Bulk Buying and Inequality

E. Mallick Hossain

August 20, 2019

Updated Version Here[*]

Abstract

I document that high-income households buy in bulk more often than low-income households, especially for storable, non-food items. Because they miss out on bulk discounts, low-income households pay higher unit prices. These differences in bulk buying may bias standard measures of inequality and price indices. At least three factors drive bulk buying: store access, unit price salience, and storage costs. Using data on warehouse club entry, I show that warehouse club entry increases bulk purchasing, but only for middle- and high-income households. Using data on state-level unit pricing regulations, I find that these regulations increase bulk purchasing, but only for middle- and high-income households. Using a demand model that incorporates unit price salience and storage costs, I find that storage costs account for a moderate portion of this difference with the majority accounted for by differences in preferences for product quantity. Removing storage costs would close 23% of the gap compared to changing preferences for product quantity, which would close 63% of the gap. These findings show that storage constraints are not the primary reason that low-income households do not buy in bulk. Addressing factors that affect quantity demanded, such as budget constraints, would be more effective at encouraging low-income households to buy in bulk.

[*]I thank my advisers, Aviv Nevo, Holger Sieg, Katja Seim, and Sarah Moshary for their sound advice and constructive criticism. Thanks to Emek Basker, Chris Cronin, Frank DiTraglia, Eileen Divringi, Jim Ferry, Ishan Ghosh, Joao Granja, Jeff Lin, Paolo Martellini, Davin Reed, Paul Sangrey, Anna Tranfaglia, and Keith Wardrip for their comments. Thanks to Dmitri Koustas for providing data on warehouse clubs. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

Bulk discounts are the most common discount available in supermarkets, covering 92% of products. I find that high-income households are more likely to buy in bulk compared to low-income households for which the marginal dollar of savings is more valuable. At least three factors affect a household's decision to buy in bulk: *store access*, *unit price salience*, and *storage constraints*. I document key facts about the bulk discounting landscape and consumer purchasing behavior using Nielsen's Retail Scanner and Consumer Panel data. Then, I provide reduced-form evidence that *store access* and *unit price salience* increase a household's bulk buying. Finally, I incorporate *unit price salience* and *storage constraints* into a formal demand model to simulate how consumers would respond to changes in unit price regulations and product assortments.

These bulk buying differences have implications for policymakers and households. First, bulk discounts are likely to understate consumption inequality, as measured by expenditures, if high-income households obtain the same products more cheaply than low-income households do. Second, if prices of large and small packages move differently, bulk discounts will affect price indices and inflation measures (Fox and Melser 2014; Triplett 2004). Not only may households experience different inflation rates, any inflation-indexed benefits may not properly account for actual price changes experienced by households. This paper establishes that differences in bulk buying exist across households and assesses the contributing factors to these differences. Decision-makers can use these findings to craft policies that encourage households to take advantage of bulk discounts.

First, I document basic facts on the importance of bulk discounts. In addition to being near-universal, bulk discounts are economically significant. At the median, a 10% increase in a product's package size is associated with a 4.4% decrease in the unit price with substantial variation across product categories. Households differ in their bulk buying with households making over $100,000 being nine percentage points more likely to buy non-food items in bulk relative to observationally similar households making less than $25,000 (i.e. same household size, age, marital status, etc.).[1] I find that most households miss out on meaningful savings, but low-income households miss out on more savings. For example, low-income households spend 36% more on toilet paper than they would if they were to purchase the lowest unit priced item available (of their preferred brand).[2]

[1] I define "bulk" sizes as packages in the top two quintiles of the size distribution for that product category. Non-food items are basic cleaning, household, and hygiene products like toilet paper, diapers, paper towels, etc.

[2] Throughout this paper, I use "high-income" to refer to households making over $100,000 and "low-income" to refer to households making under $25,000. "Middle-income" refers to households making between $25,000

At least three factors contribute to this “bulk-buying gap,” each with different policy implications: *Store access*, *unit price salience*, and *storage costs*.

First, *store access* could be a factor because stores typically do not stock the full range of brands and sizes. Most notably, warehouse clubs (like Costco) only stock “bulk” sizes while dollar stores (like Dollar General) tend to stock smaller, “budget” sizes. Recent research has explored how access and availability (e.g. food deserts) affect a household’s nutrition choices (Allcott et al. 2019). I establish that there are minor supply differences for product sizes between high- and low-income areas, but there is substantial size variation between different store types. Additionally, I find that households shop at different types of stores with high-income households more likely to shop at warehouse clubs while low-income households are more likely to shop at discount stores (like Walmart) and dollar stores. I estimate a differences-in-differences model using warehouse club entry as a shock to local product assortments to estimate the effect of club entry on bulk buying by comparing households that experienced an entry with households that never had a warehouse club nearby and households that always had a warehouse club. I find that warehouse club entry increases bulk buying by two percentage points, but these effects are primarily among middle- and high-income households. Furthermore, most low-income households still do not shop at clubs even when they are nearby, which may be due to the annual membership fees required to shop at the clubs.

Encouraging warehouse club entry would increase bulk buying, but policymakers should be aware that this effect is limited to middle- and high-income households. While low-income households may have reduced access to bulk sizes, they still do not take advantage of bulk discounts available within the stores they currently shop at. I show that even within retail chains or locations, high-income households are more likely to buy in bulk than low-income households.

Second, *unit price salience* could affect whether or not households take advantage of bulk discounts. Research has shown that consumers respond to sales taxes when they are more salient (Chetty, Looney, and Kroft 2009). Similarly, consumers may not find unit prices salient unless they are presented in a clear and easily understood format. Using variation in state-level regulations on unit pricing, I show that households are 0.4 percentage points more likely to buy in bulk in states with unit price regulations. However, even within states, this effect increases with income. By leveraging household moves between states with different regulatory regimes, I show that moving from a state without regulations to a state with unit price regulations increases bulk buying by about one percentage point while moving from

and \$100,000.

a state with regulation to a state without regulation has no significant effect. This pattern is consistent with salience because households that move from states with regulations may still purchase based on unit prices. These effects are medium-term effects so there may be a longer-term decay that I cannot capture. If high-income households compare unit prices while low-income households do not, then the bulk-buying gap could be reduced by increasing the salience of unit prices through education and reducing consumers' cognitive load by adopting unit pricing regulations that make unit prices visible and easily comparable across products.

The fourth factor is *storage* costs. "Storage" in this setting refers both to transportation and physical storage of the item, both of which are directly related to an item's physical size. If access to cars or other transportation facilitates bulk purchasing then policies that reduce transportation costs such as transit initiatives could encourage bulk buying. If storage is the issue, then purchasing cooperatives could buy in bulk and pass on savings to members or policymakers could encourage firms to increase the concentration of their products to reduce their storage footprint.

These reduced-form tests cannot precisely capture how low-income consumers would respond to different supply conditions or regulatory regimes. For example, moving between states affects households more than just changing the salience of unit pricing. Warehouse clubs not only expand the set of products available, but they impose membership fees. To precisely answer whether supply assortments or regulatory regimes affect household bulk buying, I estimate a structural demand model and simulate counterfactuals with different size assortments and unit price regulations.

This paper's main contribution to the literature is to provide a wide range of insights into why high- and low-income households differ in their bulk buying in the United States, even for basic household necessities for which the savings are most valuable. This paper contributes to research into "poverty penalties" on how poor households may pay more for food (Talukdar 2008; Chung and Myers Jr 1999) and how households differ in their propensity to buy in bulk (Kunreuther 1973; Attanasio and Frayne 2006; Griffith et al. 2009; Beatty 2010; Nevo and Wong 2018; Orhun and Palazzolo 2019). This paper expands on previous research by also analyzing bulk buying of non-food items, which do not suffer from issues of between-category substitution (Allcott et al. 2019) (e.g. laundry detergent and toilet paper do not substitute for each other while juice, soda, and water do) or with home production (Aguiar and Hurst 2005, 2007). As a result, this paper provides a fundamental analysis of differences in bulk purchasing between households, even within stores.

This paper also contributes to the literature on consumer purchasing behavior. Both marketing and industrial organization have extensively studied how consumers take advantage of a variety of discounts.[3] Griffith et al. (2009) and Nevo and Wong (2018) provide an overview of how often consumers take advantage of sales, coupons, bulk discounts, and generic brands and examine how these behaviors changed during the Great Recession. While while numerous studies have analyzed sales and coupons, bulk buying has been relatively understudied despite its many desirable properties. For example, using Nielsen's Retail Scanner data, I show that bulk discounts apply across a wide range of products and are common within stores. Compared to other discounts like coupons and sales, consumers do not have to search for bulk discounts, they are always available. Furthermore, quality differences between products are virtually non-existent because different sized packages contain identical goods. I contribute to this literature and identify the factors that affect a household's ability to use bulk discounts, arguably the most universal and largest discount available in retail.

Finally, this paper also contributes to the literature on second-degree price discrimination (see Stole (2007) for a review). A majority of this literature has focused on firms' ability to use non-linear pricing to price discriminate (Pesendorfer 2002; Nevo and Wolfram 2002; Hendel and Nevo 2006a, 2006b, 2013; McManus 2007; Cohen 2008; Hendel, Lizzeri, and Roketskiy 2014). Nonlinear pricing is harder to sustain when products are storable, but can be rationalized by heterogeneity in storage costs (Hendel, Lizzeri, and Roketskiy 2014). I provide a novel method of identifying storage costs by leveraging differences in product concentration.

The rest of the paper is structured as follows. Section 2 describes the Nielsen data used for analysis. Section 3 documents stylized facts of bulk discounting. Section 4 presents reduced-form, causal analysis of contributing factors to the bulk-buying gap. Section 5 introduces the model. Section 6 presents estimation results. Section 7 shows the counterfactual exercises and Section 8 concludes.

2 Data

In this section, I describe the two Nielsen datasets used for my analysis and give a brief overview of their respective features.[4] The Consumer Panel data provides information on

[3]For analyses of consumer responses to sales, see Pesendorfer (2002); Hong, McAfee, and Nayyar (2002); Erdem, Imai, and Keane (2003); Hendel and Nevo (2006a); Hendel and Nevo (2006b); Hendel and Nevo (2013). For analyses of coupons, see Narasimhan (1984); Nevo and Wolfram (2002); Anderson and Song (2004).

[4]Researcher's own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

what shopping and purchasing decisions households make. The Retail Scanner data provides information on the product options and prices available each week. Combining the two gives a comprehensive overview of each household's purchasing behavior and choice environment.

2.1 Nielsen Consumer Panel Data

I use the Nielsen Consumer Panel Dataset from 2004–2017. This dataset is a panel of about 178,000 unique households, of which I observe about 40,000 each year from 2004–2006 and about 60,000 households each year from 2007–2017. Households scan all items that they purchase, input quantities, prices, date of purchase, and store. Nielsen retains about 80% of its panel from year to year with the mean and median tenure of a household being four and three years, respectively.

I consider food, drink, and non-food grocery (e.g. paper towels, toilet paper, detergent, etc.) purchases made at grocery stores, discount stores, dollar stores, warehouse clubs, and drug stores. These outlets account for over 90% of household expenditures. I exclude alcohol, tobacco, health, and general merchandise products from my analysis since these products may have different consumption patterns than food and cleaning products (e.g. cigarettes, painkillers) or are not suited for bulk purchases (e.g. printers, cookware, linens). I also exclude households with a student or military head of household as well as those with an annual income of less than $5,000. Only about 3% of households are excluded and I use the remaining 173,000 households for my analysis. See Appendix for further details of sample construction.

Table 1 presents descriptive statistics for households in the sample. The following analysis uses Nielsen's projection weights to be nationally representative unless otherwise stated. Weights are computed to match moments based on household size, income, age, race, ethnicity, education, occupation, and presence of children.

2.2 Nielsen Scanner Data

The Nielsen Scanner Data contains weekly prices and volume sold of individual products at about 35,000 stores from about 90 retail chains between 2006–2016.[5] I match the Retail Scanner Data with the Consumer Panel data based on store identification numbers and purchase dates. By matching the two datasets, I can reconstruct a household's choice set and compare that set to what was ultimately purchased.

[5]This is the most recent data available because data revisions and corrections delayed the release of 2017 data.

Table 1: Nielsen Consumer Panel Summary Statistics

Variable	Mean	SD	25th Pctile	75th Pctile
Household income ($000s)	54.98	31.69	27.5	85
Household size	2.54	1.45	1	3
Hispanic	0.11	0.32	0	0
Age	52.65	14.39	41.5	63
College Educated	0.37	0.48	0	1
White	0.77	0.42	1	1
Child present	0.32	0.47	0	1
Married	0.50	0.50	0	1

Notes: Data includes 770,317 household-year observations from 2004–2017 across 173,142 unique households and are weighted for national representativeness.

3 Stylized Facts

In this section, I document two stylized facts about the shopping landscape. First, I show that bulk discounts apply to 92% of grocery categories and can be large, exceeding the savings offered by coupons for non-food items. Second, I show that households making over $100,000 are nine percentage points more likely to buy non-food items in bulk than households making under $25,000, compared to only one percentage point for food items.

3.1 Bulk Discount Prevalence

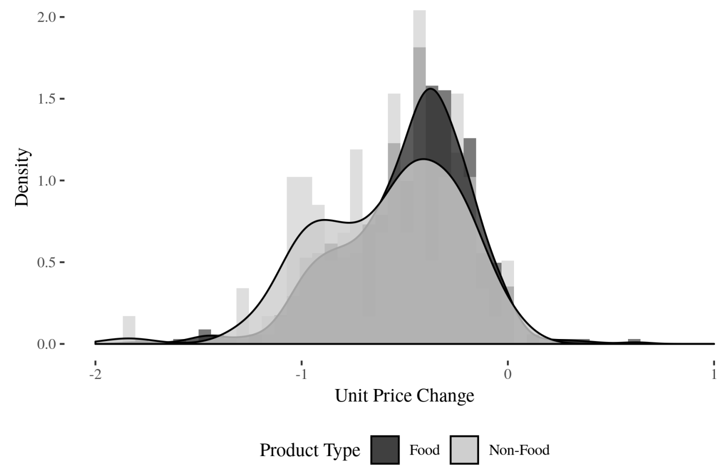
To establish the prevalence and magnitude of bulk discounts, I use Nielsen's Retail Scanner data from 2016. I estimate bulk discounts using the following regression for each of the 693 product categories.

$$\ln(P)_{ibm} = \beta \ln(Size)_{ibm} + \lambda_{bm} + \epsilon_{ibm}, \tag{1}$$

where P is the unit price (package price divided by package size) of product i from brand b purchased in market m (defined as a store-week). $Size$ is the item's package size, which is the number of units included in a UPC (e.g. quart, square feet, count, pound, etc.). To account for sales and promotions, I control for a brand-store-week fixed effect, λ. Variation in unit prices across package sizes within the same brand-store-week identify β. If retailers offer bulk discounts, then β will be negative.

Figure 1 plots the distribution of β across product categories. I find that 92% of all product modules have a statistically significant and negative β and that non-food items generally have larger discounts than food items. The median β is -0.53 for non-food products, which means that a 10 percent increase in package size corresponds to a 5.3 percent decrease

Figure 1: Distribution of Bulk Discounts by Product Type



Notes: Using Nielsen Retail Scanner data from 2016, this figure plots the distribution of coefficients from a regression of log unit price on log package size (Equation (1)) for individual product categories. "Unit Price Change" denotes the percent reduction in a product's unit price for a one percent increase in the package size. Regression controls for store-brand-week fixed effects.

in unit price. This discount is larger than the median β for food items (-0.43). The size and near-universality of bulk discounts suggest they offer substantial savings to households without sacrificing consumption.[6]

I then compare savings from bulk discounts to savings from coupons. To be conservative, I compare the savings from redeemed coupons (likely higher than the average savings of all coupons offered) to savings offered by bulk discounts (likely lower than bulk discounts actually redeemed). For each product purchased in the Consumer Panel data, households can input the value saved if they used a coupon. For each product category, I compute the average discount across all products in that category.[7]

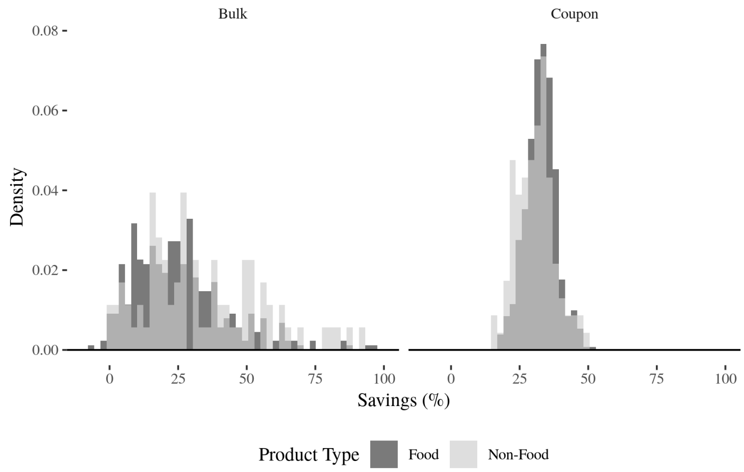
I then estimate bulk discount savings based on moving from a product in the second quintile to the fourth quintile of the size distribution (20th percentile to 60th percentile). This leaves out small product sizes that may have high unit prices due to convenience (e.g. a 20-oz soda bottle at the checkout counter) and especially large sizes that may not be widely available at all stores. This range covers sizes that households are likely to consider when making their purchase decision.

Figure 2 plots the distribution of coupon savings and estimated bulk savings for food and non-food products. Coupon savings are narrowly clustered with a median savings of 31% for

[6] Primarily for non-food items. Food items are more likely to deteriorate and some items, like milk, expire quickly, so using bulk discounts to stockpile is infeasible.

[7] I do not use projection weights since this is a product-level analysis, not a household-level analysis.

Figure 2: Percent Savings from Coupons and Bulk Discounts



Notes: The Nielsen Consumer Panel is used to compute coupon savings. For each coupon redemption, the percent savings are the ratio of the coupon value to the product's price. These savings are then averaged across all purchases in that product category. Bulk discounts are computed using coefficient estimates obtained from Equation (1) relating log unit prices to log package sizes. Bulk savings are the estimated savings obtained from moving from the second to the fourth quintile of the size distribution for each product category.

non-food products and 33% for food products. Bulk discounts have lower median savings for non-food and food products of 27% and 23%, respectively, but are more widely dispersed, even exceeding 50% savings for some non-food products.[8] Coupon savings are similar across product categories while there is substantial variation in bulk discounts with non-food products offering higher savings.

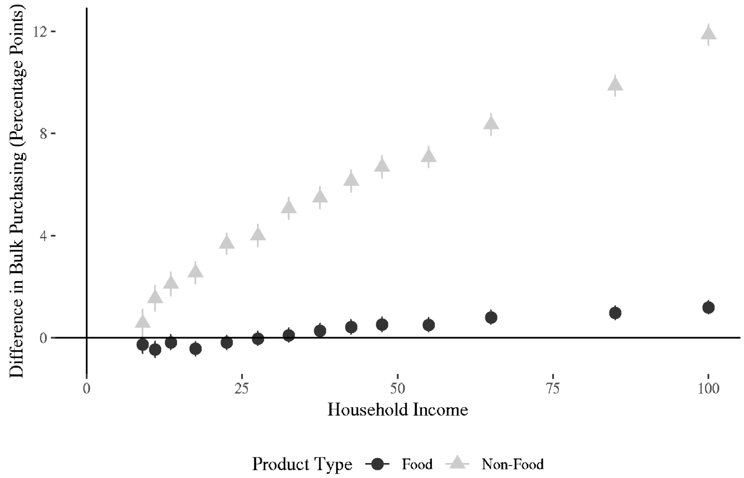
3.2 Bulk Purchasing

Next, I document the take up of bulk discounts among different types of households. I classify a product as "bulk" if it is in the top two quintiles of the size distribution for that product category. Then, for each household, I compute the expenditure-weighted share of bulk purchases of food and non-food items. Thus, for each household, I have the share of purchases that were bulk purchases. I then regress this "bulk share" on household income and other household characteristics that could affect consumption patterns and may be correlated with income, and plot the income coefficients. The equation below is estimated on food and non-food purchases separately.

$$BulkShare_{imt} = \sum_b \beta^b Income^b_{imt} + \gamma X_{imt} + \lambda_m + \lambda_t + \epsilon_{imt}, \tag{2}$$

[8]Smaller shifts, such as from the second to third quintile or third to fourth quintile generate smaller savings, but still preserve the long right tail primarily for non-food products.

Figure 3: Bulk Purchasing by Household Income and Product Type



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the income bin coefficients from Equation (2), which regresses the share of annual purchases that were bulk packages on income, household size, age, marital status, presence of children, and education, as well as market and year fixed effects. Nielsen projection weights are used to ensure national representativeness. Households making $5–8k are the reference group.

where $BulkShare$ is household i's share of bulk purchases in market m in year t.[9] $Income$ consists of dummies for each income bin b. X consists of household demographics (age, household size, marital status, presence of children, and education).[10] λ are year and market fixed effect.

The results are plotted in Figure 3. Bulk purchases compose a 12 percentage point larger share of non-food expenditures for households making over $100,000 compared to those making $5–8k. As income increases, bulk purchases comprise an increasing share of expenditures. For food items, there is a more muted increase of two percentage points across income groups.

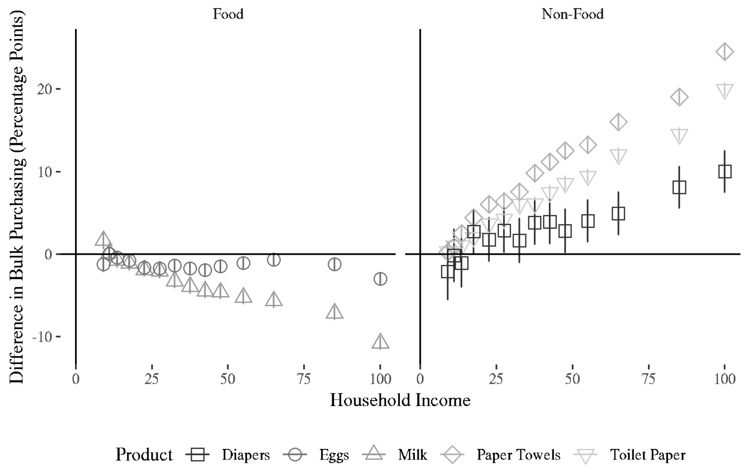
These patterns are consistent with high-income households buying at the lowest unit price and consume out of storage. Given the existence of bulk discounts, larger packages correspond to lower unit prices. The fact that low-income households are less likely to buy these storable items in bulk suggests that some obstacles may prevent them from buying and storing large packages.

Across popular spending categories, the biggest gaps are in storable categories like paper towels (20 percentage points), toilet paper (18 percentage points), and diapers (9 percentage points), while popular food categories like milk and eggs show little relationship or even a

[9] I define a market as a Designated Market Area (DMA), which are non-overlapping groups of counties originally defined by Nielsen to measure television audience share. DMA's provide a finer geographic aggregation than MSA's while capturing markets that are bigger than single counties.

[10] See Data Appendix for details of demographic variables and how they are collected.

Figure 4: Bulk Purchasing by Household Income and Product Type



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the income bin coefficients from Equation (2), which regresses the share of annual purchases that were bulk packages on income, household size, age, presence of children, and marital status as well as market and year fixed effects. This regression is estimated for milk, eggs, diapers, toilet paper, and paper towels. Nielsen projection weights are used to ensure national representativeness. Households making $5–8k are the reference group.

negative relationship (See Figure 4).[11]

3.3 Excess Spending

Given that low-income households are less likely to buy in bulk and bulk savings can be substantial, how much does this practice cost them? The easiest way to compute the extra expense would be to see what alternatives the household had when it was shopping and compare the unit price it could have paid with the unit price it actually paid. Linking the Nielsen Consumer Panel with the Nielsen Retail Scanner data provides this information.

I compute excess spending for toilet paper, diapers, milk, and eggs using the following approach. First, for each shopping trip, I compute the lowest unit price the household could have paid given its brand and store choice. Then, to get the average excess spending for a household, I expenditure-weight excess spending and aggregate to the household level. Based on this measure, Table 2 reports average excess spending by income group, computed for a family of four.

Overall, excess spending is quite substantial and it appears that most households could realize savings from buying larger packages. I estimate the differences in excess spending between

[11] The gaps for all product categories are discussed in the Appendix. This granular analysis shows that average package sizes purchased increases with income across a majority of product categories.

Table 2: Excess Spending by Household Income and Product

	Non-Perishable		Perishable	
Income	Toilet Paper	Diapers	Milk	Eggs
<$25k	0.36	0.33	0.31	0.17
$25-50k	0.35	0.33	0.30	0.17
$50-100k	0.34	0.33	0.31	0.17
>$100k	0.33	0.31	0.33	0.18

Notes: This table uses 2006–2016 Nielsen Retail Scanner and Consumer Panel data to compute average excess spending relative to the lowest unit price available given a household's brand and store choice. Average excess spending for a family of four is reported above. For example, a household making <$25k pays unit prices that are 36% higher than the lowest unit price available.

households from the following regression:

$$Y_{imt} = \beta Income_{imt} + \gamma X_{imt} + \lambda_{mt} + \epsilon_{imt}, \tag{3}$$

where Y is the excess spending of household i in market m in year t. $Income$ is the household's income bin and X consists of demographic measures (age, household size, presence of children). λ is a market-year fixed effect. Table 3 shows that low-income households miss out on two percentage points more savings than high-income households and the excess spending is primarily in non-food categories like toilet paper (36% savings) and diapers (33% savings) as opposed to food categories like milk (31% savings) and eggs (17% savings). Given the perishability of food items, these savings may not be realized if the product perishes before it can be consumed.

Overall, low-income households could benefit substantially from buying in bulk and obtaining lower unit prices. Furthermore, these savings are likely to be more important for low-income households since the marginal utility of an additional dollar of savings is likely to be higher than for high-income households. This analysis also provides evidence that all households could benefit from purchasing at the lowest unit price. Because of this, making unit prices more transparent could help households make better purchasing decisions. This hypothesis is explored in Section 4.2.

4 Reduced-Form Analysis

In this section, I estimate how important *store access*, *liquidity constraints*, and *unit-price salience* are to the bulk-buying gap. Using reduced-form tests, I explore the causal impact of warehouse club entry, income changes, and unit pricing regulation on bulk purchasing.

Table 3: Regression Results of Excess Spending Across Household Income and Products

	Diapers	Toilet Paper	Eggs	Milk
	(1)	(2)	(3)	(4)
25-50k	-0.010^{**}	-0.005^{***}	0.001	-0.002
	(0.005)	(0.001)	(0.001)	(0.001)
50-100k	-0.015^{***}	-0.013^{***}	0.004^{***}	0.002^{**}
	(0.005)	(0.001)	(0.001)	(0.001)
>100k	-0.018^{***}	-0.017^{***}	0.018^{***}	0.010^{***}
	(0.005)	(0.002)	(0.002)	(0.001)
Demographics	Y	Y	Y	Y
Market-Year FE	Y	Y	Y	Y
Observations	36,903	182,415	194,413	247,451
Adjusted R^2	0.012	0.071	0.117	0.231
Note:	*p<0.1; **p<0.05; ***p<0.01			

Notes: This table uses 2006–2016 Nielsen Retail Scanner and Consumer Panel data and reports the income coefficients of Equation (3), which regresses excess spending relative to the lowest unit price available on household demographics (household income, size, age, presence of children, and marital status) as well as a market and year fixed effect. Units are percentage points. For example, a household making over $100k have 2 percentage points lower excess spending than households making under $25k. Nielsen's projection weights are used for national representativeness.

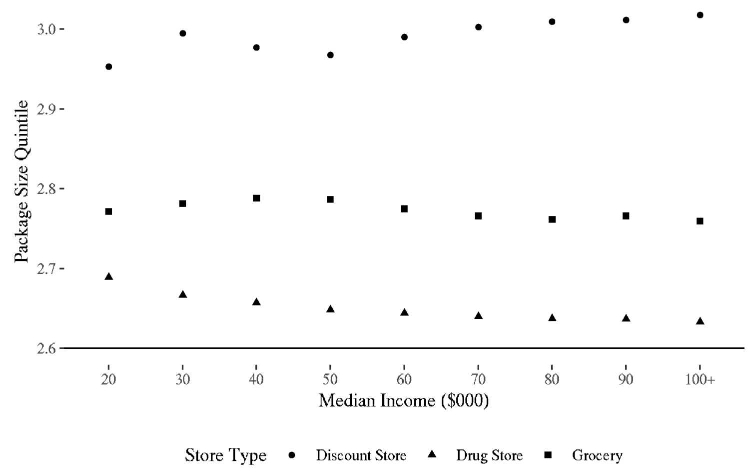
4.1 Store Access

The first factor that could drive the bulk-buying gap is *store access*, primarily access to warehouse clubs. In this subsection, I show that there are differences in package sizes offered by neighborhood income, but these differences are much smaller than the differences between different store types. I show that different income groups tend to shop at different stores, with households making over $100,000 allocating almost 20% of their shopping expenditures at warehouse clubs compared to households making under $25,000, which spend less than 5% of their expenditures at warehouse clubs. I then show that warehouse clubs are located closer to high-income households. Given that low-income households tend to be located further away from warehouse clubs, I then use an event study to examine the causal effect of warehouse club entry on bulk purchasing and how that high-income households increase their bulk purchasing by about 2 percentage points when a warehouse club opens compared to no significant effect for low-income households.

4.1.1 Package Sizes by Neighborhood Income

Supply differences could drive differences in bulk purchasing across income groups. I compute the average package quintile offered by stores based on the median income of the ZIP code

Figure 5: Average Package Quintile By Zip Code Median Income



Notes: Using 2016 Nielsen Retail Scanner data, this figure plots the average size quintile offered within a store type within categorizes of ZIP code median income. The $100,000 income bin includes all ZIP codes with incomes greater than $100,000.

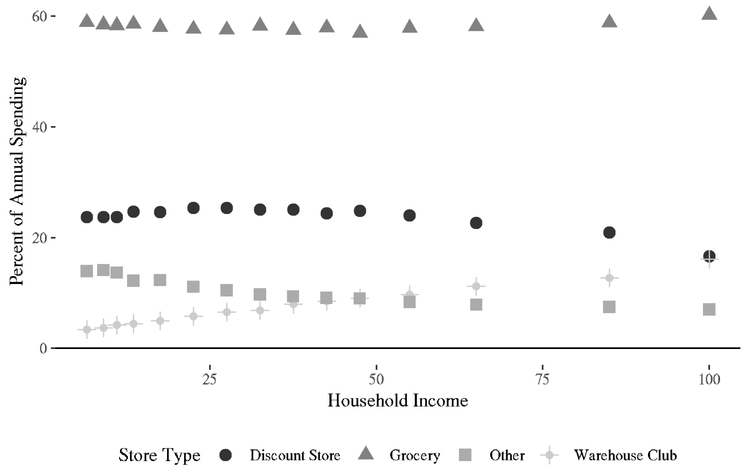
where the store is located. Since selections vary substantially based on store type (grocery stores offer different assortments than dollar stores), I compute the average quintile offered separately by store type. Figure 5 shows that within store types, there is little variation in the product sizes offered between high- and low-income areas. There are substantially larger differences between store types with discount stores like Walmart offering larger sizes than grocery or drug stores. Since these averages are only based on the stores that chose to locate in these areas, there is the possibility of selection bias. The selection of stores available within neighborhoods does vary by income with low-income areas having substantially fewer grocery and discount stores and more drug stores than high-income areas (Allcott et al. 2019). As a result, there are differences in whether or not certain store types locate in high- and low-income areas, but conditional on a store type locating in that area, there is little variation in the product sizes offered to each area.

These patterns suggest that store choice rather than a household's location may have a larger influence on bulk purchasing. This hypothesis is more fully explored in Section 4.1.3.

4.1.2 Store Preferences by Income Group

Where a household shops could heavily affect the products and sizes it ultimately purchases. For example, warehouse clubs like Costco and Sam's Club offer bulk sizes and deliver value through volume discounts. On the other hand, dollar stores like Dollar General and Family Dollar offer smaller packages and deliver value by offering lower prices. Discount stores, like Walmart and Target, offer a wider array of products and sizes, but may not have the largest

Figure 6: Annual Expenditure Shares by Store Type and Household Income



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the average share of annual spending at each store type. For each household, expenditure shares by store type are computed and then these shares are averaged across all households using Nielsen's projection weights to ensure national representativeness. "Other" includes all other store types such as convenience stores, dollar stores, etc.

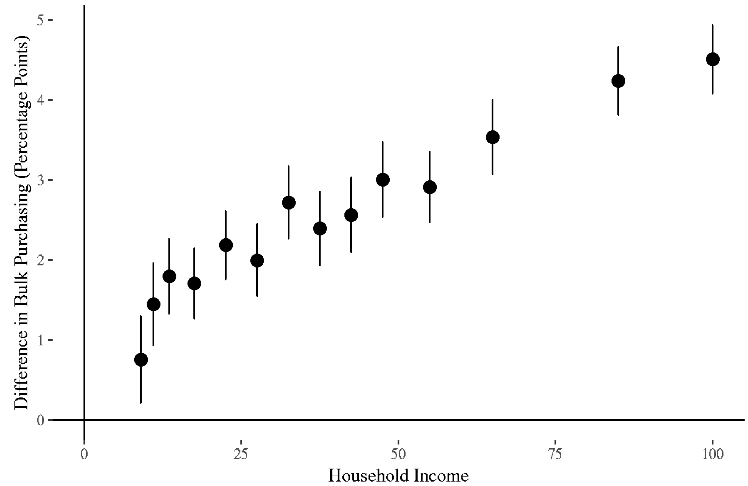
sizes that are stocked at warehouse clubs or the smallest sizes that are stocked at dollar stores. On average, warehouse clubs tend to offer larger sizes than discount stores, which in turn, offer larger sizes than dollar stores. If households sort into different types of stores, then this bulk-buying gap could be the result of differences in store access or preference instead of differences in other factors like *liquidity*, *unit-price salience*, or *storage costs*.

I first document that there are differences in where households shop and then I show that differences in package sizes still persist within store types. Figure 6 plots the average share of annual expenditures in each store type across income groups. Across all income groups, households spend about 60% of their annual expenditures at grocery stores.[12] Low-income households spend most of their remaining annual budgets at discount stores (like Walmart or Target) with the remainder at "Other" stores, primarily dollar stores, drug stores, and convenience stores. On the other hand, high-income households spend the remainder of their annual budgets at warehouse clubs and discount stores. Shopping at different store types could be due to preferences or a lack of access, whether due to distance, lack of public transit, etc.

To see how store preferences may affect the bulk-buying gap in non-food purchases, I re-estimate Equation (2) excluding households that shop at warehouse clubs, since clubs are the most likely contributor to this gap. Since warehouse clubs are more sparsely located and

[12] Annual expenditures refers to annual spending only in Nielsen-tracked categories. I do not have information on household expenditures outside of these categories.

Figure 7: Bulk Purchasing by Household Income and Product Type



Notes: Using 2004–2017 Nielsen Consumer Panel data, but excluding all warehouse club purchases, this figure plots the income bin coefficients from Equation (2), which regresses the share of annual purchases that were bulk packages on income, household size, age, presence of children, and marital status as well as market and year fixed effects. This regression is estimated for milk, eggs, diapers, toilet paper, and paper towels. Nielsen projection weights are used to ensure national representativeness. Households making $5–8k are the reference group.

charge membership fees compared to grocery and discount stores, households that shop at warehouse clubs may behave differently than households that do not. The results are plotted in Figure 7. Excluding households that shop at warehouse clubs reduces the bulk-buying gap to about three percentage points between households making over $100,000 and those making under $25,000, but the gap is still substantial given that households buy in bulk for about 30–40% of their purchases. While a portion of the gap can be attributed to warehouse clubs, other factors must still be preventing households from completely closing the gap.

4.1.3 Bulk Buying Within Locations and Store Types

I showed in the previous sections that there are minor differences in package sizes available by ZIP code, but there are substantial differences in the types of stores households shop at.

4.1.4 Effect of Warehouse Club Entry on Bulk Purchasing

The previous section suggests that a portion of this "bulk-buying gap" can be attributed to warehouse clubs. Low-income households may not have access to the large, bulk sizes available at warehouse clubs because of membership fees or the sparsity of club locations. Warehouse clubs offer access to large packages with low unit prices in exchange for an annual membership fee. The largest warehouse club chains are Costco, Sam's Club, and BJ's. If access is the primary driver of this gap, then warehouse club entry should correspond with an increase in bulk purchasing.

Over my sample period, I observe 411 warehouse club openings between 2004 and 2015.[13] Since the Nielsen data has a panel structure, I observe the changes in bulk purchases within a household before and after a warehouse club opens. Within an area experiencing an entry, households may or may not shop at the warehouse club. If warehouse clubs affect bulk buying, then only households that shop at warehouse clubs should have a change in their bulk buying while households that do not shop at warehouse clubs will not. For each household-quarter, I compute the bulk share of expenditures and identify whether a household shopped at a warehouse club in that quarter. Using a differences-in-differences model, I estimate the effect of warehouse club entry on a household's bulk purchasing. This model compares households that shop at clubs against those that do not before and after a warehouse club enters.

I estimate the following regression:

$$BulkShare_{imt} = \beta_1 * Entry_{it} + \beta_2 * ClubShopper_{it} + \beta_3 * Entry_{it} \times ClubShopper_{it} + \gamma X_{it} + \lambda_{im} + \lambda_t + \epsilon_{imt}, \quad (4)$$

where Y is the expenditure-weighted share of bulk purchases made by household i in market m (defined as a ZIP code) in quarter t. *Entry* is an indicator for whether a warehouse club opens within 15 miles of household i in quarter t.[14] *ClubShopper* is an indicator for whether or not the household shops at a warehouse club in quarter t. X is a vector of household characteristics including household size, age, marital status, and presence of children. λ is a fixed effect for each household-market and year-quarter. Conditioning on λ_{mt} ensures that changes in bulk purchasing are due to entry instead of a household moving to an area with a warehouse club. Because the households that experience an entry are not nationally representative, I do not use Nielsen's projection weights.

The parameter of interest is β_3, which is identified by variation in bulk purchasing for households that shop at a warehouse club after one enters within 15 miles of a household. Variation in bulk purchasing for households that do not shop at a warehouse club after it enters identifies β_1. Variation in bulk purchasing for households that do shop at a warehouse club before one enters within 15 miles identify β_2.[15] The hypothesis is that warehouse club

[13] Data was collected by contacting companies and searching for store opening dates and was provided by the authors of Coibion, Gorodnichenko, and Koustas (2017). 2015 is the last full year of data with club openings.

[14] The average household traveled about seven miles to "buy goods" with a margin of error of about one mile. The 15-mile cutoff is well above the average travel distance to account for the fact that households may be willing to travel further to shop at a warehouse club relative to shopping at other types of stores.

[15] Some households may travel further than 15 miles to shop at a warehouse club and hence will have a club indicator before club entry.

entry will increase bulk purchasing only for households that shop at warehouse clubs ($\beta_3 > 0$) and there should be no effect on households that do not shop at warehouse clubs ($\beta_1 = 0$).

Table 4 shows the results of estimating Equation (4) separately for food and non-food products. Column (1) shows that bulk buying of food products increases by 2.6 percentage points after warehouse club entry for households that shop at warehouse clubs with no significant change for households that do not shop at warehouse clubs post-entry. Column (2) shows a similar effect for non-food items. Finally, for households that shop at clubs before warehouse club entry (i.e. shopping at a club further than 15 miles away), they buy non-food products in bulk 17.6 percentage points more than non-club shoppers and this is higher than the 5.3 percentage point increase on food products.

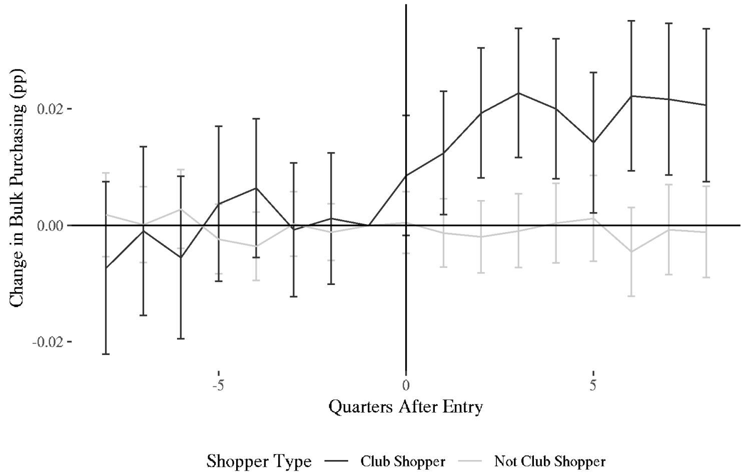
Table 4: Effect of Warehouse Club Entry on Bulk Buying

	Food	Non-Food
	(1)	(2)
Post-Entry	−0.0001	−0.001
	(0.002)	(0.004)
Club Shopper	0.053***	0.176***
	(0.003)	(0.006)
Post-Entry : Club Shopper	0.026***	0.028***
	(0.004)	(0.006)
Demographic Controls	Y	Y
Household-ZIP FE's	Y	Y
Observations	50,345	49,783
Adjusted R^2	0.628	0.441
Note:	*p<0.1; **p<0.05; ***p<0.01	

Notes: This table uses 2004–2015 Nielsen Consumer Panel data at the household-quarter level. Coefficients are reported for Equation (4) which regresses households' quarterly bulk purchase shares on an indicator for warehouse club entry, an indicator for whether the household shops at a warehouse club, and an interaction term as well as household demographics (household size, age, presence of children, and marital status). Household-ZIP code and year-quarter fixed effects are included. Standard errors are clustered at the household level and projection weights are not used.

To visualize the changes in bulk buying, I estimate a modified version of Equation (4) that includes quarterly indicators for pre- and post-entry:

Figure 8: Event Study of Warehouse Club Entry



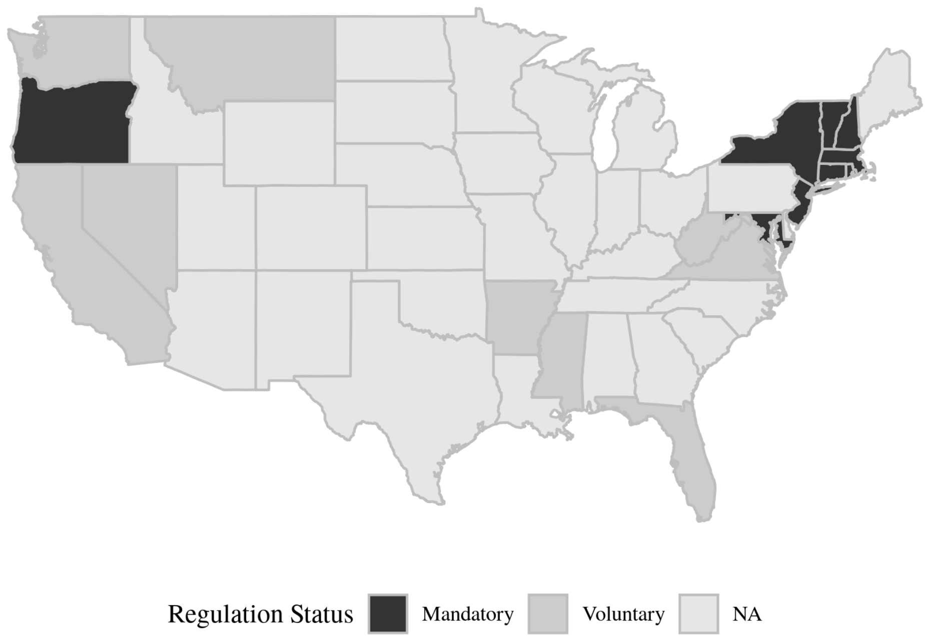
Notes: This figure plots the β_1^q coefficients from Equation (5)—the effects of warehouse club entry on bulk purchasing of households that shop and do not shop at club stores—using 2004–2015 household-by-quarter Nielsen Consumer Panel data. The regression controls for household size, age, presence of children, and marital status as well as household-ZIP code fixed effects. Standard errors are clustered by household and all coefficients are relative to bulk purchasing in the quarter before entry ($q = -1$). Standard errors are suppressed for clarity.

$$\begin{aligned} BulkShare_{imt} = & \sum_q \beta_1^q PostEntry_{it}^q + \beta_2 ClubShopper_{it} + \\ & \sum_q \beta_3^q PostEntry_{it}^q \times ClubShopper_{it} + \gamma X_{it} + \lambda_{im} + \epsilon_{imt}, \end{aligned} \tag{5}$$

where $PostEntry$ is a dummy for the eight quarters before and after a warehouse club enters with the quarter immediately preceding entry ($q = -1$) as the reference group. Figure 8 plots the coefficients on the quarterly dummies showing that only households that shop at warehouse clubs after entry increase their bulk purchasing by about two percentage points.

This analysis shows that warehouse club entry increases bulk purchasing only for households that shop at warehouse clubs after entry. Since high-income households are more likely to shop at warehouse clubs (about 70% of households making over \$100,000 compared to about 30% of households making under \$25,000), some of the gap in bulk purchasing is due to the extensive margin of whether or not households shop at warehouse clubs. Some of this difference in the extensive margin is likely due to the fact that warehouse clubs charge annual membership fees, which are more affordable for high-income households.

Figure 9: Unit Price Regulations by State (2017)



Notes: Figure plots whether or not a state has laws or regulations in place governing the display of unit prices as of 2017. Data is from NIST Handbook 130.

4.2 Salience of Unit Prices

Another possible contributor to this bulk-buying gap is the *salience* of unit prices. Consumers may not be aware of the bulk discount because it is not visible or they do not compute unit prices when making purchases. Previous research has shown that To test this hypothesis, I leverage state-level variation in laws about displaying unit-prices. Figure 9 shows that, as of 2017, 18 states have regulations on the display of unit prices and 32 have no regulations. As my measure of unit-pricing laws, I use the information compiled in the National Institute of Standards and Technology (NIST) Handbook 130. This handbook is published annually following the National Conference on Weights and Measures and consists of numerous laws and regulations that are recommended for states to adopt in order to "achieve ... uniformity in weights and measures laws and regulations among the various states and local jurisdictions in order to ... provide uniform and sufficient protection to all consumers." (National Institute of Standards and Technology 2019). Every year, NIST publishes a summary of which states have adopted unit pricing laws. I use its classification to indicate whether or not a state has laws regarding unit pricing (see Appendix for further details).

I first estimate the differences in bulk purchasing between states with and without unit pricing laws using the following regression:

$$Y_{imt} = \beta_1 Law_{imt} + \beta_2 Income_{imt} + \beta_3 Law_{imt} \times Income_{imt} + \gamma X_{imt} + \lambda_m + \lambda_t + \epsilon_{imt}, \quad (6)$$

where Y is the annual share of expenditures that were bulk purchases for household i in market m in year t. Law is an indicator for whether or not unit-price regulations are in effect. $Income$ is a household's income bin. X controls for household age, size, marital status, and presence of children. I control for market and time fixed effects through λ. If there are different patterns across income groups within states that have unit pricing regulations, then the interaction term will pick them up. I cluster my standard errors at the market (DMA) level.

Since 2004, only four states have adopted or repealed regulations on unit prices, so the coefficient on unit pricing regulation is primarily identified from cross-sectional variation between states that have regulations and those that do not.

This estimation only reveals cross-sectional differences and does not prove a causal relationship between unit pricing regulations and bulk purchasing. However, given that shopping environments are likely similar across states, this would strongly suggest that these regulations would have some effect on how much households buy in bulk. To provide causal evidence, I examine households that move between states with different regulations. To estimate the effect of unit-price laws on these movers, I use the following regression:

$$Y_{imt} = \beta_1 Law_{imt} + \beta_2 Law_{imt} \times Income_{imt} + \gamma X_{imt} + \lambda_i + \lambda_t + \epsilon_{imt}, \quad (7)$$

where the variables are the same as in Equation (6), but I control for household fixed effects.[16] With this specification, β_1 is identified by changes in bulk purchases for households that move from a state with unit-price regulations to a state without unit-price regulations (or vice versa) and β_2 allows that effect to vary by income level. Projection weights are not used in estimation because the weights are not designed for this subsample of movers.

Table 5 reports the results of estimating Equations (6) and (7) on food and non-food items separately. Columns (1) and (2) report the cross-sectional estimates which reveal that households in states with unit-price regulations are more likely to buy in bulk and this effect increases with income and is similar between food and non-food items. Columns (3) through (6) report the same estimates, but only uses households that move between states to identify

[16] Household fixed effects are actually household-income fixed effects to ensure that the interaction term is identified by differences in regulation and not by changes in a household's income.

Table 5: Association of Unit Price Laws and Bulk Buying

	Full Sample		Movers			
	Food	Non-Food	Food		Non-Food	
	(1)	(2)	(3)	(4)	(5)	(6)
Law	-0.005	-0.008	0.002^{***}	0.003	0.006^{***}	-0.001
	(0.004)	(0.005)	(0.001)	(0.002)	(0.001)	(0.004)
25-50k : Law	0.007^{**}	0.007^{*}		0.001		0.009^{**}
	(0.003)	(0.004)		(0.002)		(0.004)
50-100k : Law	0.014^{***}	0.014^{***}		-0.0003		0.008^{*}
	(0.003)	(0.004)		(0.002)		(0.004)
>100k : Law	0.015^{***}	0.015^{**}		-0.003		0.004
	(0.005)	(0.007)		(0.003)		(0.005)
Market FE	Y	Y	N	N		
Year FE	Y	Y	Y	Y		
Household-Income FE	N	N	Y	Y		
Observations	767,946	766,789	767,946	767,946	766,789	766,789
Adjusted R^2	0.097	0.082	0.781	0.781	0.668	0.668

Note: *p<0.1; **p<0.05; ***p<0.01

Notes: Using 2004–2017 Nielsen Consumer Panel data and @nist130, Columns (1) and (2) report the regression coefficients of Equation (6), which regresses household bulk share of purchases on an indicator for whether a state has unit-price regulations, household income, and their interaction along with household size, age, presence of children, and marital status. Market and year fixed effects are included. Standard errors are clustered at the market level and Nielsen projection weights are used to ensure national representativeness. Columns (3)–(6) use the same data and report the coefficients of Equation (7) which estimates the event study of bulk purchasing changes for households that move between states with and without unit-price regulations. Household bulk share of purchases are regressed on the same covariates as previously, but household-income and year fixed effects are included. Projection weights are not used in this regression the effects of unit-price regulations. Columns (3) and (5) show that there is a small increase in bulk buying when households move to states with unit pricing regulations, but this effect is larger for non-food items compared to food items. Columns (4) and (6) allow for different effects by income group and while both have much less statistical power, column (6) shows that there are small, but significant increases in bulk purchasing when households move to a state with unit pricing regulations. Since most movers are households making between \$25,000 and \$100,000, the remaining coefficients do not have enough power to rule in or rule out possible effects of unit-price laws for the richest or the poorest households.

To illustrate this effect graphically, I plot an event study by estimating a modified version of Equation (7):

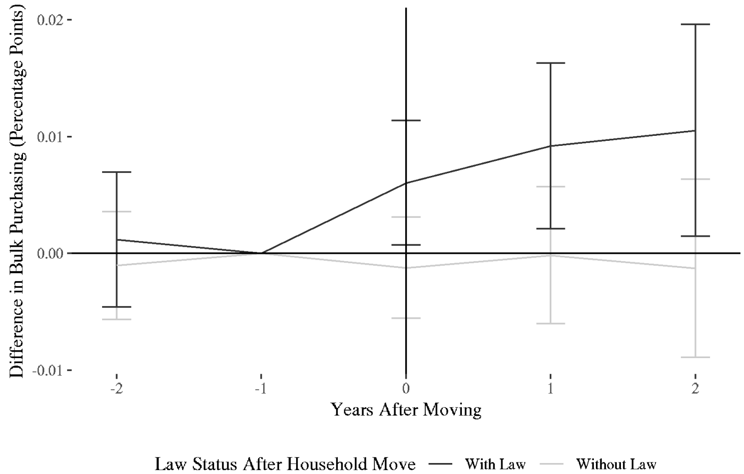
$$Y_{imt} = \sum_{T} \beta_1^t E_{imt} + \gamma X_{imt} + \lambda_i + \lambda_t + \epsilon_{imt}, \tag{8}$$

where E is a dummy for each year before or after a household moves to a state with a different unit pricing regime. The reference group is $t = -1$ so all effects are relative to the year before the household moves. I estimate this equation separately for different types of moves and plot the annual coefficients in Figure 10. Moves from states without regulations to states with regulations are likely to have different effects than moves from states with regulations to states without regulations. Figure 10 shows that households increase their bulk buying by just under one percentage point when they move from a state without unit-price regulations to a state with unit-price regulations. On the other hand, households that move from states with unit-price regulations to states without unit pricing regulations show no difference in their bulk-buying habits. The most likely explanation for this asymmetric behavior is that unit prices remain *salient* for households even after they leave a state with unit-price regulations. Therefore, they continue to shop based on unit prices. For households that lived in states without unit-price regulations, unit prices were not *salient* and the new regulations made them *salient*.

Overall, this shows that unit-price regulations, which increase the *salience* of unit prices have an effect on bulk purchasing, but the effect is relatively minor and smaller than the effect of warehouse club entry. Since unit-price regulation does not have an effect on households making under \$25,000 and this effect may be different by income, there is likely scope to improve education and messaging around understanding unit prices and how to utilize them to lower one's total shopping bill.

Even though potential improvements in bulk purchasing are relatively small, unit pricing regulations are relatively simple to implement for both policymakers and retailers. Retailers will bear some initial setup costs of redesigning their price labels, but ongoing costs will

Figure 10: Event Study of Movers and Unit Price Laws



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the β_1^t coefficients and 95% confidence intervals from Equation (8), which regresses household bulk purchasing on dummies for years before and after a household moves to a state with a different unit pricing regime than the state it moves from. The regression controls for household size, age, marital status, and presence of children as well as household and year fixed effects. "With Law" reports estimates for households that move from a state without unit price regulations to a state with unit price regulations. "Without Law" reports estimates for households that move from a state with unit price regulations to a state without regulations.

likely be similar to current menu costs that firms bear. Adopting unit pricing policies (like those recommended by the National Conference on Weights and Measures) would narrow the bulk-buying gap while imposing few costs. These findings support the broader assertion that increasing price transparency allows households to improve their decisionmaking.

5 Model

The previous analyses show that both income increases and warehouse club entry increase bulk buying. However, these changes could be related to other important factors, such as storage costs. For example, if storage costs are the limiting factor, then providing incentives for warehouse clubs to open will have limited effects for households that are storage-constrained. To isolate storage costs and determine the effects of *only* changing product concentrations or household preferences, a formal model is necessary.

In order to conduct this analysis, I estimate a discrete choice model of toilet paper purchases. Toilet paper is the ideal product for this analysis because it is a necessity item with easily observable dimensions of differentiation, namely price, quality, quantity, and package size. It is offered in a wide range of package sizes and stores often stock numerous brands and sizes (grocery and mass merchandise stores usually stock 35–40 unique brand-sizes). The top five brands and private-label store brands account for 86% of sales. I focus on the five most common package sizes, which are 4, 6, 9, 12, and 24-roll packages. I define a product as a unique brand-size combination.[17] Additionally, underlying toilet paper consumption is primarily a function of household size and age, not income.[18] Unlike other products, annual toilet paper purchases do not vary significantly with income. High-income household consume a similar amount as low-income households but make fewer purchases (Orhun and Palazzolo 2019).

The biggest challenge to identification is separately identifying storage costs from underlying demand. When the volume of a product is what households demand, these two parameters cannot be separately identified because a household that purchases a large volume could have a high demand or low storage costs. However, for items like toilet paper and detergent, households are not demanding a particular volume of the product. In the case of detergent, households demand a certain number of loads to be washed and for toilet paper, they demand a certain number of uses. Because detergent comes in varying concentrations, the same number of laundry loads could be washed using 100 ounces of regular detergent or 50 ounces of 2x concentrated detergent. Households that purchase the 100 ounce package have the same demand for laundry loads as households that purchase the 50 ounce package, but lower storage costs since they can store the bigger package.

The same reasoning holds true for toilet paper. Households do not demand a particular number of rolls (the primary determinant of package size), but choose how long they want their supply to last (i.e. purchase enough to last for 2 weeks, a month, 2 months, etc.).[19] Toilet paper comes in a variety of concentrations with "mega" rolls being more concentrated than "double" rolls which are more concentrated than "single" rolls. Therefore, households that purchase 24 "single" rolls have the same demand for toilet paper uses as households that purchase 12 "double" rolls, but the former household has lower storage costs since they can store the bigger package.

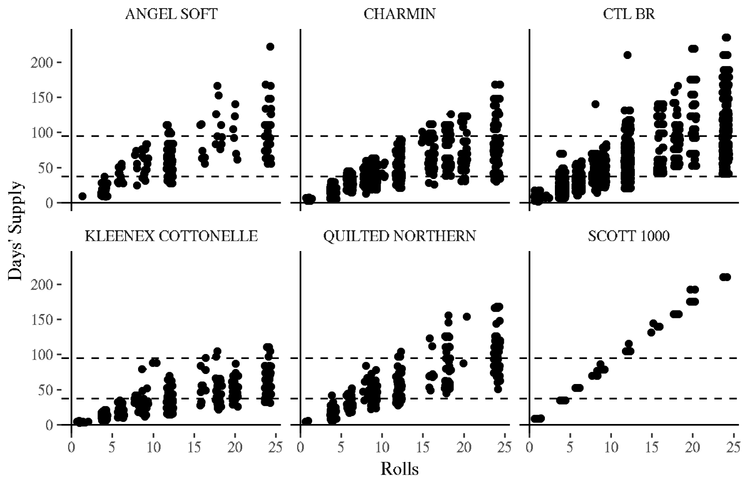
To illustrate the varying concentrations of toilet paper, Figure 11 plots the distribution of quantity (measured in number of days the supply will last for a single person) against package sizes (measured in rolls) for toilet paper products in the Nielsen data. As expected, there is an increasing relationship between the number of rolls in a package and how long the package will last, but there is substantial variation within packages containing the same number of

[17] Specifically, this is a unique brand-roll count-sheet count because packages can differ in their "concentration" due to "double," "mega," and "super mega" rolls.

[18] A 100-fold cross-validated LASSO regression of annual purchases on household characteristics rules out income as significantly predictive. See Appendix for details.

[19] According to a 2007 Charmin survey, the average person uses 57 sheets per day. I assume this consumption rate when computing how long a product will last.

Figure 11: Scatterplot of Toilet Paper Package Size and Quantity



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the package sizes and quantities of the top five toilet paper brands and private-label products. The x-axis represents the number of toilet paper rolls contained in a package while the y-axis represents the number of days a product will last a single person household assuming a consumption rate of 57 2-ply sheets per day. Noise is added horizontally to better illustrate the number of products available within package sizes since roll counts are discrete. Dashed lines indicate the 25th and 75th percentiles of the average days' supply purchased by a household.

rolls. The variation is large enough to ensure significant overlap in quantity between packages with 12 rolls versus those with 24 rolls.[20] This overlap generates the necessary variation to separate storage costs from underlying demand. The dashed lines illustrate that a majority of households purchase quantities for which there are many different sizes available.

5.1 Model Setup

A household's product choice is modeled as follows. Households consume toilet paper daily and typically consume toilet paper out of their inventory. When its inventory runs out, it makes a purchase during their next shopping trip.[21] When making a purchase, a household considers the price, quality, quantity, and size of each product and chooses the product that maximizes its utility. A household's choice is static because even though it can store toilet paper, I assume purchases are only made when its inventory is exhausted.[22] Household i's

[20]The only exception is Scott toilet paper which does not offer different roll types. All rolls have 1000 sheets.

[21]Since toilet paper is purchased jointly with other products in 99% of cases, I assume a household's store choice is based on the bundle of goods it has to purchase and toilet paper does not significantly influence store choice.

[22]Hendel, Lizzeri, and Roketskiy (2014) offers a theoretical justification for this assumption. Even with dynamics, firms will price so consumers purchase when their inventory is exhausted.

utility from product j purchased on shopping trip t is represented in the following way:

$$U_{ijt} = \underbrace{\beta_1 Price_{jt} + \beta_2 \log(SheetsPP_j) + \beta_3 Large_j + \theta_{b(j)}}_{\beta' x_{ijt}} + \epsilon_{ijt}, \tag{9}$$

where $Price$ is the price of the item, $SheetsPP$ is the number of total sheets in the package divided by household size, $Large$ is a dummy for the package having more than 12 rolls and θ is a brand fixed effect. I use brand fixed effects to capture quality differences between products. Quality differences within brands are virtually non-existent in the toilet paper category because within brands, the only differences are in package size (e.g. 4, 12, 24-roll packages) or roll type (single, double, mega rolls). Since consumption is related to household size, I divide total usable sheets by number of people in the household to generate a quantity measure that is comparable across households of different sizes. I define "large" as a package with more than 12 rolls because a 12-roll package is likely the upper limit of what can comfortably be carried in addition to other items. This serves as an approximation of storage and transportation costs because it is an indicator for a package's physical size. I assume that the ϵ is identically and independently distributed following a type I extreme value distribution.

Given these assumptions, the probability that household i chooses product j on trip t can be written in a closed form:

$$P_{ijt} = \frac{e^{\beta' x_{ijt}}}{\sum_j e^{\beta' x_{ijt}}} \tag{10}$$

As a result, the log-likelihood function can be written as:

$$LL(\beta) = \sum_t \sum_i \sum_j y_{ijt} \log(P_{ijt}), \tag{11}$$

where y indicates whether household i chose product j on shopping trip t. The preference parameters β can then be estimated using MLE. I estimate this model separately for each income quartile and year, allowing for heterogeneity in preferences between income groups, but not within income groups.

The price coefficient is primarily identified from variation between shopping trips, whether due to price differences between stores or price differences within stores, such as sales or more permanent price shifts. The "large" and "sheets" coefficients are identified from variation in product composition. Sheet counts differ between brands and also within brands because of "single," "double," and "mega" rolls. For example, two four-roll packages will differ in the number of sheets if one package is four "double" rolls and the other is four "mega" rolls, even though the physical package sizes are similar.

The objects of interest are the willingness-to-pay for package size and quantity. The willingness to pay for large packages is a measure of *storage* costs because these are a function of the package size. On the other hand, willingness to pay for quantity measures differences in preferences for buying in bulk. These differences could be due to differences in *liquidity* constraints or the *salience* of bulk discounts.

6 Estimation Results

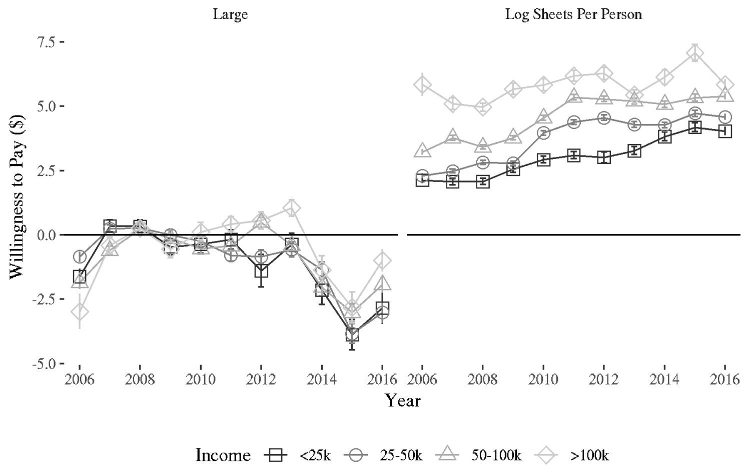
I estimate the model for each year and income group. Between 2006 and 2016, I observe about 600,000 toilet paper purchases across about 72,000 households at grocery and mass merchandisers. For interpretability, I convert all estimated coefficients into a willingness to pay measure by dividing each coefficient by the price coefficient. Figure 12 shows the willingness to pay for package size and quantity by year and income group. As expected, storage costs (the willingness to pay for package size in my framework) are negative, but they are relatively small with little difference between income groups until about 2014 when storage costs increase dramatically. The main finding is that while storage costs are significantly negative, they are similar across income groups and move in parallel over time. Therefore, while storage costs exist and are negative (as expected), the fact that they are similar across households shows that they are not the primary factor driving the difference in bulk buying across different income groups.

Willingness to pay for quantity (measured in log sheets per person) also moves in parallel over time, but there are substantial differences between income groups. Over time, all income groups have increased their willingness to pay for quantity with the lowest-income households increasing the most, from \$1.96 per log point to \$4.20 per log point.[23] However, a gap still remains between income groups with higher-income groups willing to pay more for quantity than lower-income groups. These differences across income groups generate the gap between high- and low-income households in bulk buying.

Given my specification, this quantity preference is a combination of both the *salience* of unit prices and *liquidity* constraints. This preference includes the *salience* of unit prices because households that are more aware of bulk discounts and unit prices will pay more attention

[23]Because toilet paper packages come in a wide range of quantities, the normal approximation between log differences and percent breaks down. A full log point would correspond to about a 2.7x increase in the package quantity. Put in context, this would be the willingness to pay to move from about a package of four "regular" rolls to a package of four "triple" rolls.

Figure 12: Preferences for Package Size and Quantity



Notes: This figure presents MLE estimates of Equation (9) for β_2 and β_3 converted into willingness-to-pay measures for each year of data. Willingness-to-pay is in nominal dollars and represents the amount a household is willing to pay for a large package (>12 rolls) or for an additional log-point of per-capita sheets. Standard errors are computed using the delta method.

to the underlying quantity available in a package. This preference also includes *liquidity* constraints because package prices increase in quantity, so a budget-constrained household's inability to purchase a large quantity package would manifest itself as a weaker preference for quantity than an unconstrained household.

In addition to the size and quantity preferences, this figure reveals an interesting trend across income groups. For households making under $25,000, their preference for quantity has been gradually increasing over time, with slightly accelerated increases during the Great Recession. For households making over $25,000, the story is different. Preferences were stable between 2006–2008, increased between 2009–2011, and then stabilized again since 2012. This pattern supports the finding in Nevo and Wong (2018) that household's adopted more value-conscious habits during the Great Recession. While other households have stabilized in their preferences, households making under $25,000 have continued to increase in their preference for quantity.

This model provides a foundation for understanding how preferences differ across households by capturing heterogeneity in preferences attributable to household income and household size (through the sheets per person variable). While there may be some dimensions of unobserved heterogeneity within income groups that cannot be captured by this model, the fact that storage costs are a minor contributor to the bulk-buying gap is striking.

7 Counterfactuals

With the parameter estimates from the previous section, I can run counterfactuals to predict how households will respond to the elimination of storage costs through increased product "concentration" and changes in preferences. For these counterfactual exercises, I use parameter estimates from 2016 and compare all counterfactual results to a "base case" of predicted purchases given 2016 selections. Since 2016 was a year with moderate storage costs, this provides a conservative upper bound on how the bulk-buying gap would change if storage costs were eliminated. The results of my counterfactual exercises are reported in Table 6. For expositional ease, I have translated quantity purchased per trip into a days' supply measure assuming a constant daily consumption of 57 2-ply sheets per day.[24]

Table 6: Predicted Effects on Bulk Purchasing

Income	Base	No Storage	Same Qty Pref.	Both
<25k	45.48	47.60	48.26	50.68
25-50k	46.79	48.65	47.89	49.86
50-100k	48.44	49.89	48.85	50.34
>100k	50.10	51.31	50.10	51.31

Note: This table reports predicted package quantities purchased by households using model estimates of Equation (9). Units are number of days the chosen package will last assuming average daily consumption rate of 57 2-ply sheets [@jaffe2007]. "No Storage" scenario imposes that no products incur storage costs for households. "Same Qty Pref." scenario imposes that all households have the same preferences for quantity as households making over $100k. "Both" combines both scenarios.

The "Base" case reports the average days' supply purchased per trip for each income group given the shopping environment and parameter estimates in 2016. The quantity purchased is increasing in income with the highest income households purchasing 10% larger quantities than the lowest income households. In the "No Storage" scenario, I set the storage cost of all products to 0 and as a result, all households purchase larger quantities with the largest increase happening in the poorest households.[25] Eliminating storage costs reduces the gap between high- and low-income households by 23%. Even though the poorest households substantially increase their purchase quantity, this change encourages all households to purchase larger quantities. In the "Same Qty. Pref." scenario, I set all household's preference for quantity equal to that of the highest-income households. This provides an even larger increase in quantity purchased over the base case compared to eliminating storage costs, with

[24] This comes from a survey conducted by Charmin in 2007. See Jaffe (2007) for details.

[25] This would be possible if manufacturers increased the "concentration" of their products, which allows them to offer the same quantity in a smaller package. In the toilet paper sector, this often happens when manufacturers offer "mega" or "ultra" rolls that guarantee more sheets per roll.

the gap shrinking by 63%. Combining both of these cases, the gap shrinks by 89%. The remaining gap is due to differences in brand preferences across income groups.

These counterfactuals show that eliminating storage costs would encourage low-income households to buy larger quantities, but would only moderately reduce the inequality between high- and low-income households. A more effective policy would focus on increasing low-income households preferences for quantity. One possibility is through increasing the *salience* of unit prices so that low-income households become more aware of the value of buying in bulk. In light of the cross-sectional results reported in Section 4.2, simply posting unit prices may not be sufficient enough to increase the *salience* of unit prices. An alternative policy could simply identify the lowest unit-price item within the product category to assist consumers in making value choices without requiring much cognitive resources. If preferences for quantity are being driven by budget constraints then policies that reduce these constraints, such as increased access to paychecks as wages are earned or short-term credit could improve outcomes for low-income households.

8 Conclusion

In this paper, I document the existence of a bulk-buying gap in which high-income households buy in bulk more often than low-income households, especially for storable, necessity items like toilet paper and diapers. Because they buy small quantities and do not take advantage of bulk discounts, low-income households miss out on substantial savings (paying 36% higher prices in the case of toilet paper). I provide evidence that *store access*, *unit-price salience*, *liquidity constraints*, and *storage costs* contribute to this gap.

I first study the effect of warehouse club entry on bulk buying since these stores often offer large sizes that are not available at other stores. Using an event study, I find that warehouse club entry increases bulk buying by about one percentage point, but this increase is primarily concentrated in middle- and high-income households. Even if warehouse clubs are available, low-income households may not use them because of annual membership fees or the costs of storing exceptionally large items. Lowering these other barriers to *accessing* warehouse clubs might help low-income households take advantage of the bulk-buying options that higher income households already take advantage of.

Leveraging within-household income changes, I find that households increase their bulk buying as their income increases. These income changes increase bulk buying by about one percentage point or less with larger increases in essential, non-food items. To the extent that a household's annual income is related to its intra-year liquidity, some of the differences in bulk buying could be attributable to *liquidity constraints*.

Using between-state variation in regulations on posting unit prices, I show that households in states with unit-price regulations buy in bulk more often than states without, but as with other changes, these effects are most pronounced for middle- and high-income households. Given that grocery shopping is similar across states, this suggests that unit prices are more *salient* for high-income households compared to low-income households.

My formal demand model allows me to simulate counterfactuals in which household can buy more "concentrated" products, obviating storage costs or in which preferences are the same across income groups. I find that removing storage costs would reduce the bulk-buying gap between high- and low-income households by 23% while equalizing preferences for quantity would reduce the gap by 63%. Given these findings, policies that focus on making unit prices more *salient* for consumers and help reduce their *liquidity* constraints will be more effective at helping low-income households take advantage of bulk discounts compared to policies that reduce storage costs.

This paper is one of the first to focus on consumer's take-up of bulk discounts and document the factors that contribute to this decision. It provides evidence that while bulk discounts theoretically discriminate between households that can and cannot store, empirically, bulk discounts discriminate in ways that align with income. Low-income households are less likely to buy in bulk compared to high-income households. My findings show that policies that increase *access* to bulk sizes, increase *liquidity*, or increase the *salience* of unit prices will help low-income households take advantage of bulk discounts in the same way that high-income households do.

9 References

Aguiar, Mark, and Mark Bils. 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105 (9): 2725–56.

Aguiar, Mark, and Erik Hurst. 2005. "Consumption Versus Expenditure." *Journal of Political Economy* 113 (5): 919–48.

———. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *The Quarterly Journal of Economics* 122 (3): 969–1006.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *The Quarterly*

Journal of Economics, May.

Anderson, Eric T, and Inseong Song. 2004. “Coordinating Price Reductions and Coupon Events.” *Journal of Marketing Research* 41 (4): 411–22.

Attanasio, Orazio, and Christine Frayne. 2006. “Do the Poor Pay More?” Mario Einaudi Center for International Studies.

Attanasio, Orazio, and Luigi Pistaferri. 2014. “Consumption Inequality over the Last Half Century: Some Evidence Using the New Psid Consumption Measure.” *American Economic Review* 104 (5): 122–26.

Beatty, Timothy KM. 2010. “Do the Poor Pay More for Food? Evidence from the United Kingdom.” *American Journal of Agricultural Economics* 92 (3). Oxford University Press: 608–21.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99 (4): 1145–77.

Chung, Chanjin, and Samuel L Myers Jr. 1999. “Do the Poor Pay More for Food? An Analysis of Grocery Store Availability and Food Price Disparities.” *Journal of Consumer Affairs* 33 (2): 276–96.

Cohen, Andrew. 2008. “Package Size and Price Discrimination in the Paper Towel Market.” *International Journal of Industrial Organization* 26 (2): 502–16.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas. 2017. “Consumption Inequality and the Frequency of Purchases.” Working Paper 23357. Working Paper Series. National Bureau of Economic Research. http://www.nber.org/papers/w23357.

Erdem, Tülin, Susumu Imai, and Michael P Keane. 2003. “Brand and Quantity Choice Dynamics Under Price Uncertainty.” *Quantitative Marketing and Economics* 1 (1): 5–64.

Fox, Kevin J, and Daniel Melser. 2014. “Non-Linear Pricing and Price Indexes: Evidence and Implications from Scanner Data.” *Review of Income and Wealth* 60 (2): 261–78.

Griffith, Rachel, Ephraim Leibtag, Andrew Leicester, and Aviv Nevo. 2009. “Consumer Shopping Behavior: How Much Do Consumers Save?” *Journal of Economic Perspectives* 23 (2): 99–120.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante. 2010. “Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006.” *Review of Economic Dynamics* 13 (1): 15–51.

Hendel, Igal, Alessandro Lizzeri, and Nikita Roketskiy. 2014. "Nonlinear Pricing of Storable Goods." *American Economic Journal: Microeconomics* 6 (3): 1–34.

Hendel, Igal, and Aviv Nevo. 2006a. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica* 74 (6): 1637–73.

———. 2006b. "Sales and Consumer Inventory." *The RAND Journal of Economics* 37 (3): 543–61.

———. 2013. "Intertemporal Price Discrimination in Storable Goods Markets." *American Economic Review* 103 (7): 2722–51.

Hong, Pilky, R Preston McAfee, and Ashish Nayyar. 2002. "Equilibrium Price Dispersion with Consumer Inventories." *Journal of Economic Theory* 105 (2): 503–17.

Isakson, Hans R, and Alex R Maurizi. 1973. "The Consumer Economics of Unit Pricing." *Journal of Marketing Research* 10 (3): 277–85.

Jaffe, Chuck. 2007. "On a Roll." *MarketWatch*, October. https://www.marketwatch.com/story/on-a-roll-toilet-paper-purchases-set-pace-for-weightier-matters.

Kaplan, Greg, and Guido Menzio. 2015. "The Morphology of Price Dispersion." *International Economic Review* 56 (4): 1165–1206.

Krueger, Dirk, and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *The Review of Economic Studies* 73 (1): 163–93.

Kunreuther, Howard. 1973. "Why the Poor May Pay More for Food: Theoretical and Empirical Evidence." *The Journal of Business* 46 (3): 368–83.

McManus, Brian. 2007. "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee." *The RAND Journal of Economics* 38 (2): 512–32.

Narasimhan, Chakravarthi. 1984. "A Price Discrimination Theory of Coupons." *Marketing Science* 3 (2): 128–47.

National Institute of Standards and Technology. 2014. "Unit Pricing Guide; a Best Practice Approach to Unit Pricing (Special Publication 1181)." Edited by David Sefcik.

———. 2019. "Uniform Laws and Regulations in the Areas of Legal Metrology and Fuel Quality (Handbook 130)." Edited by Linda Crown, David Sefcik, and Lisa Warfield.

Nevo, Aviv, and Catherine Wolfram. 2002. "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals." *RAND Journal of Economics* 33 (2): 319–39.

Nevo, Aviv, and Arlene Wong. 2018. "The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession." *International Economic Review (Forthcoming).*

Orhun, A. Yeşim, and Michael Palazzolo. 2019. "Frugality Is Hard to Afford." *Journal of Marketing Research* 56 (1).

Pesendorfer, Martin. 2002. "Retail Sales: A Study of Pricing Behavior in Supermarkets." *The Journal of Business* 75 (1): 33–66.

Russo, J Edward. 1977. "The Value of Unit Price Information." *Journal of Marketing Research* 14 (2): 193–201.

Stole, Lars A. 2007. "Price Discrimination and Competition." *Handbook of Industrial Organization* 3: 2221–99.

Talukdar, Debabrata. 2008. "Cost of Being Poor: Retail Price and Consumer Price Search Differences Across Inner-City and Suburban Neighborhoods." *Journal of Consumer Research* 35 (3): 457–71.

Triplett, Jack. 2004. "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes." OECD.

Yao, Jun, and Harmen Oppewal. 2016. "Unit Pricing Increases Price Sensitivity Even When Products Are of Identical Size." *Journal of Retailing* 92 (1): 109–21.

Zou, Hui, and Trevor Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20.

A Annual Consumption Analysis

In order to establish that income is not predictive of a household's toilet paper consumption rate, I first estimate basic OLS regressions and then formalize the result using a 100-fold cross-validated elastic net regression to select the most predictive variables. If income and toilet paper consumption are related, then an OLS regression will extract the correlation.

First, I estimate a household's daily consumption in the following manner. I compute the total number of sheets purchased by a household in a given year, excluding the final purchase of the year since it may not be consumed within the year. I divide this total by the number of days between the first and last purchase of the year to get a household's average daily consumption rate. This method avoids complications where end of the year inventory may be carried over to the following year or a household may start the year with some inventory.

Given a household's average daily consumption rate, I estimate an OLS regression of consumption on household characteristics:

$$Y_i = \beta X_i + \epsilon_i, \tag{12}$$

where Y is household i's average daily consumption and X is a vector of household characteristics. Figure 13 plots the income coefficients of an OLS regression including only income covariates and the coefficients when household size, age, marital status, and presence of children are included. The graph illustrates that after controlling for covariates that plausibly cause increased consumption, income is not significantly correlated with consumption.

The above specification omits many other possible covariates that could be correlated with average daily consumption. When there are many possible variables that can be included, there is a risk of over-fitting. Elastic net regularization is a machine learning method that penalizes over-fitting and selects only the most predictive variables.

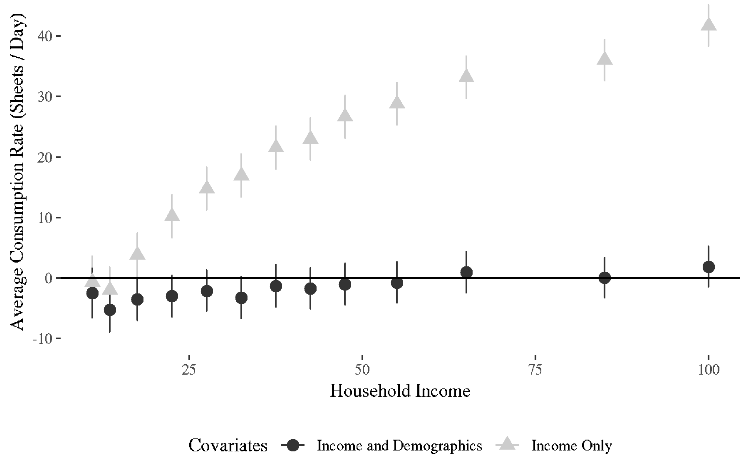
The elastic net solves the following minimization problem:

$$\min_{\beta} \|y - X\beta\|^2 + \lambda \left(\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2\right), \tag{13}$$

where $\|\cdot\|_1$ is the L1 norm and $\|\cdot\|_2$ is the L2 norm. The OLS estimate is the β that solves the minimization problem with only the first term. The second term and third term provide penalties to shrink and select for the most predictive variables.

I set the mixing parameter (α) to be 0.5. When covariates are correlated in groups, Lasso

Figure 13: Average Daily Consumption by Household Income



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the income bin coefficients from Equation (12), which regresses average daily household toilet paper consumption on household demographics (household size, age, presence of children, and marital status). Average daily consumption is computed by dividing total quantity purchased in a year by the number of days a household was active in the panel.

regression ($\alpha = 1$) tends to only select one and discard all other members of the group while ridge regression ($\alpha = 0$) tends to shrink correlated coefficients towards each other (Zou and Hastie 2005). Since some of the possible covariates form natural groups (e.g. all income bins or all markets), I chose $\alpha = 0.5$ since this tends to include or exclude groups together.

I estimate a 100-fold cross-validated elastic net regression to select the most predictive covariates. The resulting estimates selects many household characteristics including household size, age, marital status, presence of children, and race, but excludes almost all income and geographic coefficients.[26]

B Bulk-Buying Gaps Across Product Categories

The Nielsen Consumer Panel records information on individual purchases made by households. For each household, I compute the total volume and number of packages purchased over the year for each product category. Dividing the total volume by the number of packages gives the average package size purchased by that household. For example, if a household purchased 20 rolls of toilet paper across two packages, then the average package size purchased is 10 rolls per package. Therefore, I have the average package size purchased for each household-year-product category. To establish that high income is associated with larger average packages, I

[26]Elastic net results are available upon request.

estimate the following regression for each product category:

$$log(Y)_{im} = \beta_1 Income_{im} + \beta_2 X_{im} + \lambda_m + \lambda_t + \epsilon_{im}, \tag{14}$$

where Y is the average package size purchased by household i in market m. I define a market as a DMA-year.[27] *Income* is the household's income bin. X controls for other household demographics including household size, presence of children, age, and marital status. Finally, to capture changes over time within geographic regions, I include a market and year fixed effects λ.

Each product category gives a set of coefficients. These coefficients are plotted in Figure 14. For clarity, I only show product categories in which at least 5000 households made a purchase in a year (about 5–10% of Nielsen households). The figure shows that there are wide differences in average package sizes purchased across a wide variety of products. For paper towels, aluminum foil, and kitchen trash bags, households making over \$100,000 buy 40% larger packages than those purchased by households making under \$25,000. Furthermore, while standard errors are not shown, they are narrow and differences between income groups are also significant. Households making over \$100,000 purchase larger packages than those making \$50–100,000 who in turn purchase larger packages than those making \$25–50,000.

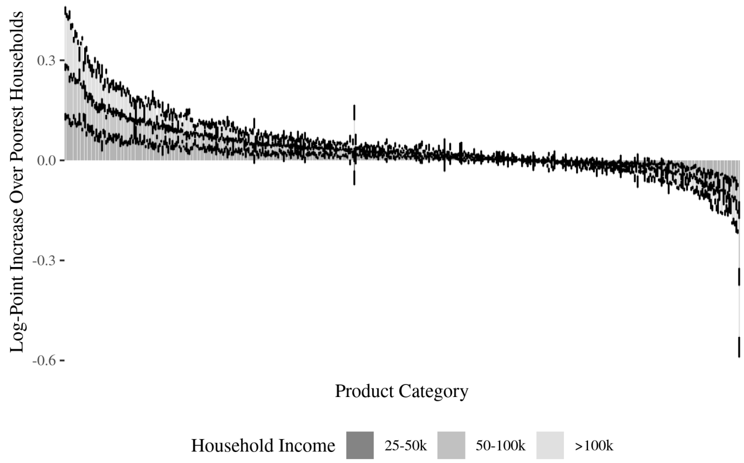
In about 70% of products, higher income households purchase larger packages.

C Unit Pricing History and Regulations

Posting of unit prices started during the consumer protection movement of the late 1960s and early 1970s. Supporters argued that unit prices would allow consumers to make better decisions by facilitating comparisions between brands and sizes of the same product. During this era, several large grocery chains began voluntarily posting unit prices and on January 1, 1971, Massachusetts became the first state to mandate posting of unit prices (Isakson and Maurizi 1973). However, Figure 15 illustrates that there was much heterogeneity in how unit prices were presented and how prominently the unit price was displayed. Research found that unit prices help households identify package shrinkage and avoid quantity surcharges where bigger packages have higher unit prices than smaller ones (Yao and Oppewal 2016). Earlier research found that middle- and high-income households are more likely to use unit pricing information compared to low-income households to lower their expenditures (Isakson and Maurizi 1973). However, most of the research on consumer response to unit pricing

[27]DMAs are non-overlapping groups of counties that define television markets, but they also provide reasonable sub-MSA geographic regions for analysis.

Figure 14: Bulk Purchasing Across Product Categories and Income Groups



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the income coefficients from Equation (14) which regresses the log average package size purchased on household income, size, age, presence of children, and marital status as well as market-year fixed effects. Each bar represents a different product category and only statistically significant (at the 5% leve) are plotted. Values are relative to households making <$25,000 (the reference category). Only categorie sfor which more than 5000 households made a purchase in a given year are plotted.

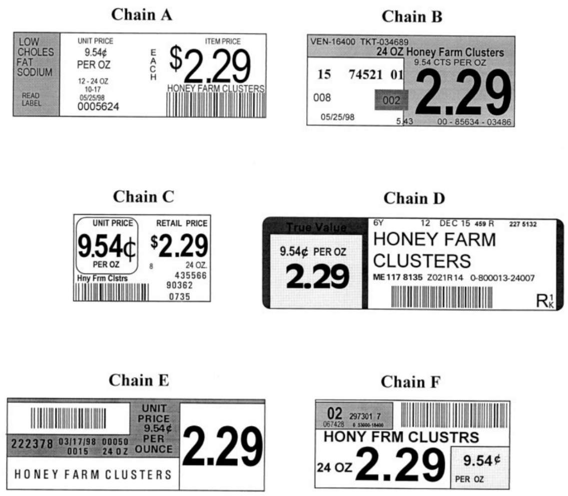
has been done through consumer surveys (i.e. self-reports of awareness and/or use of unit pricing) or limited field experiments at stores in small markets on a small subset of products (Russo 1977). To the best of my knowledge, no analysis has been conducted on a nationally representative sample of consumer purchases like the Nielsen Consumer Panel to determine consumer response to unit pricing.

To promote "uniformity in weights and measures laws," the National Conference on Weights and Measures (supported by the National Institute of Standards and Technology) publishes recommended regulations that states can adopt. One of these regulations, the Uniform Unit Pricing Regulation (UPR), standardizes how unit prices are displayed across all package sizes of products. As of October 2018, eight states have adopted the recommended regulation and 10 states have adopted some form of unit price regulations that are not based on the UPR. The remaining 32 states have no regulations or the UPR is only referenced as a guideline.

States may have varying levels of stringency regarding the display of unit prices. In particular, the UPR as specified only applies to retailers that voluntarily choose to display unit prices. However, if a retailer does display unit prices, the UPR becomes effective and requires the following (exceprts from the UPR are below):

§2 [T]he unit price of a particular commodity in all package sizes ... shall be uniformly and consistently expressed in terms of [weight, volume, count, or area].

Figure 15: Unit Pricing Across Retailers



Notes: This figure illustrates the heterogeneity in unit pricing presentation across a wide range of retailers. Image reproduces Figure 1 from [@miyazaki2000].

§4 (a) The unit price shall be to the nearest cent when a dollar or more

§4 (b) If the unit price is under a dollar, it shall be listed to the tent of a cent or to the whole cent, but not both.

§6 (a) If different brands or package sizes of the same consumer commodity are expressed in more than one unit of measure (e.g. soft drinks are offered for sale in 2 L bottles and 12 fl oz cans), the retailer establishment shall unit price the items consistently.

Within a commodity, unit prices must be specified in the same units (e.g. for drinks, unit prices can be per ounce or per liter, but cannot be expressed in different units for different packages).

An important fact is that the UPR gives retailers discretion over whether or not to display unit prices. However, as of 2015, nine states (Connecticut, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Oregon, Rhode Island, and Vermont) have more stringent rules that mandate the display of unit prices by retailers and then specifying how unit prices are calculated and displayed (National Institute of Standards and Technology 2014). In general, the requirements around how unit prices are calculated and displayed closely mirror those outlined in the UPR, so the main difference is whether retailers can voluntarily display unit prices or if they are mandatated to display unit prices.

D Data Appendix

D.1 Consumer Panel Data

The Nielsen Consumer Panel consists of about 40,000–60,000 US households that provide information on their shopping purchases using in-home scanners or Nielsen's mobile app. Panelists are geographically dispersed and demographically balanced. Households are recruited based on key demographic characteristics, primarily household size, income, age, education, presence of children, race, ethnicity, and occupation. In order to generate national averages, Nielsen assigns each household a projection factor.

Households are recruited through direct mail and online invitations. To incentivize households to remain in the panel, Nielsen provides monthly prize drawings, sweepstakes, points, and regular communication and support to panelists. Nielsen tries to ensure that incentive methods are non-biasing and regularly tests for its correlation with retention rates.

To ensure data quality, Nielsen filters out any households that are poor reporters and do not meet minimum spending thresholds based on their household size. All households in the sample meet this threshold for the full year.

Demographic variables are recorded and updated annually. For my analysis, I collapse some of the demographic variables into more aggregate categories. Household size is defined as the number of individuals residing in the home. Marital status is an indicator for whether the head of household is married or not (I do not distinguish between single, divorced, or widowed). Presence of children is an indicator for whether or not there is at least one child under the age of 18 present in the household. Education is an indicator for whether at least one head of household completed college. Finally, age is the age of the head of household. In the case of two heads, I average the two ages.

To construct my analysis sample, I remove any households where the head of household is a student or a member of the military because these households likely have different living arrangements that are not representative of a typical household's decision (i.e. on campus housing or barracks are different than traditional homes and apartments). I also remove any households making less that $5,000 and those that could not be geocoded based on their ZIP code and the 2017 Census Gazetteer which provides a correspondence between ZIP codes and the latitude and longitude of the centroid of the ZIP code. Finally, some households were dropped because they could not be matched to tract-level car ownership data. Table 7 reports how many households were removed based on this cleaning procedure.

In the purchase data, I exclude alcohol, tobacco, pet items, health and beauty items, general

Table 7: Homescan Sample Construction

Step	HH
Starting HH:	178,232
Exclude military and students:	175,102
Exclude Households under $5k:	174,106
Drop ZIPs Not Geocoded:	173,346
Cannot Be Matched to Car Access:	173,143

Notes: Using 2004–2017 Nielsen Consumer Panel data, this table reports the number of unique households in the sample after each step of data refinement.

merchandise, "magnet," and "deferred" product categories from my analysis. Alcohol and tobacco are excluded because of their addictive qualities, which may induce peculiar purchase patterns. For example, a smoker may choose to only buy one pack of cigarettes with the intention of quitting even though a full carton may deliver a better value. Pet items are excluded to focus on products intended for human consumption. I exclude health and beauty items and general merchandise because these products such as trash cans, printers, eye shadow, and antacids are unlikely to be bought in bulk or have irregular consumption patterns. "Deferred" modules are categories that Nielsen has stopped tracking, so to maintain a consistent sample of products, these are excluded from my analysis. Finally, "magnet" purchases are items which do not have a UPC codes such as fresh fruits and vegetables, deli counter items, or bakery items. Because these items are only recorded for a subset of Nielsen households and are not standardized, I also exclude them from my analysis. This process leaves me with 721 unique product categories. Since this paper focuses on bulk purchases, I also exclude 28 categories that have five or fewer sizes across all products.[28] Overall, the products analyzed are common household staples including almost all food categories, basic toiletry items, and non-food essentials like toilet paper, soaps/detergents, and diapers. See Table 8 for summary statistics of the top 20 product categories by annual spending.

In order to compare sizes across different product categories, I assign each product to its quintile in the size distribution for that product category. I assign quintiles based upon the sample quintiles of product sizes to ensure that each quintile has 20% of available products in its support. An alternative strategy would assign quintiles based on cutting the range of

[28]These excluded categories are: jelled aspic salad, sour cream sauce mix, canned roast beef, canned roast beef hash, retort pouch bags, prepared sandwiches, canned rice, canned dumplings, canned bread, frozen vegetables in pastry, frozen grapefruit juice, frozen grape juice, frozen orange juice, frozen cream substitutes, canned ham patties, bathroom accessory, packaged soap, borateem, dry starch, grease relief, bathroom brushes, miscellaneous brushes, thermometers, dustpans, feather dusters, laundry baskets, sanitary belts, gift package with candy or gum.

Table 8: Summary Statistics of Top 20 Product Categories in Nielsen Homescan Data (2017)

Product	Annual Spending	SD	Avg. Price	SD	Avg. Size	SD
Soft Drinks	79.38	139.02	4.75	4.17	85.87	53.81
Diet Soft Drinks	74.82	132.79	4.73	5.07	84.18	65.06
Milk	65.65	77.02	3.11	1.79	97.79	35.00
Cereal	57.97	68.37	4.06	2.10	18.05	8.17
Toilet Paper	56.15	49.47	11.44	7.09	17.09	10.51
Yogurt	55.00	75.68	3.28	2.17	17.25	15.22
Coffee	53.97	61.69	8.60	5.74	21.84	11.05
Bread	50.03	47.09	2.88	1.52	20.54	4.64
Cookies	46.97	57.60	3.59	3.44	13.02	6.39
Fresh Meat	46.96	62.86	7.75	5.03	30.48	24.97
Frozen Pizza	44.48	60.64	5.99	3.67	20.69	12.48
Bottled Water	44.06	73.46	4.21	3.75	261.91	181.39
Fresh Fruit	42.68	64.91	4.28	2.06	1.93	1.31
Chocolate Candy	41.05	53.83	3.91	3.67	8.64	9.15
Detergent	40.17	45.29	10.05	7.85	99.52	61.23
Shredded Cheese	39.16	42.80	4.21	2.45	13.37	10.98
Bacon	37.63	45.44	6.87	4.67	17.42	11.88
Ice Cream	37.36	50.34	4.43	2.03	46.80	24.47
Potato Chips	35.99	41.71	3.04	1.89	8.87	3.81
Canned Soup	32.39	38.36	3.21	2.22	22.07	17.33

Notes: Using 2004–2017 Nielsen Consumer Panel data, this table reports summary statistics for the top 20 product categories by total spending. Annual spending is the average spending in that product category among households that purchased in that product category over the course of the year. Average price and average size are the average prices and sizes of products purchased in their corresponding category. All estimates are weighted using Nielsen's projection weights. Prices are in nominal 2017 dollars. Sizes are reported in common units for for that category (e.g. ounces for milk).

product sizes into equal intervals. However, in some product categories, this risks generating quintiles with sparse support when there is an especially large package available. As an example, consider eggs. Most packages contain 6, 12, or 18 eggs, but there are some products that offer up to 15-dozen eggs (180 eggs). Generating quintiles by cutting the available range into equal intervals would generate quintiles of 1-36, 37-72, 73-108, 109-144, 145-180 which would assign almost all packages to the first quintile and the fifth quintile. Using the sample quintiles generates a more even distribution ensuring better support of each quintile. For products with a narrow range of sizes, I assign the product to the minimum quintile. For example, over 60% of egg products are dozens, which covers three quintiles. I assign all products with 12 or fewer eggs to the first quintile.

E Implications for Inequality

Since bulk discounts are disproportionately taken up by high-income households, consumption inequality is persistently understated. In this section, I quantify the extent of this underestimation.

I quantify consumption inequality in terms of unit prices since this is a more natural way of capturing bulk discounts. This approach abstracts away from consumption inequality attributable to differences in quantity consumed. For normal goods, adding this dimension will serve to further increase consumption inequality since high income households will consume larger quantities of the goods, compounding the inequality generated by lower unit prices. Unit prices are a conceptually simple way to approach consumption inequality. In the absence of bulk discounting, unit price differences would logically be attributed to differences in quality, such as brand or store amenities.

However, bulk discounts complicate this measurement. Consider the following example: a homogenous product has a bulk discount. If some households buy large packages with low unit prices ("bulk" buyers) and some buy small packages with high unit prices ("budget" buyers), then the bulk discount would generate a difference in the underlying unit price. Even more troublingly, the inequality would be of the wrong sign. If one did not account for the bulk discount and only assumed unit price differences reflected differences in quality, then one would conclude that the "budget" buyers were better off than the "bulk" buyers, which is untrue since the "bulk" buyers obtain the same good more cheaply.

In order to quantify how much consumption inequality is underestimated, I decompose the unit price of each product category into store, brand, and size components using the following regression (estimated separately for each product category and year):

$$\log(unitPrice)_{ibr} = \beta \log(pkgSize)_{ibr} + \lambda_b + \lambda_r + \epsilon_{ibr}, \tag{15}$$

where *unitPrice* is the unit price of a product purchased by household i of brand b at retailer r. *pkgSize* is the quantity contained within the package. Brand and retailer components are captured by λ. All coefficients are relative to the reference category which is the generic brand product sold at the most popular discount retailer.

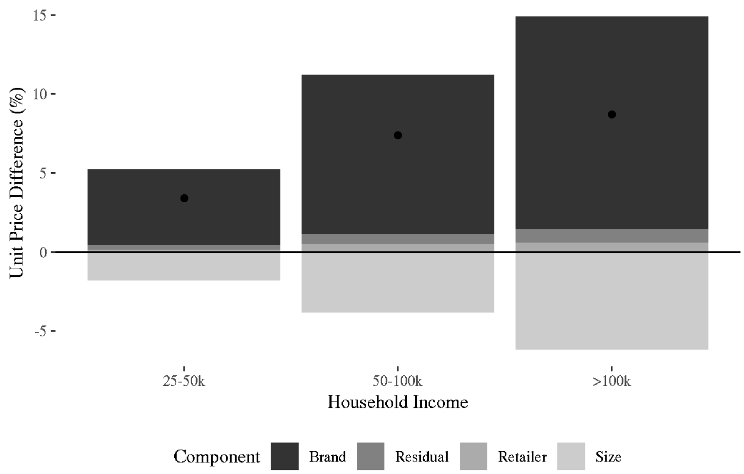
Using the estimates from Equation (15), I decompose each purchase into a brand, retailer, and size component. I then aggregate these components to the household level by taking the expenditure-weighted average of the components across all purchases made by a particular household. I then regress each of these components on household income and other demographics (household size, age, marital status and presence of children) to determine which components contribute to the difference in unit prices paid by different income groups.

$$Y_{it} = \beta Income_{it} + \gamma X_{it} + \lambda_t + \epsilon_{it}, \tag{16}$$

where Y is the average brand/retailer/size component for household i's purchases made in year t. X consists of household size, age, marital status, and presence of children. Year fixed effects are captured by λ. The regression is weighted using Nielsen's projection weights. Figure 16 plots the income group coefficients for each component on all toilet paper purchases. First, the plotted points show that the average unit price paid increases with household income. The decomposition shows that this increase is overwhelmingly due to brand composition with a small, but significantly positive contribution from retailer composition. Excluding a residual component, households making over \$100,000 would pay 14% higher unit prices based on brand and store choices (13% from brand choice and 1% from store choice). Bulk discounts are estimated to reduce unit costs by six percentage points. As a result, consumption inequality in this category would be understated by 43% (unit prices for high-income households would be 8% compared to the "true" difference of 14%). Due to a residual factor, the measured difference in unit prices is 9%, so the "true" difference is understated by a still-substantial 36%.

Bulk discounts have large implications for price measurement. Most macro-level research has focused on differences in aggregate spending between households or between various product categories (Krueger and Perri 2006; Heathcote, Perri, and Violante 2010; Attanasio and Pistaferri 2014; Aguiar and Bils 2015). Under the assumption that homogeneous goods have the same prices, increased expenditures directly correspond to increased consumption.

Figure 16: Decomposition of Unit Price Differences



Notes: Using 2004–2017 Nielsen Consumer Panel data, this figure plots the contribution of brand, store, and package size choices to the differences in unit prices paid for toilet paper for different income groups relative to households making less than $25,000. Dots denote difference in average unit price paid by each income group relative to households making under $25,000. See Appendix Table 9 for values.

Table 9: Unit Price Decomposition (Toilet Paper)

	Unit Price	Brand	Retailer	Size	Residual
	(1)	(2)	(3)	(4)	(5)
25-50k	0.034***	0.048***	0.001***	−0.018***	0.003***
	(0.001)	(0.001)	(0.0002)	(0.0004)	(0.001)
50-100k	0.074***	0.101***	0.005***	−0.038***	0.006***
	(0.001)	(0.001)	(0.0002)	(0.0004)	(0.001)
>100k	0.087***	0.135***	0.006***	−0.062***	0.009***
	(0.002)	(0.001)	(0.0002)	(0.0005)	(0.001)
Observations	710,647	710,647	710,647	710,647	710,647
Adjusted R^2	0.031	0.212	0.064	0.684	0.001

Note: *p<0.1; **p<0.05; ***p<0.01

Using 2004–2017 Nielsen Consumer Panel data, this table reports the contribution of brand, retailer, and package size to the average log unit price obtained by each household income group. Log unit prices were decomposed using Equation (15). Each household's component composition is the expenditure-weighted average of each component applied to each of its purchases. Household-level components were then regressed on household income, size, age, marital status, and presence of children. Income coefficients are reported above.

However, as Kaplan and Menzio (2015) show, prices for homogeneous goods are not the same and this non-uniformity has severe consequences for measurement. Most recently, Coibion, Gorodnichenko, and Koustas (2017) prove that consumption inequality is overstated due to bulk buying because households make fewer shopping trips and therefore measured price dispersion increases even if underlying consumption does not.[29]

Using hedonic regression, I find that unit prices are underestimated by about 10% due to bulk discounts, but this underestimation varies widely across product categories. See Appendix for analysis.

Previous research has documented that this "spikiness" has strong implications for measuring consumption inequality (Coibion, Gorodnichenko, and Koustas 2017). Furthermore, bulk discounts could have implications for inflation measurement as well. Given the prevalence of bulk discounts, if a basket price is computed using the largest package available, measured prices may underestimate the price faced by consumers that cannot afford that package, or buy the same amount in smaller shopping trips. For example, the BLS uses unit-prices when computing the price of the CPI market basket. Therefore, the choice of package size can strongly influence this price. If the prices of different package sizes fluctuate differently, bulk discounts could systematically affect our measurement of inflation. Further research would be necessary to determine if inflation rates differ across product sizes is the case and what the magnitude could be.[30]

[29]This dispersion results because the Diary portion of the Consumer Expenditure Survey only records the past 2 weeks of purchases. Bulk buying allows some households to make purchases during those 2 weeks and others will not have to purchase during those 2 weeks.

[30]The BLS specifically deals with multi-unit discounts buy using the "first multiple-unit price" as the price measured. However, the method for dealing with different package sizes is less well defined. Per correspondence with the BLS: "In general, price-per-unit calculations are the standard practice in calculating items with quantity changes. For example if a loaf of bread is $1 and you a second loaf for $0.50, you have two loaves for $1.50, and then depending on the weight of the two loaves, you divide the $1.50 by number of ounces and get a price per ounce. Due to sample rotation, there is no guarantee that any specific item and specific size – toilet paper, olive oil, etc. – will be in the sample at a given time. If it is, and the economic assistant can't find the specified size and has to substitute to a different size, price change is calculated on a per unit basis, as illustrated in the above example." Given that the BLS uses unit-prices to compute market basket prices, the choice of package size could strongly influence the computed price of the market basket.