The Effect of Online Shopping on Grocery Demand

Katherine A. Harris

Michigan State University

October 2018

Abstract

This paper utilizes novel household panel data to analyze the effect of online grocery shopping on grocery demand. In order to obtain a causal estimate of the impact of online grocery shopping on purchasing decisions, I utilize variation in the timing that an online shopping service was introduced as a source of exogenous variation in the decision to shop online. Local average treatment effects indicate that online shopping induces a 3.8, 5.9, 5.7 and 7.4 percent increase in the average budget shares for dairy, fruit, meats and vegetables, respectively. This reallocation of funds comes at the expense of drinks, oils and snacks/sweets with estimates indicating a 5.2, 4.1 and 13.6 percent decrease in the average budget shares, respectively. I further explore how shopping online influences grocery purchases by estimating a formal model of grocery demand. Comparisons of in-store and online price elasticities indicate that households are generally less price sensitive when shopping online. Specifically, I find that own-price (cross-price) elasticities are two (three) times larger in-store than they are online, on average. These insights into consumer purchasing behavior can be utilized to inform policies aimed at improving the nutritional quality of food purchases and can also inform optimal web design and online pricing strategies.

JEL: D12, L81, I12
Keywords: retail, e-commerce, consumer choice, nutrition

[0]Special thanks to Mike Conlin, Todd Elder, Jeff Wooldridge and Bob Myers for their helpful comments and guidance throughout my work on this project. I would also like to thank the retailer for providing the data and the many employees who answered numerous questions and provided invaluable insights into the purchasing environment. Additional thanks are extended to Emek Basker, participants of the UM-MSU-WU Labor Day conference and the NAREA annual meetings for their helpful comments and suggestions.

1 Introduction

"Plus, since I wasn't at the store I stuck to my list and didn't give into those random, impulse purchases [...]"

- Customer Review of Online Grocery Experience, April 2016

Online grocery purchases amounted to $20.5 billion dollars in sales and represented 4.3 percent of all groceries purchased in 2016 (Nielsen & FMI). In recent years, online grocery shopping has become more and more commonplace. Many brick and mortar retailers have begun to offer online grocery services and web based retailers, such as Amazon, have also entered the online grocery market. Due to increasing accessibility and consumer adoption of online grocery services, it is projected that within the next ten years 20 percent of all grocery purchases will be conducted online (Nielsen & FMI). Despite the large predicted growth rate, there has been relatively little research that examines how online purchasing environments influence consumer choice.

This paper evaluates how shopping for groceries online affects demand. In order to isolate the effect of an online shopping environment, I utilize grocery scanner data generated from the purchases of 34 thousand households, 25 thousand of whom have utilized an online shopping service. These data provide an attractive setting to study the effect of online grocery shopping for three reasons: first, the panel structure of the data allows for a within household comparison of purchases across the two environments; second, online and in-store purchases are fulfilled by the same retailer, alleviating concerns over differences in product selection and branding; third, the retailer of this study offers products for purchase online at the same prices as those found in the store.[1]

In a reduced form analysis, I employ panel difference-in-differences and two-stage least squares estimation strategies that utilize variation in the time the online grocery service became available, at different store locations, as a source of exogenous variation in the decision to shop online. I find that when ordering groceries online, households begin

[1]This is true only for the online grocery service featured in this study. The retailer of this study also has an online shopping delivery service. The prices when shopping via the delivery service are different than the in-store prices.

to allocate a larger share of their grocery budget toward product categories that generally contain healthier items (dairy, fruit, meats and vegetables) at the expense of product categories that generally contain more indulgent products (drinks, oils and snacks/sweets). Specifically, I estimate local average treatment effects that indicate a 3.8, 5.9, 5.7 and 7.4 percent increase in the average budget shares for dairy, fruit, meat and vegetables. Additionally, this reallocation of funds comes at the expense of drinks, oils and snacks/sweets with estimates indicating a 5.2, 4.1 and 13.6 percent decrease in average budget shares, respectively. Motivated by these findings, I estimate a formal model of grocery demand and find that households are less price sensitive when shopping online. I find that own-price elasticities and cross-price elasticities are, on average, two and three times larger in-store than they are online, respectively.

The findings of this paper are consistent with behavioral theories that suggest consumers have difficulty exercising self-control due to time inconsistent preferences (Thaler 1981, Laibson 2004), visceral influences (Loewenstein 1996), and/or consumption cues (Laibson 2001). Theories on time inconsistent preferences predict that the decisions consumers make for themselves in the future are better than the decisions they make for themselves in the present. Thus, the time delay between ordering and receiving groceries, that exists when shopping online, could lead to more healthful purchases. Visceral influences and cue theories of consumption indicate that as the level of distraction (noise, congestion, presence of children) and the level of product placement (checkout lanes, end of aisle displays) declines in a shopping environment, a consumers' ability to exercise self-control may increase. If the online shopping experience is less distracting than the in-store shopping experience, households may be able to exercise more self-control over their purchases.[2]

Policy makers can leverage the insights into consumer behavior that this paper provides to help consumers make decisions that are better aligned with their long term health goals. For example, one of the main findings of this paper is that in-store product place-

[2]The representation of products with pictures has also been theorized to improve the healthfulness of food purchases (Shiv and Fedorikhin 1999; Shiv and Fedorikhin 2002).

ment is an effective tool for influencing consumer purchases. Hence, policy makers interested in improving the healthiness of food purchases could potentially do so by implementing policies that optimally place healthier products throughout a store or in an online shopping environment.[3] Additionally, differences in price elasticities across the two purchasing environments suggest that firms may be able to increase online revenues by implementing a more sophisticated pricing strategy in the online environment.

This paper contributes to a growing literature that analyzes how an online purchasing environment, in and of itself, may influence the healthiness of consumer purchases. Huyghe et al. (2017) utilize panel data for households shopping online and in-store at the same European retailer and find that expenditure shares for unhealthy products are lower in online shopping trips relative to in-store shopping trips. However, Huyghe et al. do not address the endogeneity of the decision to shop online and their data is limited to a four month observation period over a restricted set of product categories (salty snacks, chips, chocolate, candy bars and sweets and chewing gum).[4] Milkman, Rogers, and Bazerman (2009) test whether increased delay in delivery improves the healthiness of grocery purchases utilizing online grocery orders generated from a panel of households. They find that the share of "should" items (vegetables and fruit) in an online grocery order increases the further in advance the order is placed relative to delivery. However, it is possible that the circumstances in which a consumer places an order far in advance of delivery are correlated with product choice; thus, a limitation of their work is that these findings may also not be causal.[5] In contrast, this paper utilizes a difference-in-differences and instrumental variables framework to estimate a causal effect of online grocery shopping on the healthiness of consumer purchases.

This paper also contributes to studies that have evaluated structural differences between online and in-store demand. Pozzi (2012) analyzes how online shopping lists and

[3]This type of policy could be implemented at little or no cost to the retailer.

[4]In order to address these limitations, Huyghe et al. run an experiment that randomizes the purchasing environment each participant experiences and find further evidence to suggest that when consumers face pictures of products they are less likely to purchase indulgent items.

[5]For example, buying groceries in advance of an event at which you plan to have a specific meal prepared.

the convenience of being able to pre-select your grocery basket influences decisions over brand choice within breakfast cereals. He finds that brand exploration is systematically more prevalent in-store than online and quantifies how features of the ordering website (favorites lists and difficulty of quality verification) contribute to decreased brand exploration behavior.[6] Chu et al. (2008) also study how purchasing patterns vary across in-store and online grocery shopping environments fulfilled by the same retailer. They find that households are more brand loyal, more size loyal and less price sensitive when they shop for groceries online.[7] Pozzi (2012, 2013b) and Chu et al. (2008, 2010) provide valuable insight into how and why purchasing behavior is different online; however, they only shed light on decision making processes over very narrowly defined product spaces (cereal, drinks, laundry detergent, frozen pizza, etc.). In contrast, this paper estimates structural changes in demand over a more broadly defined product space that incorporates all grocery products; analysis at this level of aggregation effectively analyzes a different set of substitution patterns than those that have been previously studied.

The remainder of this paper is structured as follows: Section 2 discusses the online purchasing service and the different ways in which an online shopping environment might influence demand. Section 3 describes the data. Section 4 presents the reduced form estimation strategy and results. Section 5 discusses robustness checks for the reduced form estimation. Motivated by the reduced form findings, the paper then estimates a more formal model of grocery demand; Section 6 presents the Linear Approximate Almost Ideal Demand System (LA/AIDS) for grocery demand and discusses the estimated

[6] Pozzi (2013b) utilizes a reduced form approach to evaluate how online convenience influences purchases over products that are inconvenient to transport, namely, bulk products. He finds that when households use an online grocery service, they increase their monthly share of expenditure for bulk items in laundry detergent and soda by 30 and 80 percent, respectively. Pozzi attributes this finding to the decreased transportation cost of purchasing bulk products online; however, as Pozzi notes, these estimates could overstate the magnitude because there is evidence to suggest that households change their mix of retailers when shopping online. That is, households may not actually be buying more bulk items then before; they may just be buying more bulk items from the retailer featured in the study than they did before.

[7] The authors do not identify the mechanism that explains this finding but suggest that time constraints, online shopping list creation, the presence of non-price information for products (nutrition facts), lack of online competition and convenience could be contributing to this result. Furthermore, in a follow-up paper (Chu et al., 2010), the authors find that the magnitude of the effect of online shopping on brand loyalty, size loyalty and price sensitivity varies with household and product characteristics.

price elasticities. Lastly, Section 7 concludes.

2 In-Store and Online Purchasing Environments

The supermarket chain featured in this study offers grocery products as well as a large variety of general merchandise items. Over the past couple of years, this retailer has begun to introduce an online shopping service which allows customers, for a small convenience fee, to select their groceries online, choose an appointment window with their local store, and pick-up and pay for their groceries at a designated "drive-through."[8] The wait time to pick up groceries depends on the size of the order and the volume of orders the retailer receives at the time of the order.[9]

Over the course of two and a half years, thirty-three store locations introduced the online purchase service. The online service was first introduced in March 2015 and was slowly rolled out to additional stores following the initial introduction of the program. Figure 1 illustrates the proportion of households that had the online shopping service available to them over time.[10] In March 2015, roughly 20% of households have access to the online shopping service. This proportion increases over time as more stores begin to

[8]The convenience fee varies by the location but is between $5 to $10 per online shopping occasion. This convenience fee changed for some stores over the time period of this study.

[9]Unfortunately, I do not have access to the average wait time in my data; however, through personal experience, it seems as if same day pick-up is probable (if you place an order in the morning) and next day pick-up is very likely. There is one idiosyncrasy of the online shopping environment that is worth noting. First, shoppers are not able to use paper coupons when they shop online, but they are allowed to use digital coupons. Paper coupon offerings are primarily composed of the coupons that print when the customer checks out at the store. According to the retailer, they rarely publish paper coupons in their weekly ads and paper coupons are rarely used.

[10]I constructed the date it was available to a given household based on the stores the household visited in the six months prior to any store having the service available (i.e. Septemer 2015-February 2015). After constructing the store footprint for each household in the six months prior to introduction, I then assigned each household an availability date based on the first store (within their pre-online service footprint) that offered the online purchasing service. Roughly three thousand in-store households and three thousand online households did not visit a store in the six months prior to introduction that later introduced the online purchasing service. Since I cannot assign these households an availability date according to the definition of availability outlined above, these households have been dropped from the main estimation results of this paper. However, Table A3 of the appendix presents estimation results that includes these households by changing the definition of online availability to be based off of the entire store footprint of the household. These results illustrate that the main findings of this paper are not sensitive to changes in the definition of online service availability.

offer the service and by March 2017, all of the households in my sample have access to the online shopping service.

The gradual roll-out of the online service lends itself nicely to a difference-in-differences framework, where the treatment group are households that have the service available to them in year-month m and the control group is the set of households for whom the service is not yet available in year-month m. In order to correctly employ this estimation strategy, I restrict the time periods of my data so that there is always a control group of households who have not yet received access to the online shopping service.[11] Explicitly, I only use data prior to October 2016, the month in which the last group received access to the online shopping service.

The identifying assumption in this framework is that the service was made available to different households at different times for reasons that are independent of consumer demand. I believe this to be a valid assumption for two reasons: (1) the first location chosen to pilot this service was close to the corporate headquarters, where it was presumably easiest to manage; and (2) the ability of a location to provide this service is highly dependent on the existing infrastructure of the store. In order to effectively implement this program a location needs to have a designated space to stage groceries for customer pickup and a convenient entrance for employees to exit and re-enter when delivering groceries to customers' cars. I explore the validity of my identifying assumption further in Section 3.

One of the most predominant differences between the two purchasing environments is search. Sales filters and price sorting options featured in the online purchasing environment reduce the cost of price comparison, which may lead consumers to have heightened price sensitivities when shopping online (Bakos 1997; Brynjolfsson and Smith 2000; Clemons, Hann, and Hitt 2002; Ellison and Ellison 2005).[12] On the other hand, product recommendations and customer favorite's lists can also reduce shopping costs by elimi-

[11] Borusyak & Jaravel (2016) show that event study estimates suffer from under identification and negative weighting when all units or groups are treated.

[12] Additionally, decreased search costs could also motivate consumers to seek out products that they have previously never purchased (Brown and Goolsbee 2002).

nating the need for search. The online favorites list allows the customer to instantly add the items they frequently purchase to their order on each shopping occasion. However, time savings come at the expense of price comparison; thus, features of website design could lead to decreased price sensitivity in the online shopping environment.

Online search functions and product recommendations change the way consumers "browse" when they are online relative to the in-store purchasing environment. While the in-store search path (or browsing experience) is dictated by the physical layout of the products in the store, the online search function generally does not impose a specific search path on the consumer.[13] The time and effort that brick and mortar retailers put into product displays, store and web design suggests that search paths play an important role in nudging customers towards purchases.[14] For example, the absence of a checkout lane when shopping online is likely to lead to decreased purchases of candy bars, mints and gum; additionally, if you are less likely to be hungry when shopping online we may expect to see less hunger driven impulse purchases. Search differences between the online and in-store purchasing environments could lead to less unplanned purchasing when shopping online.

Beyond differences in search, there are many other elements of the online purchasing experience that could influence consumer demand. For example, time delays between the point of purchase and actual receipt of the goods could also lead to differences in consumer demand across the two purchasing environments. A multiple selves framework in which our long-term selves value "should" products and our short-term selves value "want" products predicts that shoppers might purchase more healthful foods when shopping online simply because they are receiving the goods further in the future then they would if they were in the store.[15] Additionally, the valuation of goods that are generally

[13]Customers can search for products either by using the online search bar or by clicking through a hierarchy of product categories. According to the retailer, the search bar is the most popular form of search in the online purchasing environment.

[14]Laibson (2001) indicates that the placement of products in checkout lanes can be interpreted as a "cue" that increases the marginal utility of consumption when an individual is exposed to it. According to this theory, in the absence of the cue, we would expect to see different consumption decisions being made and hence different purchasing decisions being made.

[15]For more on multiple selves theory see the work of Schelling (1984), Bazerman et al. (1998) and Thaler and Shefrin (1981).

consumed immediately after purchase will likely decrease in the presence of time delays.

It could be difficult for consumers to verify the quality of a product when shopping online due to the inability to physically inspect it. An inability to verify product quality is likely to lead to decreased brand exploration when shopping online. Shiv and Fedorikhin (1999, 2002) suggest that symbolic product representation creates sensory distance, which decreases a product's vividness and makes immediate gratification less important.[16] Hence, in the online shopping environment, households may be less tempted to purchase indulgent products simply because they are represented by pictures rather than by the physical products themselves.

The literature discussed above generates predictions about how online demand should differ from in-store demand. This paper explicitly tests two of these predictions. First, differences in search, timing and the representation of products suggest that households may be less likely to make unhealthy purchases when shopping online. Second, differences in search costs and convenience tools (online favorites lists and recently purchased items lists) suggest that price elasticities will differ across the two purchasing environments. However, reductions in search costs due to the ease of price comparison online suggest heightened price sensitivities when shopping online, while differences in convenience tools suggest decreased price sensitivities when shopping online. Hence, which element of the online purchasing environment influences price elasticities more remains an empirical question.

3 Data & Summary Statistics

I have access to household level purchasing data at the day, store, universal product code (UPC) level before and after the introduction of the online purchasing service. Within these data, I have the entire purchasing history (over grocery products) for roughly 130 thousand households from September 2014 through March 2017. This sample of house-

[16]Other research on this topic includes the following papers: Hoch and Loewenstein (1991); Loewenstein (1996); Mischel and Ebbesen (1970).

holds was constructed based on two criteria: (1) all households that had used the online service in that time frame; and (2) a random sample of households that had not yet used the service but had recently visited a store that offered the online purchasing service. I limit the households in my sample based on visit and purchase requirements in order to identify households that frequently shop with the retailer.[17] The final household sample consists of 34 thousand households, 25 thousand of which have used the online service and 9 thousand of which have not used the online service (over the time frame of my data). The data also contain detailed product information; including the product name, category, nutritional content and product attribute claims made by the manufacturer (i.e. organic, gluten free, etc.). Most importantly, I can distinguish, at the household-day-store-UPC level, purchases that were made online from purchases that were made in the store.

Based on United States Department of Agriculture (USDA) classifications, I have assigned products to eleven different product categories: Dairy, Drinks, Fruits, Grains, Meat, Oils, Other, Prepared, Snacks/Sweets, Sugars and Vegetables.[18] I have also collapsed the purchasing data to the household-year-month level and defined an indicator for online service use if the service was used to buy any products in the monthly basket. I evaluate the impact of online service availability on combined (in-store and online) monthly grocery purchases because I am interested in understanding how using the online service impacts overall food demand rather than understanding how online orders differ from in-store orders. For example, suppose households use the online service only to buy healthy foods; if I were to analyze orders, I would find that online orders are much healthier than in-store orders. However, analysis at the order level ignores the fact that

[17] First, I drop households that do not visit the retailer at least once every two months (roughly 87,171 households). Next, I drop households that spend less than $20 per month on average (72 households). Additionally, there are small businesses in the data set so I drop households who spend more than $1,500 per month on average (2,290 households). I further limit the household sample to the group of households for whom I have demographic information on; this restriction drops 7% of the eligible households from my sample. Additional households were dropped based on the definition of online service availability; these restrictions were discussed in Section 2.

[18] These product categories were chosen and created based on a document authored by the United States Department of Agriculture (USDA) called, "What We Eat in America". The descriptions of the products assigned to each of these product categories can be found in Table A1 of the appendix.

the same household may be supplementing all of their healthy online purchases with unhealthy in-store purchases that could perfectly balance their overall demand (in-store and online) to where it was before the household began shopping online. Hence, in this hypothetical scenario, online service use has had no impact on demand; it has only impacted how the consumer chooses to purchase the various items in their basket.

Tables 1 and 2 compare the demographics and purchasing patterns of households who eventually adopt the online purchasing service (online households) to households who never adopt the online purchasing service (in-store only households). The comparisons between these two different types of shoppers are made over the time period in which no one had access to the online purchasing service. Table 1 illustrates that households who adopt the online purchasing service tend to be younger, are more likely to be in a higher income group, are more likely to be married and are more likely to have children.[19] Table 2 indicates that the households that eventually adopt the online purchasing service tend to spend more with the retailer per month ($448 vs. $331) and make more trips to the store each month (7.5 vs. 6.8), prior to online service adoption, relative to the households who never adopt the online purchasing service (in-store only households). Furthermore, online adoption households allocate a larger percentage of their grocery spending towards dairy (13.2% vs. 11.9%), fruit (7.5% vs. 7.3%), grains (7.6% vs. 7.2%) and other (1.8% vs. 1.7%); while in-store only households allocate a larger proportion of their budgets towards drinks (10.1% vs. 10.9%), meats (18.5% vs. 18.9%) and snacks/sweets (15.5% vs. 16.1%).[20] In the analysis that follows, I restrict the majority of my attention to the subset of households that eventually use the online purchasing service (i.e. the online households).[21] The pre-existing differences between online adopters and in-store only households suggest that the results of this paper will not be representative of the effect of online shopping for the general population of shoppers; however, the results of this pa-

[19] Most of my demographic variables are provided categorically. Table A2 in the appendix provides the specific categorical values for each demographic variable.

[20] There are no statistically significant differences among the two household types in the budget shares for oils (4.4%), prepared (10.8%) and vegetables (9.1%).

[21] The in-store households (households that do not adopt the online purchasing service over the time frame of my data) are utilized in Section 5.2 as a robustness check.

per are representative of the effect of online shopping for the early adopters of the online purchasing service.

Tables 1 and 2 also compare the demographics and pre-online purchasing patterns of the online service adopters who received access to the service at earlier dates to those who received access to the service at later dates. In order to neatly compare the households that received access to the online shopping service first to those who received access later, I assign households to three waves based on the month and year that they received access. Households that received access to the online shopping service before or during January 2016 are assigned to wave one (53% of the online household sample), households that received access to the online service after January 2016 but before July 2016 are assigned to wave two (29% of the online household sample) and households who received access after or during July 2016 are assigned to wave three (18% of the online household sample). Table 1 illustrates that households in later availability dates tend to be smaller and older. Households in the last availability wave also have a higher proportion of households in lower income categories, are less likely to be married and are less likely to have children. Table 2 indicates that households in the first wave tend to spend roughly $40 more per month, purchase roughly 15 more items and make one more trip to the store per month than households in the second or last wave. Wave 1 households also tend to allocate more of their grocery budget towards dairy, other and vegetables and less of their budget toward drinks, meat, prepared and snacks/sweets relative to households in other waves.

Table 3 presents the average price and nutrition content per ounce of food in each product category.[22] It is interesting to note the positive correlation that exists between mean budget shares and the prices of the product categories. For example, the average price of meat and snacks/sweets is $0.23/oz and $0.21/oz, while the average price of grains and fruit is $0.13/oz and $0.09/oz, respectively.[23] Unsurprisingly, there are considerable differences in the mean amount of calories and nutrients contained in an ounce of

[22]The price averages represent the average price paid by all households over all time periods, while the nutrition averages are conducted over all of the products assigned to that product category.

[23]The category of other has the largest price ($0.47/oz) because it contains spices, which are extremely expensive per ounce.

food across product categories. For example, snacks/sweets contains the highest amount of calories (120 kcal/oz), while meats contain the most protein (7.7 g/oz), grains the most carbs (18.5 g/oz), oils the most fat (7.7 g/oz) and sugars the most sugar (14.9 g/oz).

4 Reduced Form Estimation & Results

The pre-existing differences between the households assigned to different online service availability waves motivates the household fixed effects model I employ in the analysis that follows. I estimate the effect of online service use by utilizing availability of the service as an instrument for the decision to shop online. Specifically, my equation of interest is the following:

$$s_{ihm} = \alpha_i + \phi_i 1\{Online_{hm}\} + \gamma_{im} + \gamma_{ih} + \epsilon_{ihm} \tag{1}$$

where s_{ihm} is the budget share of product category i for household h in year-month m, $1\{Online_{hm}\}$ is an indicator that equals one if the online service was ever used by household h in year-month m, γ_{im} is a year-month fixed effect to control for differences across time and γ_{ih} is a household fixed effect to control for unobserved household preferences. Specifications without household and year-month fixed effects include treatment group indicators and post-online-availability time indicators in order to maintain the panel difference-in-differences framework that is featured in the two-stage least squares estimation strategy presented shortly. For specifications without household fixed effects, demographic characteristics of the households are included as covariates.[24]

I estimate the local average treatment effect of online service availability by instrumenting online use, $1\{Online_{hm}\}$, with online availability, $1\{OnlineAvail_{hm}\}$. This produces panel difference-in-differences reduced form and first stage equations that estimate the average treatment effect of online availability on the expenditure shares for product

[24]Since my demographic variables are provided categorically, I include the demographics by creating indicators for whether household h belongs to a given demographic category. Additionally, the demographic information I have access to does not vary over time within households.

category i and the probability of shopping online, respectively. Explicitly, the reduced form is:

$$s_{ihm} = \nu_i + \tau_i 1\{OnlineAvail_{hm}\} + \gamma_{im} + \gamma_{ih} + \omega_{ihm} \tag{2}$$

and the first stage is:

$$1\{Online_{hm}\} = \lambda_i + \theta_i 1\{OnlineAvail_{hm}\} + \gamma_{im} + \gamma_{ih} + v_{ihm} \tag{3}$$

Thus, the two-stage least squares estimate of ϕ_i, the average effect of online service use or the local average treatment effect of online availability, is the ratio of two difference-in-differences estimates; specifically, $\phi_{i,2SLS} = \tau_i/\theta_i$.

Table 4 presents the difference-in-differences estimates, $\hat{\tau}_i$. The estimates presented in column (1) do not include year-month or household fixed effects, column (2) presents estimates from regressions that include year-month fixed effects and column (3) presents the estimates that incorporate both year-month and household fixed effects. The results of these regressions indicate modest increases in the budget shares of dairy, fruit, meat and vegetables and modest decreases in the budget shares for drinks, oils and snacks/sweets. Specifically, the results of the full model indicate a 0.10, 0.08, 0.20 and a 0.13 percentage point increase in the budget shares for dairy, fruit, meat and vegetables (respectively) and a 0.10, 0.04 and a 0.41 percentage point decrease in the budget shares of drinks, oils and snacks/sweets (respectively). Each of these shifts represents less than a three percent change in the average pre-online service budget share, which is calculated over the six months prior to online service introduction at any store location. Table 5 presents the first-stage estimates, $\hat{\theta}_i$. These estimates indicate a strong and positive relationship between the availability of the online service and actual use of the online service.[25] The full model indicates that the introduction of the online service increases the probability of shopping online, in a given month, by 19.3 percentage points, on average.

Table 6 presents the ordinary least squares and two-stage least squares estimates of the effect of online shopping. The two-stage least squares estimates (columns 4, 5 and 6)

[25] The first stage is a linear probability model.

are aligned with the findings of the reduced-form estimates. Furthermore, the two-stage least squares estimates are stable across three different specifications and indicate a 0.5, 0.4, 1.1 and 0.7 percentage point increase in the budget shares for dairy, fruit, meat and vegetables, (respectively), and a 0.5, 0.2 and 2.1 percentage point decrease in the budget shares for drinks, oil and snacks/sweets, respectively. Interpreted relative to the pre-online service average budget shares, these estimates indicate a 3.8, 5.9, 5.7 and 7.4 percentage increase in the average pre-online service budget shares for dairy, fruit, meat and vegetables, respectively. These estimates also indicate a 5.2, 4.1 and 13.6 percentage decrease in the average pre-online service budget shares for drinks, oil and snacks/sweets, respectively.

These results suggest that when households shop online they allocate a larger share of their budget towards healthier product categories at the expense of product categories that are traditionally thought to contain less healthy items. In order to better understand the implications that online shopping may have for consumer health, I evaluate changes in the average nutrient content per ounce of food purchased by the households in my sample. I follow the same estimation strategy presented above but with nutrients per ounce as the dependent variable.[26] The results of the ordinary least squares and two-stage least squares regressions are presented in Table 7, where $1\{Online_{hm}\}$ is instrumented with $1\{OnlineAvail_{hm}\}$. The results of the two-stage least squares models indicate that the average amount of calories, carbohydrates, fats and sugars per ounce of food purchased decreases in the months that households shop online. My estimates indicate an average decrease of 1.97 kcal/oz, 0.3 g/oz, 0.11 g/oz and 0.13 g/oz for the nutrient categories of calories, carbohydrates, fat and sugars, respectively. Furthermore, there are no significant changes in the average amount of protein, cholesterol and sodium content per ounce of food purchased. Interpreted relative to the pre-online service average, these results equate to a 4.2, 5.0, 5.8 and a 5.8 percent decrease in the average amount of calories,

[26]Nutrition information is available for 90% of the items purchased by these households. Hence, these regressions test changes in the nutritional value of purchases made over the products for which I have nutrition information. The majority of the gaps in the nutrition information data occur for products that are not purchased in a package (i.e. fresh produce and meats).

carbohydrates, fats and sugars contained in an ounce of food purchased before the online service was available, respectively.

Before the online shopping service was available, households purchased 2,383 ounces of food per month, on average. Holding the amount of food purchased constant, an average decline of two calories per ounce equates to 4,766 fewer calories being purchased in the months in which a household shops online. Extrapolating this results to the individual-day level, the change in the amount of calories purchased represents a decline of 53 calories per person, per day.[27] If the declines in caloric purchases perfectly translated into declines in caloric intake, these changes would induce an average adult to lose one pound every ten weeks or roughly five pounds over the course of a year.

The magnitude and signs of the estimates presented above support the prediction that the online shopping environment reduces the incidence of impulsive and unhealthy purchases. Furthermore, estimated changes in the amount of calories purchased indicate that online shopping could contribute to slow but gradual improvements in consumer health through weight loss. However, more dramatic changes in consumer health could be made by improving the nutritional value of the purchases households *intend* to make.

4.1 Online Shopping & Retailer Substitution Patterns

If consumers change retailer substitution patterns differentially across product categories when shopping online, then consumer "crowd-in" could explain the changes we observe in grocery basket composition as well as the documented changes in the nutritional content of food purchased. For example, suppose that the households in this study, prior to using the online shopping service, purchased produce at other grocery stores (health food grocery stores, etc.) and/or a farmers market. Then, further suppose that after transitioning to the online shopping service, these households stopped buying their produce from other retailers and began purchasing more produce with the retailer in this study because the online service made these purchases more convenient. In this hypothetical

[27] This number is generated by dividing 4,766 by thirty days and three people (the size of the average online household).

scenario, we would expect to see the budget share for fruits and vegetables increase, but these shifts are the result of changing retailer substitution patterns rather than an effect of shopping online on grocery demand.

Prior research by Pozzi (2013a) documents that online grocery services can lead consumers to divert their grocery business away from other retailers, toward the online shopping service provider. Pozzi also finds that households living in areas with higher levels of retailer competition increase monthly expenditures with the online retailer at a higher rate than households living in areas with relatively lower levels of outside competition.

The households studied in this paper also exhibit increases in monthly grocery expenditure when they begin using the online shopping service. Following the estimation strategy outlined in equation (1), but with expenditures as the outcome variable, the two-stage least squares estimates presented in Table 8 indicate that households spend \$49 more per month (roughly a 10.9% increase over average pre-online service expenditures), on average, when shopping online.[28] Additionally, all product categories, with the exception of other and snacks/sweets, experience significant increases in overall expenditure.[29] The final column of Table 8 presents the percent change in average spending that the two-stage least squares estimates represent. The calculated percent changes in categorical expenditures illustrate that, in general, the estimated changes in budget shares are positive if the percent change for product category expenditure increased by more than 10.9% (the average percent change in total expenditure) and are negative if the percent change for product category expenditure increased by less than 10.9%. Assuming that shopping for groceries online does not increase a household's monthly grocery budget (across all retailers), the fact that there are relatively large estimated increases in expenditures for dairy (\$7.7), fruit (\$7.2), meat (\$11.8) and vegetables (\$8.0), while the estimated changes in expenditure for drinks (\$4.0) and snacks/sweets (\$0.76) are relatively small suggests that

[28]Note that these figures are conditional on arrival to the store in a given month. Explicitly, observations of \$0 have not been imputed for year-month occasions in which a household does not buy any grocery products from the retailer of this study.

[29]Specifically, the two-stage least squares estimates indicate a \$7.7, \$4.0, \$7.2, \$3.8, \$11.8, \$1.7, \$2.7, \$0.6 and \$8.0 average increase in monthly spending for the product categories of dairy, drinks, fruit, grains, meat, oil, prepared, sugars and vegetables, respectively.

households could be changing retailer substitution patterns differentially across product categories.

I explore whether consumer crowd-in can explain the entirety of the documented differences between online and in-store purchases in two ways. First, I assess whether or not there are heterogenous effects of shopping online on grocery expenditures that can be attributed to the level of competition surrounding the store location. The goal of this exercise is to identify households that do not change retailer substitution patterns after shopping online; once those households have been identified, any changes in these households' budget shares after shopping online can be attributed to changes in grocery demand. Second, I further explore the extent to which consumer crowd-in can explain my findings by analyzing the effect of online shopping on purchasing decisions made within narrower product spaces. Consumer decisions over product choice in a more narrowly defined product space are less likely to be influenced by changes in the mix of retailers the consumer visits. For example, it is less likely that consumers shop for different types of bread at different types of retailers than it is for them to shop for different types of grocery products at different types of retailers. Under the assumption that households buy the same variety of bread at each retailer they shop with, if consumers allocate more of their bread purchases toward the retailer of this study when shopping online, we would not expect consumer "crowd-in" to change the composition of breads purchased, only total expenditures over bread. Hence, changes in purchasing decisions within these product spaces can be attributed to differences in the purchasing environments rather than changes in retailer substitution patterns.

4.1.1 Heterogeneous Effects of Online Availability on Retailer Substitution Patterns

I begin exploring the extent to which shopping online influences retailer substitution patterns by evaluating how the effect of online shopping on total grocery expenditures varies with the level of competition the store location faces. In order to do this, I split online households in two groups: those whose online service availability was determined by a store location that faces below average competition and those whose online service availability was determined by a store location that faces above average competition. The level of competition faced by each store location was provided to me by the retailer; those in the low competition group have a competition index that is lower than the average store location and those in the high competition group have a competition index that is higher than the average store location.[30]

Tables 9 and 10 present difference-in-differences estimates, utilizing the estimation strategy presented in equation (2), for the low competition and high competition online household population, respectively. Table 9 indicates that households who adopt the online purchasing service in store locations that face lower levels of outside competition do not increase their monthly expenditure with the retailer of this study after the online purchasing service becomes available; however, these households still exhibit changes in the composition of their grocery purchases. After the online service becomes available, low competition households exhibit significant increases in the budget shares of fruit and meat. Additionally, this reallocation of funds comes at the expense of the snacks/sweets product category. The point estimates indicate a 0.10, 0.22 and -0.24 percentage point change in the budget shares of fruit, meat and snacks/sweets, respectively. These findings illustrate that even when households do not exhibit changes in substitution patterns across retailers when shopping online, they still change the composition of their purchases when shopping online.

In contrast, households in competitive purchasing environments significantly increase monthly grocery expenditures after the online purchasing service becomes available. Specifically, Table 10 indicates that households in competitive environments increase monthly grocery expenditures by $10.60 when the online purchasing service becomes available. These results suggest that the online purchasing service may be effective at poaching customer's grocery purchases from competitors. The point estimates of the effect of online availability on budget shares indicate a 0.04, 0.21 and 0.17 percentage point increase in the budget shares of grain, meat and vegetables respectively. This reallocation of funds also comes at the expense of snacks/sweets with point estimates indicating a 0.49 percentage

[30] To provide some context on what the competition index is capturing, stores that have a lower than average competition index have (on average) two fewer major competitors and twelve fewer minor competitors compared to store locations that have a higher than average competition index.

point decrease, on average.

Comparing the results of difference-in-difference estimates for low competition households (Table 9) to the difference-in-difference estimates for all households (Table 4) provides a rough indication of how much of the documented changes in grocery basket composition could be due to differences in retailer substitution patterns after households begin shopping online. The point estimates for the product categories of fruit, grains, meat, oil and other are remarkably close for the two estimation procedures.[31] In contrast, the estimates for the product categories of dairy, drinks, snacks/sweets, sugar and vegetables are not as closely aligned.[32] The differences between these estimates suggest that, on average, roughly 40 percent of the estimated effect of online shopping on the budget shares of dairy, drinks, snacks/sweets, sugar and vegetables could be due to changes in retailer substitution patterns after households begin shopping online.

4.1.2 Narrow Product Categories

I further explore the extent to which consumer crowd-in can explain the changes in online grocery basket composition by analyzing online product choice within narrower product spaces. The advantage of a more narrowly defined product space is that it is less likely that a consumer would buy different types of products, within these product spaces, at different retailers. For example, it is less likely that consumers shop for different types of yogurt at different types of retailers than it is for them to shop for different types of grocery products at different types of retailers. Under the assumption that households buy the same variety of yogurt at each retailer they shop with, if consumers allocate more of their yogurt purchases toward the retailer of this study when shopping online, we would not expect consumer "crowd-in" to change the composition of yogurts purchased, only total expenditures over yogurt.

This section evaluates whether there are significant differences in the types of bread,

[31] Note, I am comparing the estimates that utilize household and year-month fixed effects and am calculating the ratio of the low competition estimate to the all household estimate.

[32] The estimates for dairy, drinks and vegetables are also no longer significant in the regressions that incorporate the low competition households only.

breakfast cereal, salty snacks and yogurt purchased when households shop online. Specifically, I analyze whether households begin to allocate a larger portion of their expenditures, within these product categories, towards healthier options. I begin by assigning individual products into different subcategories within the product group. For example, I assign individual bread UPCs to five different bread categories: white, wheat, other, seed and grain. Cereal UPCs are assigned to the categories of kids, organic kids, standard, frosted standard and healthy. Within salty snacks I create the product classifications of chips, healthy chips, popcorn, pretzels and tortilla chips. Lastly, I assign yogurt products to the categories of probiotic, light greek, greek, indulgent, kids, organic, light traditional and traditional.[33]

Table 11 presents the average nutritional content (per ounce of food) for each product subcategory within breads, cereals, salty snacks and yogurt. In the product category of bread, wheat bread is generally considered a healthier choice. Table 11 illustrates that all breads contain roughly 70 calories per ounce; however, wheat, grain and seed breads tend to contain more protein per ounce (2.8 g vs. 2.3 g).[34] Within the product category of cereals, healthy (Grape Nuts, Kashi, Fiberone) and standard cereals (Cheerios, Chex, Cornflakes) tend to have better nutritional values than kids (Apple Jacks, Fruity Pebbles), organic kids (Annies, Cascadian Farm) and frosted standard cereals (Frosted Flakes, Corn Pops). Healthy cereals contain less calories per ounce, more protein per ounce and less sugar per ounce. Standard cereals also exhibit lower levels of fat per ounce and sugars per ounce. Healthy chips (Sun Chips, Veggie chips etc.), pretzels and tortilla chips are healthier snack options relative to regular chips; healthy chips, pretzels and tortilla chips contain fewer calories per ounce, more protein and less fat than regular chips. Lastly, Table 11 illustrates that greek, light greek and light traditional yogurt have better nutritional values than other types of yogurt in terms of calorie, protein and sugar content. These yogurts have fewer calories per ounce, higher amounts of protein and contain less sugar

[33] Table A4, in the appendix, briefly describes the subcategories of products that have been created within each product category.

[34] Breads in the other category tend to have a higher sugar content per ounce, likely due to cinnamon raisin breads, etc.

than other yogurt varieties, on average.

In order to better understand how shopping online influences the composition of the different types of products purchased within these categories, I analyze the effect of online service availability on the budget shares for each product subcategory.[35] I analyze the budget share outcomes utilizing ordinary least squares and a fractional probit model, which accounts for the fractional nature of the share outcome and alleviates concerns associated with corner solutions. The underlying data in these regressions is generated from the year-month purchase occasions in which a household buys at least one item from the parent product category (i.e. households that purchase at least one bread for the bread share outcomes, one cereal for the cereal share outcomes, etc.). I estimate regressions of the following form:

$$s_{khm} = \nu_k + \tau_k 1\{OnlineAvail_{hm}\} + \gamma_{km} + \gamma_{kh} + \omega_{khm} \tag{4}$$

where s_{khm} is the share of product category sales (bread, cereal, etc.) allocated to product sub-category k for household h in year-month m, $1\{OnlineAvail_{hm}\}$ is an indicator that equals one if the online service is available to household h in year-month m, γ_{km} is a year-month fixed effect to control for differences across time and γ_{kh} is a household fixed effect to control for unobserved household preferences.[36]

[35]Tables A5, A6, A7 and A8 present the estimated effect of online service availability on sales ($\hat{\tau}_k$) for three ordinary least squares specifications, as well as the estimated average partial effects for two tobit specifications. The results of these regressions illustrate that online service availability has a positive and significant effect on the sales of wheat ($0.09) and seed breads ($0.01), frosted cereals ($0.08), tortilla chips ($0.06), traditional light ($0.03), traditional ($0.06), organic ($0.02) and greek yogurts ($0.12). Furthermore, there is no evidence for statistically significant changes in the sales volumes for white, other and grain breads; kids, standard and healthy cereals; pretzels, popcorn, healthy chips and chips; or kids and indulgent yogurts. A preliminary analysis of the data indicates that it is not uncommon for households to purchase more than one type of bread, cereal, salty snack and/or yogurt in a given month. Hence, I chose to model the decision of product choice in a continuous framework rather than a discrete choice environment. Specifically, 50% of bread purchases, 60% of cereal purchases, 70% of salty snack purchases and 62% of yogurt purchases in a year-month purchase occasion contain more than one type of bread, cereal, salty snack or yogurt product purchased.

[36]Specifications without household and year-month fixed effects include treatment group indicators and post-availability time indicators in order to maintain the panel difference-in-differences framework. Additionally, specifications without household fixed effects include demographic characteristics of the households. Since my demographic variables are provided categorically, I include the demographics by creating indicators for whether household h belongs to a given demographic category.

Tables 12, 13, 14 and 15 present the estimates of τ_k from the ordinary least squares specifications, as well as estimates of the average partial effects for the fractional probit regressions. These tables indicate that after the online purchasing service becomes available, the only product categories that exhibit changes in the composition of products purchased are bread and salty snacks. The fractional probit estimates indicate that the budget share for wheat bread increases by 0.63 percentage points, while the budget share for other bread decreases by 0.37 percentage points. Additionally, the budget share for tortilla chips increases by 0.5 percentage points, while the budget share for pretzels decreases by 0.34 percentage points, on average. Lastly, the budget shares for product within the cereal and yogurt product categories remain unchanged.

These results suggest that shopping online does not induce households to become more health conscious within all product categories. I do not find any changes in the composition of cereals or yogurts purchased, a finding which is consistent with theories that predict decreased brand exploration in the online shopping environment due to an inability to verify product quality.[37] However, there are significant differences in the composition of bread and salty/snacks purchased when households shop online. These changes could be explained by differences in product search and (or) product placement across the two purchasing environments. For example, consider the purchasing patterns observed in bread. We observe that there are increases in wheat bread sales, as well as increases in the budget share for wheat bread. Furthermore, the increases in the budget share for wheat bread come at the expense of other bread. This result could be explained by in-store shopping behavior in which the consumer adds other breads (cinnamon raisin bread, etc.) to their cart because they see these breads and are reminded to buy them when shopping in the store. In the absence of these visual reminders when shopping online, the consumer may forgo purchases of other bread types. Differences in the location of products when shopping online could also lead to changes in purchasing patterns. For example, when in the store, the pretzels, popcorn and chips are all displayed on the

[37]However, I do find that households spend significantly more on all types of yogurt when they shop online. This could be due to households bringing their yogurt demand from other retailers to the retailer of the study and (or) yogurt could be more attractive product when shopping online.

shelves of the same aisle. However, when shopping online, tortilla chips are nested within the chips category, while popcorn and pretzels are listed in their own product categories. Tortilla chips could become more popular when households shop online because, in the online purchasing environment, they are displayed relative to regular chips only.

5 Reduced Form Robustness Checks

This section tests the robustness of the panel difference-in-difference regressions performed earlier in the paper. First, I assess the validity of the parallel trends assumption by testing for pre-treatment differences in the budget shares for treated and control households. Then I run difference-in-differences regressions over the set of households in my sample who never shop online (i.e. in-store only households) to ensure that the effects captured in the main analysis are the result of online shopping and not other changes that could be occurring at the same time the online purchasing service was introduced. Lastly, I test whether the results of my main analysis can be explained by limited online product offerings. The results of these robustness checks indicate that the parallel trends assumption is valid for all product categories with the exception of dairy. Furthermore, I find no evidence that other changes are occurring in the purchasing environment at the same time the online service was introduced. Lastly, I do not find evidence that suggests limited online product offerings are responsible for the main results of this paper.

5.1 Event Study

I estimate event study specifications to determine whether or not there were shifts in budget share allocations before the online service was introduced and to evaluate the effect of online shopping over time. Specifically, I estimate regressions of the following form:

$$s_{ihm} = \nu_i + \sum_k \tau_{ik} 1\{TimeAvail_{hm} = k\} + \gamma_{im} + \gamma_{ih} + \omega_{ihm} \tag{5}$$

where s_{ihm} is the budget share of good, i, for household, h, in year-month, m, $1\{TimeAvail_{hm} = k\}$ is an indicator that equals one when the household is k time-periods from online service introduction, γ_{im} and γ_{ih} are year-month and household fixed effects, respectively. I allow the reference time-period for $1\{TimeAvail_{hm} = k\}$ to be all time-periods that were more than five months prior to the online service introduction. Furthermore, k is discrete from $k = -5$ to $k = 5$ (5 months before online service introduction to 5 months following introduction, where time zero is the month of introduction) with one additional indicator for 6 months or more following introduction.

Figure 2 presents the estimates of τ_{ik}, with 95% confidence intervals, for each budget share outcome as well as for the outcome of online service use.[38] Eleven of the twelve graphs do not indicate a consistent pre-trend violation. However, the graph for dairy illustrates that the the point estimates of τ_{ik} were increasing in the months before online service introduction, indicating that caution should be exercised when drawing conclusions about the effect of online grocery shopping on dairy purchases.

The estimates for drinks, meat, snacks/sweets and vegetables illustrate post-treatment effects that remain fairly stable in the post-treatment time periods. However, the negative effects for drinks and snacks/sweets seem to fade six or more months after treatment. In contrast, the treatment effects on the budget shares for dairy, fruit and oil are fairly noisy, with each graph illustrating only one or two post-introduction coefficients that are significant. In summary, the graphs in Figure 2 illustrate striking discontinuities in the estimates of τ_{ik} at the period of online introduction (t=0) for the product categories of drinks, meat, snacks/sweets and vegetables.

5.2 Placebo Test

In order to test whether or not there are other changes influencing demand that occur simultaneously with the introduction of the online purchasing service, I estimate difference-in-difference regressions over the subset of households that never adopt the online service. I expect these estimates to indicate that online service availability has no effect on the budget shares of households that never use the online service. Table 16 presents the difference-in-difference regressions for in-store only households. These

[38]The estimates of these regressions are presented in Tables A17 and A18 of the appendix.

results are as one would expect; there is no strong evidence to suggest that online service availability has any significant effect on the budget share allocations of households who never use the online purchasing service. Although the difference-in-difference estimates for the prepared category are significant at the 10% significance level, regressions evaluating the effect of online service availability on in-store households' expenditures (Appendix Table A13), nutrition outcomes (Appendix Table A14) and ounce shares (Appendix Table A15) reveal no statistically significant changes over any of these outcomes. The results of these regressions suggest that there were no changes to other factors that influence demand at the time of online service introduction.

5.3 Online Product Offerings

When the online purchasing environment was launched the retailer only had select products available for online purchasing. At the time of launch, roughly 22 thousand unique items were available for purchase online. In contrast, by January 2016, roughly 78 thousand unique items were available for purchase online.[39] However, the retailer has indicated that most of the growth in product offerings came from general merchandise products rather than grocery products.

There are some challenges to understanding what products were available online. Primarily, I do not know what products were available online at any given point in time; I only have access to information regarding whether or not a product was purchased online and (or) in the store. Given that the online service was slowly introduced and that only a subset of households are able to shop online at any given point in time, there may be more products available online than are purchased online.

Figure 3 presents the percent of products that were purchased in-store and also purchased online. In March 2015, 27% of the products in the dairy category that were purchased in-store were also purchased online (the maximum of any product category); in

[39]The retailer has provided me a list of roughly 108 thousand unique products within grocery that have been available for sale in their store at any point in time. Actual product offerings in the store are almost surely lower than this at any given point in time. As of March 2017 there were 89 thousand unique products available for purchase online.

contrast, only 9% of products in the other category that were purchased in-store were also purchased online (the minimum of any product category). However, by January 2016, these proportions significantly increased to the point where at least 53% of the products (in any given category) purchased in-store have also been purchased online. Figure 4 presents the percent of in-store sales that come from products that were also purchased online in each month. The column for January 2016 indicates that the minimum percentage of in-store sales represented by products that were also purchased online is 86%. Thus, 53% of the products represent 86% of the sales in a product category, snacks/sweets, that has been shown to be less popular online.[40] Given that the most popular in-store items are also being purchased online and that only half of my household sample has access to the online purchasing service by January 2016, I assume that online product offerings are representative of in-store product offerings from January 2016 on.

Under this assumption, I test whether limited product offerings are driving the results of the previous analysis by restricting the data to all dates after January 2016 and estimating equation (1). If limited product offerings are responsible for the main results of my paper, the time-restricted regressions should indicate that shopping online has no effect on any of the budget share outcomes. Table 17 presents the ordinary least squares and two-stage least squares estimates from the time restricted regressions. The results indicate that the effect of online shopping for dairy and fruit budget shares remains positive and significant and the effect for snacks and sweets remains negative and significant. In contrast, the estimated effects for drinks, meat, oils and vegetables retain the same signs as earlier regressions, but are no longer statistically significant.[41] The magnitudes of the estimates over all time periods and the time restricted estimates are remarkably similar for the product categories of dairy, fruit and snacks/sweets with larger discrepancies occurring in the product categories of drinks, meat, oils and vegetables.

[40]Additionally, only 56% of the households in my sample have access to the online shopping service by January 2016.

[41]The estimates obtained from running regressions over later time periods of the data indicate the following percent changes in average budget shares: dairy (5.2%), fruit (5.5%), meat (3.5%), vegetables (4.4%), drinks (-1.3%), oil (-1.2%), snacks/sweets (-13.5%). For comparison, the estimates over all time periods indicated the following percent changes in average budget shares: dairy (3.8%), fruit (5.9%), meat (5.7%), vegetables (7.4%), drinks (-5.2%), oil (-4.1%), snacks/sweets (-13.6%).

Identification is driven by the households whose status of online availability changes over time. Thus, estimates of the effect of online shopping that are derived from the restricted set of time periods are identified only by the Wave 2 and Wave 3 households defined in Section 3. There is no way to parse out the differences in these results that are due to heterogeneous treatment effects and those that are due to increased availability of online product offerings.[42] However, because differences in online and offline budget shares continue to persist in time periods in which the online product offerings are representative of in-store offerings, this robustness check indicates that the results estimated over all time periods cannot be explained by limited online product offerings.

6 Modeling Grocery Demand

The reduced form findings of this paper indicate that purchasing environments have influence over consumer choice. My findings imply that, when shopping online, households may be less tempted to make unhealthy impulse purchases simply due to search and timing differences between the in-store and online purchasing environments. Shopping online also leads to monthly purchases that are of a higher nutritional quality compared to months in which all purchases were made in-store. In order to better understand how online demand differs from in-store demand, I present and estimate a formal model of grocery demand which allows me to test for differences in price elasticities across the two purchasing environments.

I estimate grocery demand by utilizing the LA/AIDS demand system, developed by Deaton and Muellbauer (1980); I also incorporate an extension from Atkin (2013) that models varying household tastes by allowing the vector of first price coefficients (τ_{ih}, in the expenditure system defined shortly) to vary by household.[43] The LA/AIDS demand

[42]Figure A8, of the Appendix, illustrates that wave two and wave three households allocated more of their budget towards drinks and less of their budgets towards vegetables compared to wave one households. On the other hand, the drinks product category exhibited a lot of growth in product offerings between March 2015 and January 2016 but it is unclear if this growth came from additional products being offered online or more people purchasing these UPCs online.

[43]Allowing the first price coefficients to vary by household is a reasonable way to model tastes for the following reason: if a household really values a particular item, price changes for that item are likely to

system is derived from an expenditure function, $e(p,u;\tau_h)$: the minimum expenditure necessary to achieve utility u, given a vector of prices p and varying household tastes, τ_h. The log expenditure function over I goods is defined as follows:

$$ln(e(p,u;\tau_h)) = \alpha_0 + \sum_{i=1}^{I} \tau_{ih} ln(p_i) + 1/2 \sum_{i=1}^{I} \sum_{i\prime=1}^{I} \gamma_{ii\prime} ln(p_i) ln(p_{i\prime}) + u\beta_0 \prod_{i=1}^{I} p_i^{\beta_i} \tag{6}$$

where i indexes good i and $i\prime$ indexes good $i\prime$.[44] Note that by Shephard's Lemma: $\frac{\partial e(p,u;\tau_h)}{\partial p_i} = q_{ih}$. This equality implies:

$$\frac{\partial ln(e(p,u;\tau_h))}{\partial ln(p_i)} = \frac{p_i q_{ih}}{e(p,u;\tau)} = s_{ih} \tag{7}$$

where s_{ih} is the share of household h's budget allocated to good i. Differentiating the right hand side of equation (6) with respect to $ln(p_i)$ and setting it equal to s_{ih} produces the following:[45]

$$s_{ih} = \tau_{ih} + \sum_{i\prime=1}^{I} \gamma_{ii\prime} ln(p_{i\prime}) + \beta_i u \beta_0 \prod_{i=1}^{I} p_i^{\beta_i} \tag{8}$$

Recall that a utility maximizing consumer sets $e(p,u;\tau_h)$ equal to total expenditure, m; thus, I re-write utility as a function of total expenditure and prices by rearranging equation (6). I use the functional form of utility to get budget shares as a function of tastes (τ_{ih}), prices (p), and total expenditure (m_h):

$$s_{ih} = \tau_{ih} + \sum_{i\prime} \gamma_{ii\prime} ln(p_{i\prime}) + \beta_i ln(\frac{m_h}{P}) \tag{9}$$

Where $ln(P)$ is defined by the following equation:

$$ln(P) = \alpha_0 + \sum_{i} \tau_{ih} ln(p_i) + 1/2 \sum_{i} \sum_{i\prime} \gamma_{ii\prime} ln(p_i) ln(p_{i\prime}) \tag{10}$$

---

change their total expenditure differently than it would for a household who has little value for that item.

[44]Goods i and $i\prime$ are both within I and are sometimes equivalent and sometimes different than each other.

[45]I use the fact that $\frac{\partial ln(e(p,u;\tau_h))}{\partial ln(p_i)} = \frac{\frac{\partial ln(e(p,u;\tau_h))}{\partial p_i}}{\frac{\partial ln(p_i)}{\partial p_i}}$ in order to get equation (8).

Following Deaton and Muellbauer (1980), I approximate $ln(P)$ with a Stone Index, $ln(P) = \sum_i \bar{s}_i log(p_i)$, making the system linear.[46] I assume weak separability between grocery products and other expenditures which allows me to replace household expenditure with total expenditure over grocery products and calculate budget shares as the share of grocery expenditure allocated to product category i.

Note that household tastes, τ_{ih}, act as budget share shifters in this setting. I assume that tastes are a function of household demographics that influence household preferences (e.g. economic status, household size, age etc.) and seasonal fluctuations that influence preferences over food groups.[47] Hence, I define τ_{iht} as follows:

$$\tau_{iht} \stackrel{\text{def}}{=} \alpha_i + \nu_{ih} D_h + \nu_{it} Z_t \tag{11}$$

where D_h is a vector of demographic variables (household size, age, income, loyalty measures, marital status and number of children) and Z_t is a vector containing month dummies (to control for seasonality) and a linear time trend.

Adding a time subscript (year-month, t) and substituting τ_{iht} into equation (9) produces the following:

$$s_{iht} = \alpha_i + \sum_{i\prime=1}^{11} \gamma_{ii\prime} ln(p_{i\prime ht}) + \beta_i ln(\frac{m_{ht}}{P_{ht}}) + \nu_{ih} D_h + \nu_{it} Z_t + \epsilon_{iht} \tag{12}$$

I allow for structural differences to exist between the in-store and online demand functions by permitting the underlying parameters to vary in months that a household shops online. Explicitly, I include $1\{Online_{ht}\}$, an indicator that equals one when the household purchases at least one good online in year-month t, and I interact prices and expenditure with $1\{Online_{ht}\}$.[48] Modifying equation (12) to allow for structural differences between

[46] In the definition of $ln(P)$, $\bar{s}_i$ is the average budget share of category i over all periods of the data. This is substituted in place of s_{iht} in order to avoid the endogeneity that arises when the dependent variable also appears as an independent variable.

[47] Atkin (2013) provides evidence that tastes are a function of what type of food an individual was fed as a child.

[48] These interactions nest two demand equations (in-store and online) within one estimating equation.

in-store and online demand produces the following:

$$s_{iht} = \alpha_i + \phi_i 1\{Online_{ht}\} + \sum_{i\prime=1}^{11}[\gamma_{ii\prime} ln(p_{i\prime ht}) + \gamma_{ii\prime o} ln(p_{i\prime ht})1\{Online_{ht}\}] \\ + \beta_i ln(\frac{m_{ht}}{P_{ht}}) + \beta_{io} ln(\frac{m_{ht}}{P_{ht}})1\{Online_{ht}\} + \nu_{ih} D_h + \nu_{it} Z_t + \epsilon_{iht} \tag{13}$$

Theoretical properties of demand systems imply the following constraints on the parameters in equation (13):

Zero Homogeneity:

$$\sum_{i\prime=1}^{11} \gamma_{ii\prime} = 0, \quad \sum_{i\prime=1}^{11} \gamma_{ii\prime o} = 0 \tag{14}$$

Slutsky Symmetry:

$$\gamma_{ii\prime} = \gamma_{i\prime i}, \quad \gamma_{ii\prime o} = \gamma_{i\prime io} \tag{15}$$

Adding Up:

$$\sum_i \alpha_i = 1, \quad \sum_i \beta_i = 0, \quad \sum_i \phi_i = 0, \quad \sum_i \beta_{io} = 0 \tag{16}$$

I jointly estimate a system of ten demand equations, as presented in equation (13), subject to the constraints presented in equations (14) and (15). The parameters of the eleventh demand equation are derived from the estimation results of the ten equation system.[49] Specifically, I calculate $\hat{\alpha}_i$, $\hat{\beta}_i$, $\hat{\phi}_i$ and $\hat{\beta}_{io}$ of the eleventh demand equation by imposing the adding up restrictions presented in equation (16); additionally, the price coefficients of the eleventh equation are defined by the symmetry restrictions presented in equation (15).

Marshallian elasticities are computed from the estimated demand parameters as follows:

Own Price Elasticity

$$\eta_{ii} = \frac{\hat{\gamma}_{ii}^*}{\bar{s}_i^*} - \hat{\beta}_i^* - 1 \tag{17}$$

[49] The iterated FGLS estimates, presented in this paper, are equivalent to MLE and are invariant to which equation is dropped.

Cross Price Elasticity

$$\eta_{ii\prime} = \frac{\hat{\gamma}^*_{ii\prime}}{\bar{s}^*_i} - \hat{\beta}^*_i \frac{\bar{s}^*_{i\prime}}{\bar{s}^*_i} \tag{18}$$

In-store price elasticities are obtained by setting $\hat{\gamma}^*_{ii} = \hat{\gamma}_{ii}$, $\bar{s}^*_i = \bar{s}_i$, $\hat{\beta}^*_i = \hat{\beta}_i$ and $\hat{\gamma}^*_{ii\prime} = \hat{\gamma}_{ii\prime}$, while the online elasticities are obtained by setting $\hat{\gamma}^*_{ii} = \hat{\gamma}_{ii} + \hat{\gamma}_{iio}$, $\bar{s}^*_i = \bar{s}_i + \hat{\phi}_i$, $\hat{\beta}^*_i = \hat{\beta}_i + \hat{\beta}_{io}$ and $\hat{\gamma}^*_{ii\prime} = \hat{\gamma}_{ii\prime} + \hat{\gamma}_{ii\prime o}$.[50]

6.1 Demand Estimation

This section outlines how I address endogeneity and efficiency concerns when estimating the demand parameters. First, prices defined at the household level are likely endogenous to the demand system due to unobserved preference shocks and/or preferences over unobserved product quality. In order to alleviate these endogeneity concerns, I instrument prices with the mean wholesale costs of products available to the household. This isolates the variation in household prices that is due to factors that shift supply.[51] I also instrument $1\{Online_{ht}\}$ with $1\{OnlineAvail_{ht}\}$ following the logic presented earlier in the paper.[52] The second estimation concern is efficiency. Since I estimate ten demand equations for grocery products, it is likely that the errors for a given budget share equation, i, are correlated with the errors of another budget share equation, $i\prime$; in other words, $\text{cov}(\epsilon_{iht}, \epsilon_{i\prime ht}) \neq 0$.[53] In order improve efficiency, and to address the endogeneity concerns

[50]In the elasticity equations, $\bar{s}_i$ is the average budget share for category i over all time periods of the data. Note that the standard errors of the elasticities are computed as in Chalfant (1987), treating the $\bar{s}_i$'s as constants.

[51]One of the challenges associated with defining instruments for household level prices is that the instrument must also vary at the household level. In order to ensure household level variation in the wholesale costs of products available, I utilize the fact that each household has a unique store footprint (i.e. the stores visited by the household) in a given year-month. Specifically, I define the price instruments as follows: first, I calculate the average cost of the products available at the category-store-year-month level, then, I average the category-store-year-month level wholesale costs over the stores the household visited in a given year-month. Explicitly, $c_{iht} = \sum_s \frac{\bar{c}_{ist}}{S_{ht}}$, where s indexes the store locations frequented by household h in year-month t, S_{ht} is the number of stores visited by household h in year-month t and $\bar{c}_{ist}$ is the mean wholesale cost of the products sold in category i, at store s in year-month t. Additionally, since household prices appear in $ln(P_{ht})$, I instrument $ln(P_{ht})$ with $\ln(C_{ht}) = \sum_i \bar{s}_i ln(c_{iht})$.

[52]All interactions of $1\{Online_{ht}\}$ and $ln(p_{iht})$ are instrumented with interactions of $1\{\text{OnlineAvail}_{ht}\}$ and $ln(c_{iht})$.

[53]The result that generalized least squares is equivalent to ordinary least squares when the set of dependent variables is the same in each equation is lost when restrictions are placed on parameters either within

outlined above, I utilize a three-stage least squares estimation strategy.[54] The three steps of the estimation procedure are as follows:

1. Perform the first stage for each endogenous variable and obtain the fitted values.
2. Perform second stage estimation (i.e. two-stage least squares), equation by equation, replacing endogenous variables with their fitted values. Obtain the estimated errors for each equation and estimate the covariance matrix for the errors.
3. Compute the feasible generalized least squares estimator utilizing the estimated covariance matrix from Step 2.

The results of the first-stage regressions are presented in Tables A21, A22, A23, A24 and A25 of the appendix. Each of the proposed instruments is strong in the sense that it is highly correlated with its corresponding endogenous variable.[55] For prices, the cost instruments indicate a strong and positive relationship between wholesale costs and prices. The coefficients on the log cost instruments range from to 0.3 to 1.0, indicating that a one percent increase in average store costs results in a 0.3 to 1 percent increase in household prices. The estimated demand parameters of the three-stage least squares procedure can be found in Tables A26, A27, A28 and A29 of the appendix.

6.2 Elasticities

The in-store and online price elasticity estimates and standard errors are reported in Tables 18 and 19, respectively.[56] The in-store own-price elasticities are overwhelmingly neg-

or across equations (Greene 2012). There are three cases in which there are no efficiency gains from generalized least squares relative to equation by equation ordinary least squares: (1) if the disturbance terms across equations are independent of each other, (2) the independent variables in each equation are identical and (3) if the regressors of one equation are a subset of the regressors in another equation (Greene 2012). My demand system satisfies the second of these conditions because each of the share equations contains the same independent variables.

[54]Note that the generalized least squares procedure only accounts for the correlation of errors across equations. Hence, the standard errors are not robust to heteroskedasticity of the errors within each equation. If this is an area of concern, the standard errors could be corrected by employing a clustered bootstrap.

[55]The t-statistics for the first-stages range from 4.6 to 60.1, in absolute value.

[56]The elasticity matrices are not symmetric because I am computing Marshallian elasticities, which hold income (as opposed to utility for Hicksian demand) constant. Hence, the estimated cross-price elasticities reflect a symmetric price effect and an income effect, which is not necessarily symmetric.

ative and statistically significant; additionally, the negative and statistically significant in-store own-price elasticities range from -1.91 (drinks) to -0.66 (sugars).[57] In contrast, the online own-price elasticities are much less likely to be statistically significant. In fact, the only own-price elasticities that are significant for the online matrix are dairy (-0.66), drinks (-0.65), fruit (-0.81) and snacks/sweets (-1.53).[58]

Figure 5 presents the ratio of the in-store price elasticity to the online price elasticity.[59] Specifically, each cell is colored black if $|\eta_{Instore}|$ is significantly different and larger in magnitude than $|\eta_{Online}|$, white if $\eta_{Instore}$ is not statistically different from η_{Online} and gray if $|\eta_{Instore}|$ is significantly different and smaller in magnitude than $|\eta_{Online}|$. Figure 5 illustrates that roughly one-third of the elasticities are significantly different for the months in which a household engages in online shopping. For the product categories of dairy, drinks, and fruits the own-price elasticities are larger (in absolute value) in-store than they are online. Many of the cross-price elasticities that are significantly different across the two shopping environments exhibit price substitution patterns across product categories are significantly different and larger in magnitude when the household is shopping in-store only. Of the own-price elasticities that are significantly different across the two purchasing environments, the in-store own-price elasticities are, on average, 2.0 times larger (in absolute value) than the online own-price elasticities.[60] Additionally, the

[57]The two exceptions to this statement are the own-price elasticity for grains (0.29) and for prepared (0.62) both of which are statistically significant. Harding and Lovenheim (2017) also find positive and significant own-price elasticities for prepared foods and cereal utilizing similar data and a similar estimation strategy. In general, the estimated in-store elasticities are aligned with the current food demand literature.

[58]Figures A13 and A14 present visual representations of the two elasticity matrices. In these figures, darker shades represent negative elasticities, lighter shades represent positive elasticities and white represents statistically insignificant elasticities. The own-price elasticities are presented in the left column and run along the diagonal of these matrices from the corner of the second to left hand column to the upper right hand corner. Notice that the in-store own-price elasticities are overwhelmingly significant and negative (with the exception of the product categories of prepared and grains), while the own-price elasticities in the online price elasticity matrix are a combination of white (statistically insignificant) and darker shades (negative). Additionally, comparing the cross-price elasticities of these two matrices, the in-store elasticity matrix illustrates many statistically significant cross-price substitution patterns, while the online price elasticity matrix illustrates that many of these substitution patterns become insignificantly different from zero in months when a household shops both in-store and online.

[59]Table A30 presents the difference between the in-store and online price elasticities as well as the standard errors for that difference.

[60]The in-store own-price elasticities are, on average, 0.15 times larger (in absolute value) than the online own-price elasticities when incorporating all own-price elasticities.

in-store cross price elasticities are, on average, 2.8 times larger (in absolute value) than the online cross-price elasticities.[61]

These findings illustrate that consumers are more price sensitive in-store than they are online. Consumers may be less sensitive to prices when shopping online due to convenience factors built into the online shopping experience and the search differences that exist across the two environments. For example, when shopping online a household may utilize features of the website that allow them to instantly generate their basket based off of their previous online order.[62] An online shopper, who values convenience, may have a tendency to re-create their previous order without checking item prices and/or prices of substitutes. Furthermore, decreased price-sensitivity when shopping online occurs for product categories that contain goods that a household would likely add to an online favorites list. For example, products like milk, coffee, and bananas sound like reasonable candidates for an instant add to cart list.

The results of this paper are also consistent with the current e-commerce literature. Specifically, Pozzi (2012) finds that in-store own-price elasticities for cereal are roughly fifty percent higher than online own-price elasticities, while in-store cross price elasticities are nearly three times as large as online cross-price elasticities. Additionally, Chu et al. (2008) find evidence, across twelve different product categories, that households tend to be less sensitive to prices when shopping online.[63]

[61] This average only includes the elasticities that are significantly different from each other and excludes two outliers. The excluded outliers are the two cross price elasticities between oil and fruit ($\eta_{oil,fruit}$ and $\eta_{fruit,oil}$). These elasticities are relatively large when households shop in-store only (0.40 and 0.77) and get quite close to zero in the months in which households shop online; hence, the ratio of the in-store to online price elasticity is quite large for these observations. The in-store cross-price elasticities are, on average, 3.6 times larger (in absolute value) than the online cross-price elasticities when incorporating all cross-price elasticities.

[62] A household may also use the search engine to find a particular brand of an item and add that item to their cart without ever looking at the prices of competing brands. Alternatively, other features of the online website may make it easier for households to find sales; specifically, households can filter their search results to items that are currently on sale or sort their search results according to price.

[63] Specifically, they find that in-store elasticities are higher (in absolute value) for the product categories of yogurt, milk, cooking oil, toilet paper, square bread, eggs, fabric softener, oranges, paper towels, potatoes and dish detergent.

7 Discussion & Conclusion

This paper utilizes novel household panel data to analyze the effect of online grocery shopping on grocery demand. In order to identify a causal effect of online grocery shopping, I utilize variation in the introduction of an online shopping service, at different store locations, as a source of exogenous variation in the household's decision to shop online. Estimates from panel difference-in-differences and instrumental variables estimation strategies indicate that households allocate a significantly larger share of their total grocery expenditures toward healthier product categories (dairy, fruit, meats, vegetables) at the expense of more indulgent product categories (drinks, oils, snacks/sweets) when shopping online. In addition, I find that the nutritional content of grocery purchases is improved in the months in which households shop online. Estimated declines in the average amount of calories purchased imply that the average adult would lose roughly five pounds over the course of a year simply from engaging in online grocery shopping.[64]

Motivated by the reduced form findings, I then estimate a formal model of grocery demand and test for differences in price elasticities across the two purchasing environments. I find that households are generally less price sensitive when shopping online. Specifically, in-store own-price (cross-price) elasticities are, on average, two (three) times larger than the estimated online price elasticities. These findings indicate that households are less price sensitive and exhibit weaker substitution patterns when shopping online.

This paper illustrates that consumer demand is sensitive to differences in search and the varying levels of product placement that exist across purchasing environments. I find that when consumers shop online, they are less likely to purchase indulgent products and they are less sensitive to prices. Differences between in-store and online search are likely a large contributing factor to this result. Relative to the online purchasing environment, there are many elements of in-store design (checkout lanes, end of aisle displays, store layout) that dictate the consumer search path, influence the intensity of price-shopping and nudge consumers towards purchases. My findings suggest that product placement

[64] This statement assumes that the declines in calories purchased perfectly translate to declines in caloric intake.

is effective in influencing purchases and initiatives that promote healthier choices (via product placement either in-store or online) could improve the quality of food purchases made. Furthermore, the structural findings of this paper suggest that the marginal cost of time dedicated to search is different between the online and in-store shopping environments. Hence, a more sophisticated online pricing strategy, that incorporates the fact that the value of convenience appears to be different across the two purchasing environments, would likely lead to increased online revenues.

8 References

Atkin, David. "Trade, tastes, and nutrition in India." The American Economic Review 103.5 (2013): 1629-1663.

Baye, Michael R., John Morgan, and Patrick Scholten. "The value of information in an online consumer electronics market." Journal of Public Policy & Marketing 22.1 (2003): 17-25.

Bakos, J. Yannis. "Reducing buyer search costs: Implications for electronic marketplaces." Management science 43.12 (1997): 1676-1692.

Bazerman, Max H., Ann E. Tenbrunsel, and Kimberly Wade-Benzoni. "Negotiating with yourself and losing: Making decisions with competing internal preferences." Academy of Management Review 23.2 (1998): 225-241.

Borusyak, K., Jaravel, X. (2016). "Revisiting Event Study Designs." Available at SSRN: https://ssrn.com/abstract=2826228.

Brown, Jeffrey R., and Austan Goolsbee. "Does the Internet make markets more competitive? Evidence from the life insurance industry." Journal of Political Economy 110.3 (2002): 481-507.

Brynjolfsson, Erik, and Michael D. Smith. "Consumer decision?making at an Internet shopbot: Brand still matters." The Journal of Industrial Economics 49.4 (2001): 541-558.

Brynjolfsson, Erik, Yu Hu, and Michael D. Smith. "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers." Management Science 49.11 (2003): 1580-1596.

Brynjolfsson, Erik, and Michael D. Smith. "Frictionless commerce? A comparison of Internet and conventional retailers." Management Science 46.4 (2000): 563-585.

Chalfant, J. A. (1987). "A globally flexible, almost ideal demand system." Journal of Business Economic Statistics, 5(2), 233-242.

Chu, Junhong, Pradeep Chintagunta, and Javier Cebollada. "Research Note - A Comparison of Within-Household Price Sensitivity Across Online and Offline Channels." Marketing Science 27.2 (2008): 283-299.

Chu, Junhong, et al. "An empirical analysis of shopping behavior across online and offline channels for grocery products: the moderating effects of household and product characteristics." Journal of Interactive Marketing 24.4 (2010): 251-268.

Clemons, Eric K., Il-Horn Hann, and Lorin M. Hitt. "Price dispersion and differentiation in online travel: An empirical investigation." Management science 48.4 (2002): 534-549.

Deaton, Angus, and John Muellbauer. "An almost ideal demand system." The American Economic Review 70.3 (1980): 312-326.

Ellison, G., Fisher Ellison, S. (2005). Lessons about Markets from the Internet. Journal of Economic Perspectives, 19(2), 139-158.

Harding, M., Lovenheim, M. (2017). The effect of prices on nutrition: comparing the impact of product-and nutrient-specific taxes. Journal of Health Economics, 53, 53-71.

Hoch, Stephen J., and George F. Loewenstein. "Time-inconsistent preferences and consumer self-control." Journal of Consumer Research 17.4 (1991): 492-507.

Huyghe, Elke, et al. "Clicks as a Healthy Alternative to Bricks: How Online Grocery Shopping Reduces Vice Purchases." Journal of Marketing Research 54.1 (2017): 61-74.

Just, D. R., Mancino, L., Wansink, B. (2007). Could behavioral economics help improve diet quality for nutrition assistance program participants? (Vol. 43). Washington, DC: US Department of Agriculture, Economic Research Service.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. The Quarterly Journal of Economics, 112(2), 443-478.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. Organizational behavior and human decision processes, 65(3), 272-292.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman. "I'll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time." Marketing Letters 21.1 (2010): 17-35.

Mischel, Walter, Ebbe B. Ebbesen, and Antonette Raskoff Zeiss. "Cognitive and attentional mechanisms in delay of gratification." Journal of Personality and Social Psychology 21.2 (1972): 204.

Nielsen & Food Marketing Institute. "The Digitally Engaged Food Shopper." (2017)

Pozzi, Andrea. "Shopping cost and brand exploration in online grocery." American Economic Journal: Microeconomics 4.3 (2012): 96-120.

Pozzi, Andrea. "The effect of Internet distribution on brick and mortar sales." The RAND Journal of Economics 44.3 (2013a): 569-583.

Pozzi, Andrea. "E-commerce as a stockpiling technology: Implications for consumer savings." International Journal of Industrial Organization 31.6 (2013b): 677-689.

Quan, Thomas W., and Kevin R. Williams. "Product variety, across-market demand heterogeneity, and the value of online retail." (2017).

Schelling, Thomas C. "Self-command in practice, in policy, and in a theory of rational choice." The American Economic Review 74.2 (1984): 1-11.

Shiv, Baba, and Alexander Fedorikhin. "Heart and mind in conflict: The interplay of affect and cognition in consumer decision making." Journal of Consumer Research 26.3 (1999): 278-292.

Shiv, Baba, and Alexander Fedorikhin. "Spontaneous versus controlled influences of stimulus-based affect on choice behavior." Organizational Behavior and Human Decision Processes 87.2 (2002): 342-370.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. Economics letters, 8(3), 201-207.

Thaler, Richard H., and Hersh M. Shefrin. "An economic theory of self-control." Journal of Political Economy 89.2 (1981): 392-406.

Zentner, Alejandro, Michael Smith, and Cuneyd Kaya. "How video rental patterns change as consumers move online." Management Science 59.11 (2013): 2622-2634.

9 Figures

Figure 1: Online Shopping Service Availability

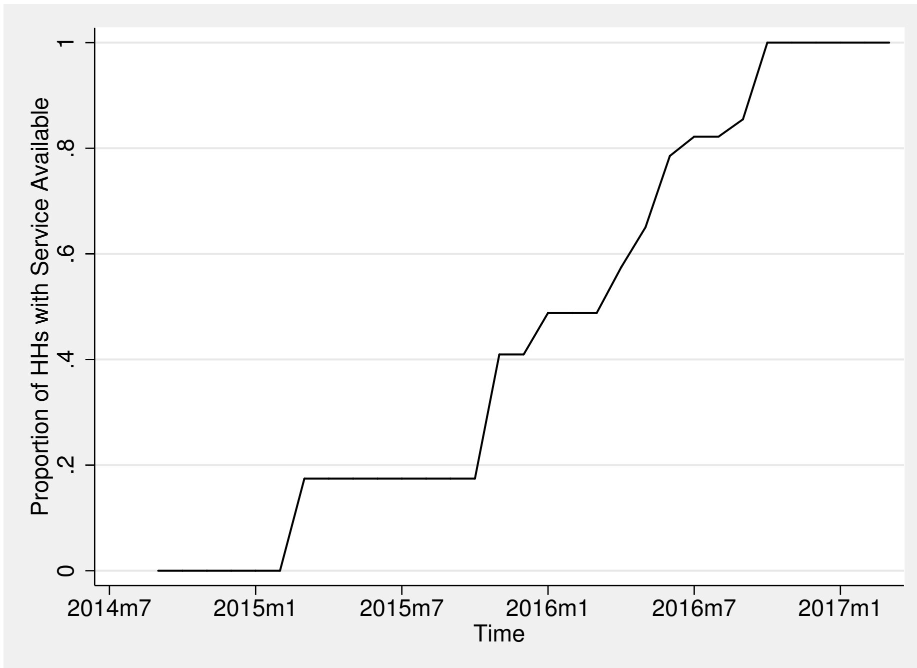

Figure 1 illustrates the proportion of households who have access to the online purchasing service over time.

Figure 2: Event Study Estimates for Online Use and Expenditure Shares

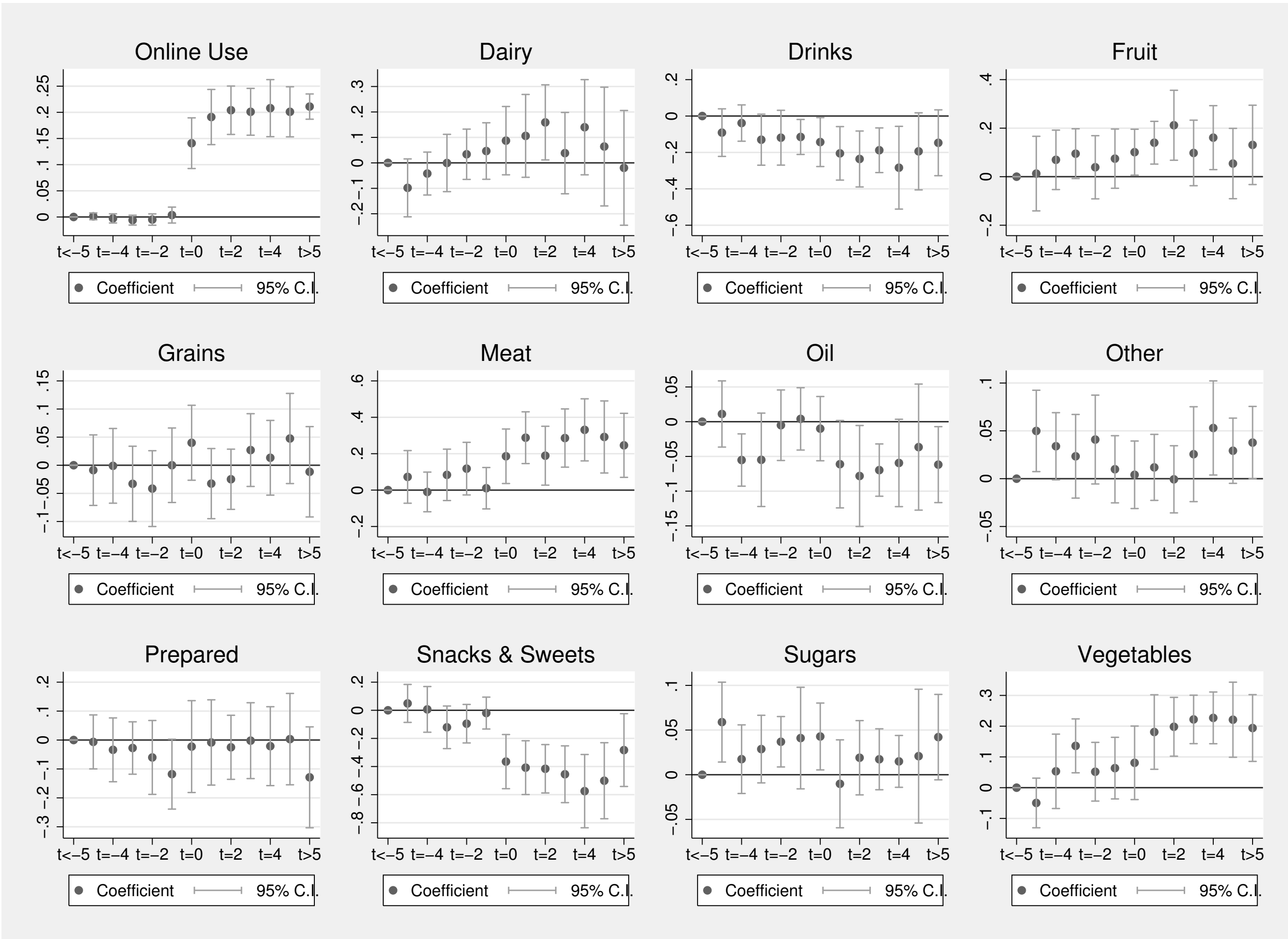

Figure 2 presents the event study estimates of the effect of online availability for online households.

Figure 3: Percent of Available Products Purchased Online

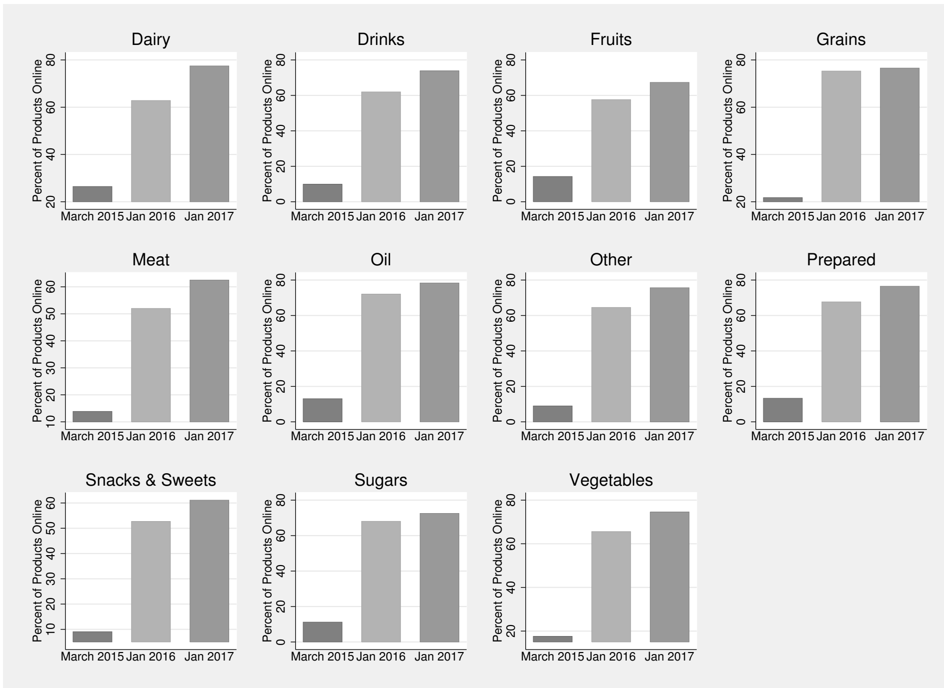

Figure 3 presents the percentage of upcs purchased both in-store and online in March 2015, January 2016 and January 2017.

Figure 4: Percent of In-Store Sales from Products Available Online

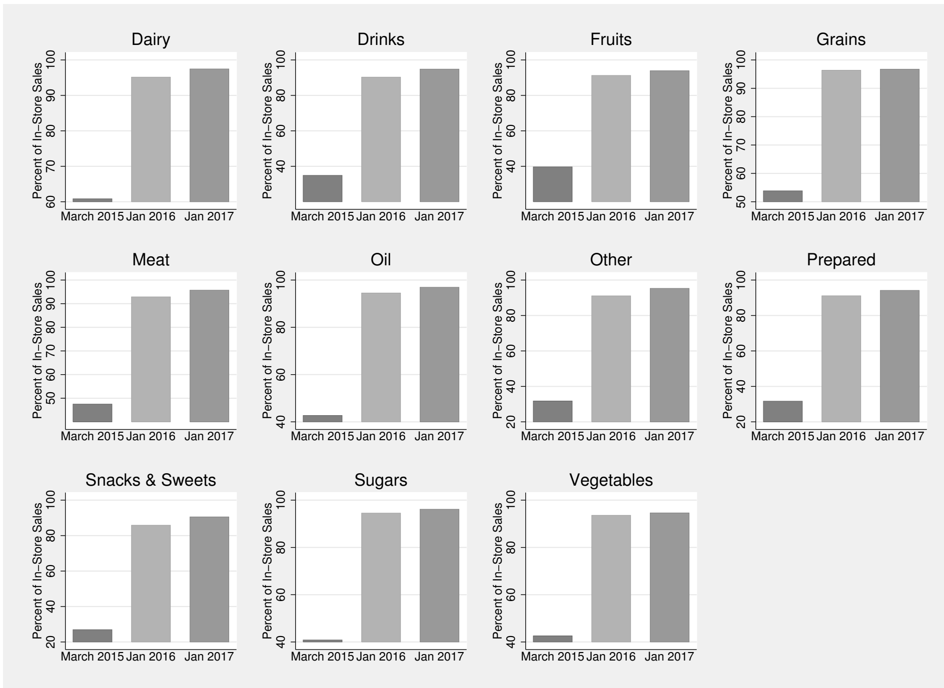

Figure 4 presents the percentage of in-store sales that are generated from upcs that have also been purchased online. This figure illustrates that by January 2016, the most popularly purchased in-store products are also available online.

Figure 5: Own and Cross Price Elasticity Ratio, $\frac{|\eta_{Instore}|}{|\eta_{Online}|}$

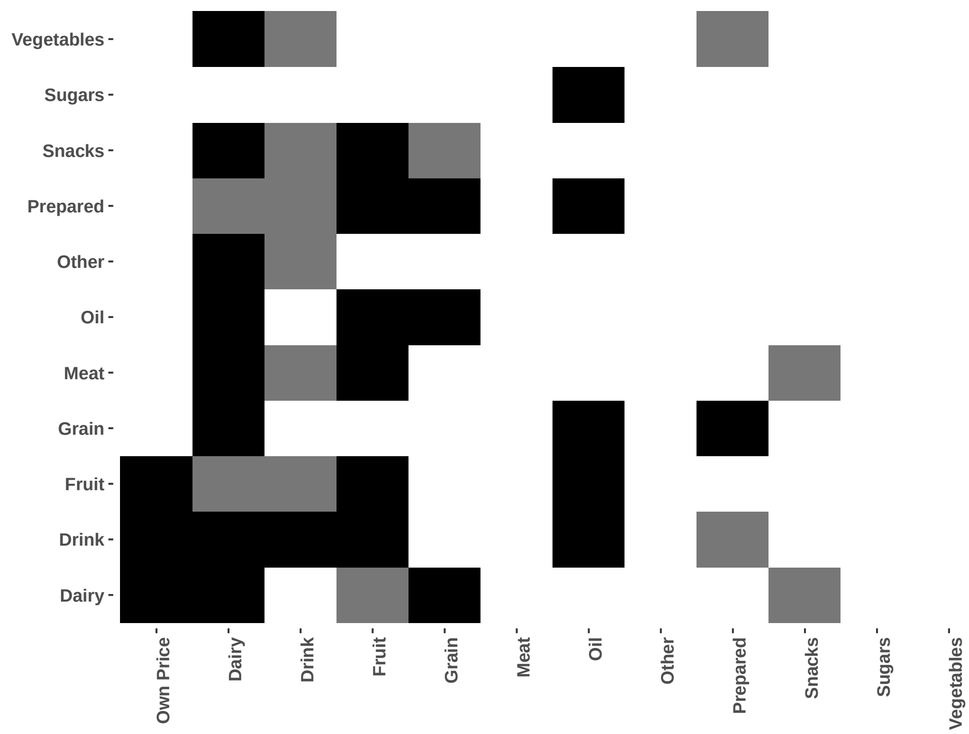

Figure 5 the ratio of the in-store and online own and cross price elasticities. Black indicates that the in-store elasticity is significantly different and greater, in absolute value, than the online elasticity. White means that the two elasticities are not significantly different from each other, at the 95% confidence level, and gray indicates that the in-store elasticity is significantly different and smaller, in absolute value, than the online elasticity.

10 Tables

Table 1: Household Demographics

Household Population	All HHs	In-Store Only HHs	Online Adoption HHs	Online Households Wave 1	Wave 2	Wave 3
1{Married}	0.68	0.67	0.69	0.70	0.68	0.67
Household Size	**All**	**In-Store**	**Online**	**Wave 1**	**Wave 2**	**Wave 3**
1{1 Person}	0.10	0.12	0.10	0.09	0.10	0.10
1{2 People}	0.22	0.22	0.22	0.21	0.21	0.23
1{3 People}	0.24	0.22	0.25	0.25	0.25	0.24
1{4 People}	0.17	0.17	0.17	0.17	0.18	0.17
1{5+ People}	0.26	0.26	0.27	0.27	0.26	0.25
Household Income	**All**	**In-Store**	**Online**	**Wave 1**	**Wave 2**	**Wave 3**
1{0-29K}	0.11	0.14	0.10	0.09	0.10	0.10
1{30-50K}	0.16	0.17	0.16	0.16	0.15	0.17
1{51-79K}	0.30	0.28	0.31	0.31	0.31	0.32
1{80-99K}	0.15	0.15	0.15	0.16	0.15	0.14
1{100-149K}	0.12	0.12	0.12	0.12	0.13	0.13
1{150K+}	0.16	0.14	0.16	0.17	0.17	0.13
Number of Children	**All**	**In-Store**	**Online**	**Wave 1**	**Wave 2**	**Wave 3**
1{0 Children}	0.43	0.49	0.41	0.40	0.41	0.43
1{1 Child}	0.33	0.31	0.33	0.33	0.33	0.33
1{2 Children}	0.14	0.13	0.15	0.15	0.15	0.14
1{3 Children}	0.07	0.06	0.08	0.08	0.08	0.07
1{4+ Children}	0.03	0.02	0.04	0.04	0.04	0.03
Household Head Age	**All**	**In-Store**	**Online**	**Wave 1**	**Wave 2**	**Wave 3**
1{18-25}	0.02	0.02	0.02	0.03	0.02	0.02
1{26-35}	0.22	0.14	0.25	0.27	0.24	0.24
1{36-45}	0.28	0.20	0.32	0.31	0.32	0.32
1{46-55}	0.20	0.24	0.19	0.19	0.20	0.18
1{56-55}	0.16	0.21	0.14	0.14	0.14	0.15
1{66+}	0.11	0.18	0.08	0.07	0.08	0.09
Household Count	34,797	9,777	25,020	13,208	7,267	4,545

Table 2: Monthly Purchasing Patterns

Time Period	Before & After Online Introduction	Before Online Introduction		Before Online Introduction		
Household Population	All HHs	In-Store Only HHs	Online Adoption HHs	Online Adoption HHs Wave 1	Wave 2	Wave 3
Monthly Shopping Habits	All	In-Store	Online	Wave 1	Wave 2	Wave 3
Grocery Expenditure ($)	437	331	448	467	420	434
Items Purchased	176	135	178	187	166	172
Visits to Store	7.7	6.8	7.5	8.0	6.9	7.1
Share of Sales Online	0.0	0.0	0.0	0.0	0.0	0.0
Share of Expenditure	All	In-Store	Online	Wave 1	Wave 2	Wave 3
Dairy	12.5	11.9	13.2	13.4	13.1	12.9
Drink	10.8	10.9	10.1	9.9	10.2	10.4
Fruit	8.3	7.3	7.5	7.6	7.5	7.1
Grains	7.5	7.2	7.6	7.6	7.6	7.6
Meats	18.5	18.9	18.5	18.3	18.5	18.9
Oils	4.4	4.4	4.4	4.4	4.3	4.4
Other	1.6	1.7	1.8	1.8	1.8	1.8
Prepared	10.1	10.9	10.8	10.7	10.8	10.9
Snacks/Sweets	15.8	16.1	15.5	15.4	15.6	15.7
Sugars	1.5	1.6	1.6	1.6	1.6	1.6
Vegetables	9.1	9.2	9.1	9.3	9.0	8.6
Observations	855,022	56,972	147,246	77,925	42,608	26,713

Table 3: Average Price and Nutrient Content Per Ounce of Food

Product Category	Price ($)	Calories (kcal)	Protein (g)	Carbs (g)	Total Fat (g)	Total Sugar (g)	Sodium (g)	Cholesterol (g)
Dairy	0.10	71	3.9	4.2	4.3	2.2	0.2	0.01
Drinks	0.07	22	0.2	4.5	0.2	3.9	0.0	0.00
Fruit	0.10	23	0.3	5.5	0.2	4.1	0.0	0.00
Grains	0.13	97	2.7	18.5	1.5	3.4	0.2	0.00
Meat	0.23	81	7.7	2.1	4.7	0.8	0.2	0.03
Oil	0.15	88	0.3	4.2	7.7	3.9	0.3	0.00
Other	0.46	73	1.4	11.5	2.5	5.4	1.9	0.00
Prepared	0.16	68	2.4	9.8	2.1	1.3	0.3	0.00
Snacks & Sweets	0.21	120	1.8	17.8	4.9	9.5	0.1	0.01
Sugars	0.11	98	1.4	18.1	3.6	14.9	0.0	0.00
Vegetables	0.09	22	0.7	4.1	0.4	1.5	0.3	0.00

Table 4: Difference in Difference / Reduced Form Estimates, (τ_i)
Online Households

Budget Shares	(1)	(2)	(3)	Avg. Budget Share (Pre-Online Service)
Dairy	0.104**	0.105**	0.0969**	13.21
	(0.0386)	(0.0387)	(0.0401)	(6.90)
Drinks	-0.103**	-0.104**	-0.100**	10.08
	(0.0380)	(0.0382)	(0.0384)	(8.31)
Fruit	0.0844**	0.0846**	0.0845**	7.48
	(0.0362)	(0.0362)	(0.0369)	(6.18)
Grain	0.0242	0.0249	0.0211	7.6
	(0.0152)	(0.0152)	(0.0148)	(4.66)
Meat	0.204***	0.205***	0.204***	18.46
	(0.0580)	(0.0581)	(0.0586)	(9.69)
Oil	-0.0350**	-0.0354**	-0.0352**	4.36
	(0.0143)	(0.0143)	(0.0150)	(3.56)
Other	-0.0121	-0.0123	-0.00970	1.81
	(0.00771)	(0.00767)	(0.00802)	(2.72)
Prepared	0.0241	0.0262	0.0305	10.78
	(0.0420)	(0.0425)	(0.0427)	(7.45)
Snacks/Sweets	-0.408***	-0.413***	-0.406***	15.52
	(0.0665)	(0.0681)	(0.0677)	(9.84)
Sugar	-0.0141	-0.0144	-0.0151	1.61
	(0.00957)	(0.00960)	(0.00976)	(2.32)
Vegetables	0.132***	0.134***	0.129***	9.09
	(0.0310)	(0.0308)	(0.0324)	(6.20)
Time Availability f.e.	X			
Treatment Cohort f.e.	X	X		
Household Demographics	X	X		
Year-Month f.e.		X	X	
Household f.e.			X	
Observations	616,357	616,357	616,357	147,246

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell represents an estimate of the effect of online availability on the budget shares.
These estimates were produced by running OLS equation by equation.
Equation by equation OLS is equivalent to SUR in this scenario.

Table 5: Difference in Difference / First Stage Estimates, (θ_i)
Online Households

1{Online}	(1)	(2)	(3)
1{OnlineAvail}	0.193***	0.193***	0.193***
	(0.0216)	(0.0216)	(0.0221)
Time Availability f.e.	X		
Treatment Cohort f.e.	X	X	
Household Demographics	X	X	
Year-Month f.e.		X	X
Household f.e.			X
Observations	616,357	616,357	616,357
R-squared	0.166	0.166	0.306
F-Statistic	79.80	79.56	76.32

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i
Online Households

Budget Shares	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Avg. Budget Share (Pre-Online Service)	Percent Change
Dairy	0.784***	0.786***	0.496***	0.539***	0.543***	0.503**	13.21	3.79
	(0.0381)	(0.0381)	(0.0544)	(0.202)	(0.202)	(0.208)	(6.90)	
Drinks	-0.709***	-0.713***	-0.561***	-0.535**	-0.538**	-0.520**	10.08	-5.16
	(0.0559)	(0.0536)	(0.0422)	(0.223)	(0.225)	(0.224)	(8.31)	
Fruit	0.317***	0.336***	0.114***	0.437**	0.438**	0.439**	7.48	5.88
	(0.0564)	(0.0608)	(0.0284)	(0.209)	(0.210)	(0.211)	(6.18)	
Grain	0.239***	0.242***	0.162***	0.125	0.129*	0.110	7.6	1.45
	(0.0232)	(0.0233)	(0.0227)	(0.0775)	(0.0772)	(0.0758)	(4.66)	
Meat	0.331***	0.330***	0.539***	1.056***	1.060***	1.060***	18.46	5.74
	(0.0883)	(0.0929)	(0.0714)	(0.311)	(0.310)	(0.314)	(9.69)	
Oil	-0.0466**	-0.0452**	-0.0112	-0.181**	-0.183**	-0.183**	4.36	-4.13
	(0.0212)	(0.0216)	(0.0202)	(0.0767)	(0.0772)	(0.0807)	(3.56)	
Other	0.00392	0.000906	0.00842	-0.0626	-0.0637	-0.0503	1.81	-2.78
	(0.0152)	(0.0146)	(0.0113)	(0.0400)	(0.0400)	(0.0416)	(2.72)	
Prepared	-0.00622	-0.00580	0.0948	0.125	0.136	0.158	10.78	1.47
	(0.101)	(0.104)	(0.0714)	(0.209)	(0.211)	(0.210)	(7.45)	
Snacks/Sweets	-1.593***	-1.623***	-1.507***	-2.113***	-2.140***	-2.106***	15.52	-13.60
	(0.124)	(0.129)	(0.128)	(0.315)	(0.316)	(0.315)	(9.84)	
Sugar	0.0384***	0.0380***	0.0149*	-0.0732	-0.0745	-0.0783	1.61	-4.86
	(0.0105)	(0.0109)	(0.00844)	(0.0499)	(0.0501)	(0.0506)	(2.32)	
Vegetables	0.643***	0.656***	0.652***	0.684***	0.694***	0.668***	9.09	7.37
	(0.0701)	(0.0695)	(0.0454)	(0.166)	(0.165)	(0.173)	(6.20)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	616,357	616,357	616,357	616,357	616,357	616,357	147,246	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate from a separate regression; note that OLS (2SLS) is equivalent to SUR (3SLS) in this scenario. The percent change is calculated utilizing the estimates provided in Column 6 and average pre-online service budget shares.

Table 7: Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i
Online Households - Nutrition Outcomes

Nutrients per Ounce	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Average (Pre-Online Service)	Percent Change
Calories	-1.548***	-1.600***	-1.374***	-2.087***	-2.122***	-1.974***	47.44	-4.16
	(0.138)	(0.133)	(0.133)	(0.350)	(0.351)	(0.345)	(13.81)	
Carbohydrates	-0.220***	-0.225***	-0.205***	-0.307***	-0.311***	-0.296***	5.99	-4.94
	(0.0246)	(0.0244)	(0.0213)	(0.0496)	(0.0504)	(0.0498)	(1.96)	
Cholesterol	0.000135**	0.000135**	2.06e-05	-0.000484	-0.000482	-0.000464	0.005	-9.28
	(5.88e-05)	(5.76e-05)	(5.58e-05)	(0.000374)	(0.000373)	(0.000387)	(0.01)	
Protein	0.00170	0.000553	0.00307	0.0226	0.0223	0.0237*	1.74	1.36
	(0.00376)	(0.00409)	(0.00308)	(0.0144)	(0.0143)	(0.0144)	(0.60)	
Sodium	-0.00370	-0.00345	-0.00229	-0.0167	-0.0169	-0.0173	0.17	-10.18
	(0.00459)	(0.00471)	(0.00349)	(0.0172)	(0.0170)	(0.0177)	(0.67)	
Total Fat	-0.0812***	-0.0843***	-0.0695***	-0.117***	-0.119***	-0.110***	1.90	-5.79
	(0.00637)	(0.00608)	(0.00697)	(0.0195)	(0.0195)	(0.0189)	(0.82)	
Total Sugar	-0.114***	-0.116***	-0.114***	-0.134***	-0.136***	-0.130***	2.26	-5.75
	(0.0158)	(0.0165)	(0.0130)	(0.0384)	(0.0391)	(0.0391)	(1.06)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	596,112	596,112	596,112	596,112	596,112	596,112	142,301	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate from a separate regression; note that OLS (2SLS) is equivalent to SUR (3SLS) in this scenario.
The percent change is calculated utilizing the estimates provided in Column 6 and average pre-online service nutrition outcomes.

Table 8: Online Households - Expenditure

Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Average (Pre-Online Service)	Percent Change
Total Expenditure	56.17***	55.94***	48.64***	51.93***	52.11***	48.87***	447.52	10.92
	(1.950)	(2.050)	(2.580)	(17.84)	(17.70)	(17.63)	(328.11)	
Dairy	10.03***	10.01***	7.631***	8.059***	8.095***	7.662***	58.58	13.08
	(0.317)	(0.317)	(0.319)	(2.605)	(2.593)	(2.586)	(47.25)	
Drinks	4.079***	4.036***	4.089***	4.346***	4.352***	3.957**	43.61	9.07
	(0.413)	(0.440)	(0.504)	(1.606)	(1.606)	(1.612)	(41.12)	
Fruit	6.807***	6.881***	4.957***	7.469**	7.495**	7.238**	33.04	21.91
	(0.533)	(0.547)	(0.257)	(3.319)	(3.300)	(3.298)	(32.35)	
Grain	4.994***	4.990***	4.109***	4.048***	4.080***	3.836***	33.86	11.33
	(0.130)	(0.127)	(0.144)	(1.313)	(1.299)	(1.288)	(28.49)	
Meat	10.07***	10.02***	9.808***	12.40***	12.44***	11.82***	86.73	13.63
	(0.588)	(0.636)	(0.646)	(4.423)	(4.386)	(4.387)	(77.43)	
Oil	2.103***	2.098***	1.946***	1.800***	1.798***	1.683***	19.67	8.56
	(0.103)	(0.101)	(0.0813)	(0.621)	(0.609)	(0.599)	(18.40)	
Other	0.926***	0.908***	0.780***	0.668	0.664	0.645	8.15	7.91
	(0.121)	(0.117)	(0.0677)	(0.414)	(0.423)	(0.443)	(10.72)	
Prepared	4.847***	4.829***	4.692***	2.947**	3.016**	2.733*	48.77	5.60
	(0.587)	(0.621)	(0.521)	(1.497)	(1.522)	(1.550)	(45.27)	
Snacks/Sweets	3.799***	3.626***	3.250***	1.386	1.287	0.764	67.4	1.13
	(0.381)	(0.386)	(0.379)	(2.535)	(2.517)	(2.536)	(58.88)	
Sugar	1.082***	1.076***	0.818***	0.633***	0.629***	0.577**	7.21	8.00
	(0.0721)	(0.0729)	(0.0454)	(0.225)	(0.226)	(0.234)	(8.93)	
Vegetables	7.428***	7.465***	6.562***	8.180***	8.246***	7.963***	40.52	19.65
	(0.429)	(0.428)	(0.336)	(1.886)	(1.870)	(1.877)	(36.55)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	616,357	616,357	616,357	616,357	616,357	616,357	147,246	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate from a separate regression; note that OLS (2SLS) is equivalent to SUR (3SLS) in this scenario. The percent change is calculated utilizing the estimates provided in Column 6 and average pre-online service expenditures.

Table 9: Difference in Difference / Reduced Form Estimates, (τ_i)
Online Households, Low Competition Stores

	(1)	(2)	(3)	Pre-Online Average	Percent Change
Monthly Exp. ($)	4.120	4.015	3.321	480.39	0.69
	(2.733)	(2.704)	(2.599)	(339.57)	
Budget Shares	(1)	(2)	(3)	Pre-Online Average	Percent Change
Dairy	0.0185	0.0190	0.0116	13.39	0.09
	(0.0757)	(0.0759)	(0.0789)	(6.55)	
Drinks	-0.0324	-0.0339	-0.0326	9.89	-0.33
	(0.0604)	(0.0601)	(0.0591)	(7.72)	
Fruit	0.0925**	0.0933**	0.101**	7.5	1.35
	(0.0361)	(0.0360)	(0.0360)	(5.84)	
Grain	0.0185	0.0187	0.0179	7.66	0.23
	(0.0232)	(0.0232)	(0.0244)	(4.41)	
Meat	0.224**	0.222**	0.216**	18.42	1.17
	(0.0802)	(0.0796)	(0.0801)	(9.14)	
Oil	-0.0283	-0.0288	-0.0296	4.4	-0.67
	(0.0276)	(0.0278)	(0.0283)	(3.26)	
Other	-0.0115	-0.0119	-0.00969	1.83	-0.53
	(0.0118)	(0.0119)	(0.0126)	(2.56)	
Prepared	-0.0857	-0.0844	-0.0824	10.67	-0.77
	(0.0497)	(0.0494)	(0.0498)	(7.03)	
Snacks/Sweets	-0.246***	-0.246***	-0.240***	15.49	-1.55
	(0.0707)	(0.0700)	(0.0696)	(9.23)	
Sugar	-0.00741	-0.00755	-0.00718	1.61	-0.45
	(0.00895)	(0.00889)	(0.00946)	(2.16)	
Vegetables	0.0599	0.0613	0.0567	9.13	0.62
	(0.0539)	(0.0538)	(0.0554)	(5.85)	
Time Availability f.e.	X				
Treatment Cohort f.e.	X	X			
Household Demographics	X	X			
Year-Month f.e.		X	X		
Household f.e.			X		
Observations	298,722	298,722	298,722	71,436	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability.
These estimates were produced equation by equation.
Percent change is derived utilizing the estimates in column (3) and the pre-online service average.

Table 10: Difference in Difference / Reduced Form Estimates, (τ_i)
Online Households, High Competition Stores

	(1)	(2)	(3)	Pre-Online Average	Percent Change
Monthly Exp. ($)	11.31***	11.51***	10.59***	420.24	2.52
	(3.346)	(3.403)	(3.385)	(314.68)	
Budget Shares	(1)	(2)	(3)	Pre-Online Average	Percent Change
Dairy	0.0970*	0.0982*	0.0856	13.09	0.65
	(0.0525)	(0.0529)	(0.0538)	(7.16)	
Drinks	-0.0826	-0.0823	-0.0805	10.21	-0.79
	(0.0513)	(0.0513)	(0.0542)	(8.78)	
Fruit	0.0613	0.0610	0.0620	7.48	0.83
	(0.0554)	(0.0556)	(0.0568)	(6.48)	
Grain	0.0489*	0.0504*	0.0434*	7.53	0.58
	(0.0253)	(0.0253)	(0.0253)	(4.85)	
Meat	0.205**	0.208**	0.205**	18.53	1.11
	(0.0882)	(0.0888)	(0.0887)	(10.13)	
Oil	-0.0318	-0.0320	-0.0306	4.32	-0.71
	(0.0221)	(0.0222)	(0.0231)	(3.75)	
Other	-0.0132	-0.0129	-0.00983	1.79	-0.55
	(0.0141)	(0.0140)	(0.0145)	(2.81)	
Prepared	0.0485	0.0520	0.0572	10.87	0.53
	(0.0631)	(0.0641)	(0.0647)	(7.76)	
Snacks/Sweets	-0.495***	-0.506***	-0.488***	15.49	-3.15
	(0.0964)	(0.0994)	(0.0978)	(10.31)	
Sugar	-0.0145	-0.0147	-0.0175	1.62	-1.08
	(0.0146)	(0.0146)	(0.0149)	(2.45)	
Vegetables	0.175***	0.177***	0.172***	9.07	1.90
	(0.0513)	(0.0505)	(0.0544)	(6.45)	
Time Availability f.e.	X				
Treatment Cohort f.e.	X	X			
Household Demographics	X	X			
Year-Month f.e.		X	X		
Household f.e.			X		
Observations	359,571	359,571	359,571	85,823	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell represents an estimate of the effect of online availability.
These estimates were produced equation by equation.
Percent change is derived utilizing the estimates in column (3) and the pre-online service average.

Table 11: Average Nutrient Content Per Ounce of Food
Narrow Product Categories

Bread Type	Healthier Choices	Calories (kcal)	Protein (g)	Carbs (g)	Total Fat (g)	Total Sugar (g)	Sodium (g)
Whole Grain	X	68	2.8	12.7	1.2	1.8	0.1
Other		74	2.4	14.0	1.0	2.3	0.1
Seed / Nut	X	67	2.8	12.2	1.2	1.5	0.1
Wheat	X	67	2.9	12.8	0.8	1.8	0.1
White		69	2.3	13.6	1.0	1.7	0.1

Cereal Type	Healthier Choices	Calories (kcal)	Protein (g)	Carbs (g)	Total Fat (g)	Total Sugar (g)	Sodium (g)
Kids		112	1.4	23.8	1.4	10.1	0.1
Organic Kids		108	2.1	22.9	1.5	6.9	0.1
Standard	X	106	2.4	23.2	0.8	2.7	0.2
Frosted Standard		108	2.3	22.9	1.3	7.3	0.1
Healthy Cereal	X	101	3.5	20.4	1.9	4.7	0.1

Snack Type	Healthier Choices	Calories (kcal)	Protein (g)	Carbs (g)	Total Fat (g)	Total Sugar (g)	Sodium (g)
Chips		150	1.8	15.9	8.9	1.0	0.2
Healthy Chips	X	132	2.0	18.1	5.7	1.7	0.2
Popcorn		148	2.4	15.9	8.7	3.1	0.2
Pretzels	X	124	2.3	20.6	3.7	2.0	0.3
Tortilla Chips	X	139	2.0	18.0	6.8	0.3	0.1

Yogurt Type	Healthier Choices	Calories (kcal)	Protein (g)	Carbs (g)	Total Fat (g)	Total Sugar (g)	Sodium (g)
Greek	X	27	2.3	3.4	0.4	2.7	0.0
Light Greek	X	20	2.3	2.5	0.1	1.7	0.0
Indulgent		36	1.3	4.6	1.4	3.7	0.0
Kids		32	1.4	5.3	0.6	4.3	0.0
Organic		27	1.5	3.2	0.9	2.7	0.0
Probiotic		22	1.6	3.6	0.2	3.0	0.0
Traditional		26	1.1	4.4	0.5	3.6	0.0
Light Traditional	X	14	0.9	2.6	0.0	1.7	0.0

Table 12: Ordinary Least Squares & Fractional Probit APE Estimates, τ_k
Bread Shares for Online Households

Shares	OLS (1)	OLS (2)	OLS (3)	F. Probit - APE (4)	F. Probit - APE (5)	Average (Pre-Online Service)	Percent Change
White	-0.167	-0.165	-0.182	-0.173	-0.167	23.56	-0.71
	(0.209)	(0.210)	(0.210)	(0.212)	(0.214)	(35.86)	
Wheat	0.615**	0.611**	0.652**	0.631**	0.627**	35.63	1.76
	(0.250)	(0.250)	(0.229)	(0.255)	(0.256)	(40.39)	
Seed	0.0487	0.0427	0.0894	0.0362	0.0292	5.47	0.53
	(0.122)	(0.121)	(0.116)	(0.114)	(0.113)	(18.35)	
Other	-0.360*	-0.360*	-0.413**	-0.367*	-0.367*	20.88	-1.76
	(0.192)	(0.204)	(0.187)	(0.190)	(0.202)	(32.79)	
Grain	-0.136	-0.128	-0.146	-0.167	-0.159	14.46	-1.10
	(0.179)	(0.177)	(0.163)	(0.194)	(0.192)	(28.63)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	478,724	478,724	478,724	478,724	478,724	112,039	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the budget share.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table 13: Ordinary Least Squares & Fractional Probit APE Estimates, τ_k
Cereal Shares for Online Households

Shares	OLS (1)	OLS (2)	OLS (3)	F. Probit - APE (4)	F. Probit - APE (5)	Average (Pre-Online Service)	Percent Change
Kids	-0.065	-0.0905	-0.144	-0.079	-0.103	37.28	-0.28
	(0.191)	(0.187)	(0.173)	(0.195)	(0.191)	(36.95)	
Org. Kids	-0.065	0.0533	0.0271	0.0362	0.0394	1.05	3.75
	(0.191)	(0.0432)	(0.0404)	(0.0392)	(0.0382)	(8.13)	
Standard	-0.065	-0.123	-0.0207	-0.141	-0.152	15.50	-0.98
	(0.191)	(0.219)	(0.194)	(0.214)	(0.209)	(27.58)	
Frosted Std.	-0.065	0.256	0.211	0.235	0.255	42.60	0.60
	(0.191)	(0.191)	(0.196)	(0.196)	(0.191)	(36.74)	
Super Healthy	-0.065	-0.0952	-0.0736	-0.119*	-0.107	3.58	-2.99
	(0.191)	(0.0735)	(0.0697)	(0.0632)	(0.0653)	(14.64)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	435,202	435,202	435,202	435,202	435,202	100,846	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the budget share.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table 14: Ordinary Least Squares & Fractional Probit APE Estimates, τ_k
Salty Snack Shares for Online Households

Shares	OLS (1)	OLS (2)	OLS (3)	F. Probit - APE (4)	F. Probit - APE (5)	Average (Pre-Online Service)	Percent Change
Tortilla Chips	0.508*	0.499*	0.568**	0.511**	0.504**	21.70	2.32
	(0.269)	(0.268)	(0.237)	(0.236)	(0.235)	(28.54)	
Pretzels	-0.388**	-0.374***	-0.353***	-0.352***	-0.338***	14.52	-2.33
	(0.106)	(0.103)	(0.110)	(0.102)	(0.0996)	(24.62)	
Popcorn	-0.0758	-0.0741	-0.0743	-0.0496	-0.0467	5.51	-0.85
	(0.0902)	(0.0908)	(0.0907)	(0.0857)	(0.0860)	(16.00)	
Veggie Chips	-0.0335	-0.00863	0.000258	-0.0262	-0.00462	7.70	-0.06
	(0.172)	(0.171)	(0.164)	(0.159)	(0.157)	(19.32)	
Chips	-0.0108	-0.0431	-0.141	-0.00854	-0.0417	50.57	-0.08
	(0.181)	(0.172)	(0.158)	(0.183)	(0.173)	(35.25)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	512,645	512,645	512,645	512,645	512,645	118,761	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the budget share.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table 15: Ordinary Least Squares & Fractional Probit APE Estimates, τ_k
Yogurt Shares for Online Households

Shares	OLS (1)	OLS (2)	OLS (3)	F. Probit - APE (4)	F. Probit - APE (5)	Average (Pre-Online Service)	Percent Change
Traditional Light	-0.099	-0.0798	-0.0121	-0.08	-0.0605	12.41	-0.49
	(0.139)	(0.136)	(0.138)	(0.143)	(0.138)	(25.99)	
Traditional	0.0283	0.0428	0.221	0.00269	0.0162	14.99	0.11
	(0.134)	(0.137)	(0.136)	(0.126)	(0.129)	(28.06)	
Probiotics	0.0725	0.076	0.0718	0.0749	0.0785	3.41	2.30
	(0.0552)	(0.0561)	(0.0639)	(0.0551)	(0.0559)	(14.84)	
Organic	0.123	0.124	0.138	0.0951	0.0972	5.20	1.87
	(0.0845)	(0.0840)	(0.0976)	(0.0925)	(0.0914)	(18.06)	
Kids	-0.0171	-0.0292	-0.0496	-0.0198	-0.0322	17.42	-0.18
	(0.241)	(0.241)	(0.224)	(0.241)	(0.241)	(29.86)	
Indulgent	-0.0313	-0.0326	-0.0728	-0.0266	-0.0278	4.09	-0.68
	(0.0890)	(0.0886)	(0.0833)	(0.0836)	(0.0832)	(15.29)	
Greek Light	0.0506	0.0587	0.0382	0.0612	0.0679	12.34	0.55
	(0.194)	(0.195)	(0.206)	(0.185)	(0.187)	(26.39)	
Greek	-0.127	-0.16	-0.334	-0.139	-0.172	30.14	-0.57
	(0.203)	(0.197)	(0.224)	(0.205)	(0.200)	(37.12)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	454,041	454,041	454,041	454,041	454,041	102,503	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the budget share.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table 16: Difference in Difference / Reduced Form Estimates, (τ_i)
In-Store Households

Budget Shares	(1)	(2)	(3)	Avg. Budget Share (Pre-Online Service)
Dairy	0.0876	0.0888	0.0855	11.95
	(0.0676)	(0.0676)	(0.0699)	(8.01)
Drinks	-0.0451	-0.0454	-0.0299	10.9
	(0.0858)	(0.0858)	(0.0845)	(10.26)
Fruit	-0.0106	-0.0114	-0.0293	7.26
	(0.0462)	(0.0463)	(0.0453)	(7.55)
Grain	-0.0291	-0.0276	-0.0203	7.2
	(0.0432)	(0.0428)	(0.0445)	(5.61)
Meat	-0.00571	-0.00406	0.0142	18.88
	(0.0890)	(0.0887)	(0.0902)	(12.09)
Oil	0.0259	0.0258	0.0228	4.36
	(0.0267)	(0.0267)	(0.0268)	(4.49)
Other	0.0105	0.0101	0.00930	1.73
	(0.0286)	(0.0287)	(0.0287)	(3.04)
Prepared	-0.116*	-0.112*	-0.111*	10.87
	(0.0644)	(0.0648)	(0.0639)	(9.19)
Snacks/Sweets	0.0295	0.0221	0.0128	16.11
	(0.0681)	(0.0666)	(0.0688)	(11.86)
Sugar	0.00935	0.00896	0.00731	1.56
	(0.0237)	(0.0238)	(0.0238)	(2.87)
Veg	0.0413	0.0432	0.0364	9.19
	(0.0547)	(0.0544)	(0.0532)	(7.77)
Time Availability f.e.	X			
Treatment Cohort f.e.	X	X		
Household Demographics	X	X		
Year-Month f.e.		X	X	
Household f.e.			X	
Observations	238,665	238,665	238,665	56,973

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the budget share.
These estimates were produced equation by equation.

Table 17: Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i
Online Households - After January 2016

Budget Shares	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Avg. Budget Share (Pre-Online Service)	Percent Change
Dairy	0.768***	0.773***	0.539***	0.726***	0.732***	0.678***	13.21	5.15
	(0.0553)	(0.0546)	(0.0577)	(0.152)	(0.154)	(0.149)	(6.90)	
Drinks	-0.729***	-0.729***	-0.564***	-0.137	-0.135	-0.126	10.08	-1.25
	(0.0596)	(0.0592)	(0.0575)	(0.242)	(0.242)	(0.245)	(8.31)	
Fruit	0.297***	0.294***	0.118***	0.424**	0.423**	0.412**	7.48	5.48
	(0.0607)	(0.0608)	(0.0370)	(0.183)	(0.183)	(0.190)	(6.18)	
Grain	0.214***	0.216***	0.148***	0.0686	0.0779	0.0286	7.6	0.38
	(0.0276)	(0.0277)	(0.0310)	(0.147)	(0.147)	(0.139)	(4.66)	
Meat	0.402***	0.404***	0.583***	0.656	0.676	0.640	18.46	3.47
	(0.0805)	(0.0821)	(0.0676)	(0.444)	(0.445)	(0.455)	(9.69)	
Oil	-0.0545**	-0.0554**	-0.0228	-0.0430	-0.0441	-0.0536	4.36	-1.23
	(0.0223)	(0.0222)	(0.0272)	(0.100)	(0.100)	(0.107)	(3.56)	
Other	-0.00758	-0.00691	0.000401	-0.0877	-0.0860	-0.0932	1.81	-5.15
	(0.0196)	(0.0196)	(0.0143)	(0.0819)	(0.0814)	(0.0847)	(2.72)	
Prepared	0.103	0.110	0.138**	0.314	0.333	0.347	10.78	3.22
	(0.0736)	(0.0749)	(0.0561)	(0.288)	(0.286)	(0.269)	(7.45)	
Snacks/Sweets	-1.679***	-1.691***	-1.623***	-2.210***	-2.281***	-2.099***	15.52	-13.53
	(0.0953)	(0.100)	(0.106)	(0.399)	(0.398)	(0.381)	(9.84)	
Sugar	0.0341***	0.0343***	0.0102	-0.138***	-0.140***	-0.136***	1.61	-8.70
	(0.00971)	(0.00978)	(0.00969)	(0.0460)	(0.0459)	(0.0451)	(2.32)	
Vegetables	0.652***	0.652***	0.674***	0.428*	0.445*	0.403	9.09	4.43
	(0.0749)	(0.0751)	(0.0502)	(0.260)	(0.258)	(0.263)	(6.20)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	222,261	222,261	222,261	222,261	222,261	222,261	147,246	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online service use on the budget share.
These estimates were produced equation by equation.
The percent change is derived utilizing the 2SLS estimates in column (6) and the pre-online service average.

Table 18: In-Store Elasticity Matrix

	Own Price Elasticity	Dairy	Drink	Fruit	Grain	Meat	Oil	Other	Prepared	Snacks	Sugars	Vegetables
Dairy	-1.08	-1.08	0.39	0.01	-0.23	0.10	0.43	0.40	-0.17	-0.55	0.05	0.47
	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.03)	(0.05)	(0.08)	(0.04)	(0.03)	(0.09)	(0.03)
Drink	-1.91	0.28	-1.91	0.75	0.18	-0.19	-0.29	0.57	-0.02	-0.05	0.41	0.29
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.06)	(0.03)	(0.02)	(0.07)	(0.02)
Fruit	-1.16	0.01	0.63	-1.16	-0.29	-0.20	0.77	-0.12	-0.52	0.25	0.50	-0.23
	(0.06)	(0.02)	(0.02)	(0.06)	(0.06)	(0.03)	(0.07)	(0.11)	(0.05)	(0.03)	(0.14)	(0.04)
Grain	0.29	-0.14	0.15	-0.28	0.29	0.05	0.62	0.03	-0.74	-0.07	-0.52	-0.18
	(0.14)	(0.02)	(0.02)	(0.05)	(0.14)	(0.04)	(0.13)	(0.18)	(0.06)	(0.04)	(0.26)	(0.05)
Meat	-1.23	0.16	-0.24	-0.43	0.17	-1.23	-0.52	-0.05	0.58	0.69	-1.00	-0.66
	(0.07)	(0.04)	(0.03)	(0.08)	(0.10)	(0.07)	(0.13)	(0.20)	(0.07)	(0.06)	(0.23)	(0.08)
Oil	-0.27	0.14	-0.11	0.40	0.36	-0.13	-0.27	-0.15	-0.38	-0.24	-0.75	0.15
	(0.14)	(0.02)	(0.02)	(0.04)	(0.08)	(0.03)	(0.14)	(0.14)	(0.04)	(0.03)	(0.20)	(0.04)
Other	-0.90	0.05	0.09	-0.02	0.01	-0.01	-0.05	-0.90	-0.03	-0.01	-0.36	-0.03
	(0.10)	(0.01)	(0.01)	(0.02)	(0.04)	(0.02)	(0.05)	(0.10)	(0.02)	(0.02)	(0.10)	(0.02)
Prepared	0.62	-0.14	0.01	-0.66	-0.99	0.30	-0.86	-0.20	0.62	-0.03	-0.17	-0.27
	(0.09)	(0.03)	(0.03)	(0.06)	(0.08)	(0.04)	(0.10)	(0.15)	(0.09)	(0.04)	(0.17)	(0.05)
Snacks	-0.89	-0.71	-0.06	0.42	-0.17	0.53	-0.87	-0.14	-0.07	-0.89	1.20	0.05
	(0.06)	(0.03)	(0.03)	(0.06)	(0.08)	(0.05)	(0.11)	(0.16)	(0.07)	(0.06)	(0.19)	(0.06)
Sugar	-0.66	0.00	0.06	0.09	-0.11	-0.09	-0.26	-0.36	-0.03	0.12	-0.66	0.05
	(0.18)	(0.01)	(0.01)	(0.03)	(0.05)	(0.02)	(0.07)	(0.10)	(0.03)	(0.02)	(0.18)	(0.02)
Vegetables	-0.76	0.34	0.29	-0.26	-0.20	-0.33	0.32	-0.18	-0.23	-0.02	0.30	-0.76
	(0.06)	(0.02)	(0.02)	(0.05)	(0.06)	(0.04)	(0.08)	(0.12)	(0.05)	(0.05)	(0.15)	(0.06)

Standard errors reported in parentheses
Standard errors clustered at the store-availability level

Table 19: Online Elasticity Matrix

	Own Price Elasticity	Dairy	Drink	Fruit	Grain	Meat	Oil	Other	Prepared	Snacks	Sugars	Vegetables
Dairy	-0.66	-0.66	0.21	0.53	-0.01	1.27	0.50	0.32	-2.02	1.06	-3.99	1.53
	(0.03)	(0.03)	(0.10)	(0.09)	(0.08)	(0.91)	(0.20)	(0.42)	(1.04)	(0.43)	(63.11)	(0.97)
Drink	-0.65	-0.04	-0.65	0.23	0.03	0.41	0.25	1.13	-0.91	0.38	2.65	1.78
	(0.22)	(0.02)	(0.22)	(0.08)	(0.06)	(0.49)	(0.19)	(1.48)	(0.37)	(0.25)	(43.09)	(1.35)
Fruit	-0.81	0.25	-0.64	-0.81	-0.66	0.53	0.02	-1.07	-0.25	-0.14	29.99	-1.59
	(0.10)	(0.03)	(0.30)	(0.10)	(0.21)	(0.46)	(0.21)	(1.43)	(0.31)	(0.24)	(458.13)	(1.23)
Grain	0.17	-0.03	-0.09	-0.51	0.17	0.27	-0.40	-0.27	0.42	-0.65	-28.67	-1.69
	(0.28)	(0.03)	(0.13)	(0.15)	(0.28)	(0.41)	(0.45)	(0.85)	(0.26)	(0.46)	(443.17)	(1.31)
Meat	1.72	-0.10	0.40	-0.09	-0.05	1.72	-0.10	1.62	1.44	-0.88	-14.25	0.87
	(2.43)	(0.04)	(0.24)	(0.12)	(0.17)	(2.43)	(0.30)	(1.99)	(0.78)	(0.49)	(218.00)	(0.83)
Oil	-0.62	0.07	-0.23	0.00	-0.18	0.19	-0.62	-1.26	-0.07	0.15	8.74	-0.18
	(0.46)	(0.02)	(0.12)	(0.07)	(0.14)	(0.38)	(0.46)	(1.34)	(0.18)	(0.23)	(137.77)	(0.31)
Other	-0.55	0.01	-0.21	-0.07	-0.02	-0.34	-0.24	-0.55	0.00	-0.02	-11.49	-0.02
	(0.76)	(0.01)	(0.07)	(0.03)	(0.06)	(0.24)	(0.10)	(0.76)	(0.09)	(0.08)	(173.14)	(0.13)
Prepared	0.48	-0.38	0.99	-0.12	0.22	-1.86	-0.11	-0.01	0.48	0.32	-0.42	0.97
	(0.81)	(0.04)	(0.39)	(0.09)	(0.14)	(1.51)	(0.22)	(0.54)	(0.81)	(0.29)	(10.64)	(0.78)
Snacks	-1.53	-0.29	0.59	0.06	0.47	-1.86	-0.28	0.20	-0.52	-1.53	-11.80	0.01
	(0.48)	(0.04)	(0.29)	(0.13)	(0.28)	(1.50)	(0.38)	(0.87)	(0.38)	(0.48)	(178.03)	(0.57)
Sugars	23.90	-0.01	-0.04	0.18	-0.23	0.30	0.16	-1.05	0.00	0.13	23.90	0.03
	(378.55)	(0.01)	(0.06)	(0.05)	(0.12)	(0.31)	(0.15)	(1.43)	(0.11)	(0.13)	(378.55)	(0.18)
Vegetables	-2.93	0.20	-1.31	-0.41	-0.59	-0.71	-0.12	-0.05	0.68	0.09	1.11	-2.93
	(1.62)	(0.03)	(0.52)	(0.13)	(0.24)	(0.84)	(0.26)	(0.56)	(0.34)	(0.25)	(16.32)	(1.62)

Standard errors reported in parentheses
Standard errors clustered at the store-availability level

11 Figure Appendix

Figure A1: Budget Allocation & Prices

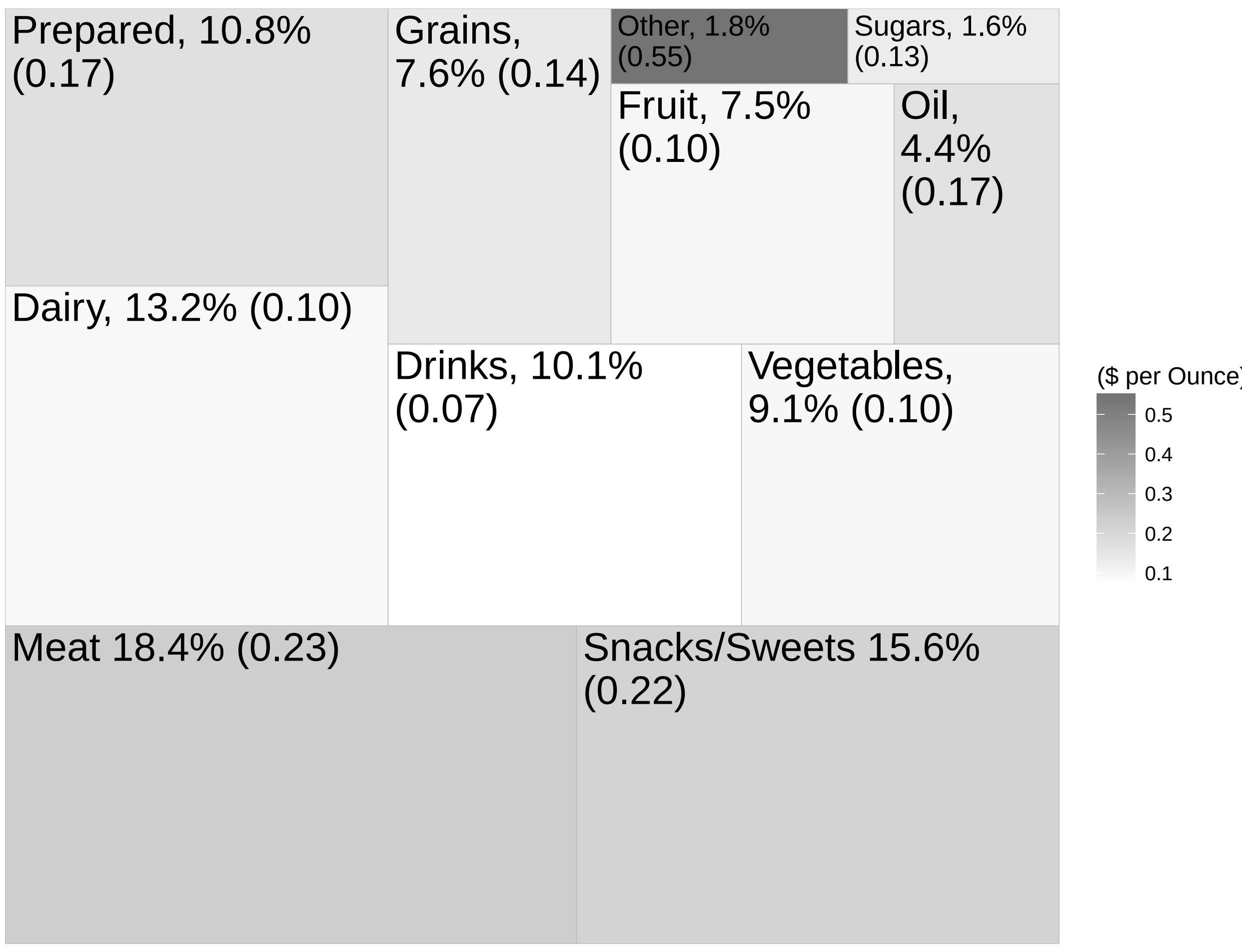

Figure A1 presents the average budget allocation, of the online households, in the six months before the online service was ever available to any household (i.e. before March 2015). The size of each box corresponds to the average budget share amount and the shade of the each box corresponds to how expensive the product category is in terms of price per ounce. Darker shade indicate more expensive product categories and lighter shades represent less expensive product categories.

Figure A2: Event Study Estimates for Online Use and Expenditure Shares
In-Store Households Only

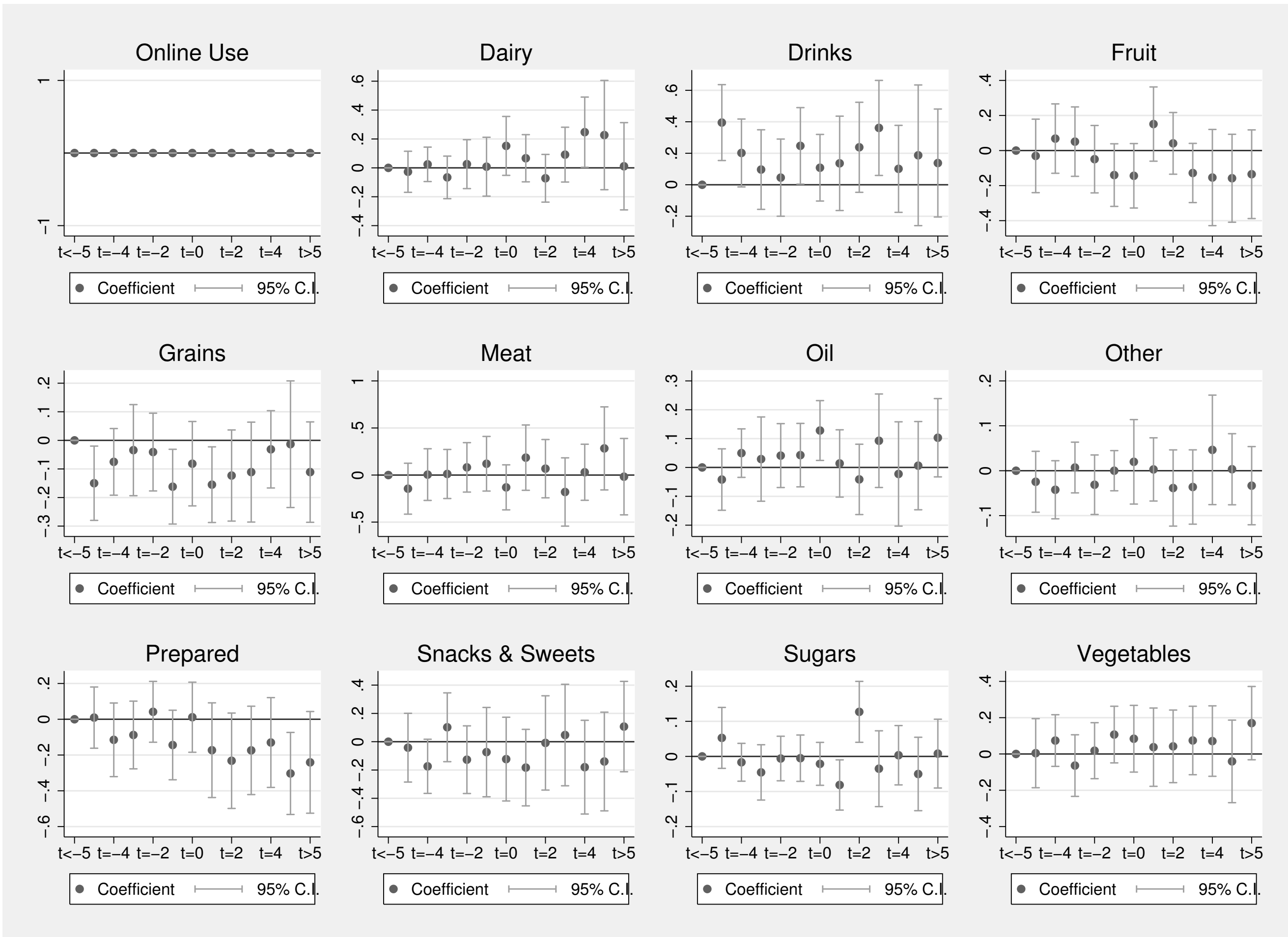

Figure A3: Mean Demographics of In-Store & Online Households

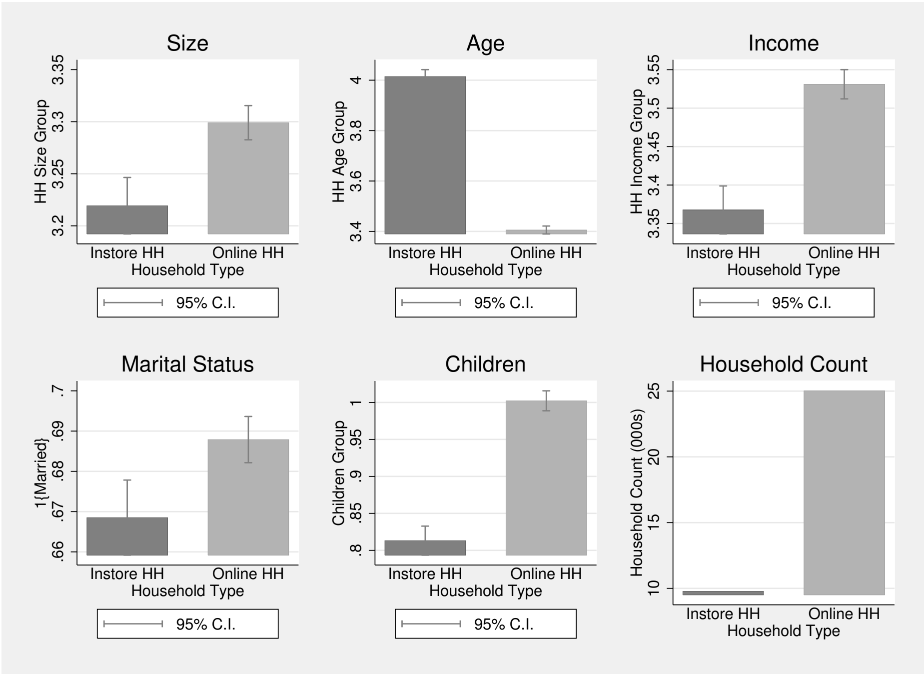

Figure A3 illustrates the demographic differences between households who have adopted the online shopping service over the time frame of my data (online households) and the households who have not adopted the online shopping service over the time frame of my data (in-store households). Specifically, this figure illustrates that early adopters of the online shopping service are, on average, larger households (3.3 vs. 3.2 people), have heads of household that are younger (36-45 vs. 46-55), have higher incomes ($80-$99k vs $51-$79k per year), are more likely to be married (69% vs. 67%) and are more likely to have a child (1 vs. 0.8). Note that my demographic variables are categorical and the specific categorical definitions can be found in Table A2 of the appendix.

Figure A4: Mean Shopping Behavior of In-Store & Online Households Pre-Online Service

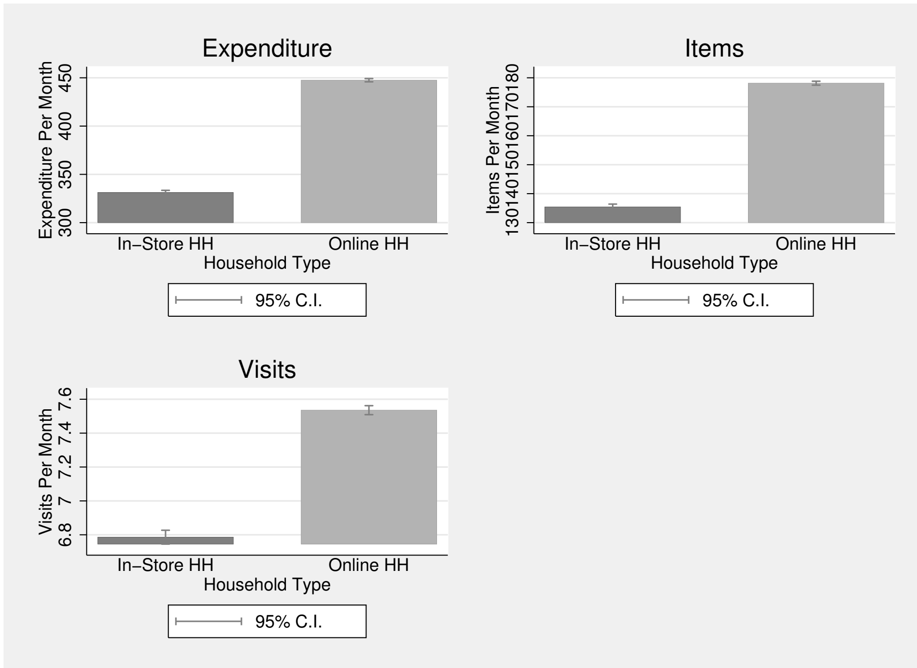

Figure A4 illustrates the differences in purchasing patterns that exist between online households and in-store households before the online service was ever introduced.

Figure A5: Mean Budget Shares of In-Store & Online Households
Pre-Online Service

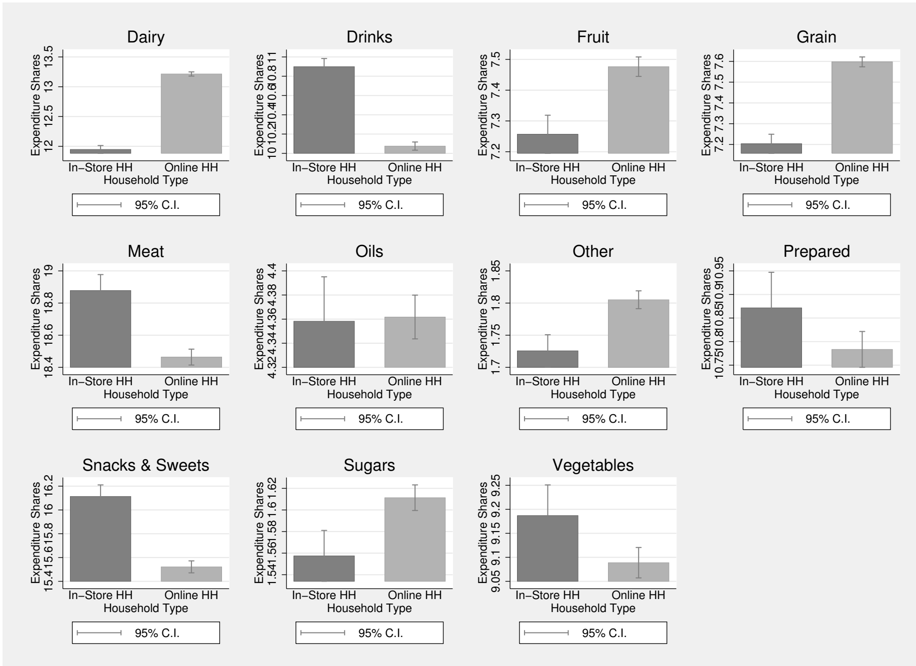

Figure A5 illustrates the differences in budget shares that exist between online households and in-store households before the online service was ever introduced.

Figure A6: Mean Demographics of Availability Waves

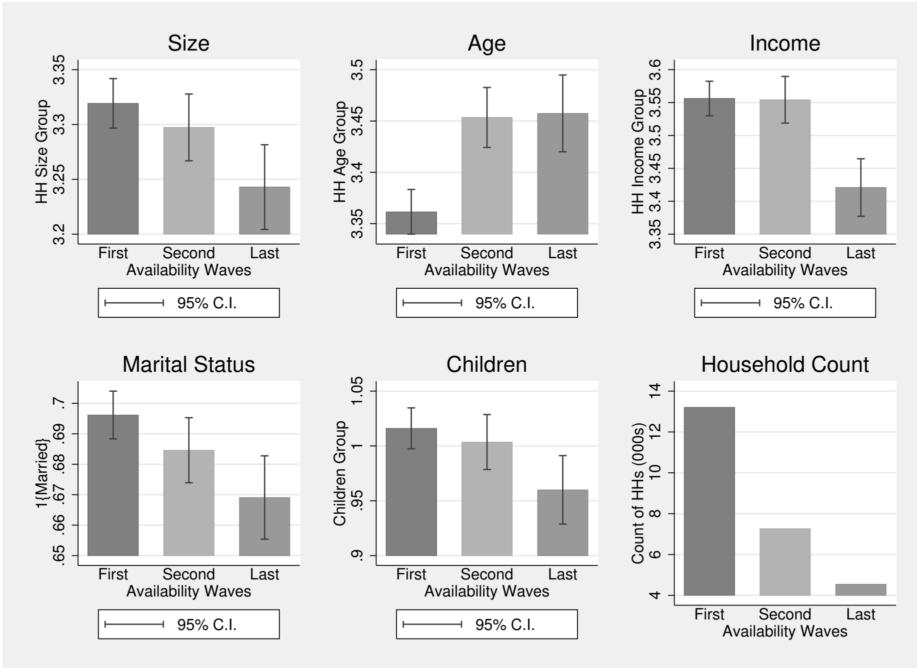

Figure A6 presents a demographic comparison of the households who received access to the online purchasing early to the households who received access to the online purchasing service later.[65] Figure A6 illustrates the following demographic information: on average, the household is composed of three individuals, the household head is between the ages of 36-45, annual household income is between $80K to $99K, roughly 70% of these households are married and they are likely to have at least one child. In contrast, households in later availability dates tend to be smaller and older. Additionally, households in the last availability wave have a higher proportion of households in lower income categories, are less likely to be married and are less likely to have children.

[65]Households that received access to the online shopping service before or during January 2016 are assigned to wave one (53% of the online household sample), households that received access to the online service after January 2016 but before July 2016 are assigned to wave two (29% of the online household sample) and households who received access after or during July 2016 are assigned to wave three (18% of the online household sample).

Figure A7: Mean Shopping Behavior of Availability Waves

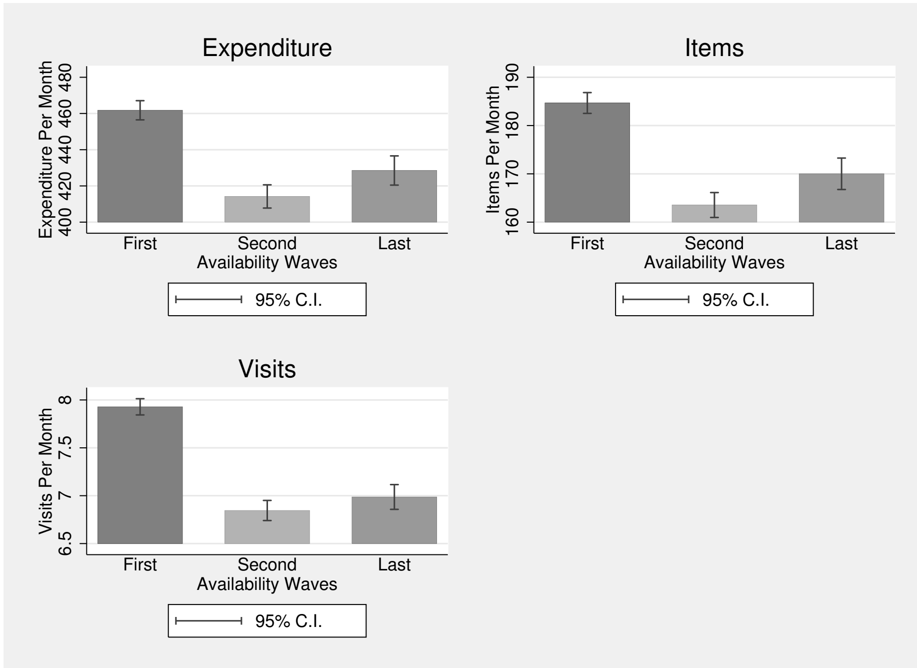

Figure A7 present averages over household shopping patterns restricted to the six months before the online service was ever available to any household (i.e. before March 2015) for the different availability waves.[66]

[66]Households that received access to the online shopping service before or during January 2016 are assigned to wave one (53% of the online household sample), households that received access to the online service after January 2016 but before July 2016 are assigned to wave two (29% of the online household sample) and households who received access after or during July 2016 are assigned to wave three (18% of the online household sample).

Figure A8: Mean Expenditure Shares of Availability Waves

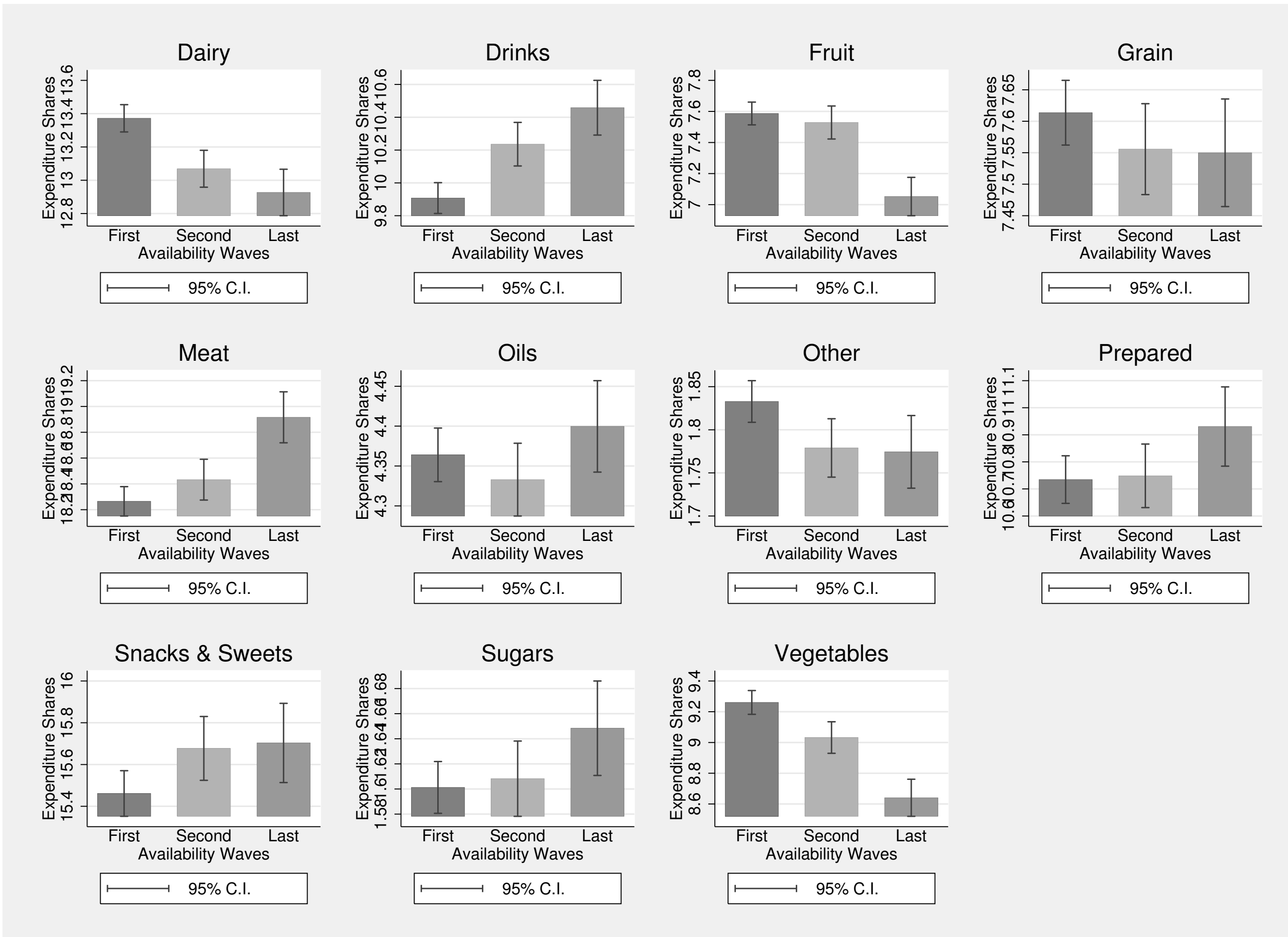

Figure A8 present averages over household shopping patterns restricted to the six months before the online service was ever available to any household (i.e. before March 2015) for the different availability waves.[67]

[67]Households that received access to the online shopping service before or during January 2016 are assigned to wave one (53% of the online household sample), households that received access to the online service after January 2016 but before July 2016 are assigned to wave two (29% of the online household sample) and households who received access after or during July 2016 are assigned to wave three (18% of the online household sample).

Figure A9: Bread Nutrition Comparison

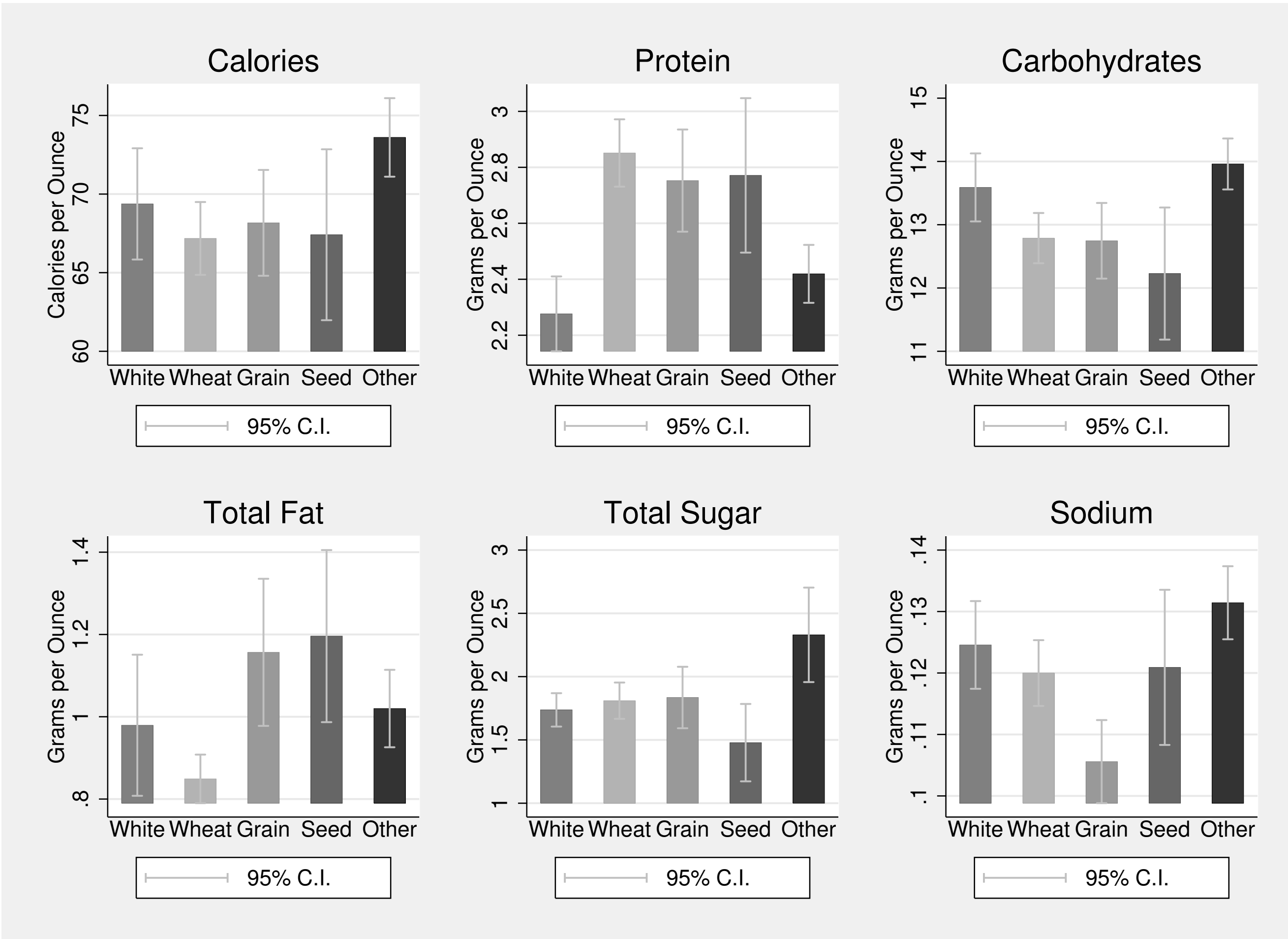

Figure A9 presents the average nutritional content (per ounce) of each type of bread.

Figure A10: Cereal Nutrition Comparison

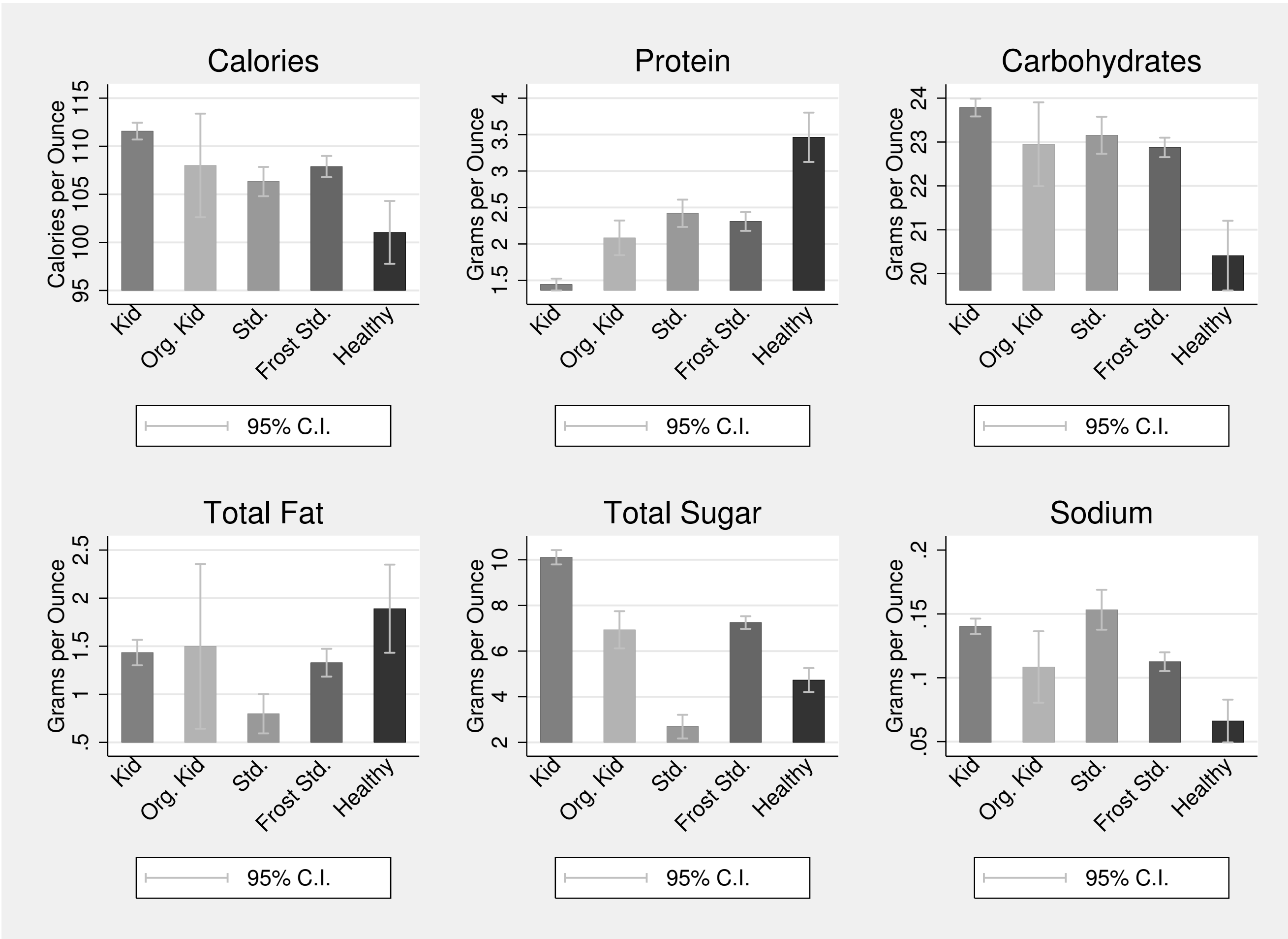

Figure A10 presents the average nutritional content (per ounce) of each type of cereal.

Figure A11: Salty Snack Nutrition Comparison

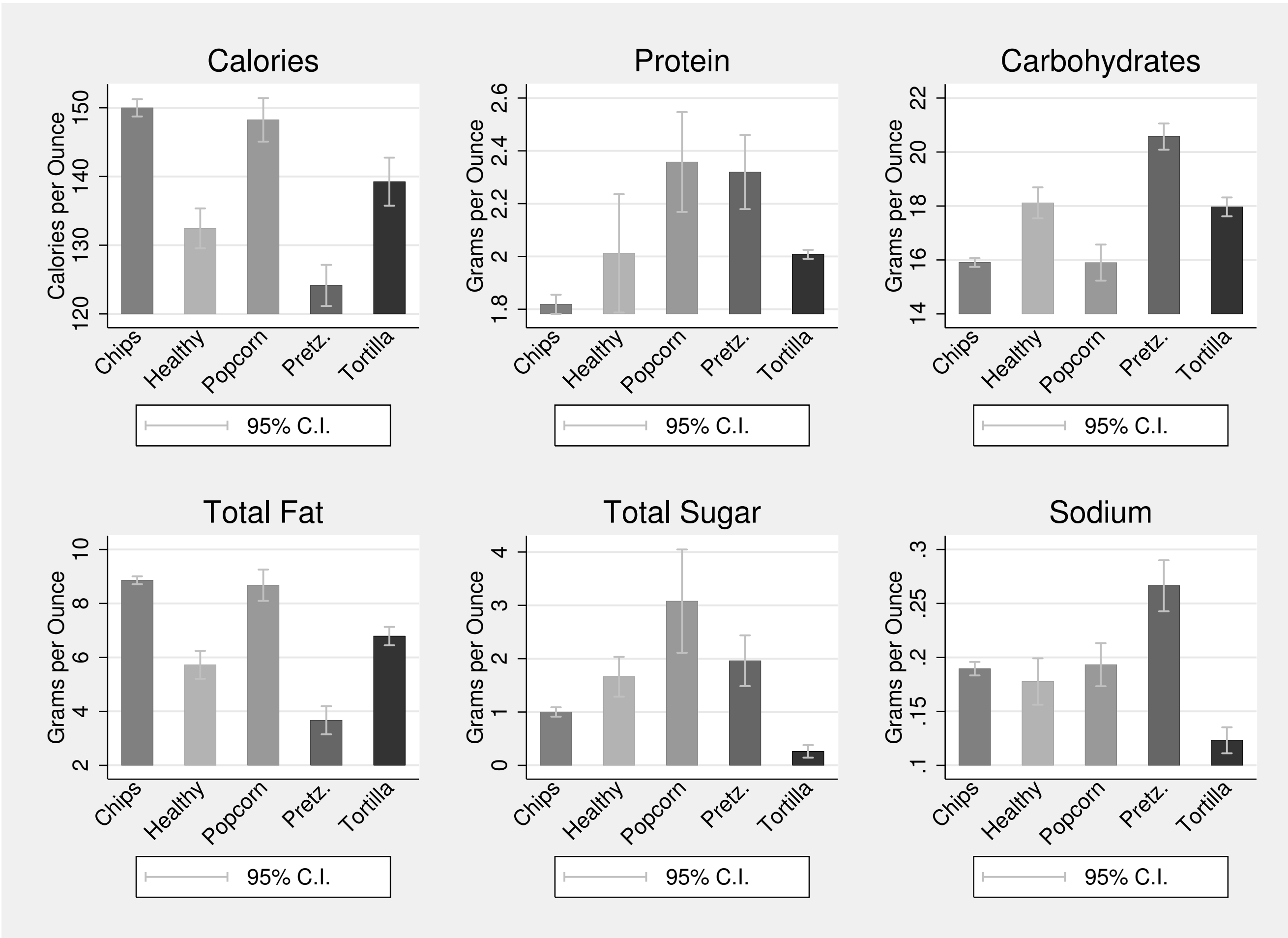

Figure A11 presents the average nutritional content (per ounce) of each type of salty snack.

Figure A12: Yogurt Nutrition Comparison

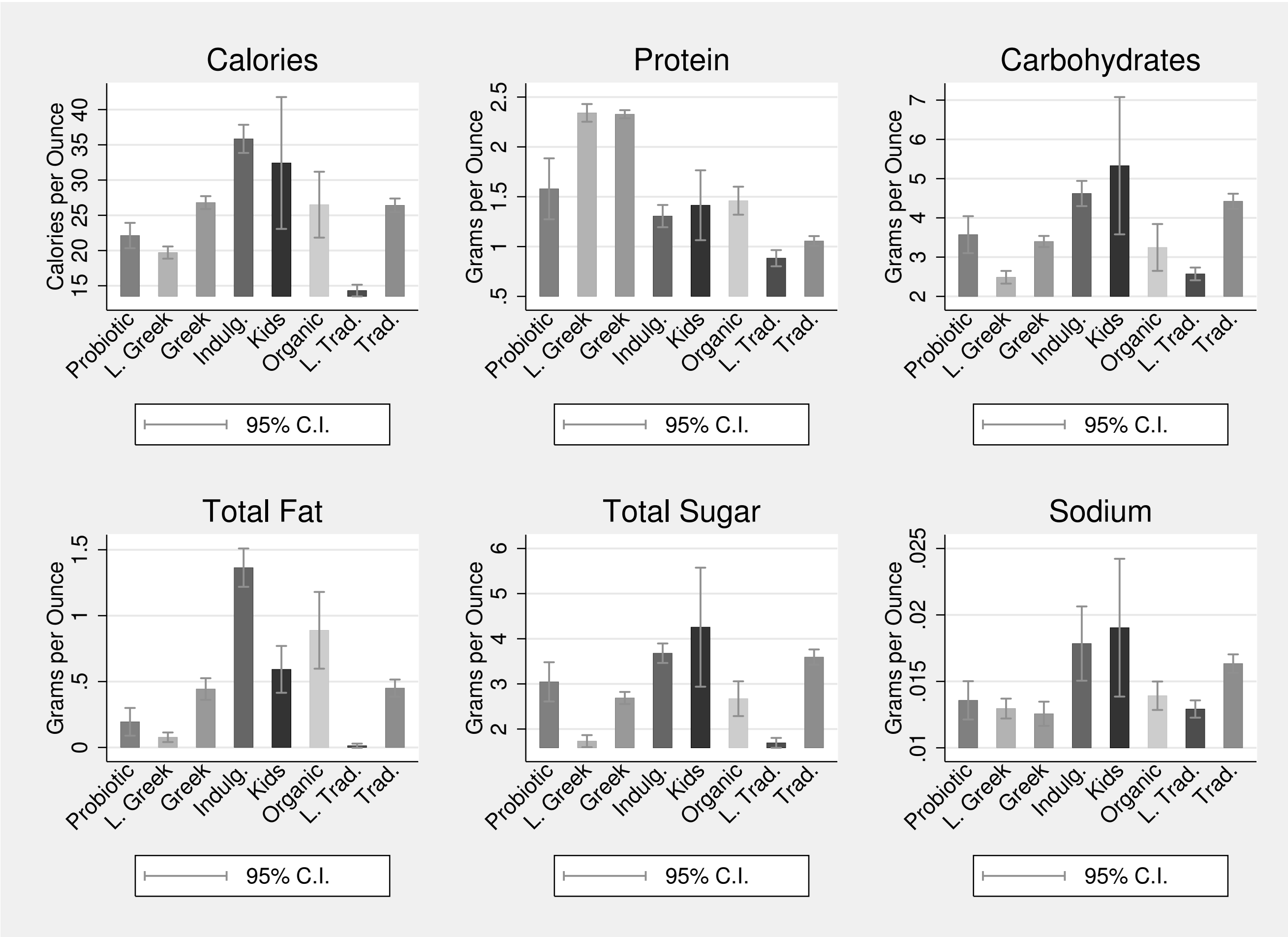

Figure A12 presents the average nutritional content (per ounce) of each type of yogurt.

Figure A13: In-store Own and Cross Price Elasticities

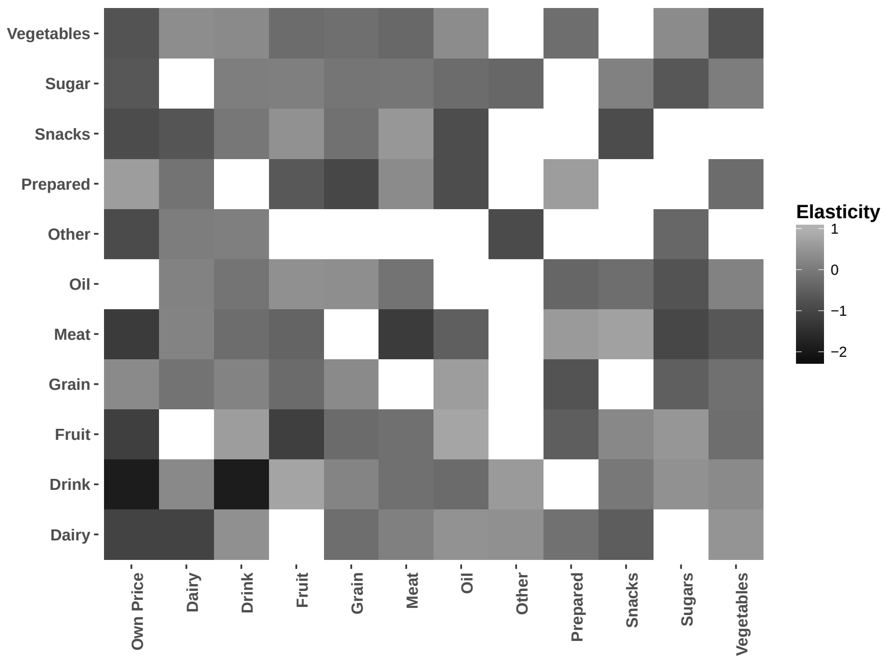

Figure A13 presents in-store own and cross price elasticities that are significant at the 95% confidence level. Darker shades represent negative elasticities (complements), while lighter shades represent positive elasticities (substitutes). White means that the elasticity is not statistically significant.

Figure A14: Online Own and Cross Price Elasticities

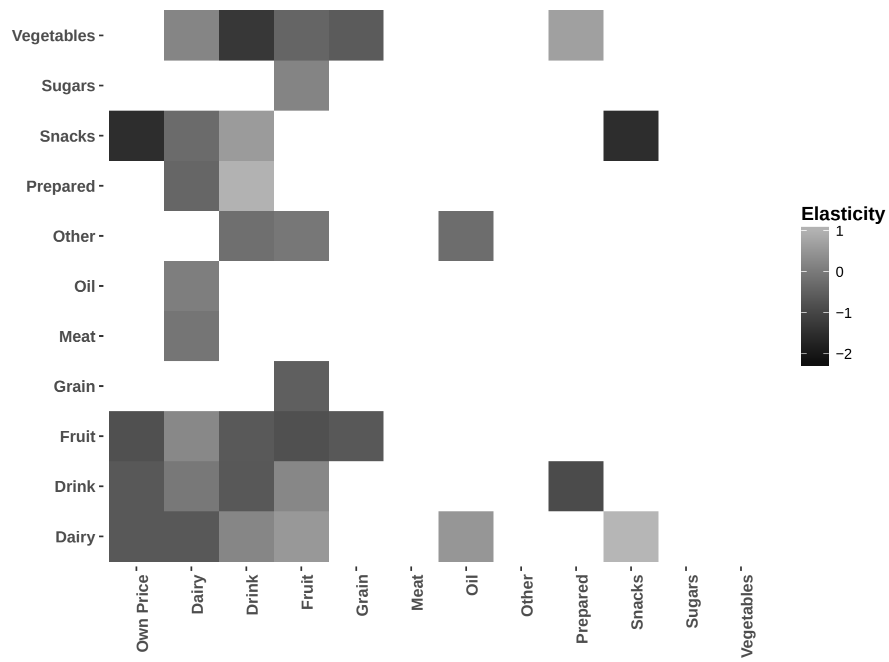

Figure A14 presents online own and cross price elasticities that are significant at the 95% confidence level. Darker shades represent negative elasticities (complements), while lighter shades represent positive elasticities (substitutes). White means that the elasticity is not statistically significant.

12 Table Appendix

Table A1: Product Category Key

Product Category	Description
Dairy	Milk & Milk Substitutes, Cheese, Yogurt, Cream Cheese
Drinks	Non-Alcoholic Beverages, Water, Soda, Juice
Fruit	Fresh, Dried and Frozen Fruits
Grains	Rice, Pasta, Bread, Cereal, Oatmeal
Meat	Beef, Poultry, Seafood, Eggs, Beans, Legumes
Oils	Butter, Mayonnaise, Salad Dressings, Vegetable Oils
Other	Flour, Gravy, Seasonings, Baking Items
Prepared	Rice Mixed Dishes, Pizza, Macaroni, Soups
Snacks and Sweets	Chips, Crackers, Granola Bars, Cakes, Candy, Ice Cream
Sugar	Sugar, Honey, Jams, Syrups
Vegetables	Fresh & Frozen Vegetables

Table A2: Demographic Variables Description

Household Size	
Variable	Category
1	1 Person
2	2 People
3	3 People
4	4 People
5	5+ People

Number of Children	
Variable	Category
0	0 Children
1	1 Child
2	2 Children
3	3 Children
4	4+ Children

Household Income	
Variable	Category
1	$0-29$k
2	$30-50$k
3	$51-79$k
4	$80-99$k
5	$100-149$k
6	$150k+$

Household Age	
Variable	Category
1	18-25
2	26-35
3	36-45
4	46-55
5	56-65
6	66+

Marital Status	
Variable	Category
0	Single
1	Married

Table A3: Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i
Online Availability Definition Based on Entire Store Footprint, Online Households Only

Budget Shares	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Avg. Ounce Share (Pre-Online Service)	Percent Change
Dairy	0.767***	0.769***	0.496***	0.811***	0.817***	0.782***	13.25	5.90
	(0.0504)	(0.0506)	(0.0755)	(0.215)	(0.217)	(0.218)	(6.96)	
Drinks	-0.758***	-0.762***	-0.594***	-0.915***	-0.922***	-0.909***	10.09	-9.01
	(0.0645)	(0.0610)	(0.0353)	(0.293)	(0.298)	(0.297)	(8.43)	
Fruit	0.335***	0.354***	0.125***	0.623***	0.626***	0.644***	7.45	8.64
	(0.0548)	(0.0566)	(0.0215)	(0.239)	(0.240)	(0.241)	(6.21)	
Grain	0.272***	0.274***	0.179***	0.222**	0.226**	0.211**	7.62	2.77
	(0.0212)	(0.0213)	(0.0223)	(0.0944)	(0.0934)	(0.101)	(4.73)	
Meat	0.288***	0.287***	0.531***	1.036***	1.038***	1.049***	18.43	5.69
	(0.0692)	(0.0728)	(0.0673)	(0.301)	(0.302)	(0.305)	(9.75)	
Oil	-0.0473*	-0.0462*	-0.00967	-0.192*	-0.195*	-0.200**	4.36	-4.59
	(0.0265)	(0.0270)	(0.0193)	(0.100)	(0.100)	(0.100)	(3.59)	
Other	0.00428	0.00116	0.00646	-0.0937*	-0.0952*	-0.0840	1.8	-4.67
	(0.0149)	(0.0140)	(0.0100)	(0.0529)	(0.0529)	(0.0545)	(2.74)	
Prepared	0.0314	0.0315	0.118*	0.186	0.199	0.221	10.8	2.05
	(0.103)	(0.108)	(0.0692)	(0.238)	(0.240)	(0.239)	(7.51)	
Snacks/Sweets	-1.584***	-1.613***	-1.527***	-2.430***	-2.462***	-2.429***	15.53	-15.64
	(0.0891)	(0.0958)	(0.139)	(0.275)	(0.274)	(0.278)	(9.97)	
Sugar	0.0407***	0.0403***	0.0124	-0.0975*	-0.0988**	-0.103**	1.62	-6.36
	(0.00613)	(0.00643)	(0.00745)	(0.0500)	(0.0501)	(0.0503)	(2.36)	
Vegetables	0.653***	0.665***	0.663***	0.853***	0.867***	0.819***	9.05	9.05
	(0.0474)	(0.0473)	(0.0241)	(0.189)	(0.189)	(0.194)	(6.23)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	690,152	690,152	690,152	690,152	690,152	690,152	164,750	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Product Subcategories

Bread	Brands/Description
White	White breads, Wonder, etc.
Wheat	Whole wheat breads
Grain	Whole grain breads, Oat breads
Seed	Seeded/nut breads
Other	Cinnamon bread, raisin bread, sourdough

Cereal	Brands/Description
Kids	Fruity Pebbles, Apple Jacks, Cocoa Puffs
Organic Kids	Annies, Cascadian Farm
Standard	Cheerios, Chex, Cornflakes, Rice Crispies
Frosted Standard	Frosted Flakes, Cornpops, Frosted Mini Wheats
Super Healthy	Kashi, Fiberone, Grapenuts

Salty Snacks	Brands/Description
Tortilla Chips	Mission, On the Border, Tostitos
Pretzels	Rold Gold, Snyder
Popcorn	Cape Cod, Skinny Pop, Smart Food
Healthy Chips	Sun Chips, Sensible Portions, Quaker
Chips	Doritos, Cheetos, Lays, Pringles

Yogurt	Brands/Description
Greek	Chobani, Dannon, Yoplait
Greek Light	Chobani, Dannon, Yoplait
Indulgent	YoCrunch, Noosa, Yoplait
Kids	Yoplait, Stoneyfield, Dannon
Organic	Stoneyfield, Wallaby, Silk
Probiotic	Dannon
Traditional	Yoplait, Dannon
Traditional Light	Yoplait, Dannon

Table A5: Ordinary Least Squares & Tobit APE Estimates, τ_k
Bread Sales for Online Households

Sales	OLS (1)	OLS (2)	OLS (3)	Tobit- APE (4)	Tobit- APE (5)	Average (Pre-Online Service)	Percent Change
White	0.0151	0.0154	0.0112	0.0178	0.0180	1.26	1.43
	(0.0159)	(0.0160)	(0.0162)	(0.0164)	(0.0165)	(2.67)	
Wheat	0.100***	0.100***	0.0972***	0.0894***	0.0895***	1.92	4.66
	(0.0175)	(0.0175)	(0.0178)	(0.0162)	(0.0162)	(3.31)	
Seed	0.0195**	0.0194**	0.0183*	0.0127**	0.0124**	0.37	3.35
	(0.00895)	(0.00886)	(0.00907)	(0.00538)	(0.00533)	(1.49)	
Other	0.00552	0.00553	0.00327	0.000964	0.000707	1.41	0.05
	(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0141)	(3.08)	
Grain	0.0138	0.0141	0.0116	0.0111	0.0112	0.96	1.17
	(0.0109)	(0.0108)	(0.0112)	(0.00850)	(0.00855)	(2.48)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	616,899	616,899	616,899	616,900	616,899	147,372	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell represents an estimate of the effect of online availability on the sales volume.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table A6: Ordinary Least Squares & Tobit APE Estimates, τ_k
Cereal Sales for Online Households

Sales	OLS (1)	OLS (2)	OLS (3)	Tobit- APE (4)	Tobit- APE (5)	Average (Pre-Online Service)	Percent Change
Kids	0.0372	0.0375	0.0300	0.0338	0.0339	3.49	0.97
	(0.0350)	(0.0349)	(0.0355)	(0.0290)	(0.0290)	(6.10)	
Org. Kids	0.0132***	0.0133***	0.0129**	0.00575	0.00559	0.10	5.59
	(0.00469)	(0.00472)	(0.00473)	(0.00351)	(0.00347)	(0.97)	
Standard	0.0314	0.0313	0.0296	0.0147	0.0139	1.36	1.02
	(0.0226)	(0.0228)	(0.0232)	(0.0168)	(0.0170)	(3.27)	
Frosted Std.	0.110***	0.111***	0.106***	0.0822***	0.0831***	4.02	2.07
	(0.0306)	(0.0305)	(0.0307)	(0.0281)	(0.0282)	(6.51)	
Super Healthy	0.00292	0.00314	0.00299	-0.00580	-0.00580	0.34	-1.71
	(0.0115)	(0.0115)	(0.0119)	(0.00715)	(0.00711)	(1.80)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	616,899	616,899	616,899	616,899	616,899	147,372	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** p<0.01, ** p<0.05, * p<0.1

Each cell represents an estimate of the effect of online availability on the sales volume.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table A7: Ordinary Least Squares & Tobit APE Estimates, τ_k
Salty Snack Sales for Online Households

Sales	OLS (1)	OLS (2)	OLS (3)	Tobit- APE (4)	Tobit- APE (5)	Average (Pre-Online Service)	Percent Change
Tortilla Chips	0.0683**	0.0676**	0.0666**	0.0627**	0.0616**	2.42	2.55
	(0.0250)	(0.0255)	(0.0265)	(0.0254)	(0.0262)	(3.89)	
Pretzels	-0.0123	-0.0125	-0.0140	-0.0232	-0.0239	1.55	-1.54
	(0.0169)	(0.0169)	(0.0172)	(0.0170)	(0.0169)	(3.05)	
Popcorn	0.00396	0.00393	0.00257	-0.00844	-0.00822	0.71	-1.16
	(0.0158)	(0.0158)	(0.0159)	(0.0130)	(0.0131)	(2.43)	
Veggie Chips	0.0253	0.0259	0.0260	0.00154	0.00181	0.92	0.20
	(0.0324)	(0.0321)	(0.0328)	(0.0209)	(0.0207)	(2.70)	
Chips	0.0569	0.0551	0.0432	0.0543	0.0520	6.54	0.80
	(0.0637)	(0.0631)	(0.0653)	(0.0627)	(0.0623)	(8.64)	
Time Availability f.e.	X						
Treatment Cohort f.e.	X	X					
Household Demographics	X	X					
Year-Month f.e.		X	X				
Household f.e.			X				
Observations	616,899	616,899	616,899	616,899	616,899	147,372	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell represents an estimate of the effect of online availability on the sales volume.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table A8: Ordinary Least Squares & Tobit APE Estimates, τ_k
Yogurt Sales for Online Households

Sales	OLS (1)	OLS (2)	OLS (3)	Tobit- APE (4)	Tobit- APE (5)	Average (Pre-Online Service)	Percent Change
Traditional Light	0.0197	0.0205	0.0187	0.0252**	0.0253**	1.11	2.28
	(0.0125)	(0.0126)	(0.0128)	(0.0127)	(0.0127)	(3.11)	
Traditional	0.0746***	0.0755***	0.0742***	0.0602***	0.0606***	1.32	4.59
	(0.0164)	(0.0165)	(0.0165)	(0.0122)	(0.0123)	(3.35)	
Probiotics	0.0210**	0.0212**	0.0204**	0.0117*	0.0117*	0.40	2.93
	(0.00915)	(0.00917)	(0.00930)	(0.00668)	(0.00667)	(2.25)	
Organic	0.0505***	0.0506***	0.0496***	0.0200**	0.0195**	0.58	3.36
	(0.0158)	(0.0159)	(0.0168)	(0.00973)	(0.00979)	(2.84)	
Kids	0.0401	0.0408	0.0367	0.0371	0.0373	1.87	1.99
	(0.0341)	(0.0341)	(0.0346)	(0.0319)	(0.0319)	(4.62)	
Indulgent	0.0125	0.0127	0.0116	0.00928	0.00948	0.46	2.06
	(0.0210)	(0.0209)	(0.0212)	(0.0105)	(0.0105)	(2.38)	
Greek Light	0.0788***	0.0800***	0.0757***	0.0534**	0.0530*	1.49	3.56
	(0.0262)	(0.0263)	(0.0270)	(0.0269)	(0.0273)	(4.68)	
Greek	0.202***	0.203***	0.199***	0.116***	0.116***	3.65	3.18
	(0.0599)	(0.0593)	(0.0604)	(0.0353)	(0.0354)	(7.64)	
Time Availability f.e.	X			X			
Treatment Cohort f.e.	X	X		X	X		
Household Demographics	X	X		X	X		
Year-Month f.e.		X	X		X		
Household f.e.			X				
Observations	616,899	616,899	616,899	616,899	616,899	147,372	

Robust standard errors in parentheses
Standard errors clustered at the specified level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Each cell represents an estimate of the effect of online availability on the sales volume.
These estimates were produced equation by equation.
The percent change is derived utilizing the APE estimates in column (5) and the pre-online service average.

Table A9: Tobit Estimates, τ_k
Bread Sales for Online Households

Sales	(1)	(2)
White	0.0609	0.0615
	(0.0560)	(0.0564)
Wheat	0.216***	0.216***
	(0.0390)	(0.0390)
Seed	0.145**	0.142**
	(0.0611)	(0.0607)
Other	0.00346	0.00254
	(0.0505)	(0.0506)
Grain	0.0551	0.0557
	(0.0421)	(0.0424)
Time Availability f.e.	X	
Treatment Cohort f.e.	X	X
Household Demographics	X	X
Year-Month f.e.		X
Household f.e.		
Observations	616,899	616,899

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A10: Tobit Estimates, τ_k
Cereal Sales for Online Households

Sales	(1)	(2)
Kids	0.0781	0.0781
	(0.0670)	(0.0668)
Org. Kids	0.372	0.365
	(0.227)	(0.226)
Standard	0.0617	0.0585
	(0.0704)	(0.0714)
Frosted Std.	0.169***	0.171***
	(0.0578)	(0.0580)
Super Healthy	-0.111	-0.111
	(0.137)	(0.137)
Time Availability f.e.	X	
Treatment Cohort f.e.	X	X
Household Demographics	X	X
Year-Month f.e.		X
Household f.e.		
Observations	616,899	616,899

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A11: Tobit Estimates, τ_k
Salty Snack Sales for Online Households

Sales	(1)	(2)
Tortilla Chips	0.138**	0.136**
	(0.0559)	(0.0575)
Pretzels	-0.0681	-0.0700
	(0.0497)	(0.0496)
Popcorn	-0.0536	-0.0523
	(0.0826)	(0.0832)
Veggie Chips	0.00796	0.00939
	(0.108)	(0.107)
Chips	0.0821	0.0785
	(0.0948)	(0.0940)
Time Availability f.e.	X	
Treatment Cohort f.e.	X	X
Household Demographics	X	X
Year-Month f.e.		X
Household f.e.		
Observations	616,899	616,899

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A12: Tobit Estimates, τ_k
Yogurt Sales for Online Households

Sales	(1)	(2)
Traditional Light	0.139**	0.140**
	(0.0697)	(0.0697)
Traditional	0.254***	0.256***
	(0.0517)	(0.0522)
Probiotics	0.235*	0.236*
	(0.134)	(0.134)
Organic	0.243**	0.237**
	(0.118)	(0.119)
Kids	0.150	0.151
	(0.129)	(0.129)
Indulgent	0.121	0.124
	(0.137)	(0.137)
Greek Light	0.283**	0.282*
	(0.143)	(0.146)
Greek	0.314***	0.312***
	(0.0952)	(0.0955)
Time Availability f.e.	X	
Treatment Cohort f.e.	X	X
Household Demographics	X	X
Year-Month f.e.		X
Household f.e.		
Observations	616,899	616,899

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A13: In-Store Households - Expenditure

Difference in Difference / Reduced Form Estimates, (τ_i)

Expenditure	(1)	(2)	(3)
1{OnlineAvail}	1.099	1.153	0.537
	(2.035)	(2.030)	(1.977)
Time Availability f.e.	X		
Treatment Cohort f.e.	X	X	
Household Demographics	X	X	
Year-Month f.e.		X	X
Household f.e.			X
Observations	238,665	238,665	238,665
R-squared	0.501	0.506	0.809

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A14: In-Store Households - Nutrient Outcomes

Difference in Difference / Reduced Form Estimates, (τ_i)

Nutrients per Ounce	(1)	(2)	(3)
Calories	0.00383	-0.00427	0.00624
	(0.118)	(0.118)	(0.121)
Carbohydrates	-0.0108	-0.0117	-0.0108
	(0.0161)	(0.0161)	(0.0172)
Cholesterol	-0.000187	-0.000187	-0.000203
	(0.000134)	(0.000135)	(0.000138)
Protein	0.00554	0.00550	0.00572
	(0.00476)	(0.00474)	(0.00487)
Sodium	0.00236	0.00238	0.00247
	(0.00605)	(0.00602)	(0.00596)
Total Fat	0.00188	0.00136	0.00200
	(0.00775)	(0.00772)	(0.00787)
Total Sugar	-0.00252	-0.00326	-0.00320
	(0.00822)	(0.00828)	(0.00883)
Time Availability f.e.	X		
Treatment Cohort f.e.	X	X	
Household Demographics	X	X	
Year-Month f.e.		X	X
Household f.e.			X
Observations	226,532	226,532	226,532

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A15: In-Store Households - Ounce Shares

Difference in Difference / Reduced Form Estimates, (τ_i)

Ounce Shares	(1)	(2)	(3)
Dairy	0.0406	0.0424	0.0323
	(0.114)	(0.113)	(0.115)
Drinks	0.0894	0.0869	0.0962
	(0.186)	(0.185)	(0.181)
Fruit	-0.0482	-0.0475	-0.0665
	(0.0713)	(0.0711)	(0.0679)
Grain	-0.0128	-0.0121	-0.00194
	(0.0431)	(0.0428)	(0.0441)
Meat	-0.0333	-0.0323	-0.0207
	(0.0556)	(0.0554)	(0.0571)
Oil	0.0218	0.0213	0.0216
	(0.0233)	(0.0233)	(0.0245)
Other	0.0117	0.0113	0.00995
	(0.0210)	(0.0210)	(0.0212)
Prepared	-0.0448	-0.0413	-0.0366
	(0.0463)	(0.0463)	(0.0466)
Snacks/Sweets	0.0119	0.00792	0.00261
	(0.0594)	(0.0597)	(0.0573)
Sugar	0.00341	0.00248	0.00225
	(0.0288)	(0.0289)	(0.0292)
Veg	-0.0397	-0.0392	-0.0392
	(0.0722)	(0.0720)	(0.0714)
Time Availability f.e.	X		
Treatment Cohort f.e.	X	X	
Household Demographics	X	X	
Year-Month f.e.		X	X
Household f.e.			X
Observations	238,563	238,563	238,563

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A16: Ordinary Least Squares & Two-Stage Least Squares Estimates, ϕ_i
Online Households - Ounce Shares

Ounce Shares	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Avg. Ounce Share (Pre-Online Service)	Percent Change
Dairy	1.420***	1.430***	0.933***	1.450***	1.460***	1.412***	17.97	7.85
	(0.0686)	(0.0688)	(0.0890)	(0.220)	(0.221)	(0.232)	(10.57)	
Drinks	-0.799***	-0.822***	-0.298***	0.0246	0.00715	-0.0362	24.63	-0.15
	(0.116)	(0.121)	(0.0995)	(0.360)	(0.364)	(0.371)	(16.99)	
Fruit	0.0619	0.0851	-0.134***	-0.127	-0.120	-0.114	9.02	-1.26
	(0.0652)	(0.0687)	(0.0339)	(0.190)	(0.191)	(0.197)	(7.39)	
Grain	0.233***	0.234***	0.160***	0.194	0.195	0.190	6.3	3.02
	(0.0221)	(0.0215)	(0.0153)	(0.122)	(0.121)	(0.126)	(4.27)	
Meat	0.107***	0.110***	0.209***	0.313**	0.315**	0.336**	9.46	3.59
	(0.0321)	(0.0340)	(0.0268)	(0.141)	(0.141)	(0.142)	(6.70)	
Oil	-0.0949***	-0.0922***	-0.0613**	-0.146*	-0.149*	-0.144*	3.36	4.17
	(0.0168)	(0.0176)	(0.0227)	(0.0763)	(0.0770)	(0.0784)	(3.33)	
Other	0.0233	0.0207	0.0184*	-0.0457	-0.0472	-0.0358	1.02	-3.51
	(0.0147)	(0.0147)	(0.00978)	(0.0376)	(0.0376)	(0.0384)	(2.22)	
Prepared	-0.0424	-0.0444	0.0215	0.0142	0.0267	0.0501	7.99	0.63
	(0.0485)	(0.0524)	(0.0318)	(0.150)	(0.152)	(0.153)	(6.03)	
Snacks/Sweets	-1.018***	-1.036***	-1.003***	-1.215***	-1.231***	-1.179***	8.83	-13.36
	(0.0926)	(0.0949)	(0.0973)	(0.220)	(0.224)	(0.221)	(7.59)	
Sugar	0.0251*	0.0221	0.0219**	-0.0452	-0.0490	-0.0528	1.57	-3.36
	(0.0133)	(0.0136)	(0.00978)	(0.0428)	(0.0433)	(0.0455)	(2.52)	
Vegetables	0.0850	0.0927*	0.132***	-0.416**	-0.409**	-0.425*	9.84	-4.37
	(0.0549)	(0.0546)	(0.0365)	(0.209)	(0.209)	(0.218)	(7.30)	
Time Availability f.e.	X			X				
Treatment Cohort f.e.	X	X		X	X			
Household Demographics	X	X		X	X			
Year-Month f.e.		X	X		X	X		
Household f.e.			X			X		
Observations	616,163	616,163	616,163	616,163	616,163	616,163	147,172	

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A17: Event Study Estimates, τ_{ik}
Online Households

	1{Online}	Dairy	Drink	Fruit	Grain	Meat
t=-5	0.00125	-0.0982	-0.0911	0.0128	-0.00870	0.0726
	(0.00329)	(0.0580)	(0.0667)	(0.0784)	(0.0320)	(0.0738)
t=-4	-0.00273	-0.0420	-0.0385	0.0695	-0.00102	-0.00999
	(0.00446)	(0.0430)	(0.0509)	(0.0625)	(0.0339)	(0.0556)
t=-3	-0.00609	-0.000427	-0.130*	0.0949*	-0.0330	0.0838
	(0.00479)	(0.0574)	(0.0715)	(0.0523)	(0.0341)	(0.0720)
t=-2	-0.00495	0.0339	-0.119	0.0389	-0.0415	0.118
	(0.00553)	(0.0504)	(0.0768)	(0.0664)	(0.0344)	(0.0737)
t=-1	0.00364	0.0465	-0.115**	0.0744	4.54e-06	0.0104
	(0.00776)	(0.0566)	(0.0490)	(0.0623)	(0.0338)	(0.0580)
t=0	0.141***	0.0873	-0.143**	0.101**	0.0400	0.186**
	(0.0247)	(0.0685)	(0.0686)	(0.0483)	(0.0340)	(0.0764)
t=1	0.191***	0.106	-0.205***	0.140***	-0.0326	0.288***
	(0.0269)	(0.0830)	(0.0750)	(0.0449)	(0.0318)	(0.0725)
t=2	0.204***	0.159**	-0.236***	0.212***	-0.0249	0.189**
	(0.0236)	(0.0753)	(0.0784)	(0.0735)	(0.0273)	(0.0825)
t=3	0.201***	0.0383	-0.188***	0.0979	0.0270	0.286***
	(0.0228)	(0.0815)	(0.0626)	(0.0689)	(0.0330)	(0.0816)
t=4	0.208***	0.140	-0.284**	0.161**	0.0133	0.331***
	(0.0278)	(0.0954)	(0.116)	(0.0672)	(0.0339)	(0.0870)
t=5	0.201***	0.0641	-0.194*	0.0542	0.0476	0.292***
	(0.0244)	(0.119)	(0.108)	(0.0739)	(0.0409)	(0.101)
t>5	0.211***	-0.0196	-0.147	0.131	-0.0116	0.246***
	(0.0123)	(0.115)	(0.0923)	(0.0834)	(0.0410)	(0.0897)
Year-Month f.e.	X	X	X	X	X	X
Household f.e.	X	X	X	X	X	X
Observations	616,357	616,357	616,357	616,357	616,357	616,357
R-Squared	0.306	0.364	0.332	0.399	0.312	0.368

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A18: Event Study Estimates, τ_{ik}
Online Households

	Oil	Other	Prepared	Snack/Sweet	Sugar	Vegetables
t=-5	0.0111	0.0499**	-0.00653	0.0489	0.0589**	-0.0497
	(0.0243)	(0.0217)	(0.0476)	(0.0689)	(0.0228)	(0.0412)
t=-4	-0.0552***	0.0339*	-0.0339	0.00647	0.0174	0.0532
	(0.0192)	(0.0179)	(0.0563)	(0.0829)	(0.0196)	(0.0617)
t=-3	-0.0549	0.0234	-0.0276	-0.121	0.0287	0.136***
	(0.0343)	(0.0223)	(0.0462)	(0.0772)	(0.0193)	(0.0447)
t=-2	-0.00510	0.0409*	-0.0602	-0.0952	0.0369**	0.0518
	(0.0259)	(0.0237)	(0.0652)	(0.0699)	(0.0144)	(0.0487)
t=-1	0.00404	0.00982	-0.118*	-0.0197	0.0410	0.0636
	(0.0229)	(0.0179)	(0.0616)	(0.0579)	(0.0290)	(0.0511)
t=0	-0.0100	0.00407	-0.0227	-0.365***	0.0428**	0.0808
	(0.0236)	(0.0180)	(0.0810)	(0.0983)	(0.0191)	(0.0610)
t=1	-0.0612*	0.0118	-0.00837	-0.408***	-0.0102	0.181***
	(0.0321)	(0.0176)	(0.0752)	(0.0976)	(0.0251)	(0.0618)
t=2	-0.0782**	-0.000670	-0.0252	-0.416***	0.0190	0.198***
	(0.0371)	(0.0179)	(0.0565)	(0.0878)	(0.0212)	(0.0488)
t=3	-0.0697***	0.0256	-0.00214	-0.455***	0.0173	0.222***
	(0.0192)	(0.0253)	(0.0669)	(0.103)	(0.0174)	(0.0402)
t=4	-0.0594*	0.0530**	-0.0212	-0.575***	0.0149	0.227***
	(0.0321)	(0.0251)	(0.0696)	(0.133)	(0.0148)	(0.0429)
t=5	-0.0366	0.0292	0.00314	-0.501***	0.0208	0.221***
	(0.0463)	(0.0174)	(0.0806)	(0.138)	(0.0382)	(0.0623)
t>5	-0.0618**	0.0377*	-0.129	-0.283**	0.0421*	0.194***
	(0.0279)	(0.0193)	(0.0890)	(0.132)	(0.0244)	(0.0554)
Year-Month f.e.	X	X	X	X	X	X
Household f.e.	X	X	X	X	X	X
Observations	616,357	616,357	616,357	616,357	616,357	616,357
R-Squared	0.183	0.172	0.353	0.319	0.166	0.414

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A19: Event Study Estimates, τ_{ik}
In-Store Households

	Dairy	Drink	Fruit	Grain	Meat
t=-5	-0.0272	0.395***	-0.0306	-0.150**	-0.145
	(0.0725)	(0.123)	(0.107)	(0.0663)	(0.138)
t=-4	0.0244	0.202*	0.0682	-0.0752	0.00469
	(0.0608)	(0.110)	(0.101)	(0.0595)	(0.140)
t=-3	-0.0659	0.0965	0.0508	-0.0342	0.0113
	(0.0752)	(0.129)	(0.101)	(0.0813)	(0.133)
t=-2	0.0257	0.0452	-0.0493	-0.0409	0.0815
	(0.0863)	(0.125)	(0.0981)	(0.0694)	(0.134)
t=-1	0.00804	0.247*	-0.140	-0.162**	0.120
	(0.104)	(0.124)	(0.0911)	(0.0667)	(0.148)
t=0	0.152	0.108	-0.144	-0.0816	-0.131
	(0.104)	(0.108)	(0.0938)	(0.0753)	(0.122)
t=1	0.0666	0.136	0.151	-0.155**	0.185
	(0.0834)	(0.153)	(0.108)	(0.0675)	(0.177)
t=2	-0.0720	0.238	0.0410	-0.123	0.0676
	(0.0843)	(0.146)	(0.0898)	(0.0813)	(0.158)
t=3	0.0915	0.361**	-0.128	-0.111	-0.180
	(0.0968)	(0.154)	(0.0861)	(0.0892)	(0.185)
t=4	0.247*	0.101	-0.154	-0.0314	0.0294
	(0.124)	(0.141)	(0.140)	(0.0690)	(0.152)
t=5	0.227	0.187	-0.158	-0.0134	0.283
	(0.193)	(0.228)	(0.128)	(0.113)	(0.225)
t>5	0.0111	0.138	-0.135	-0.111	-0.0173
	(0.154)	(0.175)	(0.129)	(0.0896)	(0.207)
Year-Month f.e.	X	X	X	X	X
Household f.e.	X	X	X	X	X
Observations	238,665	238,665	238,665	238,665	238,665
R-Squared	0.343	0.329	0.405	0.272	0.371

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A20: Event Study Estimates, τ_{ik}
In-Store Households

	Oil	Other	Prepared	Snack/Sweet	Sugar	Vegetables
t=-5	-0.0419	-0.0246	0.00916	-0.0421	0.0528	0.00442
	(0.0543)	(0.0346)	(0.0873)	(0.124)	(0.0443)	(0.0971)
t=-4	0.0497	-0.0424	-0.115	-0.174*	-0.0167	0.0740
	(0.0429)	(0.0330)	(0.105)	(0.0977)	(0.0275)	(0.0726)
t=-3	0.0290	0.00712	-0.0880	0.102	-0.0455	-0.0639
	(0.0745)	(0.0288)	(0.0965)	(0.124)	(0.0402)	(0.0866)
t=-2	0.0411	-0.0312	0.0416	-0.127	-0.00604	0.0181
	(0.0565)	(0.0338)	(0.0868)	(0.122)	(0.0324)	(0.0787)
t=-1	0.0427	1.12e-05	-0.144	-0.0735	-0.00508	0.107
	(0.0561)	(0.0228)	(0.0992)	(0.161)	(0.0337)	(0.0796)
t=0	0.128**	0.0198	0.0113	-0.123	-0.0212	0.0840
	(0.0529)	(0.0480)	(0.100)	(0.151)	(0.0312)	(0.0938)
t=1	0.0139	0.00288	-0.173	-0.183	-0.0814**	0.0375
	(0.0595)	(0.0359)	(0.135)	(0.138)	(0.0365)	(0.110)
t=2	-0.0414	-0.0385	-0.232*	-0.00862	0.127***	0.0420
	(0.0622)	(0.0433)	(0.136)	(0.170)	(0.0443)	(0.102)
t=3	0.0926	-0.0363	-0.174	0.0472	-0.0350	0.0743
	(0.0827)	(0.0422)	(0.126)	(0.183)	(0.0551)	(0.0962)
t=4	-0.0226	0.0464	-0.130	-0.180	0.00337	0.0711
	(0.0922)	(0.0622)	(0.128)	(0.169)	(0.0431)	(0.0990)
t=5	0.00623	0.00329	-0.303**	-0.140	-0.0501	-0.0408
	(0.0780)	(0.0404)	(0.117)	(0.178)	(0.0534)	(0.116)
t>5	0.103	-0.0333	-0.241	0.107	0.00797	0.170
	(0.0692)	(0.0444)	(0.145)	(0.163)	(0.0500)	(0.103)
Year-Month f.e.	X	X	X	X	X	X
Household f.e.	X	X	X	X	X	X
Observations	238,665	238,665	238,665	238,665	238,665	238,665
R-Squared	0.174	0.156	0.357	0.319	0.169	0.400

Robust standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A21: First Stage Estimates

Endogenous Variable	$\ln(p_{Dairy})$	$\ln(p_{Drink})$	$\ln(p_{Fruit})$	$\ln(p_{Grain})$	$\ln(p_{Meat})$	$\ln(p_{POil})$
1{OnlineAvail}	-0.57	0.06	-0.14	-0.12	0.19	0.25
$\ln(c_{Dairy})$	1.01	0.25	0.17	0.10	0.04	0.15
$\ln(c_{Drink})$	0.14	1.00	0.12	0.10	0.05	0.11
$\ln(c_{Fruit})$	-0.08	0.02	0.54	0.03	0.04	-0.01
$\ln(c_{Grain})$	-0.34	0.27	-0.35	0.40	-0.07	0.09
$\ln(c_{Meat})$	0.20	0.16	0.21	0.15	0.76	0.31
$\ln(c_{Oil})$	0.06	0.14	-0.04	-0.01	0.06	0.41
$\ln(c_{Other})$	-0.04	-0.15	-0.03	-0.02	0.01	-0.04
$\ln(c_{Prep})$	0.19	0.33	0.08	0.01	-0.12	-0.07
$\ln(c_{Snacks})$	-0.33	-0.43	-0.29	-0.04	-0.13	-0.09
$\ln(c_{Sugar})$	-0.02	0.05	-0.03	0.02	0.02	0.04
$\ln(c_{Veg})$	-0.07	-0.16	0.10	0.00	0.04	0.00
$\ln(c_{Dairy})$x1{OnlineAvail}	0.20	0.14	0.04	0.09	0.07	0.06
$\ln(c_{Drink})$x11{OnlineAvail}	-0.07	0.06	-0.12	0.01	0.06	0.05
$\ln(c_{Fruit})$x11{OnlineAvail}	-0.09	-0.12	-0.04	-0.06	-0.05	-0.01
$\ln(c_{Grain})$x11{OnlineAvail}	-0.33	-0.25	0.13	-0.20	-0.14	-0.26
$\ln(c_{Meat})$x11{OnlineAvail}	0.49	0.21	0.01	0.14	0.13	0.14
$\ln(c_{Oil})$x11{OnlineAvail}	-0.03	0.02	0.15	0.07	0.14	0.15
$\ln(c_{Other})$x11{OnlineAvail}	0.03	-0.02	0.01	-0.01	0.01	0.03
$\ln(c_{Prep})$x11{OnlineAvail}	-0.18	0.06	0.30	0.02	0.10	0.16
$\ln(c_{Snacks})$x1{OnlineAvail}	-0.23	-0.22	-0.26	-0.24	-0.15	-0.22
$\ln(c_{Sugar})$x11{OnlineAvail}	-0.09	-0.04	-0.01	0.00	-0.02	-0.01
$\ln(c_{Veg})$x11{OnlineAvail}	0.10	0.14	-0.13	0.09	-0.07	0.00
$\ln(\frac{M}{C})$	-0.11	-0.24	-0.09	-0.03	-0.05	-0.08
$\ln(\frac{M}{C})$x11{OnlineAvail}	0.01	0.01	0.02	0.00	0.00	0.00
Constant	X	X	X	X	X	X
Household Demographics	X	X	X	X	X	X
Month f.e.	X	X	X	X	X	X
Linear Time Trend	X	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157	616,157
R^2	0.06	0.10	0.05	0.04	0.05	0.06

Standard errors available from the author upon request

Table A22: First Stage Estimates

Endogenous Variable	$\ln(p_{Other})$	$\ln(p_{Prep})$	$\ln(p_{Snacks})$	$\ln(p_{Sugar})$	$\ln(p_{Veg})$
1{OnlineAvail}	0.06	0.17	0.13	0.36	0.15
$\ln(c_{Dairy})$	-0.11	0.05	0.12	-0.04	0.25
$\ln(c_{Drink})$	0.15	0.03	0.05	0.11	0.10
$\ln(c_{Fruit})$	0.01	0.03	0.06	0.00	-0.01
$\ln(c_{Grain})$	-0.31	-0.09	-0.19	0.15	-0.21
$\ln(c_{Meat})$	0.56	0.30	0.10	0.42	0.29
$\ln(c_{Oil})$	-0.06	0.08	0.01	0.08	-0.10
$\ln(c_{Other})$	0.30	-0.02	0.00	0.00	-0.03
$\ln(c_{Prep})$	0.16	0.67	0.02	-0.14	0.02
$\ln(c_{Snacks})$	0.33	-0.10	0.64	-0.09	-0.23
$\ln(c_{Sugar})$	0.09	0.04	-0.01	0.36	-0.02
$\ln(c_{Veg})$	-0.07	-0.04	-0.06	-0.05	0.55
$\ln(c_{Dairy})$x1{OnlineAvail}	-0.04	0.11	0.13	0.05	-0.02
$\ln(c_{Drink})$x11{OnlineAvail}	-0.22	0.15	0.03	-0.13	-0.03
$\ln(c_{Fruit})$x11{OnlineAvail}	-0.10	0.00	-0.05	-0.13	-0.08
$\ln(c_{Grain})$x11{OnlineAvail}	0.46	-0.42	-0.27	-0.34	-0.07
$\ln(c_{Meat})$x11{OnlineAvail}	-0.25	0.18	0.12	0.07	0.14
$\ln(c_{Oil})$x11{OnlineAvail}	0.11	0.00	0.08	0.29	0.10
$\ln(c_{Other})$x11{OnlineAvail}	0.01	0.01	0.00	-0.05	0.00
$\ln(c_{Prep})$x11{OnlineAvail}	0.43	0.06	0.05	0.54	0.15
$\ln(c_{Snacks})$x1{OnlineAvail}	-0.33	-0.05	-0.10	-0.44	-0.23
$\ln(c_{Sugar})$x11{OnlineAvail}	0.13	0.00	0.00	0.14	-0.05
$\ln(c_{Veg})$x11{OnlineAvail}	-0.12	-0.05	0.03	0.07	0.15
$\ln(\frac{M}{C})$	-0.34	-0.07	-0.04	-0.21	-0.09
$\ln(\frac{M}{C})$x11{OnlineAvail}	-0.01	0.00	0.00	-0.03	0.00
Constant	X	X	X	X	X
Household Demographics	X	X	X	X	X
Month f.e.	X	X	X	X	X
Linear Time Trend	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R^2	0.13	0.06	0.04	0.12	0.07
Standard errors available from the author upon request					

Table A23: First Stage Estimates

Endogenous Variable	1{Online}	$\ln(\frac{M}{P})$	$\ln(\frac{M}{P})$ x 1{Online}
1{OnlineAvail}	3.39	0.11	25.69
$\ln(c_{Dairy})$	0.01	-0.14	0.13
$\ln(c_{Drink})$	0.07	0.01	0.58
$\ln(c_{Fruit})$	-0.08	0.09	-0.66
$\ln(c_{Grain})$	-0.18	0.26	-1.42
$\ln(c_{Meat})$	0.05	-0.10	0.42
$\ln(c_{Oil})$	-0.03	0.01	-0.26
$\ln(c_{Other})$	-0.02	0.02	-0.18
$\ln(c_{Prep})$	0.06	-0.20	0.48
$\ln(c_{Snacks})$	-0.18	0.09	-1.50
$\ln(c_{Sugar})$	0.05	0.03	0.40
$\ln(c_{Veg})$	0.14	0.09	1.15
$\ln(c_{Dairy})$x1{OnlineAvail}	-0.53	-0.05	-4.22
$\ln(c_{Drink})$x11{OnlineAvail}	0.03	-0.04	0.21
$\ln(c_{Fruit})$x11{OnlineAvail}	0.52	0.02	4.21
$\ln(c_{Grain})$x11{OnlineAvail}	0.26	0.27	2.25
$\ln(c_{Meat})$x11{OnlineAvail}	-1.25	-0.13	-9.96
$\ln(c_{Oil})$x11{OnlineAvail}	0.48	-0.04	3.79
$\ln(c_{Other})$x11{OnlineAvail}	-0.09	0.02	-0.61
$\ln(c_{Prep})$x11{OnlineAvail}	1.11	-0.06	8.61
$\ln(c_{Snacks})$x1{OnlineAvail}	0.73	0.14	5.78
$\ln(c_{Sugar})$x11{OnlineAvail}	0.17	-0.01	1.29
$\ln(c_{Veg})$x11{OnlineAvail}	0.17	-0.06	1.31
$\ln(\frac{M}{C})$	0.01	1.00	0.07
$\ln(\frac{M}{C})$x11{OnlineAvail}	0.05	-0.01	0.56
Constant	X	X	X
Household Demographics	X	X	X
Month f.e.	X	X	X
Linear Time Trend	X	X	X
Observations	616,157	616,157	616,157
R^2	0.19	0.93	0.19

Standard errors available from the author upon request

Table A24: First Stage Estimates

Endogenous Variable	$\ln(p_{Dairy})x1\{O\}$	$\ln(p_{Drink})x1\{O\}$	$\ln(p_{Fruit})x1\{O\}$	$\ln(p_{Grain})x1\{O\}$	$\ln(p_{Meat})x1\{O\}$	$\ln(p_{Oil})x1\{O\}$
1{OnlineAvail}	-7.49	-8.25	-7.10	-6.19	-4.52	-5.44
$\ln(c_{Dairy})$	-0.02	-0.03	-0.02	-0.01	-0.01	-0.01
$\ln(c_{Drink})$	-0.17	-0.20	-0.17	-0.14	-0.10	-0.12
$\ln(c_{Fruit})$	0.20	0.26	0.17	0.15	0.11	0.15
$\ln(c_{Grain})$	0.43	0.53	0.39	0.35	0.27	0.31
$\ln(c_{Meat})$	-0.11	-0.14	-0.10	-0.09	-0.08	-0.08
$\ln(c_{Oil})$	0.08	0.08	0.07	0.06	0.04	0.05
$\ln(c_{Other})$	0.05	0.06	0.04	0.04	0.03	0.04
$\ln(c_{Prep})$	-0.15	-0.14	-0.14	-0.11	-0.09	-0.11
$\ln(c_{Snacks})$	0.43	0.50	0.41	0.34	0.26	0.33
$\ln(c_{Sugar})$	-0.14	-0.16	-0.13	-0.11	-0.09	-0.11
$\ln(c_{Veg})$	-0.35	-0.43	-0.31	-0.27	-0.20	-0.25
$\ln(c_{Dairy})$x1{OnlineAvail}	1.57	1.67	1.19	1.09	0.81	1.03
$\ln(c_{Drink})$x11{OnlineAvail}	-0.05	0.28	-0.03	-0.04	-0.06	-0.04
$\ln(c_{Fruit})$x11{OnlineAvail}	-1.27	-1.50	-1.07	-1.02	-0.79	-0.97
$\ln(c_{Grain})$x11{OnlineAvail}	-0.76	-1.01	-0.54	-0.42	-0.41	-0.53
$\ln(c_{Meat})$x11{OnlineAvail}	3.14	3.70	2.87	2.46	2.02	2.30
$\ln(c_{Oil})$x11{OnlineAvail}	-1.09	-1.34	-1.08	-0.91	-0.62	-0.67
$\ln(c_{Other})$x11{OnlineAvail}	0.23	0.23	0.18	0.17	0.14	0.15
$\ln(c_{Prep})$x11{OnlineAvail}	-2.74	-3.22	-2.45	-2.18	-1.61	-1.99
$\ln(c_{Snacks})$x1{OnlineAvail}	-1.79	-2.17	-1.70	-1.46	-1.13	-1.35
$\ln(c_{Sugar})$x11{OnlineAvail}	-0.42	-0.46	-0.39	-0.32	-0.24	-0.27
$\ln(c_{Veg})$x11{OnlineAvail}	-0.45	-0.47	-0.39	-0.32	-0.25	-0.30
$\ln(\frac{M}{C})$	-0.02	-0.03	-0.02	-0.02	-0.01	-0.02
$\ln(\frac{M}{C})$x11{OnlineAvail}	-0.14	-0.19	-0.12	-0.11	-0.08	-0.11
Constant	X	X	X	X	X	X
Household Demographics	X	X	X	X	X	X
Month f.e.	X	X	X	X	X	X
Linear Time Trend	X	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157	616,157
R^2	0.18	0.18	0.19	0.19	0.18	0.18

Standard errors available from the author upon request

Table A25: First Stage Estimates

Endogenous Variable	$\ln(p_{Other})x1\{O\}$	$\ln(p_{Prep})x1\{O\}$	$\ln(p_{Snacks})x1\{O\}$	$\ln(p_{Sugar})x1\{O\}$	$\ln(p_{Veg})x1\{O\}$
1{OnlineAvail}	-1.69	-5.42	-4.89	-5.06	-7.12
$\ln(c_{Dairy})$	-0.02	-0.02	-0.01	-0.02	-0.01
$\ln(c_{Drink})$	-0.05	-0.12	-0.11	-0.12	-0.17
$\ln(c_{Fruit})$	0.03	0.12	0.13	0.14	0.18
$\ln(c_{Grain})$	0.07	0.31	0.28	0.28	0.43
$\ln(c_{Meat})$	0.03	-0.07	-0.07	-0.08	-0.11
$\ln(c_{POil})$	0.05	0.05	0.05	0.05	0.07
$\ln(c_{Other})$	0.01	0.03	0.03	0.04	0.05
$\ln(c_{Prep})$	-0.08	-0.12	-0.09	-0.10	-0.14
$\ln(c_{Snacks})$	0.19	0.32	0.28	0.32	0.40
$\ln(c_{Sugar})$	-0.08	-0.10	-0.09	-0.11	-0.14
$\ln(c_{Veg})$	-0.05	-0.24	-0.23	-0.22	-0.32
$\ln(c_{Dairy})$x1{OnlineAvail}	0.30	0.96	0.89	0.95	1.27
$\ln(c_{Drink})$x11{OnlineAvail}	-0.16	-0.04	-0.02	-0.10	-0.02
$\ln(c_{Fruit})$x11{OnlineAvail}	-0.37	-0.92	-0.82	-0.95	-1.23
$\ln(c_{Grain})$x11{OnlineAvail}	-0.12	-0.54	-0.51	-0.43	-0.69
$\ln(c_{Meat})$x11{OnlineAvail}	0.62	2.23	2.00	2.09	2.96
$\ln(c_{Oil})$x11{OnlineAvail}	-0.45	-0.79	-0.71	-0.81	-1.10
$\ln(c_{Other})$x11{OnlineAvail}	0.10	0.16	0.14	0.12	0.21
$\ln(c_{Prep})$x11{OnlineAvail}	-0.36	-1.68	-1.75	-1.80	-2.54
$\ln(c_{Snacks})$x1{OnlineAvail}	-0.46	-1.35	-1.05	-1.21	-1.71
$\ln(c_{Sugar})$x11{OnlineAvail}	0.13	-0.29	-0.25	-0.09	-0.39
$\ln(c_{Veg})$x11{OnlineAvail}	-0.17	-0.32	-0.27	-0.30	-0.24
$\ln(\frac{M}{C})$	0.00	-0.02	-0.01	-0.01	-0.02
$\ln(\frac{M}{C})$x11{OnlineAvail}	-0.09	-0.10	-0.09	-0.13	-0.13
Constant	X	X	X	X	X
Household Demographics	X	X	X	X	X
Month f.e.	X	X	X	X	X
Linear Time Trend	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R^2	0.05	0.18	0.18	0.16	0.19

Standard errors available from the author upon request

Table A26: Demand Estimates
Online Households

Budget Shares	(1) Dairy	(2) Drinks	(3) Fruit	(4) Grains	(5) Oil
$\ln(p_{dairy})$	-0.00878**	0.0368***	0.00242	-0.0176***	0.0188***
	(0.00396)	(0.00227)	(0.00309)	(0.00308)	(0.00236)
$\ln(p_{drink})$	0.0368***	-0.0992***	0.0632***	0.0137***	-0.0127***
	(0.00227)	(0.00283)	(0.00237)	(0.00227)	(0.00174)
$\ln(p_{fruit})$	0.00242	0.0632***	-0.0121**	-0.0224***	0.0338***
	(0.00309)	(0.00237)	(0.00521)	(0.00432)	(0.00323)
$\ln(p_{grain})$	-0.0176***	0.0137***	-0.0224***	0.0978***	0.0274***
	(0.00308)	(0.00227)	(0.00432)	(0.0108)	(0.00589)
$\ln(p_{meat})$	0.0231***	-0.0311***	-0.0335***	0.0128*	-0.0230***
	(0.00470)	(0.00339)	(0.00626)	(0.00728)	(0.00564)
$\ln(p_{oil})$	0.0188***	-0.0127***	0.0338***	0.0274***	0.0321***
	(0.00236)	(0.00174)	(0.00323)	(0.00589)	(0.00624)
$\ln(p_{other})$	0.00649***	0.00915***	-0.00177	0.000621	-0.00228
	(0.00129)	(0.000901)	(0.00179)	(0.00291)	(0.00222)
$\ln(p_{prep})$	-0.0173***	-0.00231	-0.0533***	-0.0753***	-0.0381***
	(0.00387)	(0.00277)	(0.00489)	(0.00573)	(0.00426)
$\ln(p_{snack})$	-0.0893***	-0.0112***	0.0363***	-0.0132**	-0.0385***
	(0.00434)	(0.00307)	(0.00493)	(0.00631)	(0.00469)
$\ln(p_{sugar})$	0.000759	0.00630***	0.00777***	-0.00805**	-0.0116***
	(0.00147)	(0.00104)	(0.00211)	(0.00396)	(0.00311)
$\ln(p_{veg})$	0.0446***	0.0274***	-0.0204***	-0.0157***	0.0141***
	(0.00312)	(0.00217)	(0.00379)	(0.00472)	(0.00354)
$\ln(\frac{m}{P})$	0.0108***	-0.0322***	0.00994***	-0.00270**	-0.00134*
	(0.000609)	(0.000584)	(0.000780)	(0.00111)	(0.000795)
Constant	0.149***	0.215***	0.0638***	0.156***	0.0927***
	(0.0207)	(0.0218)	(0.0246)	(0.0250)	(0.0182)
Linear Time Trend	X	X	X	X	X
Month f.e.	X	X	X	X	X
Household Demographics	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R-squared	-0.690	-0.641	-0.873	-0.699	-0.778

Standard errors in parentheses
Standard errors clustered at the store-availability level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A27: Demand Estimates
Online Households

Budget Shares	(1) Dairy	(2) Drinks	(3) Fruit	(4) Grains	(5) Oil
1{Online}	0.455***	-0.199***	0.183***	0.117**	0.0442
	(0.0392)	(0.0327)	(0.0462)	(0.0490)	(0.0388)
ln(p_{dairy})$x1\{O\}$	0.198***	-0.0557***	0.138***	-0.00168	0.0230*
	(0.0226)	(0.0117)	(0.0183)	(0.0167)	(0.0131)
ln(p_{drink})$x1\{O\}$	-0.0557***	0.0664***	-0.00266	-0.00481	0.0350***
	(0.0117)	(0.0132)	(0.0127)	(0.0120)	(0.00925)
ln(p_{fruit})$x1\{O\}$	0.138***	-0.00266	0.0612**	-0.113***	-0.0332*
	(0.0183)	(0.0127)	(0.0305)	(0.0260)	(0.0194)
ln(p_{grain})$x1\{O\}$	-0.00168	-0.00481	-0.113***	0.122**	-0.0638**
	(0.0167)	(0.0120)	(0.0260)	(0.0524)	(0.0318)
ln(p_{meat})$x1\{O\}$	-0.0805***	-0.00716	0.0109	-0.0199	0.0143
	(0.0266)	(0.0179)	(0.0320)	(0.0332)	(0.0265)
ln(p_{oil})$x1\{O\}$	0.0230*	0.0350***	-0.0332*	-0.0638**	0.000709
	(0.0131)	(0.00925)	(0.0194)	(0.0318)	(0.0338)
ln(p_{other})$x1\{O\}$	-0.000976	0.0102**	-0.0166*	-0.00523	-0.0193*
	(0.00654)	(0.00469)	(0.00859)	(0.0126)	(0.0107)
ln(p_{prep})$x1\{O\}$	-0.208***	-0.0903***	0.0207	0.114***	0.0282
	(0.0207)	(0.0142)	(0.0267)	(0.0258)	(0.0196)
ln(p_{snack})$x1\{O\}$	-0.0795***	-0.0448**	-0.0189	0.109**	0.0148
	(0.0275)	(0.0179)	(0.0343)	(0.0442)	(0.0349)
ln(p_{sugar})$x1\{O\}$	-0.00408	-0.00259	0.0409***	-0.0362*	0.0258*
	(0.00794)	(0.00594)	(0.0120)	(0.0188)	(0.0154)
ln(p_{veg})$x1\{O\}$	0.0713***	0.0964***	-0.0874***	-0.100***	-0.0254
	(0.0187)	(0.0139)	(0.0239)	(0.0291)	(0.0227)
ln($\frac{m}{P}$)x1{O}	-0.0288***	0.0345***	-0.0118**	-0.0269***	-0.00255
	(0.00441)	(0.00412)	(0.00503)	(0.00512)	(0.00418)
Linear Time Trend	X	X	X	X	X
Month f.e.	X	X	X	X	X
Household Demographics	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R-squared	-0.690	-0.641	-0.873	-0.699	-0.778

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A28: Demand Estimates
Online Households

Budget Shares	(6) Other	(7) Prepared	(8) Snacks/Sweets	(9) Sugars	(10) Vegetables
$\ln(p_{dairy})$	0.00649***	-0.0173***	-0.0893***	0.000759	0.0446***
	(0.00129)	(0.00387)	(0.00434)	(0.00147)	(0.00312)
$\ln(p_{drink})$	0.00915***	-0.00231	-0.0112***	0.00630***	0.0274***
	(0.000901)	(0.00277)	(0.00307)	(0.00104)	(0.00217)
$\ln(p_{fruit})$	-0.00177	-0.0533***	0.0363***	0.00777***	-0.0204***
	(0.00179)	(0.00489)	(0.00493)	(0.00211)	(0.00379)
$\ln(p_{grain})$	0.000621	-0.0753***	-0.0132**	-0.00805**	-0.0157***
	(0.00291)	(0.00573)	(0.00631)	(0.00396)	(0.00472)
$\ln(p_{meat})$	-0.000450	0.0586***	0.102***	-0.0154***	-0.0581***
	(0.00305)	(0.00747)	(0.00954)	(0.00358)	(0.00706)
$\ln(p_{oil})$	-0.00228	-0.0381***	-0.0385***	-0.0116***	0.0141***
	(0.00222)	(0.00426)	(0.00469)	(0.00311)	(0.00354)
$\ln(p_{other})$	0.00154	-0.00297	-0.00200	-0.00562***	-0.00271
	(0.00162)	(0.00233)	(0.00248)	(0.00153)	(0.00190)
$\ln(p_{prep})$	-0.00297	0.164***	-0.00772	-0.00258	-0.0232***
	(0.00233)	(0.00867)	(0.00656)	(0.00268)	(0.00478)
$\ln(p_{snack})$	-0.00200	-0.00772	0.0114	0.0185***	0.00608
	(0.00248)	(0.00656)	(0.00955)	(0.00294)	(0.00555)
$\ln(p_{sugar})$	-0.00562***	-0.00258	0.0185***	0.00518*	0.00466**
	(0.00153)	(0.00268)	(0.00294)	(0.00276)	(0.00228)
$\ln(p_{veg})$	-0.00271	-0.0232***	-0.00657	0.00466**	0.0231***
	(0.00190)	(0.00478)	(0.00775)	(0.00228)	(0.00575)
$\ln(\frac{m}{P})$	0.00150***	-0.00180*	-0.0323***	0.000122	0.0122***
	(0.000526)	(0.000973)	(0.00104)	(0.000537)	(0.000772)
Constant	-0.00676	-0.0762***	-0.0380	0.0565***	0.0883***
	(0.00979)	(0.0278)	(0.0309)	(0.0116)	(0.0224)
Linear Time Trend	X	X	X	X	X
Month f.e.	X	X	X	X	X
Household Demographics	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R-squared	-0.160	-0.872	-0.343	-0.564	-0.491

Standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A29: Demand Estimates
Online Households

Budget Shares	(6) Other	(7) Prepared	(8) Snacks/Sweets	(9) Sugars	(10) Vegetables
1{Online}	0.00128	0.00274	-0.308***	-0.0139	-0.0208
	(0.0196)	(0.0491)	(0.0664)	(0.0241)	(0.0480)
$\ln(p_{dairy})x1\{O\}$	-0.000976	-0.208***	-0.0795***	-0.00408	0.0713***
	(0.00654)	(0.0207)	(0.0275)	(0.00794)	(0.0187)
$\ln(p_{drink})x1\{O\}$	0.0102**	-0.0903***	-0.0448**	-0.00259	0.0964***
	(0.00469)	(0.0142)	(0.0179)	(0.00594)	(0.0139)
$\ln(p_{fruit})x1\{O\}$	-0.0166*	0.0207	-0.0189	0.0409***	-0.0874***
	(0.00859)	(0.0267)	(0.0343)	(0.0120)	(0.0239)
$\ln(p_{grain})x1\{O\}$	-0.00523	0.114***	0.109**	-0.0362*	-0.100***
	(0.0126)	(0.0258)	(0.0442)	(0.0188)	(0.0291)
$\ln(p_{meat})x1\{O\}$	0.0281**	0.0930***	0.0316	-0.00748	0.118***
	(0.0120)	(0.0358)	(0.0506)	(0.0159)	(0.0364)
$\ln(p_{oil})x1\{O\}$	-0.0193*	0.0282	0.0148	0.0258*	-0.0254
	(0.0107)	(0.0196)	(0.0349)	(0.0154)	(0.0227)
$\ln(p_{other})x1\{O\}$	0.00608	0.00272	0.00547	-0.0124	0.00188
	(0.00596)	(0.00973)	(0.0144)	(0.00808)	(0.0100)
$\ln(p_{prep})x1\{O\}$	0.00272	-0.0128	-0.0426	0.00245	0.0925***
	(0.00973)	(0.0382)	(0.0381)	(0.0125)	(0.0270)
$\ln(p_{snack})x1\{O\}$	0.00547	-0.0426	0.0710	-0.0379*	-0.00764
	(0.0144)	(0.0381)	(0.0791)	(0.0199)	(0.0420)
$\ln(p_{sugar})x1\{O\}$	-0.0124	0.00245	-0.0379*	0.0341***	-0.00255
	(0.00808)	(0.0125)	(0.0199)	(0.0130)	(0.0139)
$\ln(p_{veg})x1\{O\}$	0.00188	0.0925***	-0.00764	-0.00255	-0.157***
	(0.0100)	(0.0270)	(0.0420)	(0.0139)	(0.0399)
$\ln(\frac{m}{P})$x1{O}	-0.00154	-0.0239***	0.0186**	0.00498**	0.00261
	(0.00212)	(0.00563)	(0.00748)	(0.00248)	(0.00510)
Linear Time Trend	X	X	X	X	X
Month f.e.	X	X	X	X	X
Household Demographics	X	X	X	X	X
Observations	616,157	616,157	616,157	616,157	616,157
R-squared	-0.160	-0.872	-0.343	-0.564	-0.491

Standard errors in parentheses
Standard errors clustered at the store-availability level
*** p<0.01, ** p<0.05, * p<0.1

Table A30: Difference in Elasticity Matrices, $\eta_{Instore} - \eta_{Online}$

	Own Price Elasticity	Dairy	Drink	Fruit	Grain	Meat	Oil	Other	Prepared	Snacks	Sugars	Vegetables
Dairy	-0.42	-0.42	0.18	-0.52	-0.22	-1.17	-0.07	0.07	1.85	-1.61	4.04	-1.05
	(0.05)	(0.05)	(0.10)	(0.10)	(0.10)	(0.91)	(0.22)	(0.45)	(1.04)	(0.43)	(63.11)	(0.98)
Drink	-1.26	0.31	-1.26	0.52	0.15	-0.60	-0.53	-0.57	0.89	-0.43	-2.25	-1.50
	(0.21)	(0.03)	(0.21)	(0.09)	(0.08)	(0.48)	(0.20)	(1.49)	(0.37)	(0.25)	(43.10)	(1.35)
Fruit	-0.34	-0.24	1.26	-0.34	0.37	-0.73	0.75	0.95	-0.28	0.39	-29.49	1.36
	(0.12)	(0.04)	(0.30)	(0.12)	(0.22)	(0.46)	(0.22)	(1.43)	(0.31)	(0.24)	(458.14)	(1.23)
Grain	0.12	-0.12	0.24	0.23	0.12	-0.21	1.03	0.30	-1.16	0.58	28.15	1.51
	(0.31)	(0.04)	(0.13)	(0.16)	(0.31)	(0.42)	(0.51)	(0.90)	(0.27)	(0.44)	(443.21)	(1.32)
Meat	-2.95	0.27	-0.64	-0.34	0.22	-2.95	-0.42	-1.66	-0.86	1.57	13.25	-1.54
	(2.42)	(0.06)	(0.24)	(0.14)	(0.20)	(2.42)	(0.32)	(2.00)	(0.79)	(0.50)	(217.96)	(0.85)
Oil	0.35	0.07	0.13	0.40	0.54	-0.33	0.35	1.11	-0.30	-0.39	-9.49	0.33
	(0.51)	(0.03)	(0.12)	(0.08)	(0.18)	(0.38)	(0.51)	(1.34)	(0.19)	(0.22)	(137.81)	(0.33)
Other	-0.35	0.04	0.30	0.05	0.03	0.33	0.19	-0.35	-0.03	0.01	11.13	-0.02
	(0.79)	(0.02)	(0.07)	(0.03)	(0.08)	(0.24)	(0.10)	(0.79)	(0.09)	(0.08)	(173.10)	(0.13)
Prepared	0.14	0.24	-0.98	-0.53	-1.20	2.16	-0.75	-0.18	0.14	-0.35	0.25	-1.23
	(0.82)	(0.05)	(0.38)	(0.12)	(0.17)	(1.51)	(0.25)	(0.58)	(0.82)	(0.28)	(10.66)	(0.79)
Snacks	0.64	-0.42	-0.65	0.36	-0.64	2.39	-0.59	-0.34	0.45	0.64	13.00	0.04
	(0.46)	(0.07)	(0.28)	(0.15)	(0.32)	(1.49)	(0.43)	(0.91)	(0.40)	(0.46)	(178.05)	(0.59)
Sugars	-24.56	0.01	0.10	-0.09	0.12	-0.38	-0.42	0.70	-0.02	-0.01	-24.56	0.02
	(378.50)	(0.02)	(0.06)	(0.06)	(0.14)	(0.31)	(0.18)	(1.46)	(0.12)	(0.13)	(378.50)	(0.19)
Vegetables	2.17	0.14	1.60	0.15	0.39	0.38	0.45	-0.13	-0.91	-0.11	-0.81	2.17
	(1.64)	(0.04)	(0.52)	(0.15)	(0.27)	(0.83)	(0.30)	(0.61)	(0.34)	(0.24)	(16.31)	(1.64)

Standard errors reported in parentheses
Standard errors clustered at the store-availability level

The Effects of the Dependent Coverage Mandates on Fathers' Job Mobility and Compensation*

Dajung Jun†

Department of Economics, Michigan State University

October 25, 2018

Click here for the most recent version

Abstract

Motivated by low rates of health coverage among young adults, some state governments began mandating health insurers to allow adult children to stay on their parents' insurance plans. First implemented in 1995, these mandates aimed to increase health coverage among young adults and reduce their dependence on employment for health insurance. In 2010, the federal government enacted a more comprehensive version of the dependent coverage mandate as part of the Affordable Care Act. These state and federal-level efforts successfully increased insurance rates for young adults, but they may have come with unintended implications for their parents. Parents who place a high value on health insurance for their young adult children may be reluctant to leave jobs with employer-provided health insurance. In addition, employers may offset the increased health care costs by reducing other types of employee benefits or earnings. To assess the extent of these impacts, I study the effects of both the state and federal dependent health insurance mandates on fathers. By analyzing the 2004 and 2008 SIPP panels, which are linked with Detailed Earnings Records and Business Registrar data from the U.S. Census, I examine the mandates' effects on voluntary job separation rates (as job-lock and job-push) and changes in fathers' compensations. After the implementation of the mandates, a 37 percent decrease in the likelihood of voluntary job separation among eligible working fathers aged 45–64 with health insurance is observed. I do not observe any significant effect associated with job-push. Additionally, the implementation of the mandates appeared to decrease total monetary compensation among eligible working fathers.

Keywords: Health Insurance, Government Regulation - Public Health, Job Mobility
JEL Classification: I13, I18, J6

*The research in this paper was conducted while the author was Special Sworn Status researcher of the U.S. Census Bureau at the Michigan Census RDC. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been approved for disclosure by the Census Bureau Disclosure Review Board (CBDRB-FY18-472; CBDRB-FY18-457).

†jundajun@msu.edu; 110 Marshall Adams, East Lansing, MI, USA 48824-1038; +1 517 582 3962.

Introduction

Historically, young adults aged 19–26 experience lower health insurance coverage rates than other groups. The main reason for this may be that young adults are generally healthy, so they may not perceive a need for health insurance (Barkowski and McLaughlin, 2018) and thereby forego health insurance purchases. For those young adults who do need health insurance, however, employer-provided health insurance (EPHI) could affect their job choice because EPHI is significantly more affordable than non-group plans. This is not always a viable option, however, as many young adults work entry level jobs where employers do not offer health insurance. Seeking to increase health insurance for this young adult population, both state and federal policymakers mandated that health insurers expand the age that children could remain covered under their parents' health insurance up to 23–26 years of age depending on the state. 30 states implemented the mandates beginning with Utah and North Dakota in 1995. In 2010, the federal government enacted the young adult dependent coverage mandate that applies to all states.[1] Although many studies confirm the positive effects of these mandates on young adults (Levine et al., 2011; Dillender, 2014; Cantor et al., 2012; Akosa Antwi et al., 2013), the literature still lacks studies detailing the implications for their parents. Because the mandates increase the value of jobs with EPHI for parents who have eligible children, parents' job mobility might be affected. The goal of this research is to determine the extent to which the state and federal mandates cause fathers to experience job-lock, remaining in their job for fear of losing access to EPHI, and job-push, seeking out jobs with EPHI that they would otherwise not have chosen.

A thorough examination of the mandates' effects on parents is needed due to the economic influence of this demographic group. Middle-aged workers (aged 45–64) are in the prime of

[1] The state-level mandates resulted in an increase in the dependent coverage rate by about 11.9 percent among the targeted group—as defined by age, state and year of implementation—compared to a non-targeted group (Burgdorf, 2014). Similarly, Monheit et al. (2011) suggest that there was a 10 percent increase in dependent coverage among the targeted group of young adults after policy implementation, relative to a control group of young adults. Due to the federal mandate, 4.5 million additional young adults who would not otherwise have had coverage have been insured (Furman and Fiedler, 2015).

their careers with longer relevant experience and higher earnings than workers who have just entered the workforce. Thus, the mobility decisions of these workers may have a critical influence on their careers and retirement savings.

Parents' mobility decisions are highly likely to be influenced by these mandates for the following reasons: (1) parents value these mandates as they provide a safeguard for their adult children's health and financial security while promoting their children's career progression, and (2) health insurance enrollment decisions in the United States are often made at the immediate family level as opposed to the individual level (Cutler and Gruber, 1996) in order to be cost-effective. For these reasons, EPHI adds additional value for parents. Thus, the cost of leaving an employer for parents with EPHI increases. Therefore, I expect workers to be less likely to leave their jobs when their children are eligible for dependent coverage mandates. Conversely, for parents without EPHI, these mandates make their current state of employment less attractive; as such, I would expect these people to be more inclined to pursue jobs with EPHI.

In addition to examining an under-studied population affected by these mandates, this paper makes two other contributions to the literature. First, I exploit both the state and the federal dependent coverage mandates to explore their effects on fathers. Second, it sheds light on a decrease in compensation by using data from the Survey of Income and Program participation (SIPP)—a publicly available, self-reported survey—linked with the administrative data from the Detailed Earnings Records (DER) and Business Registrar (BR), made available through the United States Census. While I do not observe any effect solely based on the SIPP, the combination of data from the SIPP, DER and BR provide a new perspective on the effect of the mandates on working fathers' compensation.

I find that working fathers with eligible children experienced a 37 percent decrease in the likelihood of voluntary separation from employers with EPHI. I was not able to detect any job-push effect. The results suggest that these mandates lowered total monetary compensation among eligible fathers with EPHI.

Institutional Details

In the United States, policymakers have long recognized the low rates of health insurance among young adults and the associated labor market distortions. Before the dependent coverage mandates required insurers to extend the age limit for dependents, most public (e.g., Medicaid and CHIP) and private health plans (e.g., self-insured EPHI, EPHI through an insurance company, or plans through the non-group market) removed young adults from their parents' polices when they turned 19 unless they were enrolled in a college or university as a full-time student.[2] If a dependent was a full-time student, then he or she was typically covered through the age of 22. This left many young adults who were not currently in the college without insurance. Moreover, in some states, the tax code defined coverage of dependents 19 years of age or older as a taxable benefit, deterring employers from extending coverage. Due to their limited access to health insurance, some young adults had a greater incentive to choose employment with EPHI (Hahn and Yang, 2015), rather than their ideal jobs.[3]

To increase health insurance and to weaken the link between health insurance and labor supply for young adults, state policymakers expanded access to dependent coverage. In the absence of state funds to expand public programs (Goda et al., 2016), many states required firms that offered dependent coverage to increase their age threshold (generally up to 24–26 years of age). By 2010, 30 state-level dependent coverage expansions were in effect (see Appendix Table 1).[4]

[2]The dependent, in this case, refers to biological, legally adopted, or legally fostered children or to children for whom one has been appointed legal guardian.

[3]These types of jobs, however, often required full-time employment, which may limit young adults' time and prevent them from searching for more optimal jobs (Colman and Dave, 2017).

[4]While all states with state-level mandates increased the age limitations, there were some states with no age limit (i.e., Iowa and Texas) and other states that extended the provision to age 29 (i.e., New York, New Jersey and Pennsylvania). Some states had eligibility criteria like residency within a state or with a parent policy holder. These criteria, however, are not important concerns because the requirement for residency with a parent would be difficult for insurers to enforce (Barkowski and McLaughlin, 2018). Some states might also require student status, single marital status or financial dependency to be qualified as a dependent. Moreover, the parents with EPHI from self-insured firms were exempt from the state mandates under the Employee Retirement Income Security Act of 1974. Finally, most states did not regulate the employee-paid premiums that could be levied for coverage of older dependents, potentially allowing firms to raise prices

Following the states' leads, the federal government enacted the dependent coverage expansion through the 2010 ACA, which required insurers to expand coverage to children up to the age of 26 on their parents' plans. Whereas some state mandates limited eligibility based on factors other than age, the federal law is straightforward: any insurance plan that already offers dependent coverage must offer the same level of benefits at the same price to dependents 26 years of age or younger.[5]

Although the federal mandate has fewer eligibility criteria than some states, both mandates extend the eligibility age. In addition, both the state and federal mandates pursue the same goal: increasing health insurance coverage as well as decreasing dependence on employment for health insurance among young adults. As the implicit value of being employed in a job with EPHI decreases with the mandates, the young adults who become eligible can choose a job aligned with their career aspirations rather than a job that simply provides EPHI.[6] Furman and Fiedler (2015) demonstrate that due to the federal mandate, by 2014 the uninsured rate among young adults dropped by more than 40 percent compared to that of 2009—which translates to 4.5 million additional young adults with coverage. The percent of other non-elderly adults aged 26 to 64 without insurance coverage was stable during that time.[7] Cantor et al. (2012) note that the federal mandate is a "rare public policy success in the effort to cover the uninsured [young adults]."[8] Given the similarities in the state and federal mandates, examining them together is important. If I examined the federal mandate

above what employees could afford.

[5]The requirement does not depend on co-residence with parents, student status, marital status or financial dependency. It applies to all insurance plans including self-insured EPHI, fully-insured EPHI, and plans from the non-group market. The federal law revises the IRS rules so that the benefit offered to the newly eligible young adults would be tax exempt. Dependent coverage under a parent's plan, however, does not extend to the children and spouses of those dependents.

[6]In fact, the Federal Registrar (2015) indicates that one of the most significant aims of this law was to "permit greater job mobility ... as [young adults'] insurance coverage would no longer be tied to [young adults'] own jobs or student status."

[7]These gains eliminate more than two-thirds of the gap in uninsured rates between young adults and other non-elderly adults, even as other non-elderly adults also experienced large coverage gains during 2014 (Furman and Fiedler, 2015).

[8]However, there is one caveat that is worth noting: this coverage extension might not work well for young adults living out-of-state because their parents' plan might only provide expensive, out-of-network coverage (Goldman, 2013; Reinicke, 2018).

alone, it would lead to a misclassified model since the results would not take into account the effects of the state mandates that had already been enacted. My approach primarily relies on within-state time variation because the state mandates had different ages of eligibility and timing of implementation of mandates.

While the mandates could allow young adults more freedom in choosing their jobs, they may limit job choices for their parents. By staying in their current jobs with EPHI, parents are able to provide more comprehensive plans for their children's health and financial security at a much lower cost. This also enables their children to progress professionally with less concern about their own coverage.[9] Therefore, I expect a decrease in job mobility of these parents after the implementation of dependent coverage mandates. For parents without EPHI, however, the dependent coverage mandates increase the opportunity cost of maintaining their current employment status. In this case, I expect an increase in the job mobility of parents with eligible children. In addition to job mobility, I explore whether eligible working fathers with EPHI experience a reduction in annual earnings or other types of compensation as these mandates raise the relative cost for employers to hire working fathers with eligible adult children.[10] This exploration is critical since the efficiency of these mandates largely depends on the extent to which their costs are shifted to group-specific wages.

Literature Review

A large body of empirical literature exists regarding job-lock. Most empirical and anecdotal evidence suggests that mobility constraints in the labor market stem from the fear of losing health care coverage. The range of the magnitude for job-lock is from 30 percent to 80 percent and varies depending on the identification strategy or demographic group (Gruber and

[9] For example, Brandeisky (2015) shows that an individual premium costs $486 a month for young adults, on average, in 2015. By adding two or more dependents to the parents' plan, however, a premium costs an average of $1,377 a month. When this is split by three or four, it is still less than an individual plan.

[10] In 2012, the average employer contribution for employees' family plans was about 73 percent, which is converted to $ 11,429 (The Kaiser Family Foundation, 2017). As the dependent children were covered under the fathers' family plan, the financial burden of the insurance premium and health care cost would be transferred not only to the parents, but also to their employers (Chen, 2018). In fact, employers would likely pay the majority of the costs.

Madrian, 2002). For instance, Rashad and Sarpong (2008) find that individuals with EPHI are 60 percent less likely to voluntarily leave their jobs compared to those receiving insurance elsewhere.[11] Many job-lock studies are based on the idea that a worker's unique demographic characteristics—such as proximity to retirement or health status—may lead him or her to value insurance more highly than others, making that worker more vulnerable to job-lock. For the elderly, Kapur and Rogowski (2007) and Blau and Gilleskie (2001) point out that there is unambiguous evidence that health insurance is a central determinant of retirement decisions. Additionally, Bradley et al. (2005) find that EPHI appears to create incentives to remain working and to work at a greater intensity when faced with a serious illness. Gruber and Madrian (2002) and Rashad and Sarpong (2006) have both written comprehensive literature reviews on the topic of job-lock. Compared to the abundant literature about job-lock, however, fewer researchers study job-push. Anderson (1997) wrote one paper studying job-push and suggests that EPHI encourages some workers to leave jobs that are otherwise desirable.

There is considerable literature about the state and federal mandates that focus on decreasing job-lock and job-push for young adults. Two studies by Levine et al. (2011) and Monheit et al. (2011) estimate the effect of the state-level mandates and conclude that the mandates successfully increased coverage among young adults through EPHI received as a dependent. Literature regarding federal mandates including Cantor et al. (2012), Akosa Antwi et al. (2013) and Sommers et al., (2013) finds that when the federal government enacted the ACA, it developed a pathway for more young adults to be insured. According to Colman and Dave (2017), the federal mandate also reduced job-lock among young adults.[12]

Most papers in the literature, however, examine these various outcomes by studying state- and federal-level mandates independently. So far, only one paper analyzes these mandates

[11] The main estimate in this paper—a 37 percent decrease in job mobility among parents—is comparable to theirs. Although my estimate is slightly smaller, this is to be expected because the dependent coverage mandates are not primarily targeted at health insurance coverage of parents.

[12] Colman and Dave (2017) suggest that since young adults feel less pressure to secure a full-time job with EPHI, they work fewer hours and are able to spend extra time on social, educational and job-search activities.

together (Barkowski and McLaughlin, 2018); however, this work focuses on the effects of these mandates on marriage among young adults. Most studies that only focus on the federal mandate justify not accounting for the effects of the state mandates by arguing that these effects were minimal. Instead, they only compare young adults under the age of 26—the cut-off age for the federal mandate—to those above, which is a strategy similar to that of Cantor et al. (2012).

Despite the plethora of papers regarding the effects of the dependent coverage mandates on young adults, few researchers consider other populations. There is only one paper that studies the effects of the dependent coverage mandates on parents' retirement decisions (Biehl et al., 2018), but it solely uses the federal mandate for identifying variation along with the Health and Retirement Study data. Also, this paper only considers retirement decisions without analyzing other types of voluntary job separation.

Even though little prior work explicitly investigates the link between the dependent coverage mandates and parents, there is some evidence that mandates related to child health insurance affect parents' voluntary job separation. For instance, Chatterji et al. (2016) find that *the ACA prohibition of the pre-existing condition exclusions for children* increased the likelihood of leaving an employer voluntarily by 0.7 pp among fathers of disabled children relative to fathers of healthy children. Barkowski (2017) also argues that Medicaid eligibility for household members (especially for eligible children) increased the probability of a voluntary job separation by 34 percent among working fathers with EPHI. Hamersma and Kim (2009) find that Medicaid may also have decreased job-push, suggesting that unemployed fathers or working fathers without EPHI feel less need to move to jobs that offer insurance.[13] As a caveat, Barkowski (2017) and Hamersma and Kim (2009) focus on low-income workers, while Chatterji et al. (2016) investigate job mobility of parents with disabled children. Since both groups might be systematically different from the general group of middle-aged fathers, a more comprehensive approach for this group merits discussion and is the basis of my work.

[13] This is mainly because expanded eligibility for their children through Medicaid decreases the perceived value of employment with EPHI for those fathers.

In addition to job mobility, several papers examine whether the health benefit mandates affect eligible workers' annual earnings or other types of compensation, as these mandates raise the relative cost for firms to insure them. This exploration is critical since the efficiency of these mandates largely depends on the extent to which their costs are shifted to group-specific wages. There may be several factors, however, that prevent this full group-specific shifting of the total cost of the mandated benefits to beneficiaries. For example, anti-discrimination regulations and workplace relative-pay norms make it difficult for employers to reduce the wages of the targeted group.[14] In spite of these regulations, Gruber (1994) finds substantial shifting of the costs of the group-specific mandates to the wages of the targeted group, suggesting that maternity mandates are an efficient tool of social policy. This study inspired many articles seeking to determine the effects of mandated health insurance on earnings. Monheit and Rizzo (2007) review the relevant literature regarding the costs of various mandates for employees and employers.

Alternatively, the cost of providing additional insurance may be shared by other co-workers. To investigate this, Goda et al. (2016) simulate various cases where the magnitude of the reduction in wages varies by the degree of pooling of employees. Depending on the size of pooling, the wage reduction ranges from $30 to $1,500 per worker.[15] Another strategy that employers may use would be to decrease total monetary compensation (e.g., employer contribution for deferred compensation) for eligible workers instead of directly adjusting their wages.

Many studies explore connections between health insurance and other labor market outcomes. There is also a budding literature devoted to the effects of extending health insurance coverage to young adults, specifically. My paper extends this literature by looking at more comprehensive reforms, at both the state and federal levels, and considering the potential

[14] Anti-discrimination regulations prohibit differential pay for the same job across groups and prevent differential promotion decisions by demographic characteristic. Furthermore, workplace norms that prohibit different pay across groups and union rules about equality of relative pay may have similar effects as the anti-discrimination regulations.

[15] Despite the wage reduction caused by the dependent coverage mandates, Goda et al. (2016) do not find any evidence that suggests workers reduce their labor supply in response to the lowered wages.

consequences of these reforms on parental labor force decisions.

Data

To assess the effects of policy changes on health insurance outcomes, I leverage detailed information on individuals using the 2004 and 2008 SIPP panels, which are linked to the DER and BR. The SIPP is a nationally representative household survey where each panel is divided into waves of four months. The time period covered in my data is January 2004 to December 2012.[16] This is when most state-level dependent coverage provisions and the federal mandate were implemented. The entire sample is divided into four subsamples called rotation groups. One rotation group is interviewed per month during a 4-month wave. Most SIPP questions ask the respondent to report information regarding the four months prior to the interview (SIPP Users' Guide).

Every SIPP wave includes a core set of questions about labor market outcomes, health insurance coverage and participation in government programs. In this study, I use these core questions to collect information about health insurance, demographic characteristics and employment. The SIPP provides a detailed set of information about current employment for up to two jobs in a given wave. I only include the job that is considered the 'primary job'—the job in which the individual worked the most hours. The data also shows the main reason for leaving their employer within this wave, if applicable, which allows me to separate voluntary versus involuntary job separation. In my analysis, I focus on voluntary job separation—transitioning between jobs, becoming unemployed, leaving the labor force or transitioning from working for an employer to self-employment.[17] In addition, there are topical questions included only in selected waves. For example, I use the wave 2 topical module questions that ask for the year when the respondent's last child was born. If the child was the only

[16]Because the Great Recession occurred during this time, labor force decisions might be different in my sample than in those from other periods.

[17]Involuntary job separations include layoffs, childcare problems, family/personal obligations, illness/injury, school/training, employer bankruptcy/change in ownership, termination of a temporary job, and unsatisfactory work conditions.

child, then I use the question for the year when the respondent's first child was born. My analysis is primarily based on the youngest children because fathers generally need coverage until their youngest children grow up.[18] By subtracting the birth year of the last child from the interview year, I get the youngest child's age and only include fathers with children aged 19–29. Then I link these fathers' information to the core SIPP data which provides detailed demographic characteristics, time of interview and state of residence. None of the variables that I use are imputed.

Although the SIPP provides detailed, self-reported demographic characteristics, the linked dataset between the SIPP and other administrative records (i.e., DER and BR) provides highly accurate measures of earnings and total monetary compensation. First, I use the respondents' Social Security Numbers (SSN) to link them to the DER. The DER includes their W-2 information such as wages and employer contributions to retirement benefits.[19] If the respondents had two or more W-2s that originated from the same parent company, I sum up their earnings to calculate their annual earnings.[20] Determining whether the W-2s are from the same parent company is possible because additionally linked BR data includes information like type of firms (i.e., single-unit or multi-unit) and the parent company for all companies in the United States.[21] In my research, therefore, linking SIPP data with the administrative data allows me to achieve a more comprehensive understanding of the compensation adjustments caused by the mandates. Furthermore, with this linked data, I can

[18]I use the youngest child to construct all my samples, including the sample for job-push analysis. For the job-push sample, however, any child could affect fathers' job mobility decisions. So I also run the analysis based on the oldest child for the job-push analysis. This does not change my results.

[19]The DER-SIPP linkage is only available until 2012 through the United States Census, so the observations for 2013 cannot be included from the 2008 SIPP.

[20]This analysis considers cases where employees move locations and therefore have two or more W-2s on file. For example, this could include a Walmart worker who is still part of the company but works in two locations within a year. If the W-2s did not come from the same parent company, however, I only consider the W-2 with the highest earnings to link with the corresponding SIPP data. I also run the analysis by omitting all individuals who had two or more W-2s that did not come from the same parent company (instead of using the highest earnings from the W-2s). The results from the samples with and without accounting for these special cases demonstrate no significant difference.

[21]Firms themselves sometimes change or have multiple EINs for payroll or tax purposes or for establishments with different locations within the same firms. In my analysis, about 20 percent of people who have two or more W-2s on file were from the same parent companies.

assess the accuracy of the survey data and adjust for errors in reported earnings.[22] Earnings are inflation-adjusted using the yearly CPI-U indices and 2012 as the base year.

Although most SIPP questions involve asking the respondent to report information for each of the four months prior to the interview month, I only include the responses from the interview month in order to mitigate seam bias, which is the tendency for respondents to report higher rates of events between survey waves than within survey waves (Blank and Ruggles, 1996). This means that the analysis is conducted at the father-wave level rather than the father-month level.[23]

To code the eligibility criteria for the mandates, I compile the data regarding state laws (e.g., age limit and timing of implementation) from Depew (2015), Cantor et al. (2012) and the National Conference of State Legislatures (2010). These data can be represented in Appendix Table 1. I demonstrate the change in eligibility for fathers from three of the states that introduced dependent coverage mandates in the first three rows in Table 1. The last row in Table 1 represents eligibility for fathers in states without state-level mandates. I code the fathers in these states as eligible after September 2010 when the ACA was implemented.

In Table 2, I include the sample means for the outcome variables and covariates for job-lock and job-push in two panels. For both panels, the columns titled *Always Ineligible* contain the descriptive statistics for fathers who were not affected by the state and federal mandates. In other words, this is the intersection of the fathers whose children were *ineligible* across all time periods in Table 1. *Ever Eligible* is the group of fathers who were eligible at some point in my analysis. This is the union of fathers with *eligible* children across time periods in Table 1. Both panels in Table 2 include married fathers between the ages of 45 and 64 with

[22]Bridges et al. (2003) find substantial measurement error in SIPP wage and salary data. They argue that the mean SIPP wages were understated by 7.5 percent relative to DER wages. Gottschalk and Huynh (2005) also suggest that respondents with SIPP information but no DER records had lower earnings than respondents with observed earnings in both data sets, possibly reflecting informal work arrangements.

[23]I focus on fathers only because they have more predictable labor force patterns and persistent attachment to their jobs than mothers. The wage-labor supply elasticity of fathers is often much smaller than that of mothers (Blundell and MaCurdy, 1999).

their youngest child between the ages of 19 and 29.[24] They both exclude the states that had no age limit (i.e., Iowa and Texas) and the states that extended the provision to age 29 (i.e., New York, New Jersey and Pennsylvania).

For the job-lock analysis (see Panel A), I include fathers who, in the previous wave, were compensated for their work by an employer, had EPHI under their own name and were not self-employed. The sample includes approximately 14,500 working fathers, 75 percent of whom had an *ever eligible* child. The rate of voluntary job separation within a 4-month wave, on average, is 2 percent. These rates are similar to those reported in other papers based on SIPP data (Barkowski, 2017; Chatterji et al., 2016). Hamersma and Kim (2009) report rates of voluntary employer separation ranging from 3 to 5 percent for employed, low-income parents within the 1996 and 2001 SIPP data. The small deviation may exist between the rates, in part, because I do not limit the sample to low-income respondents and the 2008 SIPP was conducted during a severe recession. Although *always ineligible* fathers within my sample are generally older, less-educated and less likely to be white, almost all other characteristics are comparable. Bansak and Raphael (2008) argue that roughly 18 percent of workers with EPHI separated from their employers within a year. This number is relatively high compared to my data even after considering that their paper focuses on the separation that happened during a year—between wave 1 and wave 4—instead of during a 4-month period. Two additional factors could drive this difference in the means. First, Bansak and Raphael (2008) are considering all separations, not just voluntary ones. Second, they consider between-wave measures for job-mobility while I look for the job separation that happens within a wave.[25]

The fathers in the job-push sample have the same demographic characteristics (e.g., their ages and the ages of their children) as those in the job-lock sample. The only important difference is that the job-push sample includes fathers who, in the *previous* wave, were unemployed or did not have EPHI from their employers. Selecting the sample in this way limits

[24] I do not extend the child's age up past 30 because adult children over 30 are systematically different from adult children in their early 20s in terms of life stage.

[25] Appendix A of Chatterji et al. (2016) explains why measuring separation using 'within-wave' would be more plausible than using 'between-wave.'

the chance that job-push and job-lock would be conflated since individuals without EPHI could not be affected by job-lock.

When I examine the effect of mandates on compensation, I use the job-lock sample subset, which contains fathers who did not voluntarily change their jobs within the wave. This is because the compensation decrease would mainly affect those who stay in their jobs and are eligible to benefit from the mandates. As expected, the descriptive statistics for this sample, which is shown in Appendix Table 3, is remarkably similar to Panel A in Table 2.

Identification Strategy

This study examines the effects of both the federal and state-level dependent coverage mandates on fathers. The primary comparison is between two groups of fathers within each state before and after the implementation of the mandates: those with youngest children whose ages are at or beneath mandate thresholds and those with youngest children whose ages are above.

The model is specified as

$$y_{ijt} = \Phi(\beta_0 + \beta_1 * Elig_{ijt} + \beta_2 * X_{it} + \beta_3 * time_t + \beta_4 * state_j + \epsilon_{ijt}) \tag{1}$$

where (1) is the probit model in a difference-in-differences framework for individual i in state j and time (wave) t. $\Phi(.)$ is a standard cumulative normal distribution function.[26] y_{ijt} is the outcome variables for voluntary job separation and health insurance for the fathers.

To investigate job-lock, y_{ijt} is equal to one if a voluntary job separation happened in that wave. For the health insurance coverage, y_{ijt} indicates whether fathers currently have health insurance that covers their young adult children.

$Elig_{ijt}$ is the main independent variable and is determined by three things: state of res-

[26] Given that I limit my sample to those who are at risk of changing their employment status, this model can be viewed as a standard discrete time hazard model (Bruce Meyer, 1990).

idence, year of interview and a youngest child's age.[27] For a given year, fathers are coded as eligible if they were living in a state with a mandate **in effect** and had a youngest child whose age was at or beneath the mandated age. For instance, in the case of a father with a youngest child who was 24 years-old living in Colorado in 2006, I would code him as eligible ($Elig_{ijt}$=1) because Colorado enacted a dependent coverage mandate at that time. For another father in Colorado in the same year but with a youngest child who was 25 years-old, I would code him as ineligible ($Elig_{ijt}$=0) because the child's age exceeded the limit of the Colorado mandate. Other requirements—most importantly, student status—are inappropriate to use for eligibility imputation because they are jointly determined outcomes.[28] I expect the coefficient of $Elig_{ijt}$, β_1, to be negative for the job-lock analysis.

X_{it} contains other covariates, including father's age and dummy variables for the following: high-school dropouts and high-school graduates, non-Hispanic white respondents, and public sector workers. I include full sets of state and year indicators, denoted with $time_t$ and $state_j$, to focus on within-state variation.[29] In addition, indicators for all the children's ages from 20–29 (the indicator for 19-year-olds serves as the baseline group) are included to account for the time invariant, behavioral difference of fathers of various aged young adults. The regression also contains state-specific, linear time trends.

The empirical strategy I rely on to detect job-push is conceptually similar to the one I use for job-lock, but with one important change: different sample selection criteria. Therefore, for job-push analysis, y_{ijt} is equal to one if fathers voluntarily left their previous jobs without EPHI (or were unemployed in the previous wave). A positive estimate of β_1 would provide evidence of job-push, meaning that fathers seeking health insurance to cover their child would be more likely to transition from their current employment status.

[27] If the father had a 22-year-old and a 16-year-old, then I do not include him in my sample.

[28] For example, a state mandate might induce individuals into or out of student status, so using it to determine eligibility would introduce bias (Depew, 2015). That being said, in states like Florida where the state mandates required student status, I would consider a father with a child who was a 21-year-old, full-time student as "eligible" **only after** the state mandate was implemented. This empirical strategy may lead to attenuation bias in the point estimate of interest (Depew, 2015).

[29] I did not include the individual-fixed effects since this Difference-in-Differences design addresses possible unobserved characteristics already.

I also examine the mandates' impact on working fathers' annual earnings or total monetary compensation based on the fathers who stayed in their jobs with EPHI—the aforementioned subset of job-lock sample. Since employers can easily identify the group of working fathers with eligible children, they may respond to the extra cost of providing dependent coverage by reducing other compensations for this group. To examine whether the cost of the mandate was transferred to working fathers with eligible children, I replace y_{ijt} with the natural log of either the annual earnings or the total monetary compensation. This total monetary compensation is the logarithmic value of the sum of annual earnings and deferred compensation. For those working fathers who were eligible for the mandates, therefore, I would expect a negative estimate of β_1. Compensation was zero for some working fathers and is automatically omitted when I run the regression since I use the natural log of compensation as the dependent variable.[30]

Results

Table 3 shows the evidence for an increase in job-lock among working fathers due to the dependent coverage mandates. All columns are estimated with a full set of covariates. The results in columns 1–2 are estimated without state-specific linear trends; those in columns 3–4 include these trends. Since multistage-stratified sampling is an important aspect of the SIPP, columns 1 and 3 are weighted while columns 2 and 4 are not.[31] The results in this table indicate that after the mandates took effect, the average probability of leaving an employer for any voluntary reason was 0.7 pp lower for working fathers with eligible children than for fathers without.[32] The 0.7 pp is a 37 percent decrease in voluntary job separation given that

[30] I treat fathers who have a valid SSN but no record in the DER as having zero earnings (Chenevert et al., 2016). Earnings may be absent from the DER because an employer may fail to report an employee's wages to the Social Security Administration. These workers without DER data are more likely to work in private households, construction, agriculture and informal occupations (e.g., street and door-to-door sales work, dancing or bartending) (Roemer, 2002).

[31] Except Table 9 and Appendix Table 2, I only report the weighted estimates for the remaining tables.

[32] Even with alternative regressions such as the linear probability model (LPM) and the logit (Appendix Table 2), similar effect sizes can be observed.

the average separation rate of *ever eligible* working fathers before the implementation was approximately 2.0 percent. This 2.0 percent rate is similar to the magnitudes reported in other papers based on the SIPP. Barkowski (2017), for example, indicates a voluntary job separation of 2.3 percent, although it is important to note that his rate is based on the 1980s and early 1990s SIPP panels. Hamersma and Kim (2009) demonstrate rates of voluntary job separation ranging from 3 to 5 percent for employed, low-income parents in the 1996 and 2001 SIPP panels. The rates in my sample are likely to be lower than those of Hamersma and Kim (2009) since I do not limit the sample to low-income respondents.

The magnitude of my results in Table 3 is comparable to the effects of similar mandates targeting children on the mobility decisions of working fathers. Barkowski (2017) suggests that Medicaid eligibility for one household member results in a 34 percent increase in the likelihood of a voluntary job exit among working member(s) in the household. However, when Barkowski (2017) further restricts the sample to exclude those workers in the top income decile—who are less likely to be affected by Medicaid—the resulting estimates are increased by up to 71 percent. Chatterji et al. (2016) also demonstrate that the *ACA prohibition on pre-existing condition exclusions* increases the probability of job exit by 35 percent for married fathers with disabled children compared to fathers with healthy children. These previous works are comparable to mine because they focus on whether mandates for children can influence parent's reliance on employment; minor differences in magnitudes or sign can be attributed to the different time periods and policies examined.[33] In Table 4, I examine the robustness of the results. To compare these estimates with the main results, column 1 is taken directly from column 3 of Table 3. In column 2, I expand the control group by including working fathers whose children were aged 27–33 to see if the result would vary depending on the range of the control group. In column 3, I also examine whether expanding the time period with the 2001 SIPP panel would alter the results. In this analysis, I omit

[33] These previous works focus on whether parents' reliance on employment for health insurance decreased, while I examine whether it increased. The primary idea, however, remains the same: health insurance mandates for children affect fathers' labor market decisions.

five more states (besides those five states that were excluded from the main analysis) because they were sampled together in the 2001 SIPP (Wyoming, Vermont, Maine, South Dakota and North Dakota). Because I was unable to verify the implementation dates for the mandates in Georgia, Nevada, South Carolina and Wyoming with more than one source (Goda et al., 2016), I exclude fathers from these states in column 4. All columns are significant and had a similar magnitude as column 1. In column 5, I treat the state mandates that have student status requirements as if they do not have mandates (i.e., Florida, Idaho, Louisiana, Massachusetts, North Dakota, Rhode Island and South Dakota). Although the result in column 5 loses statistical significance, it still has a p-value of about 0.101.[34]

In Table 5, I examine the heterogeneity of the results. Working fathers with higher education might be more responsive to the mandates given their knowledge of dependent coverage mandates. On the other hand, they might also be less likely to have children who need the dependent health coverage, because their children would be better educated and more likely to secure jobs with EPHI. Column 1 indicates the effects of the mandates on working fathers with some college education more, while column 2 shows those on working fathers with less than or equal to a high school education. The results suggest that more educated fathers have less job mobility as a result of the mandates, supporting the former hypothesis.

In Table 6, I perform falsification tests. In this table, my sample is comprised of working fathers in three groups: those with a youngest child between the ages of 8–18, 30–40 and 27–37. This allows me to implement a placebo eligibility among working fathers with youngest children unaffected by the mandates due to their ages. For example, I use a sample that is comprised of working fathers whose youngest children were between 8 and 18 years old. Then, I consider the placebo (state or federal) mandates' eligibilities by subtracting 11 from

[34] As mentioned in *Institutional Details*, full-time students aged 19, 20, 21 and 22-years-old before the mandates were implemented were often considered to be eligible under their parent's plan. In my main analysis, however, I assume all working fathers with children aged 19–22 are considered as eligible **only after** the mandates were implemented. This may raise a concern whether it is valid to consider the states that required student status as mandated states because this only applied to . By treating these states as states that had no mandate before the ACA, the results in column 5 show the consistency of my main results in Table 3.

age eligibility criteria (e.g., if the mandate expanded coverage to dependent children up to the age of 23, I would consider this state's placebo age limit to be 12). By doing this, I can examine whether those working fathers with ineligible children under 19 seem to be affected by the mandates. I repeat this process for working fathers with children aged 30 to 40. For these working fathers, I add 11 to the age eligibility (e.g., if the mandate increased the age limit up to 23, I would consider this state's placebo age limit to be 34).

One drawback of the first two falsification tests is that they do not include any fathers who are in my main sample. To rectify this, I use the sample containing working fathers with children aged 27–37 in column 3 of Table 6. With this analysis, I can examine whether there are time-effects (i.e., other circumstances that have changed over time and affect parents differentially based on the age of their children). In this column, I still include the same *Always Ineligible* fathers with children aged 27 to 29 as I use in my main analysis. However, I alter the *Ever Eligible* group by adding 11 to the original age limit, resulting in a sample of fathers with 30 to 37-year-old children. None of the falsification tests have any significant effects and the point estimates are appreciably different from my main findings in Table 3.

Fathers may not only experience job-lock due to these mandates, but also the twin phenomenon of job-push. In Table 7, I examine whether these fathers were incentivized to leave a job without EPHI due to the increase in opportunity costs of staying in their current employment status. Unlike Hamersma and Kim (2009) and Barkowski (2017) who study the Medicaid expansions, I do not find evidence of job-push for fathers. This might be because of the period in which parents are affected by the policies. While Medicaid influences parents for a long period of time, the dependent coverage mandates only affect parents for a short period when their children are in their early 20s. Parents might be less motivated to change their employment status (in this case, finding a new job with EPHI) for a short-term benefits.

To examine how employers adjust employee compensation in response to rising health insurance costs, I analyze the change in total monetary compensation in addition to the annual earnings. This is because it might be difficult for employers to directly adjust earnings for the eligible working fathers without violating non-discrimination laws. Table 8 presents the effects of this mandate on earnings and other compensation among working fathers who did not leave their jobs with EPHI, based on both the administrative data and the public data.[35] In columns 1–2, I omit the individuals who had zero compensation in the administrative data.[36] It appears that there was a modest decline in earnings and total compensation.[37]

To compare the results in columns 1–2 with those based on the public data, I include the latter results in column 3 of Table 8. Unlike the results from the administrative data, I do not find any significant effect on annual earnings with public SIPP data alone.[38] As another type of falsification test, I run the same analysis for the fathers without EPHI. Results are included in Appendix Table 5. I do not observe any significant change in compensation.

Table 9 shows the change in the dependent coverage rates.[39] The sample I use in this table is the same as the one that I use in the main job-lock analysis. The results from this table demonstrate after the mandates, there was a 32 percent increase in dependent health insurance coverage among working fathers than they would otherwise have been. In some cases, a father may experience job-lock as a result of the mandates even when he is not currently adopting EPHI for his young adult child. This might occur, for example, when a father with EPHI anticipates that his young adult child may lose his/her own insurance. The

[35]This sample is a subset of the job-lock sample, but it excludes those individuals who voluntarily separated from their jobs. I also run the same analysis on the full job-lock sample, but it does not change my results.

[36]To examine whether including fathers with zero compensation would change my results, I also add `one` to both dependent variables and convert them to natural logs (i.e., $ln(AnnualEarnings + 1)$ and $ln(TotalMonetaryCompensation + 1)$). I arbitrarily assign this *one* for the compensation outcome variable to those working fathers with zero compensation. With these adjusted dependent variables, I run the linear regression again with the same specifications as equation (1) and Tobit regressions censored at one (see columns 1-4 of Appendix Table 4).

[37]The results, however, should be interpreted with caution. As mentioned, non-discrimination laws may prevent employers from differentially paying total compensation and earnings. In addition, all workers may bear the cost of the mandate since many non-parents are potential future users of the policy and it may be difficult for firms to implement wage offsets when workers become parents. If the additional costs can be shared by more employees, then my estimates in Table 8 serve as the upper limit in absolute value.

[38]Although the earnings reported in the SIPP are monthly, I aggregate them into annual earnings for each father and use this as an outcome measure in Appendix Table 4.

[39]The outcome variable for this analysis is based on two questions: (1) whether respondents cover children 20 years or older who are living outside their household and (2) whether they cover children between the ages of 19 and 29 living in their household. If the answer to one of these two questions was 'yes,' the outcome variable is equal to one; otherwise, it is zero.

father might choose to forgo other opportunities and stay in his current job, as a safety net for his child.

Conclusions

While both the state and federal dependent coverage provisions successfully increase health insurance rates among young adults, previous research does not make it clear whether the mandates have possible effects on other sub-populations. In order to fill this gap, I explore whether fathers' dependence on employment for health insurance increased after the implementation of the mandates, as they also heavily value the dependent coverage alongside young adults. I find that eligible working fathers aged 45-64 with EPHI experienced a 37 percent decrease in the rate of voluntary job separation due to the mandates. These results suggest that the dependent coverage mandates may in fact be detrimental to some populations: while they successfully reduce uninsured rates for young adults, they increase the dependence on employment for the middle-aged, thereby weakening the justification for the law itself. This finding was robust among a variety of different specifications for the effect of these mandates.

The general pattern of job mobility for married fathers is consistent with previous findings on the effects of health insurance—specifically that targets children—on parents' job mobility (Bansak and Rapahel, 2008; Chatterji et al. 2016; Hamersma and Kim, 2009). I focus on a more broadly defined population than the other papers: fathers with and without EPHI who have children aged 19-26. This paper contributes to the little that is known about job-lock and change in compensation for eligible fathers whose responsiveness is the key determinant for the policy to be effective. On that note, this paper demonstrates that fathers are willing to adjust their labor market decisions to secure access to high quality health insurance for their children, even when these children are over 19-years-old. My results emphasize how health insurance access can have far-reaching consequences for both targeted individuals and their household members.

Although most related studies only examine the state and federal mandates independently, this paper makes use of both of these mandates as part of its identification strategy. This approach is consistent with that of Barkowski and McLaughlin (2018), which argues that to achieve more credible variation, researchers who study overlapping policies—especially those on health insurance coverage—need to account for both policies. Also, my research demonstrates that both the state and federal dependent coverage mandates may provide a potential source of exogenous variation for researchers seeking to study job mobility decisions and related outcomes among middle-aged men. Besides job-mobility and health insurance enrollment decisions, I investigate only two other possible outcomes: annual earnings and total monetary compensation; however, I believe that additional outcomes of these dependent coverage laws are open for exploration with future data.

Tables

Table 1. Examples of Childrens' Age Eligibility by State, Before and After Implementation of Mandates

States	Pre-State Law Period		State Law Period			ACA Period (from 2010)	
	Eligible	Ineligible	Beginning Year	Eligible	Ineligible	Eligible	Ineligible
Indiana	.	19–29	2008	19–23	24–29	19–26	27–29
Colorado	.	19–29	2006	19–24	25–29	19–26	27–29
Connecticut	.	19–29	2009	19–25	26–29	19–26	27–29
Michigan	.	19–29		.	19–29	19–26	27–29

Table 1: Rows indicate representative states and demonstrate the change in eligibility age of children by state of residence and time period. I choose Indiana, Colorado and Connecticut as examples of states that had implemented state-level mandates prior to the ACA. Unlike the other representative states, Michigan did not employ any state-level mandates prior to the ACA; thus, its eligibility was only affected by the federal mandate. The change in eligibility for Michigan applies to all other states without state-level mandates.

Table 2. Descriptive Statistics of Fathers

	Panel A: Job-Lock Sample		Panel B: Job-Push Sample	
	Always Ineligible	Ever Eligible	Always Ineligible	Ever Eligible
Eligible	-	.41	-	.45
		[.49]		[.50]
Age	56.30	54.09	57.29	54.03
	[4.60]	[4.60]	[4.38]	[4.75]
Highschool dropouts	.05	.04	.02	.04
	[.22]	[.18]	[.15]	[.19]
Highschool graduates	.27	.26	.16	.25
	[.44]	[.44]	[.37]	[.44]
Some college or higher	.69	.71	.81	.71
	[.46]	[.46]	[.39]	[.45]
Non-hispanic white	.81	.82	.83	.79
	[.39]	[.38]	[.38]	[.41]
African American	.07	.07	.06	.07
	[.26]	[.26]	[.25]	[.26]
Hispanic or Asian	.11	.11	.10	.14
	[.31]	[.31]	[.31]	[.35]
Public Sector worker	.21	.19	.05	.08
	[.40]	[.40]	[.23]	[.27]
Dependent Variables				
Voluntary Job Separation rates	.02	.02	.02	.02
	[.13]	[.13]	[.14]	[.15]
N. of Individuals [1,000]	.55	2.00	.10	.45
N. of Observation [1,000]	3.70	11.00	.50	1.90
Ln(Annual Earnings in the SIPP)	10.97	10.85	10.49	10.53
	[.72]	[.75]	[1.03]	[.86]
Ln(Annual Earnings in the DER)	10.84	10.94	10.00	10.32
	[.97]	[.95]	[1.46]	[1.20]
Ln(Tot. Monetary Comp.)	10.90	10.99	10.04	1.34
	[.98]	[.96]	[1.48]	[1.22]
Coverage for Dependent	.02	.07	-	-
	[.14]	[.25]	.	.
N. of Individuals [1,000]	.50	1.90	.10	.40
N. of Observation [1,000]	3.50	10.50	.40	1.70

Table 2: All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands.

Table 3. The Effects of Eligibility on Voluntary Job Separation Rates, Main Results

	[1] Weighted	[2] Unweighted	[3] Weighted	[4] Unweighted
Eligible	-.007*	-.006*	-.007*	-.006*
	[.003]	[.003]	[.004]	[.003]
Covariates	Y	Y	Y	Y
State Differential Time Trends			Y	Y
N. of Individuals [1,000]	2.50	2.50	2.50	2.50
N. of Observations [1,000]	14.5	14.5	14.5	14.5
Dependent variable means				
Ever eligible, before Mandate	.020	.017	.020	.017

Table 3: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. Standard errors are clustered at the state level. Observations are weighted using SIPP individual weight.

Table 4. The Effects of Eligibility on Voluntary Job Separation Rates, Robustness Checks

	[1] Table 3 [3]	[2]	[3]	[4]	[5]
Eligible	-.007*	-.007*	-.006*	-.008*	-.006
	[.004]	[.003]	[.003]	[.004]	[.004]
Fathers with Youngest Child Aged 19-33		Y			
Including 2001 SIPP			Y		
Excluding States with Unclear Implementation Dates				Y	
Treating States with Student-Status as Non-mandated					Y
N. of Individuals [1,000]	2.5	2.4	3.6	2.4	2.5
N. of Observations [1,000]	14.5	18.0	21.0	13.5	14.5

Table 4: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weights and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Table 5. The Effects of Eligibility on Voluntary
Job Separation Rates by Subgroups

	[1] Higher Educ.	[2] Lower Educ.
Eligible	-.009†	-.002
	[.005]	[.005]
N. of Individuals [1,000]	1.7	0.8
N. of Observations [1,000]	10.0	4.4

Table 5: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weight and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Table 6. The Effects of Eligibility on Voluntary Job Separation Rates, Falsification Tests

	[1] Eligible Age-11	[2] Eligible Age+11	[3] Eligible Age+11 (only for 19-26)
Eligible	.003	-.002	.006
	[.005]	[.007]	[.008]
N. of Individuals [1,000]	1.6	1.0	1.1
N. of Observations [1,000]	12.5	7.7	9.4

Table 6: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weight and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Table 7. The Effects of Eligibility on Voluntary Job Separation, Job Push

	[1]	[2]
Eligible	-.006	.009
	[.005]	[.008]
Covariates	Y	Y
State Differential Time Trends		Y
N. of Individuals [1,000]	0.55	0.55
N. of Observations [1,000]	2.4	2.4

Table 7: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weight. Standard errors are clustered at the state level.

Table 8. The Effects of Eligibility on Annual Earnings and Total Monetary Compensation

	[1]	[2]	[3]
	SIPP-DER-BR		Public SIPP
	ln(Earnings)	ln(Tot. Comp.)	ln(Earnings)
Eligible	-.104†	-.117†	.001
	[.062]	[.062]	[.041]
N. of Individuals [1,000]	2.4	2.4	2.4
N. of Observations [1,000]	13.5	13.5	18.5

Table 8: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.

In this table, I exclude fathers who voluntarily separated from their jobs among job-lock sample. All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weights and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Table 9. The Effect of Eligibility on Working Fathers' Health Insurance Coverage Take-up Decisions For Young Adult Dependents

	[1] Weighted	[2] Unweighted	[3] Weighted	[4] Unweighted
Eligible	.021*	.023*	.023*	.025*
	[.011]	[.012]	[.011]	[.012]
Covariates	Y	Y	Y	Y
State Differential Time Trends			Y	Y
N. of Individuals [1,000]	2.5	2.5	2.5	2.5
N. of Observations [1,000]	14.5	14.5	14.5	14.5
Dependent variable means				
Ever eligible, before Mandate	.069	.079	.069	.079

Table 9: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All columns show results of the average marginal effect based on probit regressions using the 2004 and 2008 SIPP panels. Standard errors are clustered at the state level. Observations are weighted using SIPP individual weight.

Appendices

Table A1. Implementation of the Dependent Coverage Laws

	Full Year Implemented	Max age
Federal Mandate	2010	26
States		
Colorado	2006	24
Connecticut	2009	25
Delaware	2008	23
Florida*	2008	24
Idaho*	2008	24
Illinois	2010	25
Indiana	2008	23
Kentucky	2008	25
Louisiana*	2009	23
Maine	2007	24
Maryland	2008	24
Massachusetts*	2007	25
Minnesota	2008	24
Missouri	2008	24
Montana	2008	24
New Hampshire	2007	25
New Mexico	2003	24
North Dakota*	1995	25
Rhode Island*	2007	24
South Dakota*	2005	23
Utah	1995	25
Virginia	2007	24
Washington	2009	24
West Virginia	2007	24

Appendix Table 1: This table shows the year that federal and state-level mandates were implemented along with age criteria. States that are excluded from the main analyses are ones that had no age limit defined in their dependent coverage expansions (Iowa and Texas) or that extended their dependent coverage up to the age of 29 (New Jersey, Pennsylvania, and New York). *indicates states that required student status as part of the eligibility criteria.

Table A2. Alternative Regression Results

	[1]	[2]	[3]	[4]
	Linear		Logit	
	Weighted	Unweighted	Weighted	Unweighted
Eligible	-.007	-.004	-.008*	-.006*
	[.004]	[.003]	[.004]	[.003]
N. of Individuals [1,000]	2.5	2.5	2.5	2.5
N. of Observations [1,000]	14.5	14.5	14.5	14.5
Dependent variable means				
Ever Eligible, before Mandate	.020	.019	.020	.019

Appendix Table 2: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All columns include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level. Observations are weighted using SIPP individual weight.

Table A3. Descriptive Statistics for Table 8

	Always Ineligible	Ever Eligible
Eligible	0	.41
		[.49]
Age	56.25	54.07
	[4.57]	[4.59]
Highschool dropouts	.05	.04
	[.22]	[.18]
Highschool graduates	.27	.26
	[.44]	[.44]
Some college or higher	.69	.71
	[.46]	[.46]
Non-hispanic white	.82	.82
	[.39]	[.38]
African American	.07	.07
	[.27]	[.26]
Hispanic or Asian	.11	.11
	[.31]	[.31]
Public Sector worker	.20	.19
	[.40]	[.39]
Dependent Variables		
Voluntary Job Separation rates	0	0
N. of Individuals [1,000]	.55	2.00
N. of Observation [1,000]	3.60	10.50
Ln(Annual Earnings in the SIPP)	10.85	10.97
	[.75]	[.72]
Ln(Annual Earnings in the DER)	10.85	10.94
	[.96]	[.95]
Ln(Tot. Monetary Comp.)	10.90	11.00
	[.98]	[.96]
N. of Individuals [1,000]	.50	1.90
N. of Observation [1,000]	3.60	10.50

Appendix Table 3: All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands.

Table A4. The Effects of Eligibility on Annual Earnings

	[1]	[2]	[3]	[4]	[5]	[6]
	SIPP-DER-BR				Public SIPP	
	Linear		Tobit		Linear	Tobit
	ln(Earnings+1)	ln(Tot. Comp.+1)	ln(Earnings+1)	ln(Tot. Comp.+1)	ln(Earnings+1)	ln(Earnings+1)
Eligible	-.208†	-.222†	-.213†	-.227†	-.022	-.022
	[.123]	[.126]	[.126]	[.129]	[.047]	[.047]
N. of Individuals [1,000]	2.5	2.5	2.5	2.5	2.4	2.4
N. of Observations [1,000]	14.0	14.0	14.0	14.0	19.0	19.0

Appendix Table 4: † indicates that the p-value is less than 0.1; * indicates that the In this table, I exclude fathers who voluntarily separated from their jobs among job-lock sample. All numbers of observations and individuals are first rounded according to the United States Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weights and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Table A5. The Effects of Eligibility on Annual Earnings and Total Monetary Compensation
(Job-Push Sample, Falsification Tests)

	[1]	[2]	[3]	[4]	[5]	[6]
	Linear		Linear		Tobit	
	ln(Earnings)	ln(Tot. Comp.)	ln(Earnings)+1	ln(Tot. Comp.+1)	ln(Earnings+1)	ln(Tot. Comp.+1)
Eligible	-.109	-.123	-.438	-.451	-.501	-.515
	[.243]	[.248]	[.562]	[.565]	[.630]	[.633]
N. of Individuals [1,000]	0.45	0.45	0.55	0.55	0.55	0.55
N. of Observations [1,000]	2.1	2.1	2.4	2.4	2.4	2.4

Appendix Table 5: † indicates that the p-value is less than 0.1; * indicates that the p-value is less than 0.05; ** indicates that the p-value is less than 0.01.
All numbers of observations and individuals are first rounded according to the U.S. Census disclosure rules and then are rounded to the nearest thousands. All columns are weighted by SIPP individual weights and include covariates for individual demographics and state differential trends. Standard errors are clustered at the state level.

Figures

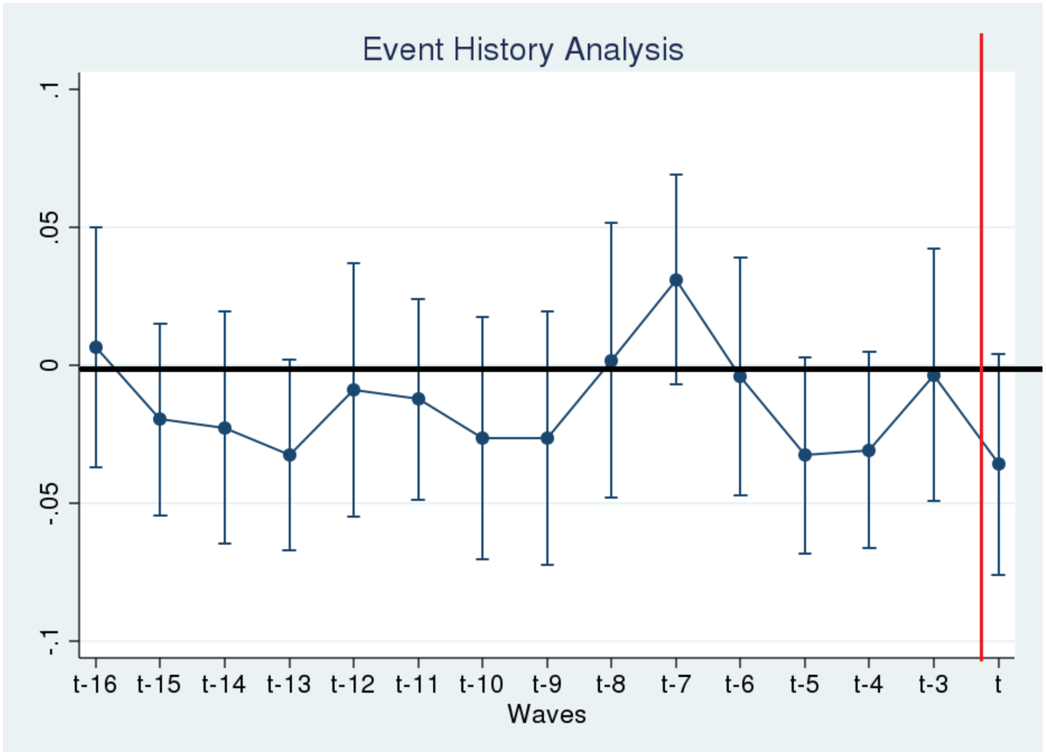

Figure 1: $t-1$ is the baseline time period for this figure. Result for $t-2$ is not reported in this graph due to the Census disclosure rule. Blue lines around the point estimates represent 95 percent confidence intervals.

References

[1] Anderson, Patricia M. (1997), *The Effect of Employer-Provided Health Insurance on Job Mobility: Job-Lock or Job-Push?,* Working Paper, Dartmouth College May 1997. 2, 11

[2] Akosa Antwi, Yaa, Asako S. Moriya, and Kosali Simon. (2013) *Effects of Federal Policy to Insure Young Adults: Evidence from the 2010 Affordable Care Act's DependentCoverage Mandate.*, American Economic Journal: Economic Policy 5 (4)

[3] Bansak, Cynthia and Steven Raphael (2008), *The State Children's Health Insurance Program and Job Mobility: Identifying Job Lock among Working Parents in Near-Poor Households*, ILR review

[4] Barkowski, Scott (2017) *Does Government Health Insurance Reduce Job Lock and Job Push?* , Working paper

[5] Barkowski, Scott and Joanne Song McLaughlin (2018) *In Sickness and in Health: The Influence of State and Federal Health Insurance Coverage Mandates on Marriage of Young Adults in the USA*, MPRA Paper No. 84014

[6] Biehl, Amelia, Tami Gurley-Calvez and Brian Hill (2018),*Child health insurance and the labor market participation of older Americans: Evidence from the young adult mandate*, Working paper

[7] Blank and Ruggles (1996), *When Do Women Use Aid to Families with Dependent Children and Food Stamps? The Dynamics of Eligibility Versus Participation*, Journal of Human Resources, vol. 31, issue 1, 57-89

[8] Blau, David, and Donna B. Gilleskie (2001), *The effect of health on employment transitions of older men*, Solomon Polachek (ed.) Worker Wellbeing in a Changing Labor Market (Research in Labor Economics, Volume 20) Emerald Group Publishing Limited, pp.35 - 65

[9] Blundell, Richard and Thomas Macurdy (1999), *LABOR SUPPLY: A REVIEW OF ALTERNATIVE APPROACHES*, Handbook of Labor Economics, Volume 3, Edited by O. AshenJelter and D. Card

[10] Brandeisky, Kara (2015),*Why young millennials are turning down health coverage at work* http://time.com/money/3821525/health-insurance-age-26/

[11] Burgdorf, James R. (2014) *Young Adult Dependent Coverage: Were the State Reforms Effective?*, Health Services Research

[12] Chatterji, Pinka, Peter Brandon and Sara Markowitz (2016),*Job mobility among parents of children with chronic health conditions: Early effects of the 2010 Affordable Care Act*, Journal of Health Economics, 2016, vol. 48, issue C, 26-43

[13] Cantor, J. C., Monheit, A. C., DeLia, D., & Lloyd, K. (2012). *Early Impact of the Affordable Care Act on Health Insurance Coverage of Young Adults*, Health Services Research

[14] Chen, Weiwei (2018), *Young Adults' Selection and Use of Dependent Coverage under the Affordable Care Act*,Front Public Health

[15] Chenevert, Rebecca, Mark Klee and Kelly Wilkin (2016), *Do imputed earnings earn their keep? evaluating sipp earnings and nonresponse with administrative records*, SEHSD Working Paper Number 2016-18

[16] Colman, Gregory and Dhaval Dave (2017) *It's about time: Effects of the Affordable Care Act Dependent Coverage Mandate on Time Use*, Contemporary Economic Policy

[17] Cutler, David and Jonathan Gruber (1996) *Does Public Insurance Crowd out Private Insurance?*, The Quarterly Journal of Economics, Volume 111

[18] Depew, Briggs. (2015) *The Effect of State Dependent Mandate Laws on the Labor Supply Decisions of Young Adults.*, Journal of Health Economics

[19] Dillender, M. (2014), *Do more health insurance options lead to higher wages? Evidence from states extending dependent coverage*, Journal of Health Economics

[20] Federal Register (2015),*Final Rules for Grandfathered Plans, Preexisting Condition Exclusions, Lifetime and Annual Limits, Rescissions, Dependent Coverage, Appeals, and Patient Protections Under the Affordable Care Act*, Authenticated United States Government Information, Vol. 80, No. 222

[21] Furman, Jason and Matt Fiedler (2015),*4.5 Million young adults have gained coverage since 2010, improving access to care and benefitting our economy*, the White House https://obamawhitehouse.archives.gov/blog/2015/01/29/45-million-young-adults-have-gained-coverage-2010-improving-access-care-and-benefitt

[22] Goldman, T.R (2013),*Progress Report: The Affordable Care Act,Äôs Extended Dependent Coverage Provision*, Health Affairs

[23] Gottschalk, Peter, and Minh Huynh (2005), *Validation study of earnings data in the SIPP – Do older workers have larger measurement error?*, Working Paper No. 2005-07. Chestnut Hill, MA: Center for Retirement Research at Boston College

[24] Gruber, Jonathan (1994), *The Incidence of Mandated Maternity Benefits*, The American Economic Review, Vol. 84, No. 3, pp. 622-641

[25] Gruber, Jonathan and Brigitte C. Madrian (2002), *Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature*, NBER Working Paper No. 8817

[26] Hamersma and Kim (2009), *The effect of parental Medicaid expansions on job mobility*, Journal of Health Economics Volume 28, Issue 4, Pages 761-770

[27] Hahn, Youjin, Hee-Seung Yang (2015), *Do Work Decisions among Young Adults Respond to Extended Dependent Coverage?*, ILR Review

[28] The Kaiser Family Foundation and Health Research and Education Trust (2017), *Employer Health Benefits - 2017 Summary of Findings*, http://files.kff.org/attachment/Summary-of-Findings-Employer-Health-Benefits-2017

[29] Levine, Phillip B, Robin McKnight, and Samantha Heep. (2011) *How Effective Are Public Policies to Increase Health Insurance Coverage Among Young Adults?*, American Economic Journal: Economic Policy

[30] Meyer, Bruce D., 1990 ,*Unemployment Insurance and Unemployment Spells*, Econometrica, Vol. 58, No. 4

[31] Monheit, Alan, and Jasmine Rizzo. (2007),*Mandated Health Insurance Benefits: a Critical Review of the Literature.* State of New Jersey: New Jersey Department of Human Services in collaboration with Rutgers Center for State Health Policy

[32] Monheit, Alan C., Joel C. Cantor, Derek DeLia, and Dina Belloff. (2011),*How Have State Policies to Expand Dependent Coverage Affected the Health Insurance Status of Young Adults?*, Health Services Research 46 (1p2):251-67

[33] National Conference of State Legislatures, (2010), *Covering young adults through their parents' or guardians' health policy*, http://www.ncsl.org/research/health/dependent-health-coverage-state-implementation.aspx

[34] Rashad, Inas; Sarpong, Eric, (2006) *Employer-Provided Health Insurance and the Incidence of"Job-Lock": Is There a Consensus?*, Georgia Health Policy Center Working Paper

[35] Rashad, Inas; Sarpong, Eric, (2008) *Employer-provided health insurance and the incidence of job lock: a literature reivew and empirical test*, Expert Review of Pharmacoeconomics & Outcomes Research

[36] Reinicke, Carmen, (2018), *It might be time to take adult kids off the family health plan*, https://www.cnbc.com/2018/07/10/it-might-be-time-to-take-your-adult-kids-off-the-family-health-plan.html

[37] Roemer, Marc, (2002) *Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation*, Longitudinal Employer-Household Dynamics Program Demographic Surveys Division U.S. Bureau of the Census

[38] Goda, Gopi Shah, Monica Farid and Jay Bhattacharya (2016), *The Incidence of Mandated Health Insurance: Evidence from the Affordable Care Act Dependent Care Mandate*, NBER Working Paper No. 21846

[39] Sommers, Benjamin D., Thomas Buchmueller, Sandra L. Decker, Colleen Carey, and Richard Kronick. (2013) *The Affordable Care Act Has Led To Significant Gains In Health Insurance And Access To Care For Young Adults.* Health Affairs