for milk products at the optimal choice bundle. Then it can be shown that $\eta_D^* \lesseqgtr -1$ whenever $x^*u'' + u' \gtreqless 0$.[5] In the following sections, we provide detailed discussions on these possible scenarios of the changes for the proportional waste measure, as well as the case of the quantity waste measure.

[5] This can be done by first applying the implicit function theorem to the first-order conditions to obtain $\frac{\partial q^*}{\partial p}$. Next, in addition to $\frac{\partial q^*}{\partial p}$, we substitute $p = u'(1-\delta)$ and the explicit expression for $|J|$ into $\eta_D^* = \frac{\partial q^*}{\partial p}\frac{p}{q^*}$. After re-arrangements, we can show that $\eta_D^* < -1$ whenever $(x^*u'' + u')(1-\varepsilon_c) > 0$, where $\varepsilon_c = \frac{(1-\delta)c''}{c'}$ is less than 1 when the cost function is convex.

3.3.1.1 Price-Inelastic Demand, Proportional Waste

As the preceding discussion shows, the direction of the policy's impact on the purchase quantity is completely determined by the price elasticity of demand at the current optimal level, while the percentage food waste always decreases with longer sell-by dates:[6]

$$
\begin{array}{llll}
\dfrac{\partial q^*}{\partial L} < 0, \dfrac{\partial \delta^{P*}}{\partial L} < 0 & \text{if} & \eta_D^* > -1 & \text{Price-Inelastic Demand} \\
\dfrac{\partial q^*}{\partial L} = 0, \dfrac{\partial \delta^{P*}}{\partial L} < 0 & \text{if} & \eta_D^* = -1 & \text{Unit Elastic Demand} \\
\dfrac{\partial q^*}{\partial L} > 0, \dfrac{\partial \delta^{P*}}{\partial L} < 0 & \text{if} & \eta_D^* < -1 & \text{Price-Elastic Demand}
\end{array}
$$

When the demand for milk products is price-inelastic, households purchase less milk and waste at a smaller percentage. As for the actual consumption $x^* = (1 - \delta^{P*})q^*$, we can infer its change through the first-order condition $u'(x^*) = p^*/(1 - \delta^{P*})$. Given the same price, when δ^{P*} decreases, $u'(x^*)$ will also decrease. The concavity of $u(\cdot)$ implies that the post-policy actual consumption is higher than its previous level. As a result, the amount of food waste $q^* - x^*$ is lowered. In short, an extended sell-by date allows households to spend less while consuming more milk and simultaneously reducing household food waste.

Figure 3.1 provides an illustration for the case of price-inelastic demand. For simplicity, we assume the price remains unchanged. After the policy implementation, the sell-by date L_0 increases to L_1. The observed demand curve $q(p, L_0)$ shifts left to $q(p, L_1)$, and the new optimal quantity decreases from q_0^* to q_1^*. The actual consumption curve $x(q, \delta_0)$ shifts right to $x(q, \delta_1)$, resulting in a higher level of actual milk consumption. Consequently, the distance between x_1^* and q_1^*, i.e., the amount of food waste, is lower than its pre-policy level.

[6] Unless the utility function $u(\cdot)$ is of the iso-elastic type (CRRA utility) for which x^*u''/u' is a constant, this is a local property at the optimal point.

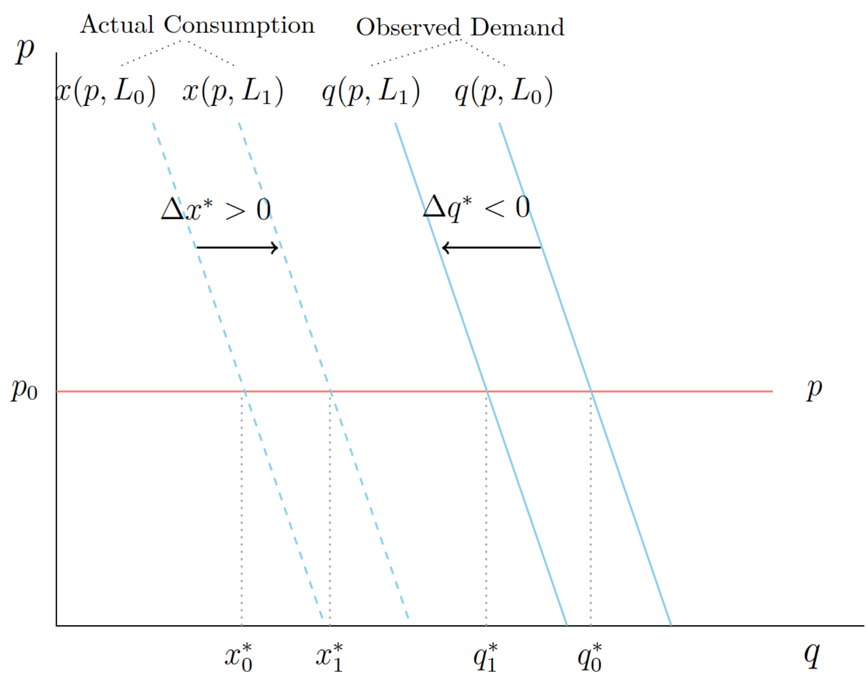


Figure 3.1: Comparative statics of an extended sell-by date, price-inelastic demand.

3.3.1.2 Price-Elastic Demand, Proportional Waste

When demand is price-elastic, an extended sell-by date increases household purchase quantities. Together with a lower percentage waste measure δ^{P*}, this implies that actual consumption increases. However, the implication on the amount of food waste is not straightforward. On one hand, because food waste is strictly reduced when $\eta_D^* = -1$, there must be a range of elasticities close to -1 where there is still reduction. On the other hand, if demand is elastic enough, then it is possible that food waste would actually increase (as suggested by Proposition 1 in Hamilton and Richards (2019)). This can be seen more clearly through the following relationship:

$$\frac{\partial(\delta^{P*}q^*)}{\partial L} = \frac{1}{|J|}\frac{\partial^2 c}{\partial\delta^P\partial L}\left[(1-\delta^{P*})qu'' + \delta^{P*}u'\right] \tag{3.6}$$

For the policy to generate more waste, we need $(1-\delta^{P*})qu'' + \delta^{P*}u' > 0$. It can be shown that this condition implies that:[7]

$$\eta_D^* < -\frac{1}{\delta^{P*}} \tag{3.7}$$

Therefore if the demand is elastic enough, then food waste could increase as a result of a longer sell-by date. For example, if households waste one-third of the purchased milk, $\delta^{P*} = 1/3$, then the elasticity has to be greater than 3. Although this is not the case for milk products, some studies show that organic fruits may exhibit such elastic demand (Lin et al., 2009). This ambiguous implication suggests that it is important to carefully examine the effect of sell-by dates on these premium types of food.

3.3.1.3 Quantity Waste

Lastly, we turn to the quantity-based waste measure. Equation 3.5 indicates that when sell-by date is extended, purchase quantity q^* and amount of food waste δ^{Q*} will decrease by the same volume. Interestingly, the change in percentage food waste, $\frac{\delta^{Q*}}{q^*}$, can be shown to reach a lower level. Hence the directions of changes in purchase quantity and food waste under the quantity-based waste measure is identical to the price-inelastic demand case of proportional waste. Additionally, from the first-order condition $p = u'(x^*)$, we know that the level of actual consumption x^* remains the same. The intuition of this result is that under the quantity waste measure, we can interpret food waste as a consumption good so that $-c(\delta^Q; L)$ is also a utility function. The household separately chooses the amount of actual consumption and food waste up to the point where each good's marginal utility is equal to the price. In sum, when facing a longer sell-by date, the quantity-based waste measure always reduces the amount of food waste and improves consumer welfare, regardless of demand elasticity.

Table 3.2 is a summary of the comparative statics analysis when the sell-by date is extended. In all cases, consumer welfare is strictly improved because the previous choices

[7] Note that this inequality implies $(1 - \delta^{P*})qu'' + \frac{1}{\delta^{P*}} * u' > 0$ because $\delta^{P*} < 1$. By the same method in footnote 5, we can show that $\left[(1 - \delta^{P*})qu'' + \frac{1}{\delta^{P*}} * u' > 0\right](1 - \varepsilon_c) > 0$ implies that $\eta_D^* < -\frac{1}{\delta^{P*}}$.

Table 3.2: Comparative Statics of an Extended Sell-by Date

	Proportional Waste			Quantity Waste
Price Elasticity of Demand	$\eta_D^* > -1$	$\eta_D^* = -1$	$\eta_D^* < -1$	any η_D^*
Purchase Volume	$\downarrow$	$=$	$\uparrow$	$\downarrow$
Actual Consumption	$\uparrow$	$\uparrow$	$\uparrow$	$=$
Food Waste (Percentage)	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Food Waste (Amount)	$\downarrow$	$\downarrow$	?	$\downarrow$
Consumer Welfare	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$

Note: "$\uparrow$": increased level; "$\downarrow$": decreased level; "$=$": unchanged; "?": undetermined.

are still available under the new policy, and a household only changes its decision if that lead to a higher level of utility.

3.4 Empirical Estimations

In this section, we empirically investigate the policy change in New York City, who eliminated its sell-by regulation of pasteurized milk products in September 2010. The new policy was expected to increase the sell-by dates from the previous 9 days to about 14 or 15 days (Hager, 2010). The preceding theoretical analysis shows that when the sell-by date of milk products is extended, the percentage of food waste decreases but the change in the amount of food waste is unclear. When the cost function is based on the quantity waste measure, households purchase less milk and reduce their food waste, while keeping the amount of actual consumption at the pre-policy level. A more complicated scenario involves the proportional waste measure. If the demand for milk is price-inelastic, then the implications on the purchase volume and food waste are similar to the case of quantity waste measure. If instead the demand is price-elastic, then the directions of the effects are quite ambiguous–

the purchase volume is higher, but the change in the amount of food waste is undetermined. In fact, when the price elasticity is high enough (equation 3.7), food waste could actually increase.

Therefore, to identify the unobserved change in food waste in a nonstructural model, it is crucial to determine the price elasticity of the demand for milk. If the estimated price elasticity is less than one and we observe a decrease in the purchase volume, then we know that the new policy lowers the amount of food waste, regardless of the specification of the waste measure in the cost function.

We first offer a quick look at the total store sales volumes at the New York City area. We compare it to the Hartford area of the neighboring state of Connecticut, for a year before and after the policy implementation. Hartford is geographically close to New York and shares similar climate conditions. In addition, many milk distributors supply their products to stores in both regions. Interestingly, before September 2010, a carton of milk sold in New York or Hartford was often stamped with two sell-by dates on its package–one for the stores in New York City and the other for those outside the city (OLR, 1999; Hager, 2010). We do not use the non-New York City metro areas of New York, Connecticut, or New Jersey as candidates for the control region, because milk purchased in these immediate surroundings can substitute for milk purchased in New York City. In other words, people who live in one region may shop in the other region. In this sense, the milk purchase volumes in metro New Jersey, for example, was also affected by the policy in New York City and, hence, not an appropriate control group. [8]

The three counties of New York City under consideration are Kings (Brooklyn), New York (Manhattan), and Queens, whereas the Hartford area includes Hartford, Middlesex,

[8] In our analysis, if we add six counties in New Jersey that are adjacent to or near New York City (Hudson, Essex, Union, Bergen, Middlesex, and Monmouth), as well as two counties on Long Island of New York State (Nassau and Suffolk), the policy's impact on New York City's milk sales is a reduction of 6-8%. This estimate is slightly smaller than those presented in our main findings. Indeed, there were likely spillover effects in the neighboring non-NYC metro areas. Further analysis shows that the indirect treatment effect on these regions is about a reduction of 4-5% in sales volume, i.e., the policy change in New York City reduced milk sales in Jersey City or Newark by 4-5%.

New London, and Tolland counties.[9] The dataset used to accomplish this task is the Nielsen Retail Scanner Data which contains weekly sales volumes and prices of pasteurized milk products for a sample of representative stores across the United States. To avoid store entry/exit issues, we only keep the stores in the two regions that lasted the entire period of analysis.

Table 3.3 lists the value of the yearly total store sales volume, in thousand gallons, and the percentage changes at the county and region levels. In all three counties of New York City, the total store sales volumes decreased, where Manhattan experienced the largest percentage drop of 8.9%. The overall change in the Hartford area is a 2.4% increase due to the rise of sales volumes in Hartford county. Figure 3.2 gives a closer look at the monthly total sales volume at the region level from September 2010 to September 2011, comparing to the corresponding months in the previous year. For New York City, the post-policy store sales volumes are more likely to be lower than the pre-policy volumes of the same months from the previous year. On the other hand, total store sales volumes in the Hartford area more frequently exceed their pre-policy levels.

3.4.1 Difference-in-Difference Estimation

3.4.1.1 The Two-Period Model

To quantitatively measure the policy impact, we first run a classic two-period difference-in-difference regression with store fixed effects (Card and Krueger, 1994). Table 3.4 contains the estimates of this model, labeled as Model A. In the second column, the dependent variable is the quantity of yearly store sales volume. It predicts the extended sell-by dates of milk products in New York City results in a 506-gallon decrease in yearly sales volumes per store. Because there is significant variation in store sizes in the data, the absolute quantity may not

[9] There are other two counties subject to New York City's regulation: Richmond (Staten Island) and Bronx. We do not include these two border counties in the analysis to reduce the possible substitution effect with neighboring counties. Later, we use the Hartford area as our main control group in the difference-in-difference estimation; however, we also construct a synthetic control group that weights milk sales from other major metro areas within a specific distance from New York City as a robustness check.

Table 3.3: Yearly Changes in Total Sales Volumes (thousand gallons)

New York City

Counties	No. of Stores	Before	After	Change	% Change
Kings	131	715.6	695.5	-20.17	-2.8%
New York	193	1012.4	922.0	-90.38	-8.9%
Queens	119	1551.2	1509.2	-41.98	-2.7%
Region Total	443	3279.2	3126.7	-152.54	-4.7%

Hartford Area

Counties	No. of Stores	Before	After	Change	% Change
Hartford	72	521.9	554.4	32.50	6.2%
Middlesex	15	101.6	97.8	-3.82	-3.8%
New London	22	170.3	170.5	0.15	0.01%
Tolland	7	63.7	55.4	-8.27	-13.0%
Region Total	116	857.5	878.0	20.56	2.4%

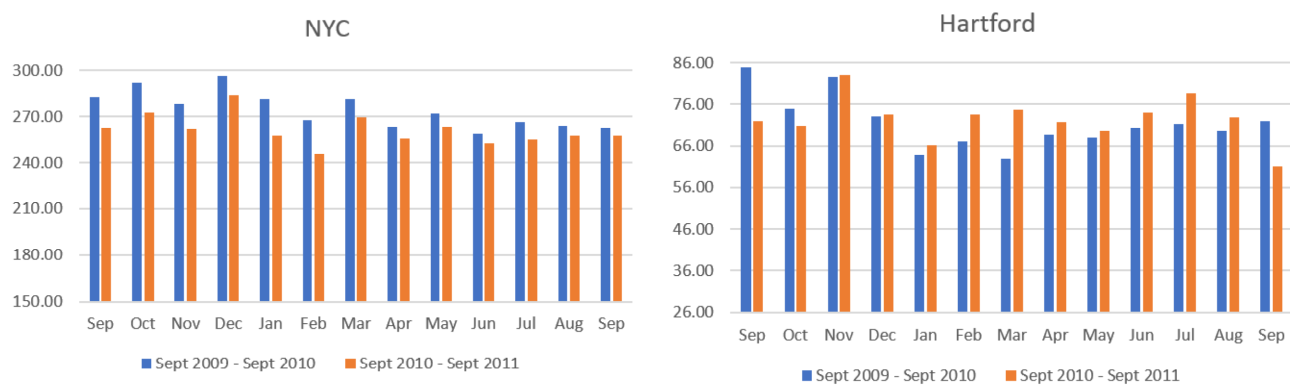


Figure 3.2: Monthly sales volumes (thousand gallons), compared to previous year.

Table 3.4: **Two-period Estimation with Store Fixed Effect**

	Model A	
Variables	Sales Volume	Log(Sales Volume)
NYC*Post-Policy	-0.506**	-0.089***
	(0.265)	(0.019)
Post-Policy	0.483*	0.082***
	(0.258)	(0.019)
Log(price)	-3.991***	-0.604***
	(1.182)	(0.168)
Constant	40.93***	5.994***
	(9.930)	(1.413)
Adjusted R^2	0.031	0.093
Observations	1,118	1,118

Note: The measurement of volumes is in thousand gallons. Both specifications control for the store fixed effect. The dummy variable of New York City is ignored due to perfect collinearity with the store fixed effect. Robust standard errors are clustered at the store level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

have a meaningful interpretation for the smallest and largest stores. Hence, we consider an additional specification using the logarithm of store sales volumes as the dependent variable to obtain a percentage estimate. The results of the second specification are contained in the last column of Table 3.4. It shows that the policy's effect is an 8.9% drop in store sales volume, on average. Moreover, the price elasticity of milk product is -0.604, indicating a price-inelastic demand.

3.4.1.2 Monthly and Quarterly Volumes

In addition, we apply the difference-in-difference method to the store monthly and quarterly sales volume in place of the two-period yearly data. Since milk consumption is highly seasonal, this specification allows for the time-specific variations within a year. Figure 3.3 displays the region-level total monthly and quarterly volumes.[10] As it suggests, the quarterly

[10] To provide a clear view of the overall trend, the figure shows 20 months before and after the policy point–a longer time frame than used in the difference-in-difference regression. For the quarterly volumes,

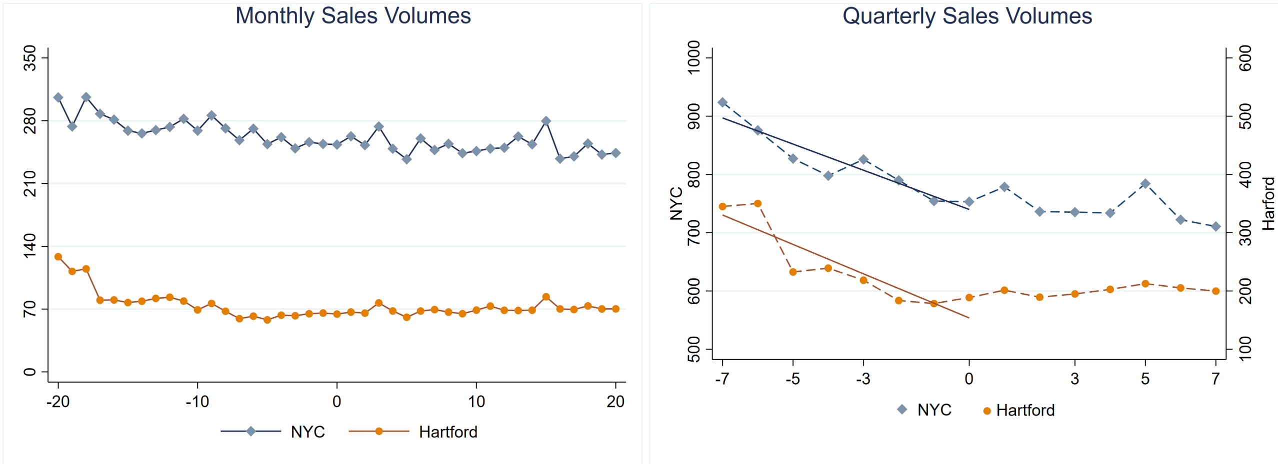


Figure 3.3: Trends of monthly and quarterly sales volumes (thousand gallons).

volume is more volatile; however, the overall trends reveal similar patterns. The linear fitting lines for the pre-policy period in the right panel illustrate the approximately parallel trends between the two regions, while the post-policy time reveals that New York City experienced relatively lower sales volumes. We check the pre-policy parallel assumption by estimating the model focusing only on the time periods before the policy implementation. We introduce an interaction term of the New York City dummy and a trend variable and find the interaction statistically insignificant. The plots in Figure 3.3 start from January 2009. We can see that in the beginning, there was a significant drop in sales volumes in both the control and treatment. We suspect that this decline was possibly due to the 2008 financial crisis. Also note that, in the later periods, the sales volumes become relatively more stable and shows a slight positive trend in Hartford area. Similar patterns can be found in other regions such as the Boston metro area.

Table 3.5 presents the results using monthly (Model B) and quarterly sales volumes (Model C). Similar to the two-period model based on the annual volumes, we use both the volumes and their logarithms as dependent variables. Before officially implementing the new

the two regions are plotted on separate axes for easy comparison. This is not re-scaling since the two axes cover the same range of values.

Table 3.5: Monthly and Quarterly Sales Volumes with Store Fixed Effect

	Model B		**Model C**	
Variables	Monthly Vol	Log(Mo.Vol)	Quarterly Vol	Log(Qr.Vol)
NYC*Post-policy	-0.030	-0.110***	-0.136*	-0.103***
	(0.034)	(0.024)	(0.074)	(0.022)
Post-policy	0.026	0.002	0.116*	0.0927***
	(0.021)	(0.022)	(0.064)	(0.020)
Hearing	-0.011	0.021	-0.003	0.003
	(0.021)	(0.015)	(0.036)	(0.015)
NYC*Hearing	0.008	-0.0858***	-0.0316	-0.0710***
	(0.031)	(0.019)	(0.047)	(0.018)
Log(price)	-0.487*	-0.658***	-1.056***	-0.682***
	(0.262)	(0.109)	(0.297)	(0.139)
Constant	4.692**	3.851***	10.72***	5.259***
	(0.219)	(0.897)	(2.488)	(1.162)
Adjusted R^2	0.041	0.107	0.043	0.101
Observations	13,950	13,950	4,464	4,464

Note: The measurement of volumes is in thousand gallons. In addition to the store fixed effects, all models control for time-specific (year and month/quarter) fixed effects. The dummy variable of New York City is ignored due to perfect collinearity with the store fixed effect. Robust standard errors are clustered at the store level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

regulation, the City Board of Health announced the schedule of a public hearing in June 2010 which was subsequently held in July. We introduce a dummy variable, *Hearing*, to indicate the months or the quarter for the hearing period preceding the policy implementation. Additionally, year and month/quarter time-specific fixed effects are controlled for in all of the four specifications. The estimated policy effects are generally in line with the two-period model. In percentage terms, the store monthly and quarterly sales volumes are estimated to decrease by 11% and 10.3%, respectively, and the demand for milk is also found to be price-inelastic.

3.4.1.3 Household Purchase Volumes

The preceding store-level analysis does not have direct implications for individual households, e.g., how much milk was consumed by a typical household, before and after the policy took effect. This is due to the lack of information on the population size that each store serves in the Retail Scanner data. Here we turn to the Nielsen Consumer Panel data that covers a nationally representative sample of households and their shopping records. While this is a useful dataset for observing household milk purchasing pattern, it also faces poor balancedness at the individual-household level. In fact, less than 5% of households reported consecutive milk consumption in all of the 24 months around September 2010, and this number is not significantly improved by considering quarterly consumption. This fact prevents us from using household-level fixed effects when we estimate the unbalanced panel of the individual household purchase volumes. We control for the county-level and time fixed effects, as well as the household demographic variables such as household size, income, and ethnic groups. The estimated effect of the policy is a decrease of 19.3 fluid ounces in volume and a statistically insignificant 3.4% decrease in percentage terms. Although we do not report the full results from this model here, we use this individual-household level data to conduct a structural estimation in a later section.

To form a balanced panel and provide a fairer check of our results, we calculate the household average purchase volumes at the county level. This model, which included county-level fixed effects, is denoted as Model D. Additionally, Model E considers the region-level average volumes as the dependent variable. Both models contain time-specific fixed effects. Estimates from these two models are shown in Table 3.6. The results show that the average household purchase of pasteurized milk dropped by about 28.6 and 22.9 ounces, respectively in the two models. The logarithm estimation reveals that the percentage volumes decrease by 13.6% and 11.4%, respectively.

Table 3.6: Average Monthly Household Purchase Volume

	Model D		Model E	
Variables	County Avg	Log(County Avg)	Region Avg	Log(Region Avg)
NYC*Post-policy	-28.63**	-0.136***	-22.90**	-0.114**
	(11.19)	(0.0454)	(10.79)	(0.048)
NYC	16.08	0.054	-39.40***	-0.206***
	(11.06)	(0.049)	(7.82)	(0.036)
Post-policy	10.00	0.039	8.55	0.043
	(23.53)	(0.096)	(24.36)	(0.110)
Hearing	19.27	0.0880	17.61	0.091
	(17.25)	(0.064)	(13.29)	(0.057)
NYC*Hearing	-21.23	-0.104*	-18.69	-0.099
	(15.18)	(0.062)	(12.77)	(0.062)
Log(price)	-212.43***	-0.842***	-161.57	-0.588
	(66.57)	(0.258)	(109.59)	(0.488)
Constant	-541.73**	2.341***	-328.80	3.396*
	(230.19)	(0.893)	(378.02)	(1.678)
Adjusted R^2	0.667	0.711	0.896	0.912
Observations	175	175	50	50

Note: The measurement of volumes is in fluid ounces. All models control for time (year and month/quarter) fixed effects. Model D controls for county fixed effects. Robust standard errors are reported in the parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

3.4.2 Discussions and Robustness

The empirical findings suggest that, due to the elimination of sell-by date regulation in New York City, the observed purchase volumes of pasteurized milk products decreased by about 10% (min. 8.9%, max. 13.6%). In addition, the estimated demand for milk is price-inelastic. The theoretical model points out that when the demand is price-inelastic, an extended sell-by date reduces both the percentage and the amount of food waste. It is important to note that the 10% change in purchase volumes only represents a lower bound for the reduction in food waste. The change in food waste is calculated as $\Delta \mathrm{FW} = \Delta q^* - \Delta x^*$, where $\Delta q^* < 0$ and $\Delta x^* > 0$. Since we only estimate the decrease in observed quantity $\Delta q^* \approx -10\%$ while the change in actual consumption Δx^* is not provided by the difference-in-difference models, the actual reduction in food waste is more than 10% of the purchase volume.

So far, we have discussed five models, each with two specifications, that are based on different choices of the dependent variable and the unit of time periods. To strengthen the robustness of the results, we cluster the standard errors and allow for heteroskedasticity, as well as include a two-period model as suggested by Bertrand et al. (2004). In the following discussion, we address several additional concerns on the robustness of the difference-in-difference estimation.

3.4.2.1 Price Endogeneity

The issue of price endogeneity in this study mainly rises from unobserved store-specific factors. First, throughout the entire time frame, individual stores may face demand shocks that are correlated with the prices of their milk products. An example might be store-wide promotions. Second, there is a particular endogeneity concern that relates to the policy itself. At the aggregate level, the correlation between price and the policy variable does not necessarily lead to inconsistent parameter estimates. However, individual stores in New York City may respond differently to the policy change due to unobserved factors, which could lead to biased estimates.

We apply two approaches to address the potential endogeneity issue. First, we use the region-level monthly and quarterly aggregate sales volumes to avoid individual store demand shocks.[11] The estimation results from these two specifications suggest that the policy reduced sales volumes by 10.3% and 9.8%, respectively. These are slightly less than but within 1% of what the store fixed-effects models predict. Second, we adopt a Hausman type instrumental variable to the store fixed-effects models, in which we use the average milk price in a store's county as an instrument for the store's price. As discussed in the preceding paragraph, the fact that the stores in the same county received the same policy treatment does not undermine the validity of the instrument because we have controlled for the treatment variable. The monthly and quarterly specifications respectively yield 13.8% and 14.7% estimated effect of the policy, compared to 11% and 10.4% in the original models. The higher estimates from using the instruments suggest that there might be some degree of variation in how stores respond to the policy.

3.4.2.2 Time Frame and Placebo Tests

The empirical estimation so far is based on data in 12 months before and after the policy initiation. We now extend the analysis to an 18-month period to check if the policy's effect is significant and persistent in a longer time frame. The two-period model now gives an estimate of the effect at 10.2%. And the store fixed-effects models of the monthly and quarterly sales volume suggest changes of 11.8% and 11.5%. In addition, we also extend the time frame in the evaluation based on the household purchase volumes, which imply percentage decreases of 14.3% and 13.7% for the county-level and region-level specifications, respectively. All of these estimates are within a 2% margin of the main results.

In a separate test, we examine the dynamic changes around the time of policy implementation. Following Autor (2003), we augment our monthly and quarterly models with

[11] For the same reason, the empirical analysis of the household purchase volumes aggregates the household-level data into county-level, therefore avoids the unobserved household-specific demand shocks and endogeneity.

leads and lags of the policy change. For leads variables, we add two dummies representing one and two months/quarters before the public hearing, which are equal to one in these time periods for observations from New York City . Similarly, we add four dummy variables for the month/quarter of policy change, and one, two, and three months/quarters after the implementation, respectively. For monthly models, there is also a lag variable that indicates the fourth month after the change and forward. The results are listed in Table 3.7. The estimated coefficients of the leads variables are either small in magnitude or statistically insignificant, suggesting that there were no major changes in sales or purchase volumes before the announcement of the policy. Starting with the period of policy implementation, New York City experienced a decline in volumes reported in both retailers' sales and households' purchases data. The estimated long-run changes are close to our previous findings, i.e., about a 10% decrease in store sales volumes and a 12-14% decrease in household purchase volumes.

3.4.2.3 Synthetic Control Method

Finally, we test the policy effect using a different estimation method—the synthetic control method (Abadie et al., 2010; Abadie et al., 2015). The main reason we do not use the synthetic control as the primary estimation strategy is that it only applies to the aggregate-level data and does not offer channels for estimating price elasticities. However, the synthetic control method generally produces better-matched pre-treatment trends between the treatment and the (synthetic) control group. Therefore, we apply it here to check whether the trend of sales volumes in Hartford is representative and whether the corresponding results are robust.

Specifically, in addition to the Hartford area, we add two types of possible control groups. One set consists of three major metro areas of the northeastern region–Boston, Philadelphia, and Baltimore that are within 200 miles radius of New York City. The second type is another metro area from New York State, i.e., Buffalo.[12] These five control groups form the candidate

[12] The nearest metro area in the state is Albany. However, its projected per capita sales in Albany city are more than four times higher than in New York City. We suspect that there is either an oversampling of stores in the urban area or the city stores are serving a much larger population than the city population. If we include Albany in the analysis, it would receive a zero weight.

Table 3.7: Estimated Policy Impact with Time Placebos

	Model B	**Model C**	**Model D**	**Model E**
Policy change $_{t+2}$	-0.014	0.054***	-0.044	-0.062
	(0.009)	(0.013)	(0.075)	(0.088)
Policy change $_{t+1}$	0.026***	-0.021	0.005	0.003
	(0.009)	(0.017)	(0.072)	(0.075)
Policy change $_{t0}$	-0.107***	-0.105***	-0.087	-0.075
	(0.016)	(0.0254)	(0.092)	(0.071)
Policy change $_{t-1}$	-0.129***	-0.080***	-0.100*	-0.068
	(0.017)	(0.025)	(0.058)	(0.064)
Policy change $_{t-2}$	-0.117)***	-0.090***	-0.152**	-0.129**
	(0.017)	(0.031)	(0.077)	(0.057)
Policy change $_{t-3}$	-0.143***	-0.092***	-0.182***	-0.141**
	(0.023)	(0.033)	(0.062)	(0.060)
Policy change $_{t-4 \text{ forward}}$	-0.099***		-0.140***	-0.123**
	(0.025)		(0.052)	(0.052)
Adjusted R^2	0.107	0.105	0.702	0.833

Note: In all four models, the dependent variable is the logarithm of sales/purchase volumes. Also note that for Model C, there is no forward lag variable because the Policy change $_{t-3}$ already includes the last several time periods in the data. Fixed effects are included and standard errors are clustered by the same rules as in Tables 3.5 and 3.6. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

pool that is used to construct a synthetic control group. We assign non-negative weights to each control group such that the sum of the weights equal to one and the sum of the weighted sales volumes best match the New York City's volumes in the 12 months before the policy change. Given the significant variation in the population sizes among the six regions, we project the region total sales volume into per capita sales as in (Abadie et al., 2010). Here the region-level aggregate sales volumes are projected according to the store coverage ratio based on Scantrack market and retail channels. Another possible treatment of the large variation in the region-level volumes is to use a modified synthetic control method which allows for an intercept so that the trend of synthetic control is parallel to the treatment up to a constant (Doudchenko and Imbens, 2016; Li, 2019).

Mathematically, let us denote the per capita sales in New York City as $Y_{1,t}$ and those in the control groups as $Y_{j,t}$ $(j = 2, 3.., 6)$. A set of weights, w_j $(j = 2, 3...6)$ are chosen to construct a synthetic control group $Y_t^{Synth} = \sum_{j=2}^{6} w_j Y_{j,t}$, so that $\sum_{t=1}^{12}(Y_{1,t} - Y_t^{Synth})^2$ is minimized.[13] The optimal set of weights assigns 0.151 to Hartford, 0.506 to Buffalo, 0.151 to Boston, 0.191 to Philadelphia, and 0 to Baltimore. Note that although Buffalo is assigned the highest weight, it is due to the fact that its per capita sales of milk reported in the data is much lower than New York City, not because its pre-treatment trend best resembles the later. The estimated policy effect is given by the average difference between the treatment and the synthetic control after the policy implementation.

Figure 3.4 plots the per capita milk sales volumes of New York City and its synthetic control, for 12 months before and after September 2010. Despite that the outcome variable is highly volatile, the synthetic control reasonably approximate New York City's pre-treatment trend. Relative decreases in the sales volumes of the treatment unit can be clearly seen in the post-treatment periods. More precisely, by taking the average difference $\frac{1}{12}\sum_{t=13}^{24}(Y_{1,t} -$

[13]The original synthetic control uses a set of covariate vectors X_j that includes some of the outcome variables, and minimize a weighted sum $(X_1 - X^{Synth})'V(X_1 - X^{Synth})$, where $X^{synth} = \sum_j w_j X_j$. When X_j only includes the pre-treatment outcome variables $Y_{j,t}$, then the optimal V calculated by the procedure proposed in Abadie and Gardeazabal (2003) would be an identity matrix (Bohn et al., 2014; Kaul et al., 2018). Also, as noted by Kaul et al. (2018), as long as all pre-treatment outcome variables are used, it is not necessary to add any additional covariates.

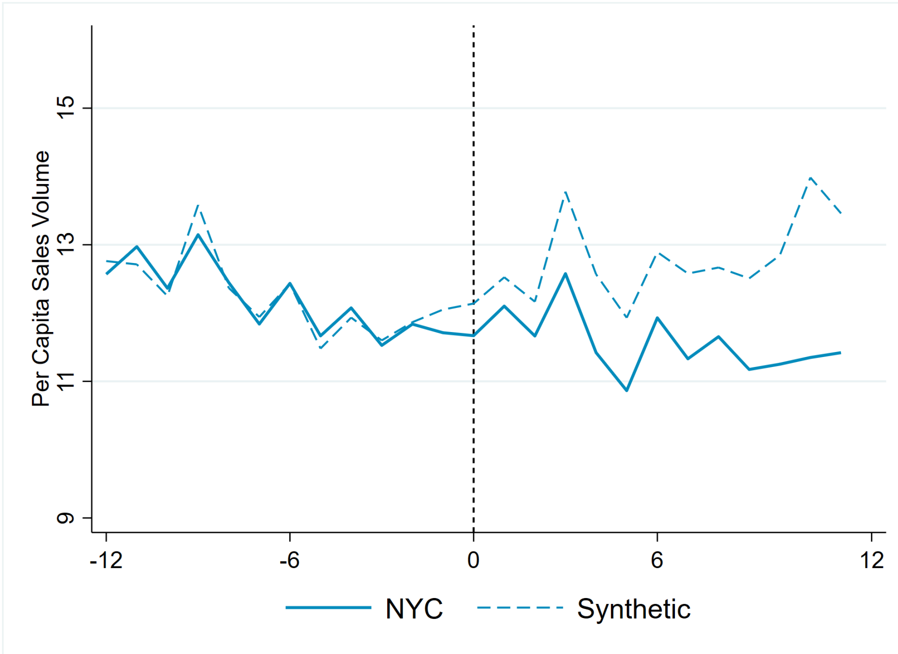


Figure 3.4: Per capita sales volume (fluid ounces): New York City vs. synthetic control.

Y_t^{Synth}), the policy is estimated to result a 1.26 ounce decrease in monthly per capita sales, which translates to a 9.7% change. We can see that the divergence of New York City's sales volume from the synthetic control started one period before the policy change. This can be explained by the fact that the policy was proposed and a public hearing was held two months before September 2010. Anticipating the resolution to be adopted, the market began to adjust in advance. In our difference-in-difference models, the effect of the public hearing consistently shows a negative impact on the purchase volumes in New York City.

To further check the robustness, we run additional synthetic control estimations where each of the five control candidates is dropped from the analysis. The regions being excluded and the estimated policy effects are, respectively: Hartford, -9.87%; Buffalo, -9.43%; Boston, -6.8%; Philadelphia, -11.0%; Baltimore, -8.98%. These estimates are close to our previous finding of the policy impact.

Overall, the results from the synthetic control method, as with results from all other checks, suggest that our main empirical finding, i.e., a significant decrease in milk sales due to the policy change, is robust to a wide range of empirical tests.

3.4.3 A Simple Structural Model

The difference-in-difference estimation and the synthetic control method all provide a clear picture of how the new policy lowers the purchase volume of milk in New York City. Additionally, we estimate that the demand for milk is price-inelastic. According to the theoretical model, a utility-maximizing household would reduce the amount of its food waste by at least 10% of the total purchase. However, these empirical methods only provide lower bounds for the reduction in food waste. In this section, we construct a simple structural model by parametrizing the utility function and the cost function. By imposing explicit parameter structures, we are able to identify the changes in the household utilization rate and food waste without relying on the theoretical implications based on price elasticities. In this sense, the structural model also serves as an empirical verification for our theoretical model developed earlier.

We assume that the parametrized utility function takes the following form for household i at period t ($t = -12, -11, ...0, ...11, 12$):

$$U(q_{i,t}, \delta_{i,t}) = A\frac{[(1-\delta_{i,t})q_{i,t}]^{1-\alpha}}{1-\alpha} - B\frac{1}{L_{i,t}}(1-\delta_{i,t})^{\beta} + w_{i,t} - p_{i,t}q_{i,t} \tag{3.8}$$

This specification is based on the case of proportional waste measure. Recall that if the cost function is based on the quantity waste measure, then the food waste decreases by exactly 10% because the actual consumption is unchanged. The model parameters satisfy these conditions: $A > 0$, $B > 0$, $\alpha > 0$, and $\beta > 0$. The sell-by date variable $L_{i,t}$ is set to be 15 days for all periods if the household lives in the Hartford area. For $t < 0$, households living in New York City were faced with $L_{i,t} = 9$ days. As for the post-policy periods $t \geq 0$, we explore 14 days and 15 days as two possible values of the sell-by dates in New York City.

Unlike the difference-in-difference estimation where the policy effect is essentially captured by dummy variables, here we can identify the cost parameter β by explicitly modelling the variation in sell-by dates. We use a general form of the iso-elastic function $A\frac{x^{1-\alpha}}{1-\alpha}$ for the utility drawn from the actual consumption. This allows for a wide range of price elasticities of the demand, which depends on whether $\alpha > 1$ or $0 < \alpha \leq 1$. On the other hand, a Cobb-Douglas function Ax^{α} artificially constrains the demand to be always price-elastic, regardless of the value of α.

The first-order conditions for the optimal purchase quantity and percentage food waste can be easily derived. In logarithm forms, the system of structural equations is presented below:

$$\log q_{i,t} = (\gamma_A^q \log A + \gamma_B^q \log B\beta) + \gamma_1^q \log L_{i,t} + \gamma_2^q \log p_{i,t} \tag{3.9}$$

$$\log \theta_{i,t} = \left(\gamma_A^\theta \log A + \gamma_B^\theta \log B\beta\right) + \gamma_1^\theta \log L_{i,t} + \gamma_2^\theta \log p_{i,t} \tag{3.10}$$

where $\theta_{i,t} = 1 - \delta_{i,t}$ is the utilization rate of food, and $\gamma_A^q = -\frac{\beta}{1-\alpha-\alpha\beta}$, $\gamma_B^q = -\gamma_1^q = -\gamma_2^\theta = \frac{1-\alpha}{1-\alpha-\alpha\beta}$, $\gamma_2^q = -\frac{1-\alpha-\beta}{1-\alpha-\alpha\beta}$, $\gamma_A^\theta = -\frac{1}{1-\alpha-\alpha\beta}$, and $\gamma_B^\theta = -\gamma_1^\theta = \frac{\alpha}{1-\alpha-\alpha\beta}$. Because the utilization rate $\theta_{i,t}$ is not observed in data, we only estimate the first equation on $\log q_{i,t}$. The estimates obtained by estimating only one of the equations are not efficient but still consistent. Once we have estimated γ_1^q and γ_2^q, we can solve for the values of α and β: $\widehat{\alpha} = \frac{\widehat{\gamma_1^q}+1}{\widehat{\gamma_1^q}-\widehat{\gamma_2^q}}$ and $\widehat{\beta} = -\frac{\widehat{\gamma_2^q}+1}{\widehat{\gamma_1^q}}$. Finally, the policy's effect on the utilization rate, $\widehat{\gamma_1^\theta}$, can be predicted by substituting $\widehat{\alpha}$ and $\widehat{\beta}$, and the change in the utilization rate can be calculated as $\Delta \log \theta_{i,t} = \widehat{\gamma_1^\theta} \cdot \Delta \log L_{i,t}$. Note that because the parameters A and B are not identified, we cannot predict the absolute value of $\log \theta_{i,t}$ but only the changes of it.

To carry out the structural estimation, we apply the parametrized, single-equation model to the micro-level data without aggregation across households or over time periods. We first use the individual household-level monthly purchase volume as the dependent variable. Household size, income, and race/ethnicity are included in the estimation. We also analyze the individual stores' monthly sales volumes with store fixed effects. Both models control for county fixed effects, as well as time (year and month) fixed effects.

Table 3.8 shows the relevant estimates of the structural estimation. First, the predicted values of α are greater than 1, and all of the estimated price elasticities, represented by $\widehat{\gamma_2^q}$, are smaller than 1 in their absolute values. These results confirm the theoretical model's implication that $xu'' + u' < 0$ whenever $\eta_D > -1$. Second, the effect of the new policy on the purchase volume can be measured by $\widehat{\gamma_1^q} \cdot \Delta \log L_{i,t}$, where $\Delta \log L_{i,t}$ is either $\log 14 - \log 9$ or $\log 15 - \log 9$. For the model of store sales volumes, the two specifications give estimates around 8%, which are close to the 10% given by the difference-in-difference estimation. As for the household purchase model, the effect is a decrease of about 4%, which is similar to the result described above when we conduct the analysis of the unbalanced household panel. Lastly and most importantly, the last row of Table 3.8 shows that the improvement in the logarithm of utilization rate is about 13-15% in the household purchase model and 17-19% in the store sales model. This means that if the households previously consume 67% of their purchased milk (1/3 food waste), now this number is increased to about 77%-81%, an increase of 10%-14% of total purchase volume. Hence, the reduction in food waste is more than 10% as the theoretical model predicts. Interestingly, though the household purchase model and the store sales model give different estimates for the decreases in volumes (4% vs. 8%), they both suggest that food waste decreases by at least 6% more than the purchase volumes.

3.5 Market-Level Policy Implications

In this section, we provide a larger picture of the implications of an extended sell-by date by looking beyond the consumer behavior and exploring the policy's impact on the supply side. We first look at how the market price of milk would react to the policy. Next, we provide several possible explanations as to why some retailers and milk distributors support

Table 3.8: Structural Estimation of Utilization Rate

	Household Purchase		Store Sales	
Parameters	$L_{i,0}^{NYC} = 14$	$L_{i.0}^{NYC} = 15$	$L_{i,0}^{NYC} = 14$	$L_{i,0}^{NYC} = 15$
$\widehat{\gamma_1^q}$	-0.0996	-0.0862	-0.1788	-0.1547
$\widehat{\gamma_2^q}$	-0.7374	-0.7374	-0.6697	-0.6697
$\widehat{\alpha}$	1.4117	1.4033	1.6728	1.6414
$\widehat{\beta}$	2.6365	3.0464	1.8473	2.1351
$\widehat{\gamma_1^\theta}$	0.3415	0.2999	0.4445	0.3959
$\Delta \log \theta$	0.1509	0.1325	0.1964	0.1749

Note: All models control for county fixed effects and time (year and month) fixed effects. The two specifications based on the household purchase volumes control for household sizes, income, and ethnic groups. The two store-sales volume specifications include store fixed effect. The estimates in the first two rows are statistically significant at the 1% level under robust standard errors clustered at the county level, except for $\widehat{\gamma_1^q}$'s in the household models whose p-values are about 0.13.

extending the sell-by dates.

3.5.1 Price Changes

Upon the empirical findings of the case of New York City and the theoretical discussions, our model suggests that the policy change results in a downward shift of the consumer demand for milk products. Regardless of the degree of competitiveness of the retail market of milk, a reduced demand for a product generally leads to a lower price of that product even when the retailers exhibit some market power. At first glance, it may seem that the retailer profits would drop as a result of lower price and quantity and, hence, it is not rational for the retailers to support the extension of milk sell-by dates. However, we point out three possible reasons as to why there has been some support from the supply side in the recent regulation changes. First and foremost, the upstream milk market in the U.S. is highly regulated with precisely-set prices by market orders, while in some states, the downstream retail markets also face with price ceilings and floors. Therefore, discretionary price changes are very limited so firm profits may not drop as much as in an unregulated market.

3.5.2 Loss-Leading Strategy

In reality, milk is often a loss-leading product for retailers, and hence, lower prices and quantities may not necessarily harm retailer profits. More importantly, the use of loss-leading strategy helps explain why small retailers have strong incentive to advocate for extended sell-by dates as seen in the recent efforts in appealing Montana's 12-day rule (United States Court of Appeals, 2017).

Note that loss-leading is an outcome of competition between large and small retailers (Johnson, 2017; Chen and Rey, 2012). The intuition is that large retailers price staple products such as milk below their costs to attract consumers who might otherwise buy such products in small stores, and that these consumers may also purchase other products during their shopping trips. On the other hand, small retailers may follow the strategy into lowering their staple products' prices, as well. However, as they often do not offer the full range of products, the benefit of using the loss-leading strategy is not as significant to them as to the large retailers. Hence the source of market power of the large retailers resides not only in their capacity to absorb the loss in profits but also in their ability to offer a larger set of products.

Theoretically, the variation in the costs of milk products is largely impacted by inventory cost because of the market order and price ceilings/floors mentioned earlier. This fact constrains the small retailers in utilizing pricing strategies to compete with large retailers. Thus, a longer sell-by date not only reduces their inventory cost but also offers an extra dimension to compete. Intuitively, milk products offered by the large retailers may not seem as "attractive" as they were before. Even though the eventual equilibrium is likely to be the one in which both large and small retailers set extended sell-by dates, this new outcome is initiated by the small retailers exercising their new options and therefore, indicates a reduced market power of the large retailers. In sum, a longer sell-by date may reduce the effectiveness of the loss-leading strategies and benefit the small retailers.

3.5.3 Market Power and Competition

Our empirical findings suggest that a longer sell-by date results in lower sales volumes and possibly lower prices for milk. Therefore, if the milk market is controlled by a monopoly, it would not increase the sell-by date and may even shorten it, given that the monopoly has the sole power in setting sell-by dates in the absence of regulation. However, the milk market is not a monopolized market, neither vertically nor horizontally. Vertically, there is an upstream market where manufacturers sell milk to retailers and a downstream market where retailers re-sell to consumers. Horizontally, there are different manufacturers and retailers competing at the upstream and downstream levels. Even for the case of private-label products where the retailer, not the manufacturer, sets sell-by dates, it still faces competition from other brands.

It is logical to infer that the policy's effect will be transmitted from the retail market to the upstream market, and hence, the manufactures will face a downward shift in the demand from the retailers as well. To explain why the manufacturers may still choose to extend the sell-by dates, we need to dig deeper into their decision mechanism. More specifically, consider an agent–a manufacturer or a private-label retailer–to whom setting the sell-by dates is a decision variable. Theoretically, its revenue function and profit function, conditional on its competitors' strategies, are generally non-linear in the product's sell-by date. A near-zero sell-by date would effectively eliminate the demand for its product, while a very long sell-by date that still ensures quality may lead to a decreased sales volume and profit. Then the "optimal" sell-by dates would depend on the curvature of its profit function. There are two sources that may affect its curvature: the competition among manufacturers and the demand they face. In particular, as the literature on imperfectly competitive markets of homogeneous goods shows, if the demand of a product is inelastic, then it is likely that the decisions of the manufacturers are strategic complements (Bulow et al., 1985). In other words, if some of the manufacturers decide to increase sell-by dates, others would follow.

There are a number of possible reasons why some manufacturers find it profitable to increase sell-by date, and we provide two examples. First, there are manufacturers that are disadvantaged in transportation distance, e.g., those who transport processed milk from another state or a location far from the market. For these firms, a very short sell-by date might leave their products only a few days left for consumers, less the time spent on transportation. Therefore, the marginal benefit of increasing the sell-by date is a significant improvement in their competitiveness. This can be seen in the recent case of "Core-Mark vs. Montana Board of Livestock" (United States Court of Appeals, 2017). Core-Mark is a milk distributor that delivers its milk products from Washington and California to Montana and therefore, may have a strong incentive to seek a longer sell-by date. It is also noteworthy that the number of such interstate distributors is not small–among the 79 licensed distributors in Montana, 49 are out-of-state.

Second, a short sell-by date imposes substantial inventory management cost to small grocery and convenience stores. While the larger retail chains have higher efficiency in product ordering, inventory reviewing and adjusting, shelf-space monitoring, and re-allocating products among stores, smaller retailers do not enjoy such efficiency from the economies of scale. Another petitioner in appealing the Montana 12-day rule is the Friends of Montana Retailers and Consumers, Inc. As some convenience-store owners testify, a large amount of milk is disposed of every month due to sell-by dates, imposing even more burden to the stores for a product that is already priced under the cost. Though the sell-by dates are set by the manufacturers, small retailers' demand for a longer sell-by date may be partially conveyed through the bargaining process that influences the manufacturers' decisions.

The above discussion shows that, theoretically, it is possible for manufacturers to prefer a longer sell-by date. Then the empirical question is whether this is the case in New York City's policy change. The answer is yes. The reason is that the previous 9-day regulation was binding—it was strictly less than the Nash equilibrium outcome that could be reached, had there been no regulation. As seen in the neighboring states of New Jersey and Connecticut, as well as other regions in New York state where there were no regulations, the typical sell-by dates are around 15 days. Moreover, the pre-policy sell-by dates in New York City were generally set at exactly 9 days. If it was profitable to shorten the sell-by date, the manufacturers would likely do so since that still complies with the regulation.

3.6 Conclusions

In this study, we show how extending the sell-by dates of New York City's milk products reduces food waste and improves consumer welfare. We find that the new policy reduces food waste by at least 10% and that the consumers now drink more milk while spending less money on milk products. The theoretical model and a structural estimation show that this pattern of change is likely to be a result of price-inelastic demand for milk in New York City. In addition, the structural model suggests that the policy resulted in a 13-19% increase in household food utilization rate. The findings presented in this study do not undermine other related policy proposals, including those that attempt to resolve confusion between sell-by and use-by dates. Instead, our findings emphasize the importance of keeping regulations and consumer education up-to-date.

The results provide the first empirical verification on the influence of sell-by dates on consumption behavior and shed light on the studies of other perishable food. However, we would like to note that a similar policy change in other regions and for other goods may not have impacts as significant as in the case of New York City because its 9-day was considered rather stringent. In addition, as we show in the discussion of the proportional waste specification, it is possible that some premium food types may exhibit different patterns of change due to their elastic demand. In this case, extending the shelf life of premium food may not necessarily reduce food waste. For premium milk products, there exists another important reason why they may not receive the same policy impact. An increasing number of premium milk products have adopted the ultra-high temperature pasteurization technology. For example, many organic and value-added milk products are ultra-pasteurized, which generally leads to shelf lives that are longer than a month. In the case of New York City's policy, these products fell outside of the scope of the regulation and were not directly affected by the policy change.

Lastly, although this study largely focuses on the consumer side, the long-term supply-side effects of sell-by date are of considerable interest. In this chapter, we provide a detailed discussion of two important aspects of the milk market that explain the supply-side support for extended sell-by dates. First, small grocery stores could benefit from a longer shelf life because milk is often a loss-leading product. When considering other products that are non-loss-leaders, an extended sell-by date may provide less benefit to small retailers. Second, high transportation and inventory costs render the out-of-state distributors and small stores disadvantaged when facing competition from the local distributors and large retailers. From this view, their incentives to seek a longer shelf life may hold true for all perishable foods, including non-loss-leading products.

Bibliography

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.

Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, 6(1):21–37.

Alvarez, A., Amsler, C., Orea, L., and Schmidt, P. (2006). Interpreting and Testing the Scaling Property in Models Where Inefficiency Depends on Firm Characteristics. *Journal of Productivity Analysis*, 25(3):201–212.

Amsler, C., Prokhorov, A., and Schmidt, P. (2016). Endogeneity in Stochastic Frontier Models. *Journal of Econometrics*, 190(2):280–288.

Angrist, J. D. and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

Archer, E., Hand, G. A., and Blair, S. N. (2013). Validity of U.S. Nutritional Surveillance:

National Health and Nutrition Examination Survey Caloric Energy Intake Data, 1971-2010. *PLoS ONE*, 8(10):1–12.

Archer, E., Thomas, D. M., McDonald, S. M., Pavela, G., Lavie, C. J., Hill, J. O., and Blair, S. N. (2016). The Validity of US Nutritional Surveillance: USDA's Loss-Adjusted Food Availability Data Series 1971-2010. *Current Problems in Cardiology*, 41(11-12):268–292.

Asmild, M., Bogetoft, P., and Leth Hougaard, J. (2013). Rationalising Inefficiency: Staff Utilisation in Branches of a Large Canadian Bank. *Omega*, 41(1):80–87.

Asmild, M. and Matthews, K. (2012). Multi-Directional Efficiency Analysis of Efficiency Patterns in Chinese Banks 1997-2008. *European Journal of Operational Research*, 219(2):434–441.

Atkinson, S. E. and Tsionas, M. G. (2016). Directional Distance Functions: Optimal Endogenous Directions. *Journal of Econometrics*, 190(2):301–314.

Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, 21(1):1–42.

Banker, R., Natarajan, R., and Zhang, D. (2019). Two-stage Estimation of the Impact of Contextual Variables in Stochastic Frontier Production Function Models Using Data Envelopment Analysis: Second Stage OLS versus Bootstrap Approaches. *European Journal of Operational Research*, 278(2):368–384.

Banker, R. D. and Natarajan, R. (2008). Evaluating Contextual Using Data Variables Envelopment Affecting Analysis. *Operations Research*, 56(1):48–58.

Bellemare, M. F., Çakir, M., Peterson, H. H., Novak, L., and Rudi, J. (2017). On the Measurement of Food Waste. *American Journal of Agricultural Economics*, 99(5):1148–1158.

Beretta, C., Stoessel, F., Baier, U., and Hellweg, S. (2013). Quantifying Food Losses and the Potential for Reduction in Switzerland. *Waste Management*, 33(3):764–773.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bogetoft, P. and Hougaard, J. L. (2003). Rational Inefficiencies. *Journal of Productivity Analysis*, 20(3):243–271.

Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population? *Review of Economics and Statistics*, 96(2):258–269.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American statistical Association*, 90:443–450.

Briefel, R. R., Sempos, C. T., Mcdowell, M. A., Chien, S., and Alaimo, K. (1997). Dietary Methods Research in the Third National Health and Nutrition Examination Survey : Underreporting of Energy Intake. *American Journal of Clinical Nutrition*, 65(suppl):1203S–1209s.

Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). The IRI Marketing Data Set. *Marketing Science*, 27(4):745–748.

Bulow, J. I., Geanakoplos, J. D., and Klemperer, P. D. (1985). Multimarket Oligopoly: Strategic Substitutes and Complements. *Journal of Political Economy*, 93(3):488–511.

Buzby, J. C., Wells, H. F., Axtman, B., and Mickey, J. (2009). *Supermarket Loss Estimates for Fresh Fruit , Vegetables , Meat , Poultry , and Seafood and Their Use in the ERS Loss-Adjusted Food Availability Data Data.* U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

Buzby, J. C., Wells, H. F., and Hyman, J. (2014). *The Estimated Amount, Value, and Calories of Postharvest Food Losses at the Retail and Consumer Levels in the United States, EIB-121.* U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4):772–793.

Chambers, R. G., Chung, Y., and Färe, R. (1998). Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications*, 98(2):351–364.

Chapagain, A. and James, K. (2011). *The Water and Carbon Footprint of Household Food and Drink Waste in the UK*. Waste & Resources Action Programme (WRAP).

Chazan, D. and Gal, S. (1977). A Markovian Model for a Perishable Product Inventory. *Management Science*, 23(5):512–521.

Chen, Z. and Rey, P. (2012). Loss Leading as an Exploitative Practice. *American Economic Review*, 102(7):3462–3482.

Danaher, P. J. and Smith, M. S. (2011). Modeling Multivariate Distributions Using Copulas: Applications in Marketing. *Marketing Science*, 30(1):4–21.

Dashti, H. S. (2016). *Is January 1st the Best Date to Kick-Start a Weight Loss Diet?* American Society for Nutrition.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. *NBER Working Paper*.

Drewnowski, A. (1997). Taste Preferences and Food Intake. *Annual Review of Nutrition*, 17(1):237–253.

Duan, Q. and Liao, T. W. (2013). A New Age-based Replenishment Policy for Supply Chain Inventory Optimization of Highly Perishable Products. *International Journal of Production Economics*, 145(2):658–671.

Dubé, J.-P. (2004). Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks. *Marketing Science*, 23(1):66–81.

Ellison, B. and Lusk, J. L. (2018). Examining Household Food Waste Decisions: A Vignette Approach. *Applied Economic Perspectives and Policy*, 40(4):613–631.

FAO (2013). *Food Wastage Footprint: Impacts on Natural Resources. (Summary Report)*. Food and Agriculture Organization of the United Nations.

FAO/WHO/UNU (1985). *Energy and Protein Requirements : Report of a Joint FAO/WHO/UNU Expert Consultation [Held in Rome from 5 to 17 October 1981]*. World Health Organization, Geneva.

Färe, R., Pasurka, C., and Vardanyan, M. (2017). On Endogenizing Direction Vectors in Parametric Directional Distance Function-Based Models. *European Journal of Operational Research*, 262(1):361–369.

Feng, G. and Serletis, A. (2014). Undesirable Outputs and a Primal Divisia Productivity Index Based on the Directional Output Distance Function. *Journal of Econometrics*, 183(1):135–146.

FLW Protocol (2016). *Food Loss and Waste Accounting and Reporting Standard*. Food Loss and Waste Protocol.

Frankenfield, D., Roth-Yousey, L., and Compher, C. (2005). Comparison of Predictive Equations for Resting Metabolic Rate in Healthy Nonobese and Obese Adults: A Systematic Review. *Journal of the American Dietetic Association*, 105(5):775–789.

Fried, H. O., Lovell, C. A. K., and Schmidt, S. S. (2008). *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, New York.

FUSIONS (2016). *Estimates of European Food Waste Levels*. European Union, Brussels, Belgium.

Goodwin, B. K. and Hungerford, A. (2014). Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts. *American Journal of Agricultural Economics*, 97(3):879–896.

Greene, W. (2004). Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems. *Health Economics*, 13(10):959–980.

Griffin, M., Sobal, J., and Lyson, T. A. (2009). An Analysis of A Community Food Waste Stream. *Agriculture and Human Values*, 26(1-2):67–81.

Hager, E. B. (2010). *Removing Mysterious Sell-By Date From Milk Cartons*. The New York Times.

Hall, K. D., Guo, J., Dore, M., and Chow, C. C. (2009). The Progressive Increase of Food Waste in America and Its Environmental Impact. *PLoS ONE*, 4(11):9–14.

Hamilton, S. F. and Richards, T. J. (2019). Food Policy and Household Food Waste. *American Journal of Agricultural Economics*, 101(2):600–614.

Heckman, J. J. (1976). Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and A Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5(4):475–492.

Hendel, I. (1999). Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns. *Review of Economic Studies*, 66(2):423–446.

Hendel, I. and Nevo, A. (2006). Measuring the Implications of Sales and Consumer Inventory Behavior. *Econometrica*, 74(6):1637–1673.

Institute of Medicine (2005). *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*. National Academies Press, Washington, D.C.

Johnson, J. P. (2017). Unplanned Purchases and Retail Competition. *American Economic Review*, 107(3):931–965.

Johnston, C. S., Day, C. S., and Swan, P. D. (2002). Postprandial Thermogenesis Is Increased 100% on a High-Protein, Low-Fat Diet versus a High-Carbohydrate, Low-Fat Diet in Healthy, Young Women. *Journal of the American College of Nutrition*, 21(1):55–61.

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *Journal of Econometrics*, 19(2-3):233–238.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2018). Synthetic Control Methods: Never Use all Pre-Intervention Outcomes Together with Covariates. *Working Paper*.

Kumbhakar, S. C., Ghosh, S., and McGuckin, J. T. (1991). A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. *Journal of Business and Economic Statistics*, 9(3):279–286.

Kumbhakar, S. C. and Tsionas, E. G. (2006). Estimation of Stochastic Frontier Production Functions with Input-Oriented Technical Efficiency. *Journal of Econometrics*, 133(1):71–96.

Kurkalova, L. A. and Carriquiry, A. (2003). Input- and Output-Oriented Technical Efficiency of Ukrainian Collective Farms, 1989-1992: Bayesian Analysis of a Stochastic Production Frontier Model. *Journal of Productivity Analysis*, 20(2):191–211.

Kutlu, L. (2010). Battese-Coelli Estimator with Endogenous Regressors. *Economics Letters*, 109(2):79–81.

Landry, C. E. and Smith, T. A. (2019). Demand for Household Food Waste. *Applied Economic Perspectives and Policy*, 41(1):20–36.

Leib, E. B., Gunders, D., Ferro, J., Nielsen, A., Nosek, G., and Qu, J. (2013). *The Dating Game: How Confusing Food Date Labels Lead to Food Waste in America*. The Natural Resources Defense Council (NRDC).

Li, K. T. (2019). Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods. *Journal of the American Statistical Association*.

Lin, B. H., Yen, S. T., Huang, C. L., and Smith, T. A. (2009). U.S. demand for organic and conventional fresh fruits: The roles of income and price. *Sustainability*, 1(3):464–478.

Linna, M. (1998). Measuring Hospital Cost Efficiency with Panel Data Models. *Health Economics*, 7:415–427.

Lusk, J. L. and Ellison, B. (2017). A Note on Modelling Household Food Waste Behaviour. *Applied Economics Letters*, 24(16):1199–1202.

Malikov, E., Kumbhakar, S. C., and Tsionas, M. G. (2016). A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of us Banks in 2001-2010. *Journal of Applied Econometrics*, 31:1407–1429.

McArdle, W. D., Katch, F. I., and Katch, V. L. (1986). *Exercise Physiology: Energy, Nutrition, and Human Performance*. Lea & Febiger, Philadelphia, 2nd edition.

Muth, M. K., Karns, S. A., Nielsen, S. J., Buzby, J. C., and Wells, H. F. (2011). *Consumer-Level Food Loss Estimates and Their Use in the ERS Loss-Adjusted Food Availability Data*. U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

Neff, R. A., Spiker, M. L., and Truant, P. L. (2015). Wasted Food: U.S. Consumers' Reported Awareness, Attitudes, and Behaviors. *PLoS ONE*, 10(6):1–16.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition.

New York City Board of Health (2010). *Notice of Adoption of A Resolution Repealing Articles 111 and 117 of the New York City Health Code*. New York.

OLR (1999). *Office of Legislative Research Report 99-R-0678*. Connecticut General Assembly.

Park, S. and Gupta, S. (2012). Handling Endogenous Regressors by Joint Estimation Using Copulas. *Marketing Science*, 31(4):567–586.

Porpino, G., Parente, J., and Wansink, B. (2015). Food Waste Paradox: Antecedents of Food Disposal in Low Income Households. *International Journal of Consumer Studies*, 39:619–629.

Qi, D. and Roe, B. E. (2016). Household Food Waste: Multivariate Regression and Principal Components Analyses of Awareness and Attitudes among U.S. Consumers. *PLoS ONE*, 11(7):1–19.

Quested, T. and Parry, A. (2011). *New Estimates for Household Food and Drink Waste in the UK*. Waste & Resources Action Programme (WRAP).

Reinhard, S., Lovell, K., and Thijssen, G. (1999). Econometric Estimation of Technical and Environmental Efficiency: An Application to Dutch Dairy Farms. *American Journal of Agricultural Economics*, 81(1):44–60.

Reynolds, C. J., Mavrakis, V., Davison, S., Høj, S. B., Vlaholias, E., Sharp, A., Thompson, K., Ward, P., Coveney, J., Piantadosi, J., Boland, J., and Dawson, D. (2014). Estimating Informal Household Food Waste in Developed Countries: The Case of Australia. *Waste management & research*, 32(12):1254–8.

Richards, T. J., Mancino, L., and Nganje, W. (2012). Nutrient Demand in Food Away from Home. *Forum for Health Economics & Policy*, 15(2).

Roe, B. E., Apolzan, J. W., Qi, D., Allen, H. R., and Martin, C. K. (2018a). Plate Waste of Adults in the United States Measured in Free-living Conditions. *PLoS ONE*, 13(2):1–16.

Roe, B. E., Phinney, D. M., Simons, C. T., Badiger, A. S., Bender, K. E., and Heldman, D. R. (2018b). Discard Intentions Are Lower for Milk Presented in Containers without Date Labels. *Food Quality and Preference*, 66(December 2017):13–18.

Roza, A. M. and Shizgal, H. M. (1984). The Harris Benedict Equation Reevaluated: Resting Energy Requirements and the Body Cell Mass. *American Journal of Clinical Nutrition*, 40(1):168–182.

Scrimshaw, N. S., Waterlow, J. C., and Schürch, B. (1996). Energy and Protein Requirements: Proceedings of an IDECG Workshop Held in London, UK, October 31-November 4, 1994. *European Journal of Clinical Nutrition*, 50:Supplement 1.

Secondi, L., Principato, L., and Laureti, T. (2015). Household Food Waste Behaviour in EU-27 Countries: A Multilevel Analysis. *Food Policy*, 56:25–40.

Shee, A. and Stefanou, S. E. (2014). Endogeneity Corrected Stochastic Produciton Frontier and Technical Efficiency. *American Journal of Agricultural Economics*, 97(3):939–952.

Simar, L., Lovell, C. A. K., and Vanden Eeckaut, P. (1994). *Stochastic Frontiers Incorporating Exogenous Influences on Efficiency*. Discussion Paper No. 9403. Institute of Statistics, Catholic University of Louvain.

Simar, L. and Wilson, P. W. (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, 136(1):31–64.

Simoes, E. J., Byers, T., Coates, R. J., Serdula, M. K., Mokdad, A. H., and Heath, G. W.

(1995). The Association Between Leisure-Time Physical Activity and Dietary Fat in American Adults. *American Journal of Public Health*, 85(2):240–244.

Sklar, A. (1959). Fonctions de repartition à n dimensions et leursmarges. *Publ. l'Institut Stat. l'Université Paris*, 8:229–231.

Stefan, V., van Herpen, E., Tudoran, A. A., and Lähteenmäki, L. (2013). Avoiding Food Waste by Romanian Consumers: The Importance of Planning and Shopping Routines. *Food Quality and Preference*, 28(1):375–381.

Subar, A. F., Freedman, L. S., Tooze, J. A., Kirkpatrick, S. I., Boushey, C., Neuhouser, M. L., Thompson, F. E., Potischman, N., Guenther, P. M., Tarasuk, V., Reedy, J., and Krebs-Smith, S. M. (2015). Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. *The Journal of Nutrition*, 145:2639–45.

Tran, K. C. and Tsionas, E. G. (2015). Endogeneity in Stochastic Frontier Models: Copula Approach without External Instruments. *Economics Letters*, 133:85–88.

Trivedi, B. (2009). *The Calorie Delusion: Why Food Labels are Wrong*. New Scientist.

Tsionas, M. G. and Izzeldin, M. (2018). A novel model of costly technical efficiency. *European Journal of Operational Research*, 268(2):653–664.

United States Court of Appeals (2017). *CORE-MARK INTERNATIONAL, INC., Plaintiff-Appellant, v. THE MONTANA BOARD OF LIVESTOCK, in its official capacity as head of the Montana Department of Livestock; et al., Defendants- Appellees .No. 15-35705. Decided: June 26, 2017.* United States Court of Appeals, Ninth Circuit.

USDA-ERS (2016). *Fluid Beverage Milk Sales Quantities by Product*. U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

Van Donselaar, K. H. and Broekmeulen, R. A. (2012). Approximations for the Relative

Outdating of Perishable Products by Combining Stochastic Modeling, Simulation and Regression Modeling. *International Journal of Production Economics*, 140(2):660–669.

Venkat, K. (2011). The Climate Change and Economic Impacts of Food Waste in the United States. *International Journal on Food System Dynamics*, 2(4):431–446.

Wang, E. Y. (2015). The Impact of Soda Taxes on Consumer Welfare: Implications of Storability and Taste Heterogeneity. *RAND Journal of Economics*, 46(2):409–441.

Wang, H. J. and Schmidt, P. (2002). One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18(2):129–144.

Wilson, N. L., Rickard, B. J., Saputo, R., and Ho, S. T. (2017). Food Waste: The Role of Date Labels, Package Size, and Product Category. *Food Quality and Preference*, 55:35–44.

World Bank (2005). Introduction to Poverty Analyses. In *Poverty Manual*. World Bank Institute.

Yu, Y. and Jaenicke, E. C. (2020a). Estimating Food Waste as Household Production Inefficiency. *American Journal of Agricultural Economics*, 102(2):525–547.

Yu, Y. and Jaenicke, E. C. (2020b). The Effect of Sell-by Dates on Purchase Volume and Food Waste. *Food Policy*.

Appendix

Supplementary Information for Chapter 1

A.1 Equations for Calculating Energy Expenditure

The Revised Harris-Benedict Equations (Roza and Shizgal, 1984):

For Male: BMR=88.362 + 13.397*weight(kg) + 4.799*height(cm) - 5.677*age(year)

For Female: BMR = 447.593 + 9.247*weight(kg) + 3.098*height(cm) - 4.33*age(year)

The Mifflin-St Joer Equations (Frankenfield et al., 2005):

For Male: BMR=5+9.99 *weight(kg) + 6.25*height(cm) - 4.92*age(year)

For Female: BMR = -161 + 9.99*weight(kg) + 6.25*height(cm) - 4.92*age(year)

A.2 Avoidable and Unavoidable Food Waste

As mentioned in Chapter 1, our measure of inputs in practice are the edible portions of food reported in FoodAPS and the food waste estimates represent avoidable waste. Nonetheless, if we choose to use total food acquisitions that include both edible and inedible parts, the waste estimates remain the same.

To see this, suppose that the portion of edible food is $\theta_i \in (0,1]$ for category i and total food acquisition is $\tilde{x}_{i,h}$. Then our previous input measure, the logarithm of edible food, can be written as $\log x_{i,h} = \log \tilde{x}_{i,h} + \log \theta_i$. When the production function is consistently estimated, we can substitute this expression into equation (1.3) of Chapter 1 to obtain a production function in terms of total food acquisitions: $\log y_h = \alpha_0^{new} + \sum_{i=1}^{I} \alpha_i^{new} \log \tilde{x}_{i,h} + \sum_{i=1}^{I} \sum_{j \leq i} \beta_{i,j}^{new} \log \tilde{x}_{i,h} \log \tilde{x}_{j,h}$. It is easy to verify that in this transformed production function, the new constant term α_0^{new} and the first-order coefficients α_i^{new} are different while the coefficients on the second-order interactions $\beta_{i,j}^{new}$ remain the same. However, the intermediate values that lead to food waste calculation, $\hat{A}$ and $\hat{B}_h$, remain unchanged, meaning the percentage waste estimate $\hat{\delta}_h$ is identical to the previous model. To see this, note that $\hat{A}$ is unaffected because $\beta_{i,j}^{new}$ has the same value as $\beta_{i,j}$. As for $\hat{B}_h$, it can be expressed as $\hat{B}_h = \sum_{i=1}^{I} \partial \log y_h / \partial \log x_{i,h}$. Since $\log x_{i,h} = \log \tilde{x}_{i,h} + \log \theta_i$, we have $\partial \log y_h / \partial \log x_{i,h} = \partial \log y_h / \partial \log \tilde{x}_{i,h}$, and hence $\hat{B}_h^{new} = \hat{B}_h$.

The intuition of this result is that the inedible portion does not contribute to producing the output and, hence, has zero marginal productivity. Therefore, when a household is wasting 31% of total food, what it really implies is that the household is wasting 31% of the edible portion. More importantly, because θ_i, the edibility of food category i, does not vary across households, it works as a scale parameter to the input quantities.[1] In other words, the percentage food waste in equation (1.8) of Chapter 1 is input-scale-invariant. We note that this property also depends on our assumption that the percentage waste is the same for all food categories. When modeling each category with a distinct waste measure, the formulation of food waste will be different from equation (1.8) and the scale-invariant property may not hold.

[1]In other words, the result holds when considering the first and the third types of the WRAP avoidability definitions, which represent systemic information. On the other hand, the second type, "possibly avoidable" food waste, is determined by household-specific taste preference and cooking habits which are not explicitly captured by the production function introduced in Chapter 1.

A.3 Likelihood Function in the LIML Estimation

Let us denote $\psi_h = (v_h, \eta_h)$ and assume its distribution as follows:

$$\psi_h \sim N(0, \Omega),\ \Omega = \begin{bmatrix} \sigma_v^2 & \sigma_{v\eta} \\ \sigma_{\eta v} & \sigma_\eta^2 \end{bmatrix} \tag{A.1}$$

The presence of endogeneity corresponds to the case when $\sigma_{\eta v} = \sigma_{v\eta} \neq 0$. The likelihood function is the joint density of ε_h and η_h, which can be derived analytically by change of variables integration. The key assumptions needed to derive this density are independence between η_h and u_h, normality of the distribution, and independence between v_h and u_h (Amsler et al., 2016):

$$\begin{aligned} f_{\varepsilon_h,\eta_h}(\varepsilon_h, \eta_h) &= \text{constant} \cdot \sigma_\eta \cdot \exp\left(-\frac{\eta_h^2}{2\sigma_\eta^2}\right) \cdot \sigma_h^{-1} \\ &\cdot \phi\left(\frac{\varepsilon_h - \mu_{ch}}{\sigma_h}\right) \cdot \Phi\left(-\frac{\lambda_h(\varepsilon_h - \mu_{ch})}{\sigma_h}\right) \end{aligned} \tag{A.2}$$

where $\mu_{c,h} = (\sigma_{v\eta}/\sigma_\eta^2)\eta_h$, $\sigma_h^2 = \sigma_{u_h}^2 + \sigma_{c,h}^2$, $\sigma_{c,h}^2 = \sigma_v^2 - \sigma_{v\eta}^2/\sigma_\eta^2$, and $\lambda_h = \sigma_{u_h}/\sigma_{c,h}$. Next, we can predict the inefficiency term u_h by its mean conditional on ε_h and η_h:

$$\hat{u_h}^{LIML} = E(u_h|\hat{\varepsilon}_h, \hat{\eta}_h) = \hat{\sigma}_h^*[\Lambda(\hat{h}_h) - \hat{h}_h] \tag{A.3}$$

where $\hat{\varepsilon_h}$ and $\hat{\eta_h}$ are residuals from the LIML estimation, $\hat{\sigma}_h^* = \frac{\hat{\sigma}_{u_h}\hat{\sigma}_{c,h}}{\hat{\sigma}_h}$, $\hat{h}_h = \frac{\hat{\lambda}_h}{\hat{\sigma}_h}(\hat{\varepsilon}_h - \hat{\mu}_{c,h})$, and $\Lambda(\hat{h}_h) = \phi(\hat{h}_h)/[1 - \Phi(\hat{h}_h)]$. The percentage food waste is carried out the same way as the baseline model (equation (1.8) of Chapter 1).

A.4 Additional Data Descriptions

Here we provide more details on the FoodAPS data and the food-input variables used in our analysis.

Random-Weights and FAFH Items

For random-weight food products purchased for at-home consumption, the survey participants report either the variable weight or count of the items, e.g., how many apples. In case variable weight is not reported, the total weight is imputed by multiplying the count by standard gram weights in the USDA database.

The FoodAPS data contains the gram weights information for almost all the away-from-home food items (98.3%), either recorded by the households or through data imputation. The data documentation reports that initially, about 51% of the total reported items directly include unit size or amount information. The rest was done by manually reviewing the item description, editing to a standardized unit, matching the information of "quantity" (another data entry) reported by the household, and when not available, assuming size as 1. After determining the unit size of a food item, the items' total gram weights are calculated. Only about 16% of items initially contain information of grams or ounces per unit that is reported along with their unit size information. For the rest, a variety of methods were used by the ERS to impute gram weights. Examples include: (1) using grams information from MenuStat (directly or match to a similar food item); (2) matching similar food items to those reported by TOP restaurants in the Quick Service Restaurant Magazine; (3) using the median grams of the same item in the food-away-from home data of FoodAPS; and (4) matching to the median grams reported in FNDDS and NHANES by food code.

Free or Donated Food

Since we use total energy expenditure as the output, it is important to include all sources of food intakes, including free or donated food. FoodAPS contains a comprehensive list of where the food was acquired. For instance, it reports whether the food item was received from a food bank. For at-home consumption, food acquisitions from food banks represent 0.49% of the whole sample, and the number is 0.13% for away-from-home consumption. In principle, the survey requires households to report all of their food acquisition events, even when the food is not paid for. This includes free school lunches, eating at a friend's place, food from hunting/fishing, vegetables grown on a household's garden, and so on. These "free food" events represent 5.96% of the at-home occasions and 41.96% of the away-from-home occasions. Hence, most of the free events take place outside the households' residence, which makes good sense. Overall, we do not notice strong evidence suggesting that donated or free food is under-reported in FoodAPS.

Storable Food

Another important aspect regarding the input variables is that some categories consist of storable food items. For examples, storable products include beverages, condiments, and some grain products that generally have purchase cycles longer than seven days (Bronnenberg et al., 2008). Given that our data only covers a period of seven days, it is possible that some households purchase more than what they would consume in a week while other households who do not purchase food simply draw from their current inventory. We assume that the occurrence of the inventory fill-ups of storable food is entirely random. This assumption implies that a household's decision to replenish its inventory is independent of other environmental factors, including its participation in the FoodAPS survey. On average, the reported purchase quantities are assumed to reflect the actual consumption.[2]

[2]Mathematically, if actual consumption $\overline{x}_{i,h}$ is smooth, it can be written as $\overline{x}_{i,h} = \Pr(x_{i,h} > 0) \cdot x_{i,h} + [1 - \Pr(x_{i,h} > 0)] \cdot 0$. Then our assumption is equivalent to saying that $\Pr(x_{i,h} > 0) = P_i$ only depends on the food category, not the particular household that is making the decision. Note that this assumption is only a necessary condition for the estimation. The precise condition that guarantees consistency of the parameter estimates depends on complicated mechanisms such as properties of the production function and correlations among $x_{i,h}$'s; therefore, we do not explicitly explore it here.

A.5 Additional Robustness Checks

A.5.1 Dieting Behavior and SNAP Payment

One of the main assumptions of our analysis is that the individuals maintained a state of energy balance during the survey period. It is reported that many people in the United States undergo some diet during the months of January and February (Dashti, 2016). The survey of FoodAPS covers the period from April 19, 2012 to January 22, 2013. Hence majority of the sample falls outside of the January-February time frame. In addition, the survey asked participants to report whether they were on any kind of diet. Among 14,317 participants, 1,853 (12.9%) indicated some dieting behavior. In theory, these individuals' reported food intakes might be less than the energy requirement that is needed to maintain body mass, which artificially makes their estimated productivity more efficient. Hence, we test if the presence of these participants leads to underestimated food waste estimates for their households. For the 733 households that have some members undergoing dieting, the average percentage food waste is 32.9%, higher than the 31.6% estimate for the rest of the sample. And the difference is not statistically significant in a student-t two-sample test. Therefore, at the household level, we do not see an underestimation of food waste for those practicing dieting.

Additionally, we examine whether receiving SNAP benefit payments during or shortly before the survey period would affect the waste estimates. For households that received SNAP payments during the survey period, we observe that their average food expenditure is about $60 higher than other SNAP-enrolled households. This observation raises a concern about whether these higher-than-average food purchase would lead to overestimated food waste. The results show that, despite their higher food expenditures, our models do not produce higher waste estimates for these households . On the contrary, they are predicted to have an average percentage waste about 1.3% lower than the rest of the SNAP households, although not statistically significant. A plausible explanation is that the date (week) of SNAP payment is uncorrelated with other household characteristics, and hence it can be regarded as a random component captured by the error term. As an extra check, we re-estimate the models without households that received payments during the survey (268 observations). The average percentage wastes are within 1% difference of the previous results, and the impacts of three demographic variables are similar. The conclusions hold true if we extend the sub-sample to those received payments one week prior to the survey, in addition to payments during the survey.

A.5.2 Copula Estimation

The estimation procedure starts with the definition of a copula. Let us denote the marginal cumulative distribution functions of x_i's and ε as $U_1 = F_1(x_1)$, $U_2 = F_2(x_2)$, ..., $U_I = F_I(x_I)$, and $U_\varepsilon = F_\varepsilon(\varepsilon)$, where I is the number of food categories and $\varepsilon = v - u$. We also write their joint cumulative distribution function as $F(x_1, x_2..., x_I, \varepsilon)$. Then the copula function $C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon)$ is defined as:

$$C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = \Pr\left(U_1 \leq \xi_1, U_2 \leq \xi_2, ..., U_I \leq \xi_I, U_\varepsilon \leq \xi_\varepsilon\right) \tag{A.4}$$

Therefore, a copula is a distribution function of some uniformly distributed random variables.[3] Moreover, because $U_i \leq \xi_i$ is equivalent to $X_i \leq F_i^{-1}(\xi_i)$, the copula function can be also written as $C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = F(x_1, x_2, ...x_I, \varepsilon)$ for $x_i = F_i^{-1}(\xi_i)$ and $\varepsilon = F_\varepsilon^{-1}(\xi_\varepsilon)$. This relation, called Sklar's theorem, shows that although the copula function has different arguments than the original distribution function of x_i's and ε, they contain the identical information that describes the distributional properties of these variables (Sklar, 1959). This result is central to copula analysis because it guarantees that estimating the copula is equivalent to estimating the original distribution.

There are many analytical forms of the copula function $C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon)$ that could

[3] Note that $F_1(x_1)$, $F_2(x_2)$, ..., $F_I(x_I)$, and $F_\varepsilon(\varepsilon)$ are uniformly distributed.

satisfy equation (A.4). We use a Gaussian copula, which is robust and flexible in most cases, and suitable for higher dimension analysis. Other choices include the Frank copula, the Clayton copula, and the Farlie–Gumbel–Morgenstern copula. A comprehensive treatment of these functional forms and their properties is contained in Nelsen (2006). Consider $\Phi_{R,I+1}$, the $(I+1)$-dimension normal distribution with the correlation matrix $R \in [-1,1]^{(I+1)\times(I+1)}$ as its covariance matrix, and the standard one-dimension normal distribution Φ. The Gaussian copula is specified as the following:

$$C(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = \Phi_{R,I+1}\left(\Phi^{-1}(\xi_1), \Phi^{-1}(\xi_2), ...\Phi^{-1}(\xi_I), \Phi^{-1}(\xi_\varepsilon)\right) \tag{A.5}$$

The copula approach comes in handy when we use its density function, $c(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) = \partial^{(I+1)}C/\partial\xi_1\partial\xi_2...\partial\xi_I\partial\xi_\varepsilon$. For the Gaussian copula, the density is given by:

$$\begin{aligned} c(\xi_1, \xi_2, ..., \xi_I, \xi_\varepsilon) \;&=\; (\det(R))^{-1/2} \cdot \exp\{-\frac{1}{2}[\Phi^{-1}(\xi_1), \Phi^{-1}(\xi_2), ...\Phi^{-1}(\xi_I), \Phi^{-1}(\xi_\varepsilon)]^T \\ &\quad (R^{-1} - \mathbb{I}_{I+1})[\Phi^{-1}(\xi_1), \Phi^{-1}(\xi_2), ...\Phi^{-1}(\xi_I), \Phi^{-1}(\xi_\varepsilon)]\} \end{aligned} \tag{A.6}$$

Denote the joint density function of x_i's and ε as $f(x_1, x_2..., x_I, \varepsilon)$ and the marginal densities as $f_1(x_1)$, $f_2(x_2)$, ..., $f_I(x_I)$, $f_\varepsilon(\varepsilon)$. By applying the chain rule to the Sklar's relation, it can be easily shown that

$$f(x_1, x_2..., x_I, \varepsilon) = c(F_1(x_1), F_2(x_2), ..., F_I(x_I), F_\varepsilon(\varepsilon)) \cdot f_\varepsilon(\varepsilon) \cdot \prod_{i=1}^{I} f_i(x_i) \tag{A.7}$$

The above density is a likelihood function that will be maximized to obtain parameter estimates. In its essence, copula estimation is also a method of maximum likelihood. The difference is that typical applications of maximum likelihood are actually conditional likelihoods with the independence assumption imposed on the explanatory variables and the error term, e.g., we typically maximize $f_\varepsilon(\varepsilon|x) = f_\varepsilon(\varepsilon)$. In fact, if x_i's and ε are independent, we have $c(F_1(x_1), F_2(x_2), ..., F_I(x_I), F_\varepsilon(\varepsilon)) = 1$ and the objective is also to maximize $f_\varepsilon(\varepsilon)$. Note that the copula density function includes a set of additional parameters in the correlation matrix R. These parameters measure the dependence between x_i's and ε, as well as the dependence among x_i's themselves. In practice, we take the logarithm of the density function and maximize the sample sum of $\log c(F_1(x_1), ..., F_I(x_I), F_\varepsilon(\varepsilon)) + \log f_\varepsilon(\varepsilon))$ because the last term $\prod_{i=1}^{I} f_i(x_i)$ does not contain any parameters. The optimization routine involves replacing $F_i(x_i)$'s by their empirical distribution from the sample, and the density function $f_\varepsilon(\varepsilon)$ is the same as equation (1.5) of Chapter 1. The cumulative distribution of the error term $F_\varepsilon(\varepsilon)$ is solved using numerical integration of its density function.

The traditional approaches to tackling endogeneity issues include finding instrumental variables for food inputs and constructing structural models that describe the decision-making mechanism of food choices. Comparing to these approaches, the advantages of using copula are appealing in the context of household productivity and efficiency analysis.

First, the instrumental-variable methods require identifying good instruments that are correlated with the endogenous variables but not with the error term. As discussed earlier, the sources of endogeneity involve a large set of unmodeled household demographic variables as well as market and product-specific characteristics. These variables influence many aspects of household life, which means it is difficult to find instruments that are uncorrelated with them but still correlated with the food choices.

Second, the results of Chapter 1 point out that household food waste amounts to a significant portion of total food purchased (about 30% on average) and the individual-household estimates range from 10% to more than 50%. Therefore the variation in food waste explains a nontrivial part of the variation in purchase quantities. This suggests that even if we find some statistically valid instruments, they are likely to be "weak instruments" (Angrist and Krueger, 1991; Bound et al., 1995).[4]

Third, both the instrumental-variable and the structural approaches require us to in-

[4] In Richards et al. (2012), instruments are constructed as linear combinations of the same variables from other households in the data, which are likely to be strong instruments. Nonetheless, this method faces the challenge of estimating 54×54 covariance matrix mentioned in the following.

troduce additional equations for the endogenous variables. In our case, what needs to be estimated is a system of 54 additional equations (9 food inputs and their interactions) and a 54×54 covariance matrix, which is a very difficult computational task.[5] On the other hand, copula estimation does not require additional equations but only estimating a 10×10 correlation matrix whose diagonal elements are already constrained at 1 (see equations (A.6) and (A.7)).

Finally, structural approaches typically need us to specify an optimization objective such as utility maximization or expenditure minimization. Modeling such mechanisms requires extra decision-making assumptions and additional data on market and product-specific information which are beyond the scope of this chapter.

A.5.3 Contextual Variables

Here we discuss the recent literature on incorporating contextual variables into stochastic frontier models and the challenges in obtaining consistent parameter estimates. We also specifically test whether it is appropriate to add more food-waste determinants as additional contextual variables in our model.

Earlier studies have used a simple two-step procedure in which contextual variables are omitted in the first step. The second step involves regressing the estimated output inefficiency $\hat{u}_h$ on the contextual variables. This procedure has been criticized as they often generate biased estimates (Simar et al., 1994; Simar and Wilson, 2007; Banker and Natarajan, 2008; Banker et al., 2019). Simar and Wilson (2007) construct a consistent second-stage estimation method that is based on a truncated regression model and its underlying data-generating process. Banker and Natarajan (2008) derive specific conditions for a second-stage ordinary least squares (OLS) estimation to be consistent. However, these two-step methods usually

[5]It is possible to construct a structural system using only 9 additional equations, for instance, the first-order conditions of the 9 inputs. But this approach requires estimating a proper Jacobian transformation of the production equation so that the second-order translog specification is written in a first-order relation (see Malikov et al. 2016). Amsler et al. (2016) also show a control function approach and a LIML approach that would only require 9 instruments. However, these approaches typically impose stronger assumptions, e.g., the endogenous variables are not correlated with the inefficiency u but only with the white noise v.

require stronger assumptions that are unlikely to be met in our study. For instance, Banker and Natarajan (2008) add the assumption of independence between d_h and $x_{i,h}$'s, which does not hold for many household-specific variables as we have already mentioned.

Scholars have identified one particular class of specification that follows a one-step procedure and also allows for the correlation between d_h and $x_{i,h}$'s. This type of specification satisfies the "scaling property" such that the output inefficiency can be written as $u_h = h(d_h) \cdot u^*$, where $h(d_h) > 0$ is a scaling function and u^* is a systemic output inefficiency term whose distribution is independent of $h(d_h)$ (Wang and Schmidt, 2002; Alvarez et al., 2006). Under this specification, d_h is allowed to be correlated with $x_{i,h}$'s, which is similar to the typical heteroskedastic error terms seen in least-squares estimations.[6]

Nonetheless, two problems remain to be solved under the scaling property condition. First, it is possible that food inputs $x_{i,h}$'s are correlated with the random variable u^*. In this case, we still need to treat this as an endogenous-variable problem. Second, perhaps more importantly, the scaling property itself is an assumption imposed on the structure of output inefficiency. Note that our half-normal specification $u_h \sim N^+(0, \sigma_h^2(d_h))$ satisfies the scaling property because it is equivalent to $u_h = \sigma_h(d_h) \cdot u^*$, where $u^* \sim N^+(0, 1)$. However, adding more contextual variables into d_h implicitly makes the scaling property a harder condition to meet. This can be tested using hypothesis testing methods outlined in Alvarez et al. (2006).

Following their procedure, we re-estimate our stochastic frontier models with a more general specification: $u_h \sim N^+(\mu, \sigma_h^2(d_h))$, a truncated normal distribution. We test if the truncation parameter $\mu = 0$, which corresponds to the scaling property condition. When d_h only contains the previous three demographic variables, we estimate $\hat{\mu} = -0.07$. The Wald test on $\mu = 0$ shows a $\chi^2(1)$ statistic at 0.44 with a p-value 0.5073. Therefore we are far from rejecting the scaling property, suggesting that even the three variables we use in d_h may correlate with food inputs, our estimates are still consistent. On the other hand, if we add either household size or SNAP benefit, the estimated $\hat{\mu}$ are larger than 5 with very large

[6] As Simar et al. (1994) show, with the scaling property, stochastic frontier models can be estimated by nonlinear least squares.

$\chi^2(1)$ statistics in the Wald test, and the p-values are practically zero. In short, adding more contextual variables in our model will most likely violate the scaling property, which makes the food waste estimates biased.

A.6 Notes on Data Access and Replication

We use the National Household Food Acquisition and Purchase Survey (FoodAPS data) via a third party agreement with the U.S. Department of Agriculture's Economic Research Service. Recently, a public version of FoodAPS has been provided by the ERS (available at https://www.ers.usda.gov/foodaps). However, the public data does not contain information on individual height and weight, which are needed to calculate basal metabolic rate and to impute physical activity levels. For researchers interested in replicating our results using the public-use data, we suggest an alternative output measure based on the Body Mass Index (BMI), which is reported in the public data. The output is calculated as a sum of household members' age-weighted BMI and gives very close estimates to our original results. We suggest assigning the actual ages to each household member under 18 years old; and for the adults, setting the weights at 18 regardless of ages. The rationale of this alternative output measure is related to the physiological outcome of consuming food, whereas age and weights are intended to capture the intra-household allocation. The second dataset used, the 2011-2012 National Health and Nutrition Examination Survey (NHANES) is a publicly available dataset published by the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services (https://www.cdc.gov/nchs/nhanes).

A.7 Tables

Table A.1: Summary Statistics–Continuous Variables

	Mean	S.D.	5% percentile	95% percentile
d_1, income	1.716	1.393	0.345	4.611
$\widetilde{PA}_h$, employment status	0.649	0.291	0.250	1.000
z_h, weekend shopping frequency	0.294	0.343	0.000	1.000
$\widehat{PA}_{m,h}$, imputed physical activity	1.605	0.149	1.356	1.849

Note: Income is household monthly total income divided by adult equivalent household size, in thousand dollars. Employment Status: for each working-age household member, 1= not working; 2= looking for work; 3= with a job but not at work; 4= working. Weekend Shopping Frequency is the percentage share of household shopping trips that occurred during weekends.

Table A.2: Summary Statistics–Dietary Healthfulness

	1	2	3	4	Total
Frequency	1093	1730	986	263	4072
Percentage	26.84%	42.49%	24.21%	6.46%	100%

Note: First row: higher values represent healthier diet.

Table A.3: Summary Statistics–Food Security

	Low	Medium	High	Total
Frequency	1105	798	2169	4072
Percentage	27.14%	19.60%	53.27%	100%

Table A.4: Summary Statistics–Household Size

	1	2	3	4	5	6	7+	Total
Freq.	904	1199	726	629	347	151	116	4072
Pct.	22.19%	29.43%	17.82%	15.44%	8.52%	3.71%	3.31%	98.38%

Note: First row: number of household members.

Table A.5: Summary Statistics–Education

	1	2	3	4	5	Total
Frequency	361	1243	2707	2513	1671	8495
Percentage	4.25%	14.63%	31.87%	29.58%	19.67%	100%

Note: For each individual of age 20 and above, value ranges from 1 to 5, representing different levels of highest degrees, with 1= up to 9th degree, 2= 9-11th grade, 3=high school, 4=associate degree, 5=college graduate or above. The variable is normalized to 0-1 in the estimation.

Table A.6: First-Stage Regression on NHANES

	Age<20	Age>=20
Weight	-0.0007	-0.0003
	(0.0006)	(0.0003)
Height	0.0036**	0.0025***
	(0.0016)	(0.0008)
Age	-0.0107**	-0.0038***
	(0.0050)	(0.0003)
Male	0.0963***	0.0738***
	(0.0248)	(0.0141)
Hispanic	0.0262	-0.0338**
	(0.0252)	(0.0148)
Non-Hispanic White	0.0857***	0.0248**
	(0.0265)	(0.0115)
Survey Period (Nov-Apr)	-0.0145	-0.0010
	(0.0211)	(0.0103)
Health Status	0.0293**	0.0291***
	(0.0126)	(0.0058)
Employment Status		0.0649***
		(0.0160)
Education		0.1066***
		(0.0245)
Income		0.0010***
		(0.0004)
Marital Status		-0.0460***
		(0.0108)
Presence of Children		-0.0036
		(0.0119)
Smoking		-0.0348**
		(0.0136)
Constant	1.1905***	1.1506***
	(0.2246)	(0.1260)
Adjusted R-squared	0.0601	0.1564
Number of Observations	990	3581

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.7: Food Waste Determinants–Calorie Contents

	Baseline	Proxy-instrument	Fitted $PA_{m,h}$
Income	0.3534***	0.3983***	0.3214***
	(0.0682)	(0.0603)	(0.0704)
Healthy Diet	1.0068*	0.7319	1.1140*
	(0.5507)	(0.4847)	(0.6404)
Food Security	1.4681	2.0207**	1.7937
	(0.9540)	(0.9801)	(1.2280)
Constant	-5.7880***	-6.1601***	-6.1472***
	(1.6364)	(1.3552)	(2.0000)
Number of Obs.	4072	3465	4049

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.8: Food Waste Determinants–Mifflin-St Joer

	Baseline	Proxy-instrument	Fitted $PA_{m,h}$
Income	0.3311***	0.3852***	0.2954***
	(0.0611)	(0.0568)	(0.0624)
Healthy Diet	1.5278***	1.1352**	1.7230**
	(0.5857)	(0.4914)	(0.6936)
Food Security	1.7733*	2.4579**	2.2641
	(1.0774)	(0.1.1289)	(1.4744)
Constant	-6.2860***	-6.7580***	-6.8742***
	(1.7442)	(1.4847)	(2.2235)
Number of Obs.	4072	3465	4049

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.9: Food Waste Determinants–Sum of Body Weights

Description	Baseline	Proxy-instrument
Income	0.3143***	0.3387***
	(0.0621)	(0.0563)
Healthy Diet	1.0630***	1.7310***
	(0.6555)	(0.5064)
Food Security	1.4652*	2.0195**
	(0.8944)	(0.9221)
Constant	-6.2000***	-6.5460***
	(1.6364)	(1.3552)
Number of Obs.	4072	3465

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.10: Full Estimation Results

Description	Baseline	Proxy-IV	Fitted $PA_{m,h}$
Production Equation			
α_1, (Milk and Dairy)	-0.0719***	-0.0913***	-0.0770***
	(0.0136)	(0.0152)	(0.0145)
α_2, (Protein Foods)	-0.0250*	-0.01518	-0.02236
	(0.0144)	(0.0167)	(0.0154)
α_3, (Mixed Dishes)	-0.0937***	-0.0994***	-0.1054***
	(0.0142)	(0.0175)	(0.0152)
α_4, (Grains)	-0.0706***	-0.0664***	-0.0778***
	(0.0147)	(0.0170)	(0.0157)
α_5, (Snacks)	-0.0228	-0.0160	-0.0238
	(0.0145)	(0.0169)	(0.0155)
α_6, (Fruit and Vegetables)	-0.0276**	-0.0410**	-0.0245*
	(0.0141)	(0.0165)	(0.0151)
α_7, (Beverages)	-0.0671***	-0.0809***	-0.0705***
	(0.0128)	(0.0152)	(0.0137)
α_8, (Condiments)	0.0114	0.0132	0.0136
	(0.0138)	(0.0168)	(0.0148)
α_9, (Infant formula & Uncoded)	-0.0021	0.0067	-0.0094
	(0.0388)	(0.0410)	(0.0414)
$\beta_{1,1}$	0.0100***	0.0119***	0.0107***
	(0.0012)	(0.0015)	(0.0013)
$\beta_{2,1}$	-0.0015	-0.0013	-0.0015
	(0.0011)	(0.0012)	(0.0012)
$\beta_{2,2}$	0.0030*	0.0032*	0.0027*
	(0.0016)	(0.0017)	(0.0017)
$\beta_{3,1}$	-0.0006	-0.0009	-0.0008
	(0.0011)	(0.0013)	(0.0012)
$\beta_{3,2}$	0.0015	0.0025	0.0017
	(0.0014)	(0.0016)	(0.0015)
$\beta_{3,3}$	0.0153***	0.0144***	0.0171***
	(0.0015)	(0.0018)	(0.0016)
$\beta_{4,1}$	0.0006	0.0004	0.0005
	(0.0010)	(0.0010)	(0.0010)
$\beta_{4,2}$	0.0012	0.0016	0.0012
	(0.0013)	(0.0015)	(0.0014)
$\beta_{4,3}$	0.0018	0.0009	0.0019
	(0.0014)	(0.0015)	(0.0014)
$\beta_{4,4}$	0.0065***	0.0072***	0.0069***
	(0.0016)	(0.0016)	(0.00167)

$\beta_{5,1}$	0.0017*	0.0016	0.0019*
	(0.0010)	(0.0012)	(0.0011)
$\beta_{5,2}$	-0.0016	-0.0025	-0.0019
	(0.0014)	(0.0016)	(0.0015)
$\beta_{5,3}$	0.0002	0.0019	0.0001
	(0.0014)	(0.0015)	(0.0013)
$\beta_{5,4}$	0.0004	0.0009	0.0005
	(0.0016)	(0.0014)	(0.0013)
$\beta_{5,5}$	0.0008	0.0024	0.0008
	(0.0015)	(0.0017)	(0.0016)
$\beta_{6,1}$	0.00004	-0.00056	0.0001
	(0.0012)	(0.0014)	(0.0013)
$\beta_{6,2}$	0.0019	0.0012	0.0019
	(0.0014)	(0.0016)	(0.0015)
$\beta_{6,3}$	0.0013	0.0000	0.0017
	(0.0014)	(0.0018)	(0.0015)
$\beta_{6,4}$	-0.0015	-0.0013	-0.0013
	(0.0015)	(0.0017)	(0.0016)
$\beta_{6,5}$	-0.0005	-0.0013	-0.0004
	(0.0015)	(0.0017)	(0.0016)
$\beta_{6,6}$	0.0031**	0.0050***	0.0023
	(0.0015)	(0.0016)	(0.0016)
$\beta_{7,1}$	0.0005	0.0010	0.0005
	(0.0011)	(0.0013)	(0.0013)
$\beta_{7,2}$	0.0010	0.0011	0.0011
	(0.0014)	(0.0017)	(0.0015)
$\beta_{7,3}$	-0.0014	-0.0016	-0.0013
	(0.0013)	(0.0016)	(0.0014)
$\beta_{7,4}$	0.0013	0.0010	0.0015
	(0.0014)	(0.0016)	(0.0015)
$\beta_{7,5}$	0.0013	-0.0005	0.0014
	(0.0014)	(0.0015)	(0.0014)
$\beta_{7,6}$	0.0001	0.0023	-0.0001
	(0.0015)	(0.0019)	(0.0016)
$\beta_{7,7}$	0.0070***	0.0070***	0.0075***
	(0.0012)	(0.0017)	(0.0013)
$\beta_{8,1}$	-0.0008	-0.0003	-0.0008
	(0.0009)	(0.0009)	(0.0009)
$\beta_{8,2}$	-0.0017	-0.0037***	-0.0020
	(0.0013)	(0.0014)	(0.0013)
$\beta_{8,3}$	-0.0015	-0.0002	-0.0015
	(0.0013)	(0.0014)	(0.0014)
$\beta_{8,4}$	-0.0004	0.0009	-0.0002
	(0.0011)	(0.0012)	(0.0012)
$\beta_{8,5}$	0.0009	0.0004	0.0009

	(0.0011)	(0.0012)	(0.0012)
$\beta_{8,6}$	0.0007	-0.0000	0.0006
	(0.0013)	(0.0015)	(0.0014)
$\beta_{8,7}$	0.0005	0.0005	0.00067
	(0.0014)	(0.0015)	(0.0015)
$\beta_{8,8}$	-0.0007	0.0005	-0.0011
	(0.0013)	(0.0017)	(0.0014)
$\beta_{9,1}$	-0.0001	0.0006	-0.0001
	(0.0016)	(0.0016)	(0.0017)
$\beta_{9,2}$	-0.0047*	-0.0046	-0.0051*
	(0.0029)	(0.0029)	(0.0031)
$\beta_{9,3}$	-0.0038*	-0.0029	-0.0039
	(0.0023)	(0.0025)	(0.0025)
$\beta_{9,4}$	-0.0017	-0.0022	-0.0022
	(0.0023)	(0.0023)	(0.0025)
$\beta_{9,5}$	-0.0021	-0.0028	-0.0024
	(0.0023)	(0.0024)	(0.0025)
$\beta_{9,6}$	0.0019	0.0062	0.0024
	(0.0029)	(0.0032)	(0.0030)
$\beta_{9,7}$	0.0079***	0.0062*	0.0084***
	(0.0030)	(0.0032)	(0.0032)
$\beta_{9,8}$	0.0028	0.0020	0.0035*
	(0.0019)	(0.0020)	(0.0020)
$\beta_{9,9}$	-0.0021	-0.0010	-0.0018
	(0.0039)	(0.0040)	(0.0041)
α_{PA}, (Employment Status)		0.2510	
		(0.2049)	
α_0, (Constant)	8.1370***	8.3356***	8.5472***
	(0.0826)	(0.2272)	(0.0874)
White Noise σ_v^2	0.4845***	0.2255***	0.5198***
	(0.0080)	(0.0184)	(0.0078)
Inefficiency $\log \sigma_{u_h}^2 = \gamma_0 + \gamma' d_h$			
Income	0.3458***	0.3980***	0.3102***
	(0.0601)	(0.0572)	(0.0617)
Healthy Diet	1.5147***	1.0986*	1.7552***
	(0.5871)	(0.5050)	(0.7000)
Food Security	1.9550*	2.6549**	2.6025
	(1.1395)	(1.2281)	(1.6317)
Constant	-6.5729***	-7.0176***	-7.372***
	(1.7434)	(1.5743)	(2.3028)
Number of Obs.	4072	3465	4049

Vita

Yang Yu

Education

Ph.D., Agricultural, Environmental and Regional Economics, Penn State University, 2020.

M.A., Economics, Boston University, 2013.

B.A., Economics, Wuhan University, 2010.

Research Interests

Food Economics, Industrial Organization, Applied Econometrics, Quantitative Methods

Publications

Yu, Y. and E. C. Jaenicke. 2020. “The Effect of Sell-by Dates on Purchase Volume and Food Waste.” Food Policy.

Yu, Y., and E. C. Jaenicke. 2020. “Estimating Food Waste as Household Production Inefficiency,” American Journal of Agricultural Economics 102(2): 525-547.

Working Papers

Yu, Y. “Optimal Government Subsidy for Achieving Healthy Eating.”

Yu, Y. “Rational Food Waste and Consistent Estimation of Consumer Demand.”

Work in Progress

"Calibrating Actual Consumption in Demand Models: A Toolkit."

"Endogenous Shelf-Lives of Perishable Products: Market Power and Regulation."

Awards and Honors

Outstanding Dissertation Award, College of Agricultural Sciences, Penn State, 2019.

Travel Grant, Agricultural and Applied Economics Association, 2019.

Graduate Student Development Funds, Department of Agricultural Economics, Sociology, and Education, Penn State, 2019.

Graduate Student Travel Award, College of Agricultural Sciences, Penn State, 2019.

Conference Scholarship, Northeastern Agricultural and Resource Economics Association, 2017, 2018.

Thomas C. and Jackie M. Floore Memorial Scholarship, College of Agricultural Sciences, Penn State, 2016.