The Pink Tax: (Why) Do Women Pay More?*

Kayleigh Barnes† Jakob Brounstein‡

January 16, 2024

Abstract

We evaluate the existence of the pink tax: the hypothesized price premium on women's consumer goods. Using scanner data, we find that women pay 4% more for consumer packaged goods in the same product-by-location market. This price differential is generated by women paying 17% higher average prices for explicitly gendered products, like personal care items, and 3.8% higher average prices for ungendered products. We estimate demand differences by gender, structurally decomposing price differences into markups and marginal costs. We find that the pink tax is not sustained by higher markups, but by women sorting into goods with higher marginal costs.

*We would like to thank Ben Handel, Jon Kolstad, Ben Faber and Nano Barahona for their mentorship and guidance. We thank Matt Backus, Mathilde Bombardini, Nina Buchmann, Bryan Chu, Giovanni Compiani, Stefano DellaVigna, Thibault Fally, Jonathan Holmes, Xavier Jaravel, Leticia Juarez, Kei Kawai, Peter Levell, Sarah Moshary, Mathilde Muñoz, Nina Rousille, and Damian Vergara, as well as participants in the UC Berkeley IO and Trade lunch seminars and University of Washington PHEnOM seminar for helpful comments. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Federal Reserve Board of Governors; email: kayleigh.n.barnes@frb.gov

‡Institute for Fiscal Studies; email: jakob.brounstein@gmail.com

1 Introduction

Is it more expensive to be a woman? The notion that there exists a price premium on women's consumer goods relative to those of men is colloquially referred to as the "pink tax". The concept has received considerable attention in popular media and has spurred recent legislation in US states to prohibit gender-differential pricing of goods and services. Existing studies of the pink tax find mixed evidence of its scope and magnitude, but typically tend to focus on a narrow set of goods (Moshary, Tuchman, and Vajravelu 2023; Guittar et al. 2022; NYCDCA 2015; Duesterhaus et al. 2011; Manzano-Antón, Martinez-Navarro, and Gavilan-Bouzas 2018).

This paper explores the existence and underlying mechanisms of the pink tax by describing consumption baskets for men and women, analyzing how they vary by quantity, price, and diversity of products consumed. We then decompose observed price differences into markups and marginal costs. We consider a broad definition of the pink tax[1], considering any channel through which women may face higher markups in the retail consumer packaged goods (CPG) space. This definition allows us to capture the role of differential sorting between men and women and second degree price discrimination, or versioning, in generating the pink tax. We find that, averaged across the entire grocery consumption basket, women pay 4% higher per unit prices than do men for products in the same product-by-location market. We find that this price difference is sustained not just by purchases of gendered products, like men's and women's razors, but also by differences in purchasing habits between men and women for food and household items. This finding could be driven by three economic mechanisms that determine pricing: (i) women could exhibit less elasticity of demand than men, (ii) women could consume products with greater market power or from less competitive markets than men, or (iii) women could consume products with higher marginal costs. Disentangling the mechanisms driving an observed price premium on women's products is important to inform economic understanding and policy alternatives.

To characterize the pink tax and gender differences in consumption habits, we employ several data sets that contain detailed information on individuals and their purchases, store-level product offerings, and retail prices. The Nielsen Consumer Panel Survey features a

15-year rotating panel of households and the near-universe of their purchases at big box retailers and grocery stores. Importantly, the data includes rich household demographic information as well as highly detailed product and purchase characteristics—including deal or sale usage, prices paid and quantities consumed, and a hierarchy that aggregates products into tractable market definitions. By restricting the bulk of our analysis to single-member households, we are able to attribute each purchase made to a specific gender. We augment the Consumer Panel with the Nielsen Retailer Scanner data which contains store level data on prices and quantities sold in any given week.

We begin by establishing the existence of systematic gender differences in consumption and pricing along two margins: consumer behavior and the product space. To document consumer behavior, we describe consumption bundles for men and women, documenting differences in their unit prices and composition. We find that women spend about 6% more annually than men do on retail CPG consumption and that their consumption bundles are larger and more diverse. The products that women purchase are on average 4% more expensive per unit than those purchased by men in the same product-by-location market. In the product space, we document a significant share of products that are exclusively bought by one gender, with the majority of these products gendered towards women. These products are particularly common in markets with explicit gender differentiation in marketing and product design, such as in beauty and personal care goods. We categorize products bought at least 90% of the time by one gender as "gendered" products, categorizing all other products as "ungendered" (with alternate cutoffs demonstrating the robustness of our results). We then decompose the average 4% price premium paid by women into a contribution from differential sorting into ungendered products and from purchases of explicitly gendered products, finding that women pay an average of 3.8% higher prices on *ungendered* products relative to men and that women pay an average of 17% higher prices on gendered products relative to those

[1]Heuristically, we identify three scenarios through which the pink tax may operate: 1) different prices for goods with identical inputs: e.g. without changing anything else, by coloring a product pink, retailers and producers can charge a higher price; 2) different prices for goods with identical uses but non-identical inputs: i.e. the price difference between goods purchased by men or women is attributable to differences in the cost of production; 3) expense differences driven by goods that are almost exclusively purchased by a single gender: e.g. the purchase of makeup or feminine hygiene products. In some instances, the pink tax refers to the luxury, sales, or value added taxes statutorily placed on women's hygienic products. Our analysis focuses on the more general case of price differences between men's and women's consumer goods.

of men. While gendered items carry large price premiums for women, they make up a small share of actual purchases; the bulk of the price premium is driven by women buying more expensive ungendered items than men.

We then turn our attention to understanding the demand and supply mechanisms that give rise to women paying higher prices. Profit maximizing firms set prices as a function of own-price elasticities, market shares, cross-price elasticities of products owned by the same parent company, and marginal costs. To assess the relative importance of these different channels, we model demand and supply, attributing differences in pricing and product choice to markups and marginal costs. We begin by estimating demand elasticity differences between men and women across the entire retail grocery consumption basket. We develop a simple, tractable model assuming constant elasticity of substitution that allows us to estimate demand by gender in the aggregate population. We aggregate individual-level purchase data to the gender-product-location-market level and we find that, on average, women consume products more elastically than do men. This finding is implies that women are charged lower markups rather than higher markups, on average, under price discrimination.

We corroborate this central finding by implementing several additional designs that leverage complementary data and identification techniques. We combine the scanner data with data on wholesaler prices paid by retailers from PriceTrak. Wholesale prices represent the cost of the product charged to the retailer. We construct retailer markups and observe that conditioning on wholesaler costs largely eliminates the observed pink tax; we also find no significant difference in retailer markups paid by men and women or along the product-gender spectrum.

Finally, we estimate overall markups and marginal costs of production for disposable razors and yogurt using a differentiated products demand model (Berry, Levinsohn, and Pakes (1995)). We incorporate product gender as a characteristic over which consumers can have heterogeneous tastes and find that products disproportionately consumed by women are associated with higher marginal costs of production and lower markups.

We find that women *do* pay higher prices than do men for similar goods, but that the pink tax is not driven by price discrimination, but rather marginal costs. Current legislation is largely focused on banning price differences for products that differ only in gendered marketing. Our paper suggests that these laws are likely to be ineffective at addressing price disparities between men and women, as the majority of our observed pink tax can be explained by men and women sorting into products that differ by more than just gender.[2] Our findings have important implications for other policy relevant issues, like potential disparities in the incidence of inflation between men and women.

In spite of its prevalence in popular discourse and policy, there are few studies that rigorously substantiate the pink tax. Much of the direct evidence on the pink tax comes from government reports or academic articles that consider either a limited set of products that gender matched in a subjective manner and document differences in list prices rather than actual prices paid. (NYCDCA 2015; Duesterhaus et al. 2011; Manzano-Antón, Martinez-Navarro, and Gavilan-Bouzas 2018; Manatis-Lornell et al. 2019) These studies find that women's goods have about a 5-7% price premium but do not attempt to allocate this price difference to differences in markups or differences in the cost of production. Recently, Moshary, Tuchman, and Vajravelu (2023) assess the prevalence of the pink tax for personal care items, improving on prior studies by directly studying controlling for brands and ingredients as a proxy for marginal costs. They find that, when comparing nearly identical products there does not exist a price premium on women's products. We explicitly study differences in the prices, markups and marginal costs of the entire range of retail goods that are bought by men and women, capturing the role of men and women sorting into different products in generating the pink tax.

Our paper contributes to the literature on gender disparities, consumption inequality and price discrimination. There is a large literature on the gender wage gap and its implications. (Blau and Kahn 2017) Taking into account differences in the cost of consumption prompts us to re-frame the widely-studied difference in wages between men and women as a *nominal* wage gap. Moretti (2013) has shown that population specific price indices have important implications for wage inequality in real terms. In line with this, the presence of an aggregate

[2]The state of New York has banned pricing on the basis of gender through bill S2679 which took effect in 2020. A similar bill, AB 1287, was signed into law in California by Governor Gavin Newsom on Sept. 27, 2022. The pink tax Repeal Act has been presented in Congress four times and aims to put national law in place similar to the New York and California policy.

pink tax on women's consumption augments these wage inequalities by reducing women's purchasing power.

Our work is closely related to research on inequality in consumption and product offerings. The consumption literature has documented that inflation, price indices and product offerings exacerbate inequality between rich and poor households (Jaravel 2019; Argente and Lee 2017; Faber and Fally 2022). Our work on gender explores a new angle through which price index inequality may shape wealth inequality at large and our findings suggest that women may experience inflation and product innovation in different ways from men. Our finding that women and men sort into inherently different products suggests that their preferences are systematically different, which may be a result of differences in social norms or market experience. Bronnenberg et al. (2015) demonstrate that market and professional experience affect product choice, where pharmacists and chefs less frequently purchase more expensive brand name items (as opposed to generic-brand equivalents) in CPGs.

Finally, we contribute to the literature on price discrimination and optimal pricing. The work on gender based price discrimination focuses on first degree discrimination in bargaining contexts, finding that women often pay higher markups (Ayres and Siegelman 1995; Goldberg 1996; Trégouët 2015; Castillo et al. 2013). Our work investigates the existence of second degree price discrimination (also known as *versioning*) against women in CPG markets. Product differentiation and second degree price discrimination are sometimes thought of as separate phenomenons but Stigler (1987) defines price discrimination as any markup difference between consumers groups. There is precedent for price discrimination among CPG retailers: Hendel and Nevo (2013) finds that grocery store chains utilize promotional sales as a way to intertemporally price discriminate against consumers. However, other work finds that retail chains do not necessarily engage in optimal pricing decisions: DellaVigna and Gentzkow (2019) find substantial price mis-optimization for retail chains, where stores typically implement uniform prices across locations irrespective of local demand and cost factors. Our work examines how optimal pricing of differentiated products could be a form of second degree price discrimination with certain consumer groups being charged systematically higher markups.

2 Data

We combine data from three main sources to conduct our analysis.[3] Our main analyses rely on data from NielsenIQ including the HomeScan Panel (HMS) and the Retailer Scanner Data (RMS). The HMS data contains purchase histories of for a rotating panel of households from 2004 to 2019. In brief, we use the HMS to assign gender to products (detailed in Section 3.2) and to directly study differences in consumption patterns by gender. The RMS data contains anonymized purchases of products aggregated to the UPC-store-week level from 2004 to 2018. We use the RMS to more accurately observe product prices and study differential demand sensitivity along the UPC-gender measures we construct with the HMS data. Lastly, we incorporate data from National Promotion Reports' PriceTrak database (PromoData), which features data on wholesaler prices charged to retailers for certain products from 2006-2011. While we discuss these data in turn, see Bronnenberg et al. (2015) and Allcott, Lockwood, and Taubinsky (2019) for further discussion of the NielsenIQ data.

The entire HMS features data on the shopping trips and transactions of approximately 60k households per year. Households remain in the panel for on average 54 months, with approximately 200,000 distinct households rotating through the HMS in total. The data report purchases made by households on the 20 million shopping trips from 2004 to 2019 made by the panelists. For each individual item purchase, we observe transaction metadata such as date, store/retailer-info, and panelist identifier, as well as granular data on product and transaction details including prices paid, amounts and units of quantities purchased, deal or sale usage, and detailed nests of product identifiers.

We primarily use the HMS data to document differences in the purchasing behavior of men and women and understand how product markets differ for men and women. To confidently assign product purchases to consumer gender demographics, we restrict our consumer panel to single-individual households that report at least 12 shopping trips per year, which eliminates approximately 75% of the panelists in the HMS. This leaves us with a panel of 47,012 households which we use to study differences in consumer behavior. We report sum-

[3] We also supplement the NielsenIQ data with the Consumer Expenditure Survey public use micro data (CE PUMD) to document descriptive evidence of differences in consumption spending across the entire consumption basket in the Appendix.

mary statistics for the sample in Table A.1. Our final sample is skewed toward women, with about 70% of our panelists identifying as a woman. In terms of balance, the men in our sample tend to have higher income[4] and be more educated, which we will control for in the analysis.[5] The second component of our analysis focuses on how the product market space varies by gender. For this analysis, we restrict our data to products to which we can confidently assign a UPC gender. We describe our methodology in detail in Section 3.2. The NielsenIQ data covers approximately 1.8 million products. We are able to confidently assign gender to 700,000 UPCs that comprise 97% of the purchases made in our singles panel.

The RMS data contain UPC-store-week level prices and volumes of products purchased by consumers from 2004 to 2018. This dataset is not tied to the consumer identifiers; rather, the strength of the RMS data lies in its relative comprehensiveness of US sales. We use the RMS data to model demand in select markets that have a high level of gendered products (as identified in the HMS data).[6] While the HMS tracks all retail purchases for a household from any store, the RMS contains a select set of stores. For our main analysis, we keep only stores that are part of a larger retail chain rather than independent stores.

Both components of the NielsenIQ data feature a highly detailed product hierarchy classification that organizes all goods into smaller nests with increasing degrees of specificity. Products in the NielsenIQ are identified with their Universal Product Code (UPC) which corresponds to a unique barcode. All UPCs fit into one of ten *departments* (the broadest category, e.g. "Health and Beauty" and "Dry Grocery"). From here, products in a department are allocated to *product groups*—of which there are 120 total—such as "Shaving Needs". Finally, UPCs in the same Product Group are assigned to *product modules*—the

[4]The HMS reports panelist household income in discrete buckets. All results referring to HMS panelist income make use of the midpoint of each discrete income buckets used for the household income field.

[5]There is considerable discussion on the representativeness of the HMS panel. Bronnenberg et al. (2015) summarize this discussion that argues in favor of the representativeness of the panel of US consumers. While applying the included HMS projection weights render the sample much more representative of the US, the raw using-sample departs significantly from basic US demographics. Our sample skews significantly more female than male, by a ratio of 3:1, and the in-sample median age of 53 is significantly older than the US median age of 38. The panelist's income demographics appear slightly more representative, with the median single-individual household earning approximately $37,000 USD per year and the median household, unconditional, reporting approximately $55,000 USD . Nonetheless, applying the projection weights yields demographics that much more closely align with those of US consumers. We therefore always use the Nielsen projection weights in our analyses.

[6]Additionally, the prices paid as observed in the HMS are constructed using prices recorded in the RMS.

most granular grouping of multiple products—e.g. "Disposable Razors". The Nielsen data identifies over 1300 distinct product modules. Brand description represents an alternate grouping that features the brand name for a given set of UPCs, not strictly contained in any single product module or group contained, such as "Venus" (a division of razors marketed to women by Gillette), for the brand of razors. We consider product modules as constituting a self-contained goods market; for certain reduced-form analyses, we further divide product modules into module-unit groups (modules composed of goods all with the same counting units: e.g. the coffee product module contains bagged coffee measured in weight (ounces) and Keurig cup coffee measured as a count (number of K-cups)).

Lastly, the PriceTrak PromoData data allow us to validate retailer markups relative to wholesaler prices. While this data does not feature information on production costs, it does provide information on intermediary costs to retailers (i.e. distributor prices). The PriceTrak data features retailer cost-data of individual UPCs for a variety of time- and geographic-denominations from 2006 and 2011, with geographic disaggregations covering 55 markets (coinciding with the metropolitan areas around large US cities). The match rate of UPCs in the Promodata to the NielsenIQ datasets is relatively low. Only about 10% of the 430,000 distinct UPCs in the RMS data matching to PromoData; however, these UPCs account for 40% of purchase volume observed in the HMS. We combine the data from PriceTrak on wholesaler prices with Nielsen data on post-deal consumer prices to compute retailer markups relative to wholesaler prices.

3 Price Disparities Across the Consumption Bundle

3.1 Consumer Behavior by Gender

We begin by analyzing how women's and men's annual retail CPG consumption baskets differ.[7] We find that women's retail consumption baskets are larger, more expensive, and

[7]We use the CE PUMD to analyze differences in full consumption baskets in Figure A.1 by plotting women's yearly spending as a percentage of men's. We find no significant differences in total spending by gender but do find that women spend significantly more of their income on housing, clothing, health and personal care, while men spend relatively more on food, alcohol and cigarettes, and transportation. The

filled with a greater number of unique UPCs. Figure 1 plots levels of female activity as a proportion of male activity for annual spending, unique product purchases, and total product purchases controlling for demographic factors such as year, county, income, age, race and education. We find that women's yearly spending is greater than that of men by about 6%, their product diversity is greater than men's by about 27% and their consumption baskets are larger than men's in terms of items purchased by about 9%.[8] This pattern is primarily driven by differences in behavior in consumption of Health and Beauty products, where women spend 51% more than men, consume 53% more unique products, and consume 49% more items. However, we observe similar results for all products after excluding Health and Beauty; such spending categories include are food grocery products, household products and alcohol. Among these products women spend about 2% more, have 25% greater product diversity and 7% more items than men.

Our documented total spending differences in Figure 1 could arise from differences in prices paid for similar goods or from differences in quantities purchased. As a clarifying example, consider consumption habits for shampoo. Women tend to have longer hair than men, which may lead them to buy more bottles of shampoo over the course of a year. We conceptualize this occurrence as driving up total spending on an *extensive* margin, that is, buying more product. It is also possible that women have preferences for higher-priced shampoos, we refer to this occurrence as the *intensive* margin, where women pay higher per unit prices. Figure 1 indicates that the "extensive" margin is an important contributor to overall differences in spending. While total items purchased captures the differences both in the intensity and variety of products purchased, information on unique products captures only this latter element, and could be driven by both greater taste for variety by women within shared-gender product spaces as well as a greater volume of products typically intended for exclusive consumption by women (e.g. feminine hygiene products, medication and beauty products).

finding that men spend more on food suggests that men are more likely to substitute food expenditure to eating out than women.

[8]We compare differences in yearly spending subsequently adding in controls in Table A.5. We find that the raw price gap is about 1.6%, while the gap between demographically similar men and women is about 6.8%.

Popular discussion of the pink tax is often focused on differences in prices paid between men and women, i.e. the intensive margin contribution to the overall spending gap. We compare per unit prices paid by men and women for products in the same market with the following specification:

$$log(P_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}, \tag{1}$$

where i denotes the individual, j denotes the product purchased and t denotes the market. Table 1 Panel A presents the results. Column (1) regresses log unit UPC price on a woman indicator and includes fixed effects for the interaction of product module, units the good is sold in and the year of purchase. One can think of the 2.3% result as the raw difference in prices paid between single men and women in the USA, not accounting for other demographic factors or location and retail chain sorting. Column (2) runs the same specification but adds in controls for age, income, and race. The increase in the coefficient, from 2.3% to 4.67% highlights the importance of demographic differences between single men and single women because older and lower income people tend to buy lower priced products. Columns (3) and (4) subsequently add in county and retailer fixed effects. Column (3) can be interpreted as the contribution of women sorting into more or less expensive locations, because the coefficient change is small, the contribution is minimal. Similarly, Column (4) can be thought of as the contribution of sorting into more or less expensive retail chains, i.e. Whole Foods vs. Walmart. Controlling for the retail chain lowers our price premium estimate to 4.19%, suggesting that retail chain sorting plays a small but significant role. Finally, in Column (6) we add in fixed effects for month rather than year. The results indicate that women spend about 4.02% more than do men per unit of goods in the same product market, bought in the same retail chain, county, and month. We consider this our preferred specification because it controls for a wide variety of potential differences that could arise between the two groups other than gender.[9] We refer to this 4% finding as our observed pink tax on the intensive margin.

[9]We estimate our preferred specification for each department in Table A.10. We find that the only departments in which men pay higher per unit prices than do women are Alcohol and General Merchandise.

Table 1 Panel B estimates Equation (1) while including product-level fixed effects instead of module-level fixed effects. The interpretation of the coefficient becomes the difference in prices paid between men and women for the same exact product. Differences in prices paid for the same good can be attributed to differences in price shopping behavior, like coupon usage and sale shopping, consistent with being a more elastic consumer. We sequentially add in fixed effects in the same manner as Table 1 Panel A, so the coefficients can be interpreted as a raw difference between men and women in column (1) and then iteratively making comparisons between demographics, location, retail chain and month. While we find that women, on average, buy more expensive products than do men, we find that they consistently spend less than men on the *same* product. In column (2) we find that controlling for demographics attenuates this gap, likely driven by differences in use of coupons by age and income. Controlling for retail chain in Column (4) increases the gap, which is consistent with women sorting into higher price chains. Column (6) shows that women pay 0.8% less for the same product than do men that are demographically similar shopping in the same retail chain-location-month market. Combining our results from Panels A and B suggests that women are buying higher-priced goods while also exhibiting behaviors associated with being more elastic consumers. Hendel and Nevo (2013) study promotional sales as a form of intertemporal price discrimination, our results would indicate that women are likely to comprise a larger share of the consumer base that benefits from this type of price discrimination (also substantiating previous related findings that women engage in price shopping to a greater degree than do men, e.g. Aguiar and Hurst (2005)).

3.2 Gender in the Product Space

We now shift our focus from consumer behavior to understanding how the the product space varies by gender. Our descriptive evidence above shows that women buy more expensive and larger consumption bundles and that the products they buy are more expensive relative to similar products bought by men. However, these observations could be driven by differences in purchase intensity of otherwise "ungendered" products or by purchases of products that are exclusively bought one gender. To fully characterize the pink tax, we document the existence of goods that are gendered, that is they are only ever bought by one gender, and decompose our observed pink tax of 4% into into its respective contributions from gendered products and differential purchasing of ungendered products.

First, we assign values of gender-stratification to each UPC. We begin by calculating a woman purchase share for each UPC in our data as the fraction of overall purchase volume by women within our panel of single individuals. We define the estimator for the time-invariant woman purchase share of UPC j (the "UPC-gender") as

$$\hat{w}_j = \frac{\sum_{i \in \mathcal{I}} Purchase_{ij} \mathbb{1}\{woman_i = 1\}}{\sum_{i \in \mathcal{I}} purchase_{ij}}$$

This fraction assigns $\hat{w}_j \in [0, 1]$ where 0 denotes a good that is only bought by men and 1 denotes a good that is exclusively bought by women. We assign goods with $\hat{w}_j \leq .1$ as men's products, and those with $\hat{w}_j \geq .9$ as women's products.[10]

To reduce measurement error in our measure of UPC-gender, we only assign an observed women purchase share to products that are observed to be bought with sufficient frequency.[11] Conceptually, each UPC in our data has a *true* UPC-gender, w_j, that we do not observe. We observe an empirical UPC-gender $\hat{w}_j$ as well as the UPC's number of unique purchasers, n_j. An observed UPC-gender represents a draw from a binomial distribution; the probability that the value lies more than .05 percentage points away from its true value is:

$$P(w_j \notin (\hat{w}_j - .05, \hat{w}_j + .05)) = \int\limits_{x \geq |\hat{w}_j \pm .05|} \binom{n_j}{\lceil \hat{w}_j n_j \rceil} x^{\lceil \hat{w}_j n_j \rceil} (1-x)^{n_j - \lceil \hat{w}_j n_j \rceil} f(x) dx,$$

where $f(x)$ is the empirical pdf of woman purchase share. We calculate thresholds for discrete bins of UPC-genders of radius 0.00025 from 0 to 1, $n^*_{jb(\hat{w}_j)}$, such that the probability

[10] For robustness, we provide an alternate set of results for our analysis of UPC-gender that implements a less restrictive UPC-gender cutoff of .25 and .75 in Section A.1

[11] To illustrate the necessity of this decision, approximately two-thirds of the UPCs purchased by Nielsen panelists are only ever observed to be purchased once (although these UPCs represent less than one percent of the overall purchase volume reported in the HMS); by merit of only seeing a single purchase, these UPCs would always be assigned to having an explicit gender of 0 or 1.

that an individual UPC's observed UPC-gender deviates from its true UPC-gender by a value less than 0.05 is 95%. We map each observed UPC-gender to its nearest bin and only keep UPC-gender observations when the underlying number of unique purchasers n_j exceeds $n^*_{jb(\hat{w}_j)}$.[12]

Figure 2 plots the distribution of woman purchase share for all products, Health and Beauty products, and all products excluding Health and Beauty. We observe significant excess mass at the right tail of the distribution where goods are bought exclusively by women, but only mild excess mass at the left tail of the distribution where goods are bought only by men.

We now describe how prices vary along our measure of UPC-gender. We map each UPC-gender $\hat{w}_j$ to a ten-percentile bin $b = 10 \cdot (\lfloor 10 \cdot (\hat{w}_j + .05) \rfloor)$ and estimate the following regression:

$$log(P_{jt}) = \phi_{t(j)} + \sum_{b \in \mathcal{B}} \beta_b \mathbf{1}_{d(j)=b} + \epsilon_{jt}.$$

Figure 3 plots the coefficients from estimating this equation, taking the 50th percentile bin as the reference point. The regression contains fixed effects for the product module of the UPC, county and half-year of purchase. The coefficients can be interpreted as averages across comparisons made of products in the same market and bought in the same location and time frame relative to products bought equally by men and women.

We observe a significant price premiums of between $10 - 40\%$ for goods purchased exclusively by either women relative to similar goods purchased at gender parity. We observe no outsized price for goods purchased purely by men. However, beyond the tails of the graph, a striking pattern emerges: prices tend to monotonically increase in woman purchase share. This increase in prices along woman purchase share suggests that our overall price premium

[12]Figure A.2 displays the gender-composition of UPC by each Nielsen department. First, we find that the majority of UPCs are unassigned because their unique purchase count falls under its exclusion threshold. The median UPC in our sample is purchased by 4 unique individuals and 63% of UPCs are purchased by less than 8 individuals. In our sample, we observe 1.8 million UPCs across 155 million purchases. While we are only able to confidently assign UPC-gender to 700,000 unique products, Figure A.3 shows those we are able to assign gender to account for greater than 95% of all purchases made in the data by expense. We find that gendered products make up a small share of purchases, 3.6% for men and 4.6% for women. Within Health and Beauty products though, gendered products make up 20% of women's purchases and 10% of men's purchases.

of 4% from Table 1 is likely explained not just by explicitly gendered products (i.e. pink products and blue products) but also by differences in preferences for non-overtly gendered products. This finding is consistent with women having preferences for higher (perceived) quality items like, for example, organic products (Ureña, Bernabéu, and Olmeda (2008)).

To explore the interaction of consumer and UPC-gender, we run the same regression specification as in Table 1 column (5), but now include an indicator for whether a good is gendered and an interaction between the woman indicator and the product gender indicator:

$$log(P_{ijt}) = \phi_{t(j)} + \beta_1 \mathbf{1}_{w(i)} + \beta_2 \mathbf{1}_{g(j)} + \beta_3 \mathbf{1}_{w(i)} \cdot \mathbf{1}_{g(j)} + \gamma X_i + \epsilon_{ijt}.$$

Table 2 presents the results of these regressions. We find that women pay a price premium of 3.83% on ungendered products relative to ungendered products bought by men. Across all departments, men pay lower prices on gendered products than they do ungendered products by about 1%. The interaction coefficient shows that women pay about 12% more for gendered goods relative to ungendered goods. Overall, we find that women pay approximately 17% higher prices on gendered products than do men. Columns (2) and (3) of Table 2 demonstrate that these findings are not driven by heath and beauty products.

However, while the magnitude of the price difference for gendered goods purchased by women is large, its contribution to the overall price premium is small. Figure A.3 indicates that gendered products make up an overall small share of a consumption bundle. Indeed, the 4% price premium from Table 1 closely aligns with the purchase-weighted price averages that women pay on ungendered items. While we find evidence that female-gendered products see significantly higher unit prices, the vast majority of our observed pink tax is in fact driven by differential sorting between men and women on less-overtly-gendered products.

While our descriptive results indeed substantiate the existence of an aggregate pink tax, they do not speak to its underlying mechanisms. We now turn to estimating differences in supply, demand, and competition between men and women and their respective goods in explaining the forces that generate the pink tax.

4 Gender differences in demand elasticity

To estimate demand elasticity differences between men and women, we augment the constant elasticity of substitution (CES) model used in Faber and Fally (2022). This approach allows us to aggregate elasticities and make comparisons of the purchasing habits of men and women across a wide range of products. If the per-unit price premium observed on women's goods and on goods purchased by women more broadly is attributable to differences markups, we should observe that women exhibit lower price-sensitivity on average as a consumer demographic.

The model characterizes a representative consumer in a geography-retailer-time combination, l, that varies in gender g. The consumer allocates their income between a vector of retail goods G and additively separable consumption of the outside option:

$$U_{gl} = U(V_{gl}(G), C_{gl}).$$

We assume that the basket of goods comprising the outside option $C_{g,l}$ is consumed normally.

The model aggregates products in two tiers: the consumer allocates consumption across product modules with Cobb-Douglas elasticity and substitutes between goods with module-specific constant elasticity of substitution. We index product modules as n and employ the term "market" to refer to a geography-retailer-time tuple (c, r, y), indexed l. Additionally, let t index module-market combinations as $t := (n, c, r, y) = (n, l)$. The consumer maximizes their utility subject to their budget constraint by choosing a vector of quantities, G, that represents their consumption bundle across all goods:

$$V_{gl}(G) = \max_{G:=\{q_j\}} \prod_{n \in \mathcal{N}_l} \left[\sum_{j \in G_{nl}} \Big(q_j \varphi_j(g) \Big)^{\frac{\sigma_t(g)-1}{\sigma_t(g)}} \right]^{\alpha_t(g) \frac{\sigma_t(g)}{\sigma_t(g)-1}}. \tag{2}$$

$\mathcal{N}_l$ refers to the set of product modules that the representative consumer purchased in market l; G_{nl} is the vector of UPCs, indexed j, contained in module n in market l.[13] φ_j refers

[13]Some studies that estimate demand elasticities with the Nielsen data study products at the brand-level (e.g. Faber and Fally (2019)), whereas we consider the UPC-level due to inconsistencies in the gender-marketing of products within brands. To illustrate, in the disposable razors market, all product brands

to the product quality of a product j, which varies as a function of the consumer gender. σ_t represents the elasticity of substitution within a module-market, and α_t denotes the share of expenditures allocated to a module $n \in \mathcal{N}_l$ within market l.[14]

Specifying the upper tier as Cobb-Douglas implies that comparisons of consumption amounts between products within the same module depend on their relative quality-adjusted prices:

$$\frac{b_{jt}(g)}{b_{kt}(g)} = \left(\frac{p_j/\varphi_j(g)}{p_k/\varphi_k(g)}\right)^{1-\sigma_t(g)}, \tag{3}$$

where $b_{jt}(g)$ is the budget share spent on product j in module-market $t := (n, c, r, y)$. From Equation (2), we derive our estimating equation:

$$\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}, \tag{4}$$

where differences are taken from one time period to the next and η_{gt} captures the change in the price index. We derive this estimating equation from a CES demand model, but this estimating equation can also be interpreted as an average (over module-markets) of heterogeneous price responses within a market. In our estimation we define a module-market t as a product module $\times$ county $\times$ retail chain $\times$ half-year combination, and alternatively discarding the retailer distinction depending on specification. We estimate our model at the half-year level, as many product categories are prone to stockpiling, which when observed in shorter time intervals, would bias our demand estimates towards greater price elasticity; this bias may further confound our estimates of $\sigma_f - \sigma_m$ if men and women exhibit differential stockpiling behavior. To address auto-correlation in the error term, we cluster standard errors at the UPC-county level.

We face the standard issues of simultaneity in demand estimation where price changes may be correlated with demand shocks. To address this issue, we rely on two identifying assumptions employed frequently in empirical works estimating product demand. First, we

produced by Gillette map to the gender of the product (e.g. Gillette Venus marketed toward women versus Gillette Fusion marketed toward men), but other razor brands like Bic do not always have brand names that map to one gender (e.g. Bic Plus razors have both female- and male-marketed UPCs under the same brand).

[14] Under a Cobb-Douglas upper nest it is the case that $\sum_{n \in \mathcal{N}_l} \alpha_t(g) = 1$ for set of modules $\mathcal{N}_l$

assume that local demand shocks are uncorrelated and idiosyncratic across localities while supply shocks are correlated across space and retailers (e.g. Hausman (1999)). Second, we assume that retail chains set prices at the national or regional level and that these prices are set independent of local demand shocks following evidence presented in DellaVigna and Gentzkow (2019). From these assumptions, we estimate $(1 - \sigma(g))$ using two instruments. The first are Hausman instruments, which we construct as national leave-out means in price changes at the county level, $H_{jcy} := \frac{1}{N^c_{jy}-1} \sum_{l=(c'y)|c'\neq c} \Delta \log(P_{jl})$ where N^c_{jy} refers to the number of observations of UPC j across counties in half-year y.[15] The second are instruments that follow DellaVigna and Gentzkow (2019) and further developed by Allcott, Lockwood, and Taubinsky (2019) which are constructed as national leave-out means of price changes at the county-retailer chain level, $DV_{jcry} := \frac{1}{N^{cr}_{jy}-1} \sum_{l=(c',r',y)|r',c'\neq r,c} \Delta \log(P_{jt})$ where N^{cr}_{jy} refers to the number of observations of UPC j across retailer-counties in the half-year.[16][17]

Section A.2.1 gives additional detail to the model. We derive own-price elasticity of demand:

$$\varepsilon_{jt}(g) = \sigma_t(g) - (\sigma_t(g) - 1) \cdot s_{jt}(g) \tag{5}$$

Where $s_{jt}(g)$ is the market share of product j in market t. Thus, we can calculate $\varepsilon_{jt}(g)$ as a function of known and estimated parameters. In the special case of monopolistic competition, all market shares are approximately zero and $\varepsilon_{jt}(g)$ collapses to the elasticity of substitution, $\sigma_n(g)$. To map elasticities to markups, we assume single product firms compete on prices and maximize firm profits given the demand that they face.

[15] County level prices of good j in half-year y are constructed as simple means over observations of good j across retailers in the half-year within each county.

[16] County-retailer level prices of good j in half-year y are constructed as simple means over observations of good j in the half-year within each retailer-county.

[17] Much of the variation in the DellaVigna-Gentzkow instrument is driven by variation in how often a product is placed on a promotional sale. The timing of these sales is driven by a bargaining process between the retailer and the manufacturer and typically only one manufacturer is put on promotional sale at a time. If competition among manufacturers is strong enough, then promotional sale decisions are largely independent of demand shocks as well.

4.1 CES Model Results

We begin by estimating differences in the elasticity of substitution, $\sigma_t(g)$, between men and women. Table 3 presents results of estimating Equation (4) and pooling the elasticities across all departments. The main coefficient of interest is $\hat{\sigma}_m - \hat{\sigma}_w$, the average difference in elasticity of substitution between men and women. In column (1) We include a UPC-market fixed effect and estimate differences in demand elasticities between men and women for the same price change for the same product. If we assume that demand shocks affect men and women in the same way, this regression does not need to be instrumented since the endogenous portion is differenced out.[18] We find that for the same UPC in the same market, women are about 4.45pp more elastic than men. Column (1) restricts only to UPCs purchased by both men and women in the same market, columns (2)-(4) include market-level fixed effects and the results correspond to our full CES model, incorporating differing product choices between men and women. Column (2) includes market fixed effects at the module, county, half-year level and instruments with Hausman instruments only. We find that women are 11.6 pp more elastic than men. Columns (3) and (4) define markets at the module, county, retail chain, half-year level. Column (3) instruments for price with the DellaVigna-Gentzkow instruments and finds similar results that women are 11 pp more elastic consumers than men. Finally, column (4) includes both instruments and finds that women are 6.9 pp more elastic consumers than men.

We now turn our focus to how elasticities of substitution vary across product departments. We find that women are either more elastic than men are or are not significantly different than men in terms of elasticity. Table 4 presents elasticity of substitution results pooled to the department level.[19] We find that across almost all food products women are significantly more elastic consumers than are men, with $\hat{\sigma}_m - \hat{\sigma}_w \in [-0.15, -0.46]$. Among non-food retail products we find no significant differences in the elasticities of substitution between men and women; the magnitude of the coefficient for Health and Beauty products suggests the

[18]Table A.12 reports analogous OLS results. Table A.13 and Table A.14 present the first stage and reduced form results respectively.

[19]Here we define markets at the retail chain × designated marketing area (DMA) × half-year level. DMAs are more aggregated geographic areas than counties but less aggregated than states. Using DMAs does not significantly change our point estimates, but improves power by reducing the amount of sparsity in the data.

possibility that women are less elastic within that specific market space.[20] The vast majority of purchases that constitute the retail consumption basket in the Nielsen data are food purchases, so our finding that women are more elastic applies to the bulk of the consumption basket. However, the majority of gendered products exist in non-food purchases, particularly Health and Beauty products. We interpret this finding as evidence that women demonstrate greater price elasticity across markets even with little explicit gendering. But, we cannot reject that women are less elastic in markets with significant gendering.

Up to this point, we have estimated elasticities of substitution, $\sigma(g)$, whereas actual price elasticities of demand are given by Equation (5) and are a function of the elasticity of substitution and market shares. Under this derivation, price elasticities of demand will range from $\sigma_t(g)$ (in case of monopolistic competition where market shares are approximately 0) to 1 (in case of monopoly where the market share of the single good is 1). Because we have found that women generally substitute more elastically than men, the primary remaining channel for women to be less elastic on average as a consumer demographic is through lower market competitiveness for women's markets than for those of men. From Figure 1 in the descriptive analysis, we know that women a greater number of distinct products than do men by about 27%. This suggests *prima facie* that women's markets are more diverse than men's and are also likely more competitive.[21]

[20] We find that Health and Beauty and General merchandise products tend to exhibit lower price elasticity than other departments. The finding that consumption of Health and Beauty products is more inelastic than that for other types of products is consistent with findings in Faber and Fally (2022) as well as with our findings in Section 6. General Merchandise contains many products which can either be purchased or have substitutes sold at retailers not included in the Nielsen data and thus many of the purchase habits from this department cannot be considered complete. Examples include tools, automotive, household appliances, photographic supplies and stationary.

[21] As additional evidence that womens' consumer goods markets are actually more competitive than those of men, Figure A.5 plots histograms of log market shares for the men and women in our sample. The entire distribution of market shares for women is shifted to the left, indicating that their markets are more competitive. Figure A.6 also plots the distribution of Hirschfield-Herfindahl Index (HHI) observations over modules by consumer gender, illustrating a similarly greater level of competitiveness in markets faced by women. In particular, the modal man's UPC exhibits a market share about .75 log points greater than the modal woman's UPC—i.e. taking up twice as much market share. Additionally, market shares in our data are very small, on the order of 0.05% for the median UPC. This means that, in our setting, elasticities of substitution are close approximations of price elasticities of demand. On average, we can conclude that women are more price elastic consumers than are men. Abstracting from the role of pricing in the context of multiproduct firms, this finding strongly indicates that our average documented price differences paid by men and women—the pink tax—do not reflect differences in markups paid, but rather differences in marginal costs on average.

5 Wholesale Prices and Retail Markups

We link our scanner data environment to data on wholesale prices from distributors to retailers from PriceTrak. These data consist of wholesale price information on the UPC-geography-year level from 2006-2011, from which we construct retailer markups.[22]

Although these data do not represent direct manufacturing or production costs, they allow us to directly observe a portion of the markup-setting process. Consider a four-stage production-to-consumer setting with a manufacturer, wholesaler or distributor, retailer, and final consumer.[23] Let c represent the per-unit manufacturing cost of a good. The manufacturer sets a manufacturing markup μ^m so that the wholesaler or distributor pays a marginal cost of $\mu^m c$. The distributor adds a markup μ^d so that the retailer pays a marginal cost of $\mu^d \mu^m c$. Finally, the retailer adds a markup μ^r so that the consumer pays a final unit price of $p = \mu^r \mu^d \mu^m c$, which we observe in the Nielsen data. In this setting, the PriceTrak data specifically allow us to observe retailer cost $c^r = \mu^d \mu^m c$ and infer retailer markups μ^r.

Our inference on gender differences in *retailer* markups $\mu^r_f - \mu^r_m$ will yield unbiased inference on gender differences in *overall* markups $\mu_f - \mu_m$ under the following condition:

$$\mathbb{E}[\log(\mu_f) - \log(\mu_m) \mid \log(\mu^r_f) - \log(\mu^r_m)] \approx \mathbb{E}\Big[\Delta\%c + \Delta\%\mu^m + \Delta\%\mu^d \mid \Delta\%\mu^r\Big] = 0 \quad (6)$$

for a locally approximate proportion difference between women and men $\Delta\%x := \log(x_f) - \log(x_m)\%$. The condition requires that conditional on observing the proportion gender difference in retailer markup $\Delta\%\mu^r$, the sum of the proportion difference in 1) manufacturing cost $\Delta\%c$, 2) manufacturing markup $\Delta\%\mu^m$, and 3) distributor markup $\Delta\%\mu^d$ introduce no *additional* outsized proportion difference in overall markup (i.e. all of the informational content in proportion difference in *overall* markup between men and women is captured by the proportion difference in retailer markup). As a sufficient but not necessary condition, it could be the case that there are no conditional gender differences in any of these three

[22] We align PriceTrak markets with the ScanTrack market codes used in Nielsen based on market name. We link 55 out of 67 PriceTrak markets; for the remaining 12 markets in PriceTrak that do not correspond with a ScanTrak market code, we use national-level wholesale prices (also reported by PriceTrak).

[23] Our discussion is largely un-impacted by having distinct wholesaling and distribution entities.

left-hand-side objects, and all of the difference emerges at the retail markup setting stage. Note that the difference between "women and men" here can be interpreted equally as the difference between women and men as consumers (i.e. the average difference in markups faced by men and women) as well as the difference between women's and men's goods (the average difference in markups by UPC-gender).

There are additional important caveats to using the PriceTrak data. First, these data only cover a subset of the Nielsen data. Within the 2006-2011 timeframe, only 9% of the UPCs observed in the Nielsen data map to a PriceTrak wholesale price observation.[24] This matched subsample accounts for 37% of purchase volume we observe in the HMS panel during this timeframe. Several UPCs have multiple observations on the upc-geography-year level featuring multiple unique wholesale price values; in this case we use the lowest-observed per-unit wholesale price. Within this sample of Nielsen purchases that successfully matches to a wholesale price, around 8% of transactions exhibit a negative markup (i.e. observed wholesale prices lower than unit prices), which we discard.[25] Because we find little evidence of gender bias in PriceTrak coverage and because these data give us a unique insight into a component of the markup-setting process, we view these data as acutely informative in understanding the mechanisms underlying the pink tax.[26]

The PriceTrak data reveal several stylized facts that further substantiate our finding that the pink tax is not attributable to higher markups charged on goods women consume than on those than men consume.[27]

[24]We report the following match rate by department: 1) Health and beauty (5.45%), 2) Dry grocery (12.5%), 3) Frozen foods (14.24%), 4) Dairy (12.76%), 5) Deli (8.25%), 6) Packaged meats (16.5%), 7) Produce (3.25%), 8) Non-food grocery (10.28%), 9) Alcohol (0.2%), 10) General merchandise (2.7%).

[25]We find no evidence of a differential presence of negative markups based on consumer gender. A transaction-level regression of a binary variable for the presence of a negative markup on a binary indicator for female purchaser gender yields a coefficient of 0.0007 (standard error 0.001, p-value 0.488) off of a male baseline of 0.079. Including module fixed effects yields a female dummy coefficient of 0.0016 (standard error 0.001, p-value 0.104).

[26]Another central limitation of the PriceTrak data is that they inform retailer costs only for the set of goods purchased from wholesalers. It is possible that other goods see other vertical production structures, including goods sold directly to consumers by manufacturers or goods sold from a manufacturer directly to a retailer. Goods produced in either of these cases would not be covered by PriceTrak data.

[27]Section A.3 presents additional conceptual and empirical evidence on markups as inferrable from PriceTrak data. See Section A.3.2 for additional figures and tables on retailer markups and costs. The section includes information on costs as directly observed in PriceTrak and unconditional comparisons of markups and costs by UPC-gender.

First, Table 5 Panel (a) reproduces the descriptive results on average unit price differences paid by women and men as in Table 1 on the sample matching to the PriceTrak data. Columns (1) and (4) reproduce the least- and most-saturated specifications from Table 1 (columns (1) and (6)). Columns (2) and (5) run these same specifications on the PriceTrak sample. Lastly, columns (3) and (6) control for log wholesale price as observed in the PriceTrak data. The female coefficient in column (3) is significant and negative, indiciating that comparing goods in the same product module (and purchased in the same year), women actually pay a *lower* unit price than do men. I.e., conditioning on this measure of wholesale price, there is no pink tax. The coefficient in column (6) is positive and significant but very close to zero. The estimated coefficient represents an approximately 85% reduction in the gender-differential unit prices paid relative to as reported in column (5)—after conditioning on location, age, race, retailer, and income demographic.

Figure 4 displays the coefficients of regressions analogous to Figure 3 projecting markups onto decile-bins of female purchase share (relative to the decile of UPCs with near gender parity in purchase share). Panel (a) shows the results of this regression with no differential weighting across UPCs. The figure follows a U-shaped pattern in female purchase share, where goods purchased at near-gender-parity see the lowest markups, and highly gendered goods see greater markups. More striking however, is that the markups exhibited by male-goods are significantly higher than female goods. At the extreme ends, goods purchased nearly exclusively by men see 40% higher markups than goods at near gender parity, whereas goods purchased nearly exclusively by women see only 30% higher markups. Panel (b) estimates this same regression while including analytic weights on the amount UPC expense recorded among HMS panelists (including Nielsen sample adjustments); the graph illustrates a similar shape in markup evolution, with even greater relative markups for male goods than female goods: considering the amount expense for each UPC, male goods see 60% higher markups than the gender-parity good, visibly increasing with male purchase intensity, whereas female goods see only a 30% greater markup and a much more shallow increase in female purchase intensity.

Lastly, Table 5 Panel (b) displays the coefficients from various specifications of transaction-level regressions of retail markups (as implied by PriceTrak wholesale prices and Nielsen final sale prices) on indicators for female purchaser gender. The coefficients illustrate minimal difference in average markup faced by women and men.[28]

6 Differentiated Products: Markups and Marginal Costs

Having demonstrated that women tend to be consume more price-elastically and that they tend to purchase products with higher wholesale prices, we now turn our attention to estimating total markups and marginal costs of production. In Section 4, we estimated differences in demand elasticities between men and women across their retail consumption baskets using a constant elasticity of substitution model. To do this, we leveraged individual level purchase data aggregated to the by-gender market level. This method allowed us to capture consumer level average demand differences across a broad scope of products, but at the cost of model complexity in terms of flexible substitution patterns and market structure. Additionally, individual level purchase data faces sparsity issues in markets where purchases are relatively infrequent, like Health and Beauty products. To structurally decompose prices into markups and marginal costs, we will now allow for significantly more model complexity at the cost of narrowing our focus to fewer markets. We use weekly store-level data that does not face the same sparsity issue that the aggregated individual level data does. This lack of sparsity comes at the cost of no longer being able to attribute purchases to a specific gender. To overcome this limitation, we rely on our observed woman purchase share, $\hat{w}_j$ that we calculate using the individual level purchase data and map to the products in the weekly store level data.

We model demand in two markets: yogurt and disposable razors. Both yogurt and razors have a high level of dispersion of $\hat{w}_j$ across their product spaces. Specifically, we select yogurt because prices and consumer behavior look similar to descriptive results of

[28]Table A.15 estimates analogous set of regressions with analytic weighting for each observation equal to the consumer budget share represented by the specific good transaction. Alternating between significant negative coefficients and insignificant coefficients on the female dummy, these results imply that women either spend a lower proportion of their budgets on markups or that there is little difference in relative budget share allocation to markups per transaction.

the entire grocery consumption basket. We think of the yogurt market as representative of grocery food markets generally. While yogurt seems to have significant heterogeneity in preferences across gender, its marketing and advertising is less explicitly gendered than the market for disposable razors. We selected razors because they are commonly referred to as the canonical pink tax item; razors exhibit near complete gender segregation, and they also feature observable product characteristics.[29]Figure 5 panel (a) plots histograms of woman purchase share along with observed per unit prices for yogurt and disposable razors. Yogurt follows a similar normal distribution to what we see across the data at large, but disposable razors have a bimodal distribution reflecting its high gender segregation.

We plot the median per unit price of a product within a woman purchase share decile along with the interquartile range in Figure 5 Panel (b). We find that prices increase in woman purchase share. The average men's razor in our data priced at about \$1.2, while the average women's razor is priced at about \$1.5. We find that women's yogurt is generally priced about 5 cents higher per ounce than ungendered yogurt, this corresponds to about a 30 cent price difference for a standard six ounce cup of yogurt.

6.1 Differentiated Products Demand Model and Estimation

We follow the standard differentiated products market demand model presented in Berry, Levinsohn, and Pakes (1995) (BLP). Our main departure is that instead of typical product characteristics, we include our measure of the woman purchase share of a product, $\hat{w}_j$ and allow for heterogeneity in preferences for how gendered a product is. For each product module, consider $t = 1, ..., T$ markets defined as a retail store-month combination each with $i = 1, ..., I_T$ customers. The indirect utility that customer i receives from choosing product j in market t is:

$$u_{ijt} = \alpha p_{jt} + \beta_{\mathbf{i}} \mathbf{x_j} + \xi_{jt} + \epsilon_{ijt}, \tag{7}$$

[29]To support our main analysis we present results for three other markets: protein bars, shampoo and deodorant in Section A.3.

where p_{jt} is the price of product j in market t, x_j is vector of a constant term and the woman purchase share of the product, $\xi_{jt} = \xi_{jr(t)} + \xi_{m(t)} + \Delta\xi_{jt}$ are product-retail chain fixed effects, month fixed effects, unobservable product characteristics, and ϵ_{ijt} is a mean-zero idiosyncratic error term that assumes a Type I Extreme Value distribution. The key deviation from our CES model or a logit demand is that the coefficients on the product characteristics, β_i, are individual-specific coefficients. We can parameterize these individual coefficients as a population mean preference parameter that is absorbed by the fixed effects and an individual random taste shock that captures unobserved heterogeneity in preference for the outside option and the woman purchase share of the product:

$$\beta_{\mathbf{i}} = \mathbf{\Sigma} \cdot \mathbf{v_i}, \qquad \mathbf{v_i} \sim N(0, \mathbf{I_2})$$

Heterogeneity in preferences for product gender may generate more reasonable substitution patterns than under our CES demand model. Under CES demand, price increases on a woman's razor will lead to equal levels of substitution from the women's razor into other women's razors and men's razors. Now, the random coefficient on women purchase will generate substitution patterns that have men's razors substituting to men's razors and women's razors substituting to women's razors. Allowing for heterogeneity in preferences for the outside option is important as the value of the outside option likely differs between men and women in many of these of these markets. For example, the value of the outside option for disposable razors depends on the social stigma attached to shaving for men versus women.[30]

The resulting market share for product j in market t can be written as:

$$s_{jt} = \int \frac{\exp(\alpha p_{jt} + \beta_i \mathbf{x_j} + \xi_{jt})}{1 + \sum_k (\exp(\alpha p_{kt} + \beta_i \mathbf{x_k} + \xi_{kt}))} d\beta_i \tag{8}$$

We estimate the model using the Python package, *pyBLP* (Conlon and Gortmaker (2020)), which solves for the parameters of interest using a two-step generalized method

[30] Many papers that estimate differentiated products demand models include demographic moments as in Nevo (2001), here we do not because our product characteristic is effectively a demographic moment and will be mechanically correlated.

of moments.

We instrument for prices with the same instruments we use for our constant elasticity of substitution analysis: Hausman instruments that are a national level leave out mean of prices and DellaVigna-Gentzkow instruments that are a retail level leave out mean of prices. The Hausman instruments rely on the assumption that demand shocks are uncorrelated across markets while supply shocks are correlated across space and time. The DellaVigna-Gentzkow instrument's validity relies on retail chain level pricing being largely exogenous from local demand shocks. In addition to price instruments, we identify substitution patterns across products with quadratic differentiation instruments developed by Gandhi and Houde (2019). The instruments take the form $Z_{jt}^{diff} = \sum_k d_{jkt}^2$, where $d_{jkt} = x_{kt} - x_{jt}$ and x_{jt} is the woman purchase share of product j. We utilize two versions of this instrument: one with differences summed over products that are rivals; that is, products that are owned by other firms, and one for products produced by the same firm. The instrument captures proximity in the product space in terms of woman purchase share and is rooted in the idea that substitution likely occurs among products that are similar in gender.

We fit the supply side of the model by assuming firms, f, maximize their profits across the set of products they produce, $\mathcal{J}_f$ given the demand that they face.

$$\pi_{ft} = \sum_{j \in \mathcal{J}_f} (p_{jt} - mc_{jt}) s_{jt},$$

We construct an ownership matrix, Ω, that maps each product in our data to a common owner so that element jk is 1 if product j and product k are owned by the same firm and 0 otherwise.[31] Let J be the matrix of estimated demand derivatives, so that element jk is $\frac{\partial s_j}{\partial p_k}$. The price-cost markup is then given by:

$$\frac{p^* - mc}{p^*} = -(\Omega J)^{-1} \frac{s(p^*)}{p^*} \tag{9}$$

Because we observe price, identified markups also identify marginal costs.[32].

[31] We construct the ownership matrix through manual search, Capital IQ, and newspaper articles.
[32] The estimated parameters are presented in Table A.18

6.2 Differentiated Products Demand Model Results

We plot the interquartile range and median estimates of own price elasticities, markups and marginal costs in Figure 6. Panel (a) plots own price demand elasticities given by $\varepsilon_{jt} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}}$. Generally we find that women's products are more elastic than men's or ungendered products.[33] These findings are inconsistent with a price discrimination mechanism driving the pink tax, where we would expect to find that women's products exhibit less-elastic demand. Instead we find that women's products are much more elastic and men's products are no differently elastic than ungendered products. Our results are consistent with our CES demand estimation and suggest that women as a consumer base seem to be generally more elastic consumers than men across both gendered and ungendered products.

Firm's base product pricing on own price elasticity of the product, the cross elasticities with other products owned by the firm and with rival products, and marginal cost. Even though women's products exhibit more elastic demand, they could face higher markups through substitution patterns or the competitive structure of the market. We capture this structure through our ownership matrix, Ω, which maps products to a common owner. Multiproduct firms face incentives to price products higher because some of the lost demand is funneled into other products that they own. Women's products could still face higher markups if they are more likely to be owned by large multiproduct firms and consumers strongly substitute to other products owned by the firm.

We plot median estimated markups along with interquartile range by woman purchase share decile in Figure 6 panel (b). We find that markups are lower for women's products for both yogurt and disposable razors.[34] From the markups we directly calculate marginal costs and present them in Figure 6 panel (c). We find that marginal costs increase in woman purchase share; that is, the products that women sort into are more expensive to produce.

[33] We also find that women's products are more or no differently elastic than men's products for shampoo, deodorant, and protein bars in Figure A.13.

[34] Figure A.14 presents the estimates for shampoo, deodorant and protein bars. Protein bars are the only product market where we find that women pay significantly higher markups than men. This result is entirely driven by substitution patterns and common ownership between Clif and Luna bars, highlighting the important role that competition can play.

We estimate negative marginal costs for disposable razors.[35] This issue stems back to our estimates of downward sloping but inelastic demand in Figure 6 panel (a), where the average razor has an elasticity of -0.75. There are many reasons why this could arise, related to both supply and demand side behavior. Our partial equilibrium model assumes firms maximize profits statically in each period and that consumers are rational in their decisions. Deviations from our assumed competitive structure as well as behavioral demand factors may result in equilibrium prices and elasticities that are lower than what is rationalizable in the standard static setting. Dubé, Hitsch, and Rossi (2009) find that when consumers have brand loyalty and firms price to maximize their future stream of profits, equilibrium prices can be lower than in the static case.[36] It is also possible that other dynamic competitive factors may drive prices down, like threat of entry of other firms or products. While these factors may be basing our estimates of marginal costs down, they are unlike to change the trendline that matters for our results.[37]

7 Conclusion

We evaluate the existence of a "pink tax" on women's consumer goods relative to men's. We document a robust price premium on women's goods compared to similar men's goods of 4% on average. Further corroborating this descriptive result, we find that within markets of similar goods, unit price increases nearly monotonically in women purchase share, relative to a gender-parity baseline. Simultaneously, we observe similar prices for men's goods as for goods purchased at gender-parity. Not only do we observe a consistent women's price premium of 17% on overtly gendered goods , but we also observe that women sort into purchasing less-overtly-gendered products with higher prices 3.8%, such as organic foods.

We proceed by studying the causal components of this pink tax. We distinguish three

[35]Our results for shampoo, deodorant and protein bars in Figure A.15 demonstrate the consistency of this result across health and beauty products.

[36]In Section B, we discuss in detail brand loyalty and dynamic, forward looking firms.

[37]We validate our results for razors by looking at the number of blades and prevalence of moisture strips and ergonomic handles for women's and men's razors. We find that women's razors have more blades and are over 50% more likely to have an ergonomic handle, which is consistent with women's razors being more costly to produce. We present our results in Table D.1.

broad potential mechanisms at play: price elasticity of demand, competitive structure, and marginal costs. We estimate a CES model of demand and find that women as a consumer demographic are consistently *more* price elastic than are men. On average, women are about 11pp (30%) more price sensitive than are men.

We then link our data to data on wholesaler prices to retailers, allowing us to directly construct retailer markups. Under mild assumptions, these retailer markups are informative of overall markups. We find that controlling for wholesaler price eliminates the descriptive pink tax and that women's consumer goods see persistently higher wholesaler price (i.e. retailer cost) and that there exist no systematic difference in markups paid by men and women. Lastly, we model supply and demand in a model of differentiated products demand model (Berry, Levinsohn, and Pakes (1995)). We incorporate product gender as characteristic over which consumers can have heterogeneous tastes and find that women's products have higher marginal costs and lower markups than men's products.

We conclude from our analysis a novel set of facts to frame the discussion of the pink tax: women pay around 4% more per unit for similar goods than do men; when we study overtly gendered goods, this price difference rises to 17%. Taking consumption habits as fixed, the pink tax represents a *real* cost of living difference that exacerbates measures of the *nominal* gender wage gap by around 15-20% (Blau and Kahn (2017)). Contrary to popular discussion that attributes the pink tax to price discrimination, we find the pink tax is driven by women sorting into goods of higher marginal cost. However, it almost certainly the case that preferences are *not* exogenous to gender; it is likely that the sorting processes we identify reflect societal expectations of women's and men's consumption behaviors in addition to personal taste. Nonetheless, this result suggests that current legislation aiming to prohibit price differences for gendered products are likely to prove ineffective in improving outcomes, and may in fact induce increased product exit.

References

Aguiar, Mark and Erik Hurst (2005). "Consumption versus Expenditure". In: *Journal of Political Economy* 113.5, pp. 919–948.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky (2019). “Regressive sin taxes, with an application to the optimal soda tax”. In: *The Quarterly Journal of Economics* 134.3, pp. 1557–1626.
Argente, David and Munseob Lee (2017). “Cost of living inequality during the great recession”. In: *Kilts Center for Marketing at Chicago Booth–Nielsen Dataset Paper Series*, pp. 1–032.
Ayres, Ian and Peter Siegelman (1995). “Race and gender discrimination in bargaining for a new car”. In: *The American Economic Review*, pp. 304–321.
Berry, Steven, James Levinsohn, and Ariel Pakes (1995). “Automobile prices in market equilibrium”. In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.
Blau, Francine D. and Lawrence M. Kahn (Sept. 2017). “The Gender Wage Gap: Extent, Trends, and Explanations”. In: *Journal of Economic Literature* 55.3, pp. 789–865.
Bronnenberg, Bart J et al. (2015). “Do pharmacists buy Bayer? Informed shoppers and the brand premium”. In: *The Quarterly Journal of Economics* 130.4, pp. 1669–1726.
Castillo, Marco et al. (2013). “Gender differences in bargaining outcomes: A field experiment on discrimination”. In: *Journal of Public Economics* 99, pp. 35–48.
Conlon, Christopher and Jeff Gortmaker (2020). “Best practices for differentiated products demand estimation with pyblp”. In: *The RAND Journal of Economics* 51.4, pp. 1108–1161.
DellaVigna, Stefano and Matthew Gentzkow (2019). “Uniform pricing in us retail chains”. In: *The Quarterly Journal of Economics* 134.4, pp. 2011–2084.
Dubé, Jean-Pierre, Günter J Hitsch, and Peter E Rossi (2009). “Do switching costs make markets less competitive?” In: *Journal of Marketing research* 46.4, pp. 435–445.
Duesterhaus, Megan et al. (2011). “The cost of doing femininity: Gendered disparities in pricing of personal care products and services”. In: *Gender Issues* 28.4, pp. 175–191.
Faber, Benjamin and Thibault Fally (2019). *Firm heterogeneity in consumption baskets: Evidence from home and store scanner data*. Tech. rep. National Bureau of Economic Research.
— (2022). “Firm heterogeneity in consumption baskets: Evidence from home and store scanner data”. In: *The Review of Economic Studies* 89.3, pp. 1420–1459.
Gandhi, Amit and Jean-François Houde (2019). “Measuring substitution patterns in differentiated-products industries”. In: *NBER Working Paper* w26375.
Goldberg, Pinelopi Koujianou (1996). “Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey”. In: *Journal of Political Economy* 104.3, pp. 622–654.
Guittar, Stephanie Gonzalez et al. (2022). “Beyond the pink tax: gender-based pricing and differentiation of personal care products”. In: *Gender Issues* 39.1, pp. 1–23.
Hausman, Jerry (1999). “Cellular telephone, new products, and the CPI”. In: *Journal of business & economic statistics* 17.2, pp. 188–194.
Hendel, Igal and Aviv Nevo (2013). “Intertemporal price discrimination in storable goods markets”. In: *American Economic Review* 103.7, pp. 2722–51.
Jaravel, Xavier (2019). “The unequal gains from product innovations: Evidence from the us retail sector”. In: *The Quarterly Journal of Economics* 134.2, pp. 715–783.
Manatis-Lornell, Athena J et al. (2019). “Gender-related cost discrepancies in a cohort of 110 facial moisturizers”. In: *Journal of Cosmetic Dermatology* 18.6, pp. 1765–1766.
Manzano-Antón, Roberto, Gema Martinez-Navarro, and Diana Gavilan-Bouzas (2018). “Gender Identity, Consumption and Price Discrimination”. In: *Revista Latina de Comunicación Social* 73, pp. 385–400.
Moretti, Enrico (2013). “Real Wage Inequality”. In: *American Economic Journal: Applied Economics* 5.1, pp. 65–103.

Moshary, Sarah, Anna Tuchman, and Natasha Vajravelu (2023). "Gender-Based Pricing in Consumer Packaged Goods: A Pink Tax?" In: *Marketing Science* 0.0, null. eprint: `https://doi.org/10.1287/mksc.2023.1452`.

Nevo, Aviv (2001). "Measuring market power in the ready-to-eat cereal industry". In: *Econometrica* 69.2, pp. 307–342.

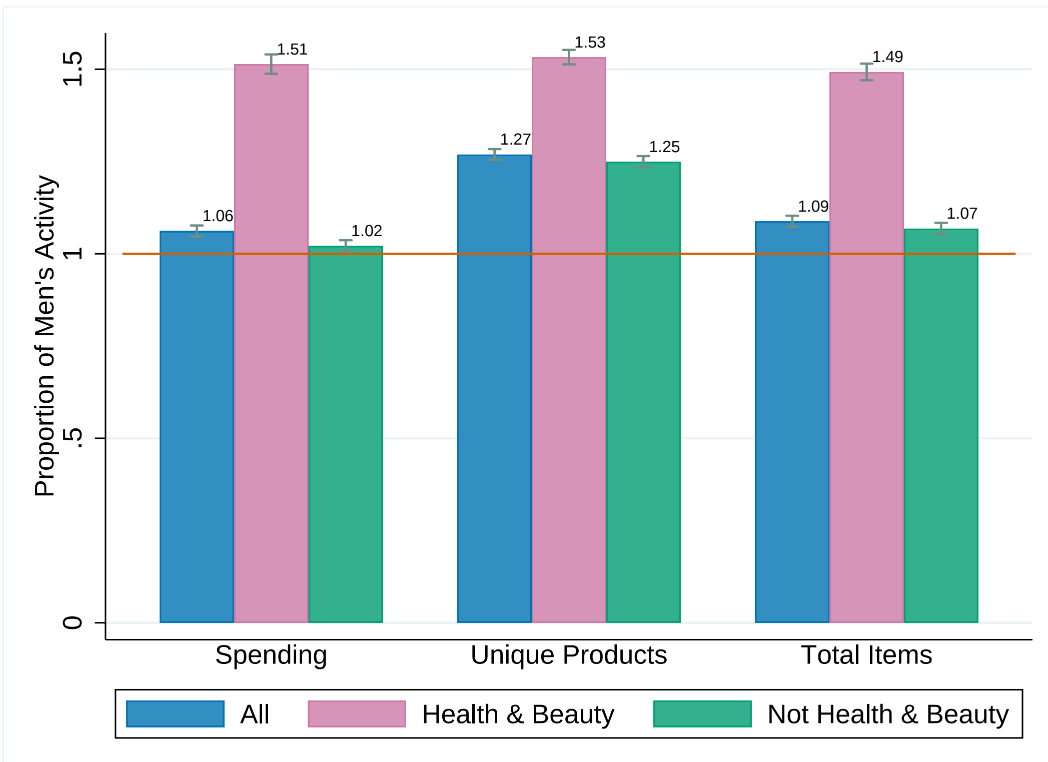
NYCDCA (2015). *From Cradle to Cane: The Cost of Being a Female Consumer*. Tech. rep.

Stigler, George J. (1987). *The Theory of Price*. Macmillan Publishing Co., Inc.

Trégouët, Thomas (2015). "Gender-based price discrimination in matching markets". In: *International journal of industrial organization* 42, pp. 34–45.

Ureña, Félix, Rodolfo Bernabéu, and Miguel Olmeda (2008). "Women, men and organic food: differences in their attitudes and willingness to pay. A Spanish case study". In: *international Journal of consumer Studies* 32.1, pp. 18–26.

Figure 1: Women's yearly retail consumption spending relative to men's



Note: This figure plots the coefficients estimated from a regression of log expenditure on an indicator for the individual being a woman and demographic controls: $\log y_{it} = \alpha + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_{it} + \varepsilon_{it}$, for dependent variables including yearly spending, unique products purchased, and total items purchased. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, and X_{it} is a vector of time- and time-id-varying controls including income, county, age, race and education. Standard errors are clustered at the individual-level.

Table 1: Gender differences in unit prices

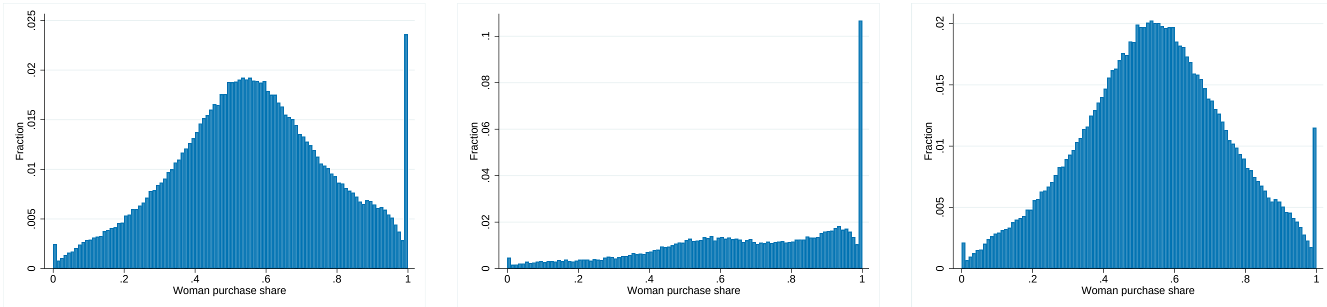
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Unit prices in same product module						
Women	0.0230***	0.0467***	0.0512***	0.0419***	0.0288***	0.0402***
	(0.0035)	(0.0033)	(0.0028)	(0.0020)	(0.0019)	(0.0018)
Men's Average	\$0.218	\$0.218	\$0.217	\$0.219	\$0.222	\$0.222
Mod. X Units FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.83	0.83	0.87	0.89	0.88	0.88
N	153,333,409	153,333,409	150,059,493	143,532,160	139,739,839	139,739,839
Number of clusters	49,256	49,256	49,256	49,256	49,256	49,256
Panel B: Unit prices for same product						
Women	-0.0089***	-0.0055***	-0.0060***	-0.0075***	-0.0089***	-0.0080***
	(0.0017)	(0.0017)	(0.0015)	(0.0010)	(0.0010)	(0.0010)
Men's Average	-1.531	0.891	0.895	0.897	-1.485	0.901
UPC FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.95	0.84	0.86	0.88	0.96	0.88
N	151,188,750	151,191,277	139,671,522	138,165,657	135,152,438	135,154,990
Number of clusters	49,256	49,256	49,256	49,255	49,256	49,256
Year FE	Yes	Yes	Yes	Yes	No	No
Month FE	No	No	No	No	Yes	Yes
County FE	No	No	Yes	Yes	Yes	Yes
Retailer FE	No	No	No	Yes	Yes	Yes
Demographic FE	No	Yes	Yes	Yes	No	Yes

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Panel A of this table presents estimates from the regression: $log(P_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. Panel B of this table presents estimates from the regression: $log(P_{ijt}) = \phi_{jt} + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a UPC-market-time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. All standard errors are clustered at the individual-level. Column 1 can be thought of as a raw gap between single men and single women, each subsequent column demonstrates the contribution of controlling for an additional market or demographic factor.

Figure 2: Distribution of Woman Purchase Share (UPC-gender) Across UPCs



(a): All Departments (b): Health and Beauty (c): Non-Health and Beauty

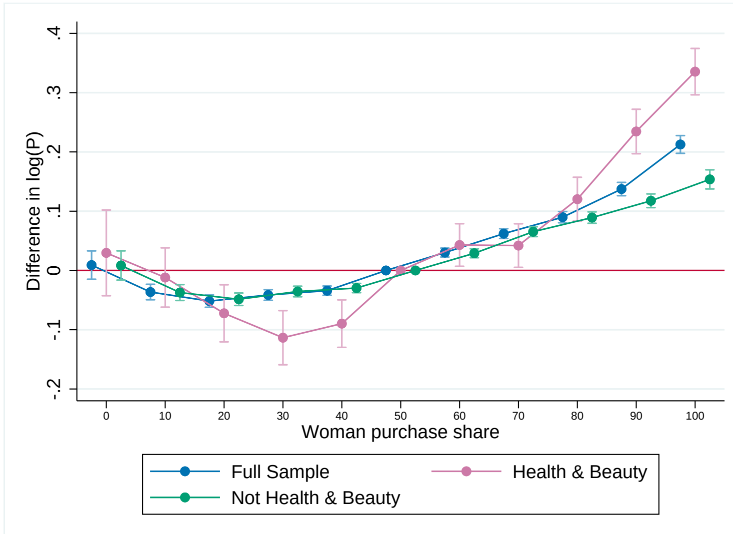
Note: This figure plots a histogram of the share of times a UPC is bought by women. We restrict to UPCs that have above a varying cutoff number of purchases by unique individuals over the panel, this cutoff number corresponds to 95% confidence that a product's true purchase share is within a 10 percentile bin centered around its observed share.

Table 2: Unit prices by gender of product and consumer

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
Women	0.0384***	0.0432***	0.0391***
	(0.0019)	(0.0032)	(0.0019)
Gendered Product	-0.0150**	0.0402***	-0.0307***
	(0.0060)	(0.0064)	(0.0075)
Women × Gendered Product	0.1198***	0.1002***	0.1125***
	(0.0064)	(0.0067)	(0.0085)
Men's Ungendered Average	\$0.217	\$0.368	\$0.212
MURLM FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes
Adj. R-squared	0.88	0.84	0.89
N	131,501,221	9,279,574	122,221,628
Number of clusters	49,256	49,128	49,256

Note: This table presents estimates from the regression: $log(P_{ijt}) = \phi_{t(j)} + \beta_1 \mathbf{1}_{w(i)} + \beta_2 \mathbf{1}_{g(j)} + \beta_3 \mathbf{1}_{w(i)} \cdot \mathbf{1}_{g(j)} + \gamma X_i + \epsilon_{ijt}$. $\phi_{t(j)}$ is a vector of fixed effects for the interaction of product module, units denomination, retailer chain, county, and half-year. X_i includes with demographic controls for income, age, race and education. Gendered products are defined as UPCs purchased exclusively 90% or more (by amount) by one gender. Columns 2 and 3 separate out Health and Beauty products. "MURLM FE" refers to Module × Unit × Retailer × County × Month fixed effects. Standard errors are clustered on the individual-level.

Figure 3: Prices of UPCs by Woman Purchase Share



Note: This figure presents plots of the results of the regression $\log P_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Bins $b \in \mathcal{B}$ include ten-percentile-width bins centered at and two bins for pure gender stratification at the tails partitioning the interval $[0, 1]$. The regression includes fixed effects for product module, county and half-year. Results are presented for the whole sample and also separating out Health and Beauty and Dry Grocery. Standard errors are clustered at the UPC-county level.

Table 3: Elasticities of Substitution

	(1) County	(2) County	(3) County-Retailer	(4) County-Retailer
$\sigma_m - \sigma_w$	-0.0445***	-0.1161***	-0.1097***	-0.0686***
	(0.0091)	(0.0221)	(0.0257)	(0.0209)
$1 - \sigma_m$		0.3055***	0.2777***	0.2548***
		(0.0193)	(0.0219)	(0.0181)
Observations	1,054,187	18,271,669	11,007,333	12,431,472
F-Statistic on first stage	N/A	12,764	8,184	5,397
UTCG FE	Yes	No	No	No
MTCG FE	No	Yes	No	No
MTCRG FE	No	No	Yes	Yes
Hausman IV	No	Yes	No	Yes
DellaVigna Gentzkow IV	No	No	Yes	Yes

UPC-County level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents the results of estimating elasticities of substitution by regressing changes in the log budget share of a product on changes in log price for men and women controlling for the location, retail chain, and half-year corresponding to the following regression: $\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$. Column (1) estimates differential price responses for men and women on the same price change for the same UPC. Columns (1) and (2) do not control for retail chain, taking the market definition to be a county-module-half-year. Column (2) utilizes only Hausman instruments. Columns (3) and (4) control for retail chain in the definition of market. Column (3) instruments for price with DellaVigna-Gentzkow instruments only. Column (4) instruments for prices with both Hausman and DellaVigna-Gentzkow instruments. "UTCG FE" refers to UPC×half-year×county×gender fixed effects. "MTCG" refers to module×half-year×county×gender and "MTCRG" refers to module×half-year×county×retailer×gender.

Table 4: Elasticities of Substitution by Department

	(1) H&B	(2) Dry Groc.	(3) Frozen	(4) Dairy	(5) Deli	(6) Meat	(7) Produce	(8) Non-food Groc.	(9) Alcohol	(10) Gen. Merch.
$\sigma_m - \sigma_w$	0.1037	-0.2682***	-0.4578***	-0.2709***	-0.1488	-0.2688**	-0.2145***	0.0161	0.7583	-0.0384
	(0.0974)	(0.0369)	(0.0907)	(0.0456)	(0.1650)	(0.1204)	(0.0798)	(0.0744)	(0.6057)	(0.1251)
$1 - \sigma_m$	0.4347***	0.2619***	0.4946***	0.1788***	0.1447	0.1667*	0.0001	0.2238***	-0.5720	0.4893***
	(0.0867)	(0.0315)	(0.0717)	(0.0357)	(0.1310)	(0.0953)	(0.0974)	(0.0672)	(0.5679)	(0.1156)
Observations	718,302	5,335,802	1,314,605	1,680,282	401,229	467,441	1,084,136	1,144,523	63,143	265,534
Adjusted R^2	-0.256	-0.287	-0.280	-0.192	-0.278	-0.212	-46.792	-0.275	-0.337	-0.333
MTDRG FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DMA IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Retailer IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

UPC-DMA level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents the results of estimating elasticities of substitution by regressing changes in the log budget share of a product on changes in log price for men and women controlling for the location, retail chain, and half-year corresponding to the following regression: $\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$. Results are pooled at the department level. Markets are defined at the product module-retail chain-DMA-half-year level. "MTDRG FE" refers to module×half-year×DMA×retailer×gender fixed effects.

Table 5: PriceTrak prices, costs, and markups

Panel A: Log prices, controlling for retailer cost

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0230***	0.0143***	-0.0071**	0.0402***	0.0354***	0.0056**
	(0.0035)	(0.0045)	(0.0034)	(0.0018)	(0.0031)	(0.0027)
Log wholesale cost			0.7603***			0.7173***
			(0.0018)			(0.0021)
Men's mean (levels USD)	0.22	0.20	0.20	0.22	0.21	0.21
PriceTrak sample	No	Yes	Yes	No	Yes	Yes
Module	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	No	No	Yes	Yes	Yes
County	No	No	No	Yes	Yes	Yes
Retailer	No	No	No	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	No	No	Yes	Yes	Yes
Adj. R-squared	0.83	0.75	0.86	0.88	0.87	0.92
N	153,333,409	17,901,327	17,901,327	139,739,839	14,342,604	14,342,604
Number of clusters	49,256	28,412	28,412	49,256	28,403	28,403

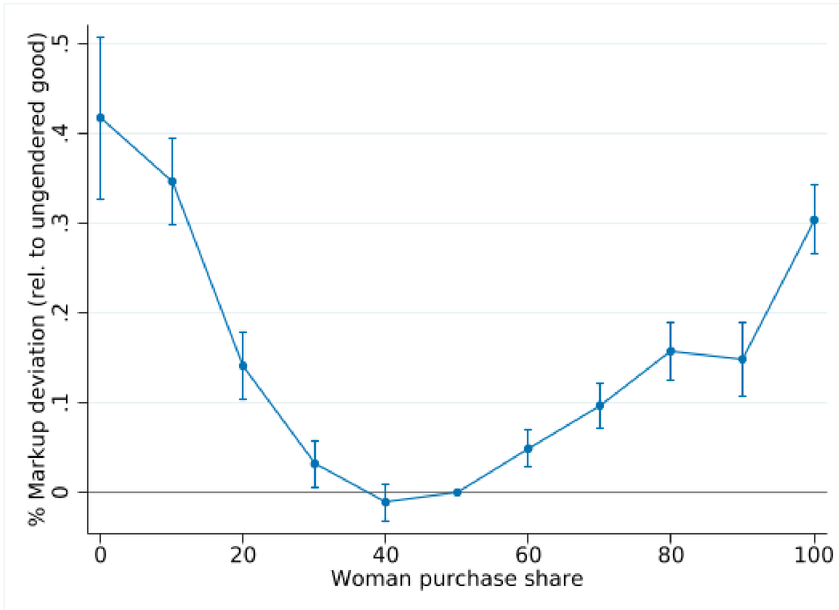
Panel B: Log markups

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0086	0.0035	0.0090	0.0062**	0.0076***	0.0034
	(0.0073)	(0.0061)	(0.0061)	(0.0032)	(0.0028)	(0.0028)
Men's mean (percent markup)	23%	23%	23%	23%	23%	24%
Demographics	No	Yes	Yes	Yes	Yes	Yes
Module	No	No	Yes	Yes	Yes	Yes
County	No	No	No	Yes	Yes	Yes
Retailer	No	No	No	No	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	No	No	No	No	Yes
Adj. R-squared	0.00	0.47	0.47	0.63	0.70	0.72
N	18,076,261	18,076,169	18,076,169	17,262,606	15,741,612	14,512,531
Number of clusters	28,412	28,412	28,412	28,406	28,400	28,403

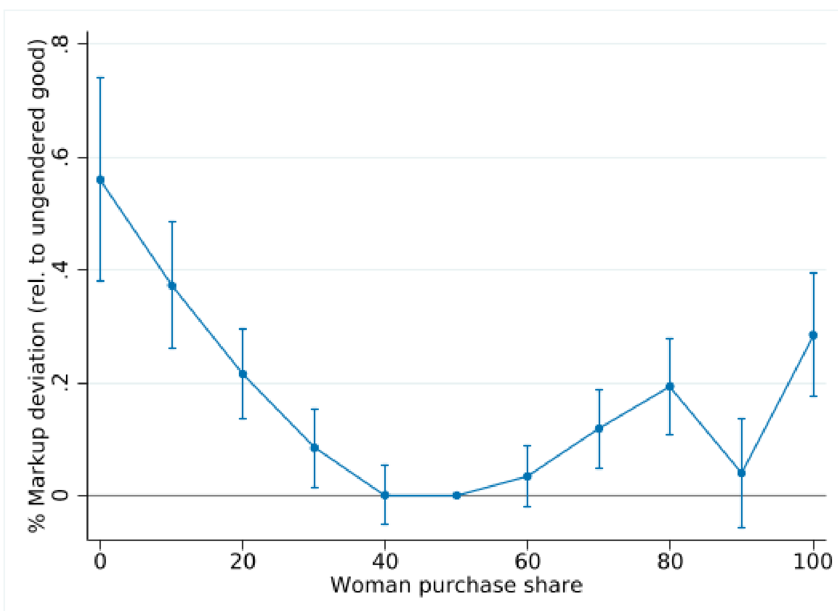
This table presents estimates from transaction-level regressions. Panel (a) estimates the form: $log(P_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \gamma C_{jt} + \Gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect, C_{jt} is the wholesale price of UPC j in year t as observed in PriceTrak (included only in columns (3) and (6)), and X_i is a vector of demographic controls including income, county, age, race and education. Panel (b) estimates a similar set of regressions, however with log markup as the dependent variable. Each column restricts to the set of UPCs matching to the PriceTrak data and varies the level of fixed effects. Standard errors are clustered at the individual-level.

Figure 4: Markup by woman purchase intensity

(a) Unweighted



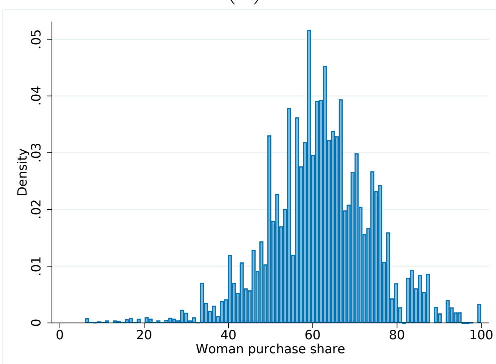
(b) Weighted by expense recorded in Nielsen HMS



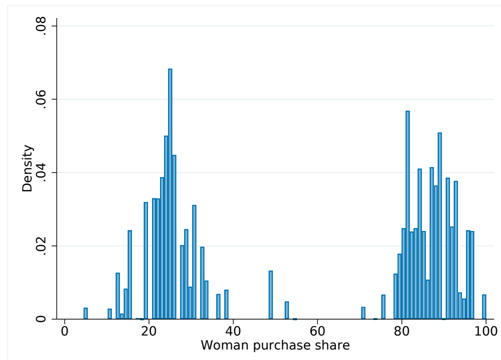
Note: These figures display the coefficients estimated from the following regression on the UPC-year level: $\log \mu_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Markup μ is constructed using PriceTrak data on wholesale prices and Nielsen final sale prices. Bins $b \in \mathcal{B}$ represent ten-percentile-width bins centered at multiples of 10 (truncated at 0 and 100) partitioning the interval $[0, 1]$; these bins reflect the aggregate amount of a UPC purchased by single women (as opposed to single men). The regression includes fixed effects for product module, county and half-year. Coefficients γ_b are estimated relative to goods in the same product module purchased at approximate gender-parity (between 45 and 55%). Panel (a) estimates this regression with equal weighting for all observations. Panel (b) presents the coefficients estimated from an analogous regression with analytic weights on UPC-year expenditure as recorded in the Nielsen HMS data. Standard errors are clustered at the UPC level.

Figure 5: Distribution and Prices of Yogurt and Razors

(a) Distribution of Woman Purchase Share

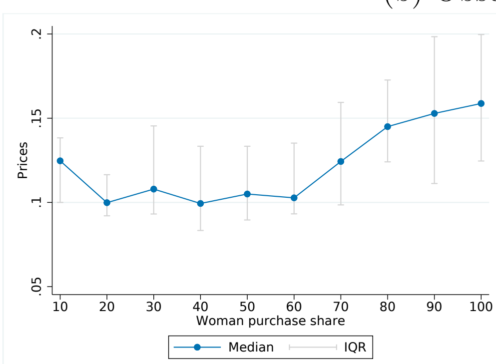


Yogurt

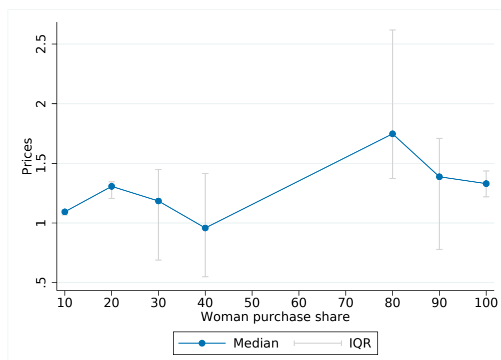


Disposable Razors

(b) Observed Prices



Yogurt

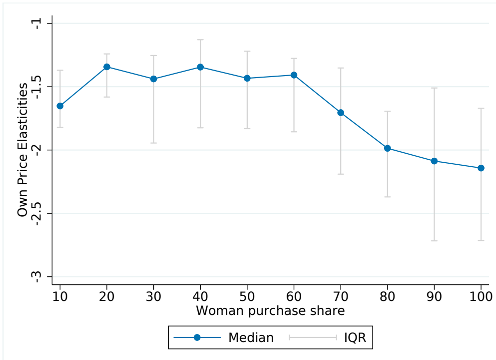


Disposable Razors

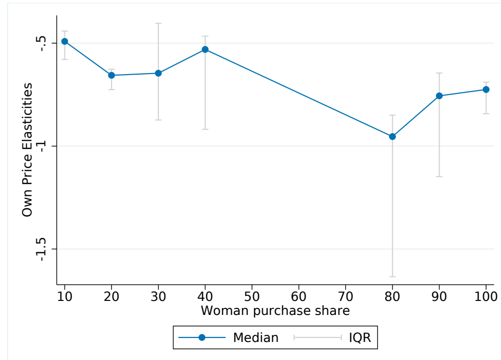
Note: Panel (a) presents the distribution of woman purchase share, $\hat{w}_j$, for yogurt and disposable razors. Panel (b) presents observed prices of yogurt and disposable razors in the markets over we estimate our differentiated products demand model on. Median and interquartile range of prices are presented for each woman purchase decile. Yogurt is plotted as price per ounce while while disposable razors are priced per count i.e. the price per razor included in the pack.

Figure 6: Differentiated Products Model Estimates

(a) Own Price Elasticities

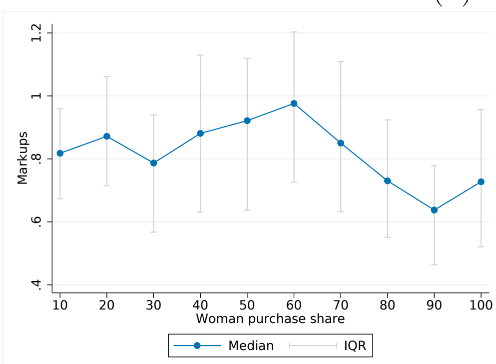


Yogurt

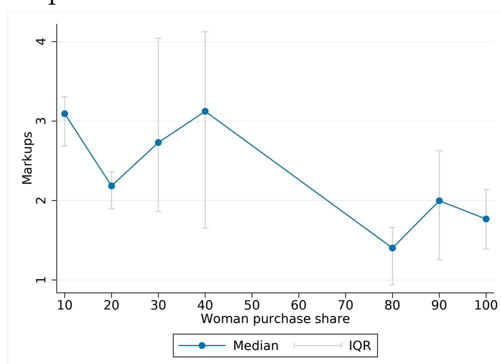


Disposable Razors

(b) Markups

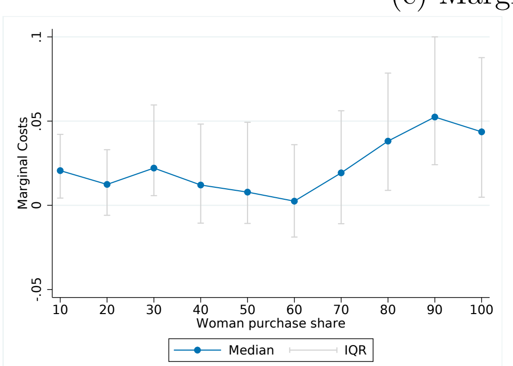


Yogurt

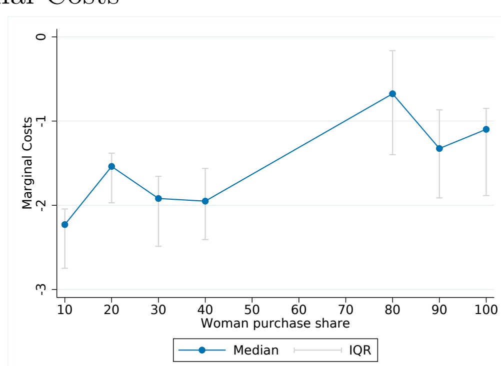


Disposable Razors

(c) Marginal Costs



Yogurt



Disposable Razors

Note: This figure presents the median and interquartile range of estimated own-price elasticities, markups and marginal costs of yogurt and disposable razors across woman purchase share. These estimates are produced from our differentiated products demand model described in Section 6.

Appendix A Additional figures and tables

A.1 Additional descriptive figures and tables on the pink tax

Table A.1: Demographics of HMS panelists sample of single-member households

	Total	Women	Men	Difference
Income	44687	39514	50682	-11167.86**
	(37202.4)	(34048.25)	(39718.72)	(340.2182)
Age	53.47	53.21	53.77	-.556**
	(16.4528)	(17.223)	(15.5078)	(.1522)
High school	0.602	0.637	0.562	.074**
	(.4894)	(.481)	(.4961)	(.0045)
College	0.238	0.206	0.275	-.069**
	(.4258)	(.4044)	(.4464)	(.0039)
Post-grad	0.120	0.115	0.127	-.012**
	(.3255)	(.3187)	(.3332)	(.003)
White	0.785	0.767	0.805	-.038**
	(.4111)	(.4228)	(.3962)	(.0038)
Black	0.133	0.157	0.106	.051**
	(.3399)	(.3636)	(.308)	(.0031)
Asian	0.0250	0.0220	0.0270	-.005**
	(.155)	(.1479)	(.1627)	(.0014)
Hispanic	0.0660	0.0670	0.0650	0.00200
	(.2485)	(.2503)	(.2463)	(.0023)
No. households	47012	33628	13384	20244

This table displays demographic data of men and women constituting single-member households as well as their differences. These figures and their corresponding gender-differences were computed using the proprietary analytic household weights included in the Nielsen Consumer Panel Survey. Dollar amounts are expressed in USD 2016.
$*p < .05$, $**\,p < .01$

Table A.2: Demographics of CE PUMD single-member households

_	Total	Women	Men	Difference
Income	30530	26950	34665	-7715.418**
	(42896.3)	(36923.05)	(48568.25)	(335.0263)
Age	54.72	58.93	49.86	9.071**
	(20.2861)	(20.2295)	(19.2376)	(.1516)
High school	0.482	0.478	0.486	-.008*
	(.4997)	(.4995)	(.4998)	(.0038)
College	0.284	0.278	0.291	-.013**
	(.4508)	(.448)	(.4541)	(.0035)
Post-grad	0.0980	0.103	0.0920	.01**
	(.2971)	(.3035)	(.2894)	(.0023)
White	0.792	0.788	0.797	-.009**
	(.4058)	(.4086)	(.4024)	(.0031)
Black	0.146	0.152	0.140	.012**
	(.3536)	(.3591)	(.3469)	(.0027)
Asian	0.0400	0.0390	0.0410	-0.00200
	(.1957)	(.1937)	(.198)	(.0015)
Hispanic	0.0830	0.0750	0.0920	-.017**
	(.2761)	(.2636)	(.2895)	(.0021)
No. observations	67950	36417	31533	4884

This table displays demographic data of men and women constituting single-member households as well as their differences. Dollar amounts are expressed in USD 2016.
$*p < .05$, $**p < .01$

Table A.3: Nielsen panelist behavior per month

	Total	Women	Men	Difference
Months in Panel	53.35 (48.378)	50.85 (46.675)	56.26 (50.1261)	-5.407** (.4468)
Trips	9.395 (6.5983)	9.018 (6.0547)	9.833 (7.1526)	-.815** (.0609)
Spending	258.8 (177.0685)	259.6 (175.8798)	257.9 (178.4388)	1.644 (1.6378)
Spending inc. share	0.0120 (.0208)	0.0140 (.0235)	0.0100 (.017)	.004** (.0002)
Purchases	53.95 (32.122)	55.78 (32.2948)	51.84 (31.7906)	3.941** (.2966)
Unique products	25.67 (14.7973)	28.44 (15.2127)	22.45 (13.6116)	5.985** (.1341)
Unique modules	6.597 (15.3426)	7.516 (16.422)	5.531 (13.9114)	1.986** (.1416)
Unique groups	3.500 (7.0203)	3.955 (7.3166)	2.973 (6.6215)	.982** (.0648)
Coupon value	11.65 (15.3496)	12.80 (15.6305)	10.31 (14.9068)	2.487** (.1415)
Coupon use	8.229 (5.4355)	9.159 (5.6248)	7.150 (4.995)	2.009** (.0494)
Deal use	2.972 (2.1307)	3.223 (2.2144)	2.682 (1.9902)	.541** (.0196)

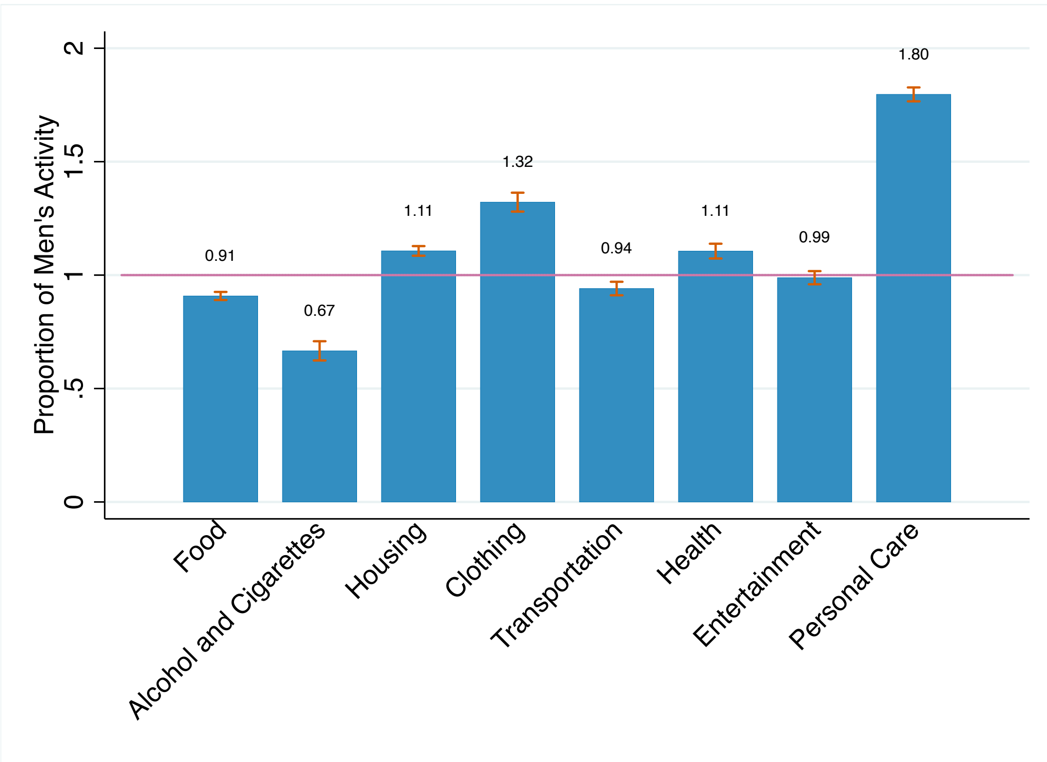
This table features shopping behavior of single-individual household Nielsen panelists per month and unconditional differences between genders. Monetary values are expressed in 2016 USD.
$*p < .05$, $**p < .01$.

Table A.4: Nielsen panelist behavior per shopping trip

	Total	Women	Men	Difference
Spending	25.61	26.82	24.46	2.357**
	(34.2295)	(35.1908)	(33.2481)	(.013)
Spending inc. share (%)	0.104	0.123	0.0860	.037**
	(.2522)	(.2911)	(.207)	(.0001)
Purchases	5.402	5.851	4.974	.877**
	(6.7014)	(7.1709)	(6.1916)	(.0025)
Unique products	5.183	5.613	4.773	.84**
	(6.341)	(6.806)	(5.8349)	(.0024)
Unique modules	4.507	4.869	4.163	.707**
	(5.2263)	(5.6165)	(4.8006)	(.002)
Unique groups	3.884	4.160	3.622	.538**
	(4.0665)	(4.3455)	(3.7633)	(.0015)
Coupon value	0.731	0.873	0.596	.277**
	(3.321)	(3.7914)	(2.7942)	(.0013)
Coupon use	0.398	0.470	0.330	.14**
	(1.5169)	(1.6698)	(1.3519)	(.0006)
Deal use	1.347	1.530	1.173	.357**
	(3.0739)	(3.333)	(2.7942)	(.0012)

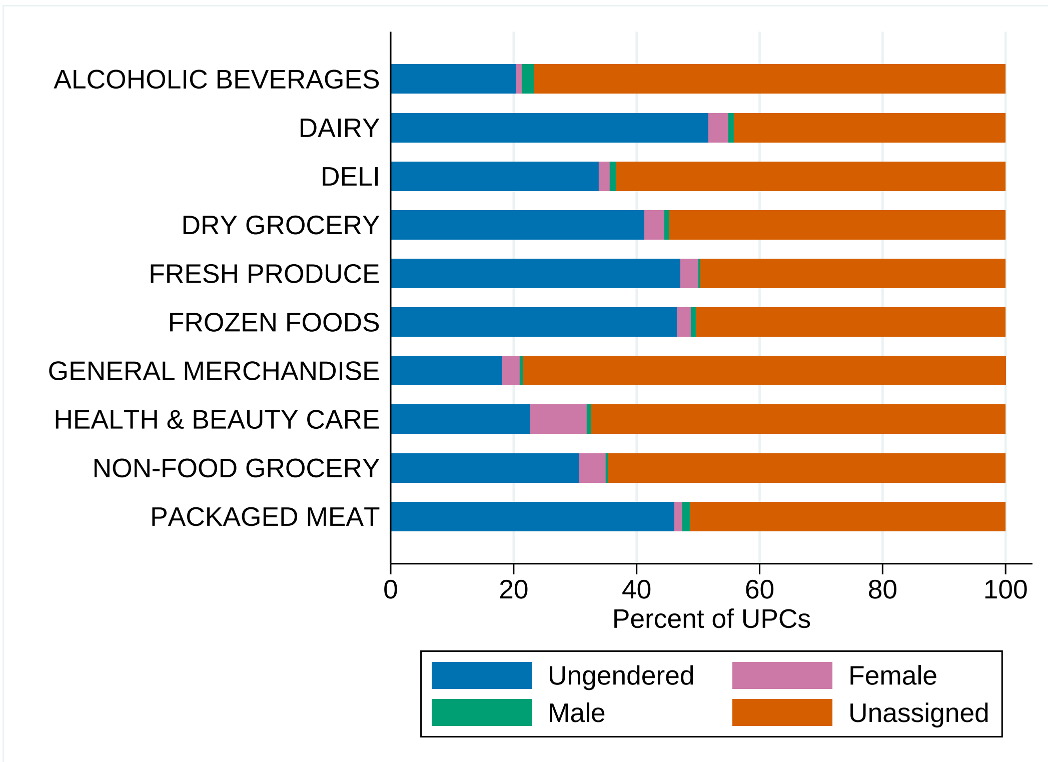
This table features descriptive statistics of shopping behavior of single-individual household Nielsen panelists per trip and unconditional differences between genders. Monetary values are expressed in 2016 USD.
$*p < .05$, $**p < .01$.

Figure A.1: Women's yearly consumption spending relative to men's



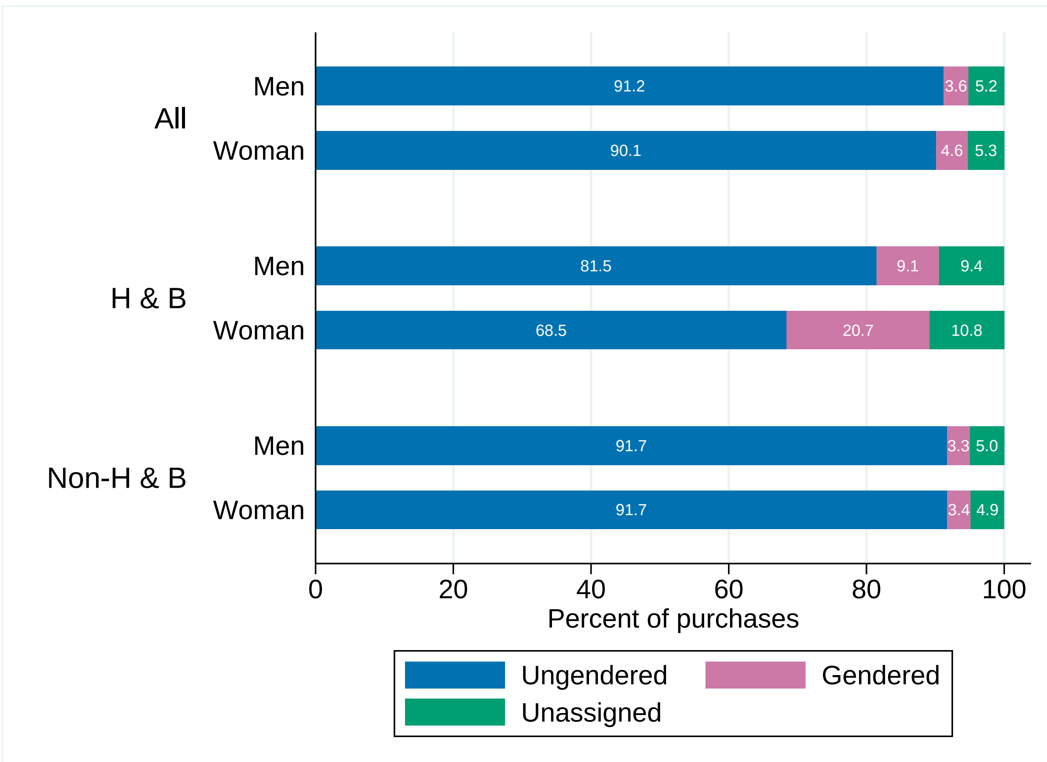
Note: This figure plots the coefficients estimated from a regression of log expenditure on an indicator for the individual being a woman and demographic controls: $\log y_{it} = \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_{it} + \varepsilon_{it}$, for spending categories food, alcohol and cigarettes, housing, clothing, transportation, health entertainment and personal care using the CE PUMD from 2010 to 2017. $\mathbb{1}\{female_i = 1\}$ is an indicator for whether the individual is a woman, and X_{it} is a vector of time- and time-id-varying controls including income, age, race and education. Standard errors are clustered at the individual-level.

Figure A.2: Assigned UPC-gender Across Departments



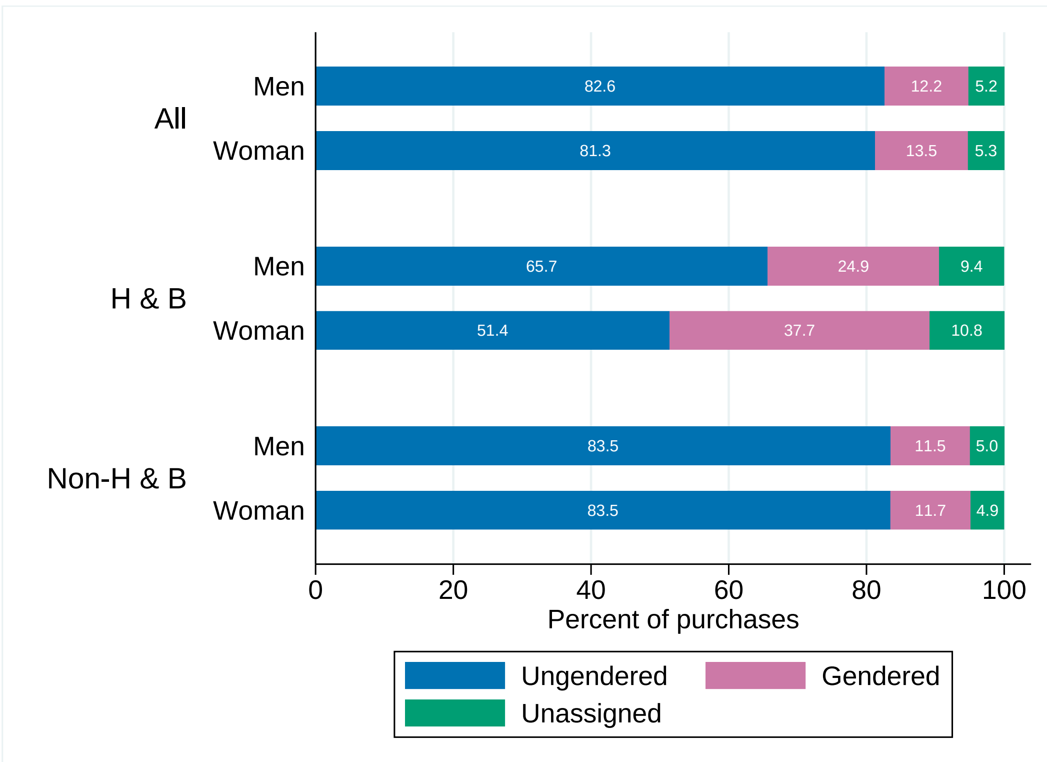
Note: This figure plots the percentage distribution of UPCs assigned to Ungendered, Female, and Male across departments. We restrict to UPCs that are observed with great enough purchase frequency to be assigned a UPC-gender with false positive probability of 5% . Unassigned UPCs are those excluded by the purchase cutoff.

Figure A.3: Consumption Basket Composition by Product Gender



Note: This figure presents plots the decomposition of purchases made by men and women into gendered, ungendered and unassigned products. The first rows show this for all product departments while the next two separate out health and beauty products.

Figure A.4: Consumption basket composition as share of purchases, 75-25 Cutoff



Note: This figure presents plots the decomposition of purchases made by men and women into gendered, ungendered and unassigned products. The first rows show this for all product departments while the next two separate out health and beauty products.

Table A.5: Yearly spending differences between men and women

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0162**	0.0248***	0.0444***	0.0616***	0.0678***	0.0678***
	(0.0080)	(0.0077)	(0.0076)	(0.0076)	(0.0075)	(0.0075)
Men's Average	\$2,423.58	\$2,423.58	\$2,423.58	\$2,423.58	\$2,423.58	\$2,423.58
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	Yes	Yes	Yes	Yes	Yes
Income FE	No	No	Yes	Yes	Yes	Yes
Age FE	No	No	No	Yes	Yes	Yes
Race FE	No	No	No	No	Yes	Yes
Education FE	No	No	No	No	No	Yes
Adj. R-squared	0.02	0.10	0.11	0.13	0.13	0.13
N	216,890	216,743	216,743	216,742	216,742	216,742
Number of clusters	46,968	46,852	46,852	46,851	46,851	46,851

Note: This table presents estimates of the percent difference in yearly spending between men and women using the following regression: $\log y_{it} = \phi_t + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_i + \varepsilon_{it}$, where y_{it} is yearly spending. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. Standard errors are clustered at the individual-level. Column 1 can be thought of as a raw gap between single men and single women, each subsequent column demonstrates the contribution of controlling for an additional demographic factor.

Table A.6: Yearly differences in number of unique products by consumer gender

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.2678***	0.2654***	0.2655***	0.2754***	0.2760***	0.2760***
	(0.0077)	(0.0073)	(0.0074)	(0.0074)	(0.0074)	(0.0074)
Men's Average	5.620	5.620	5.620	5.620	5.620	5.620
	(0.583)	(0.583)	(0.583)	(0.583)	(0.583)	(0.583)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE		Yes	Yes	Yes	Yes	Yes
Income FE			Yes	Yes	Yes	Yes
Age FE				Yes	Yes	Yes
Race FE					Yes	Yes
Education FE						Yes
Adj. R-squared	0.06	0.15	0.15	0.16	0.16	0.16
N	216890	216743	216743	216742	216742	216742
Number of clusters	46968	46852	46852	46851	46851	46851

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: This table presents estimates of the percent difference in total unique items purchased between men and women using the following regression: $\log y_{it} = \phi_t + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_i + \varepsilon_{it}$, where y_{it} is yearly spending. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. Standard errors are clustered at the individual-level. Column 1 can be thought of as a raw gap between single men and single women, each subsequent column demonstrates the contribution of controlling for an additional demographic factor.

Table A.7: Yearly differences in total items purchased by consumer gender

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.2678***	0.2654***	0.2655***	0.2754***	0.2760***	0.2760***
	(0.0077)	(0.0073)	(0.0074)	(0.0074)	(0.0074)	(0.0074)
Men's Average	5.620	5.620	5.620	5.620	5.620	5.620
	(0.583)	(0.583)	(0.583)	(0.583)	(0.583)	(0.583)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE		Yes	Yes	Yes	Yes	Yes
Income FE			Yes	Yes	Yes	Yes
Age FE				Yes	Yes	Yes
Race FE					Yes	Yes
Education FE						Yes
Adj. R-squared	0.06	0.15	0.15	0.16	0.16	0.16
N	216890	216743	216743	216742	216742	216742
Number of clusters	46968	46852	46852	46851	46851	46851

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: This table presents estimates of the percent difference in total items purchased between men and women using the following regression: $\log y_{it} = \phi_t + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_i + \varepsilon_{it}$, where y_{it} is yearly spending. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. Standard errors are clustered at the individual-level. Column 1 can be thought of as a raw gap between single men and single women, each subsequent column demonstrates the contribution of controlling for an additional demographic factor.

Table A.8: Price paid per good unit by department

	Total	Women	Men	Difference	Log difference
All departments	1.737	1.859	1.601	.258**	.091**
	(21.7607)	(27.6795)	(12.0798)	(.0035)	(.0003)
Health and beauty	5.907	7.442	3.541	3.901**	.261**
	(76.3566)	(95.0937)	(29.4796)	(.0488)	(.0013)
Dry grocery	0.302	0.317	0.286	.031**	.109**
	(3.0293)	(1.6098)	(4.0436)	(.0007)	(.0003)
Frozen foods	0.983	0.993	0.972	.021**	.056**
	(2.7548)	(2.7258)	(2.7834)	(.0015)	(.0007)
Dairy	0.419	0.432	0.405	.027**	.142**
	(1.0206)	(1.0247)	(1.0158)	(.0005)	(.0006)
Deli	3.101	3.011	3.188	-.176**	-.004**
	(5.5958)	(5.5005)	(5.6842)	(.005)	(.0015)
Packaged meat	0.606	0.617	0.597	.021**	.071**
	(1.3595)	(1.3252)	(1.388)	(.0014)	(.001)
Fresh produce	1.474	1.473	1.476	-0.00200	.002*
	(2.2024)	(2.2308)	(2.1655)	(.0014)	(.0008)
Non-food grocery	1.210	1.243	1.164	.079**	-.058**
	(17.1235)	(17.4589)	(16.6564)	(.0099)	(.001)
Alc. beverages	2.092	1.997	2.143	-.146**	-.283**
	(4.7644)	(4.3439)	(4.9772)	(.0072)	(.0039)
General merch.	9.850	8.777	11.12	-2.348**	-.238**
	(31.9754)	(32.0002)	(31.899)	(.0247)	(.0015)

This table displays per-unit prices within each department as well as the descriptive difference in per-unit prices calculated for men's and women's purchases separately. Level units are expressed as 2016 USD per unit-amount.
$*p < .05$, $**p < .01$.

Table A.9: Most Popular Brands by Product Gender - Deodorant

Ungendered	**Woman Gendered**	**Man Gendered**
Arrid	Secret	Mennen Speed Stick
Sure	Mennen Lady Speed Stick	Right Guard Sport
Ban Classic	Degree	Old Spice High Endurance
Arm & Hammer UltraMax	Dove	Gillette
Suave	Mitchum for Women	Old Spice

Table A.10: Prices paid across departments

	(1) H&B	(2) Dry Groc.	(3) Frozen	(4) Dairy	(5) Deli	(6) Pack. Meat	(7) Produce	(8) Non-food Groc.	(9) Alcohol	(10) Gen. Merch.
Panel A: Per unit prices within product module										
Women	0.0554***	0.0613***	0.0527***	0.0415***	0.0451***	0.0516***	0.0191***	0.0349***	-0.1594***	-0.0489***
	(0.0031)	(0.0023)	(0.0028)	(0.0024)	(0.0060)	(0.0035)	(0.0040)	(0.0022)	(0.0227)	(0.0040)
Men's Average	$0.391	$0.135	$0.239	$1.30	$0.807	$0.280	$.736	$0.191	$0.120	$2.694
MURLM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.83	0.85	0.90	0.93	0.90	0.79	0.78	0.87	0.68	0.78
N	10,474,703	57,155,850	14,244,360	16,588,106	5,261,920	4,090,568	10,753,409	12,456,085	1,989,976	6,712,521
Number of clusters	49,155	49,247	48,965	49,073	47,770	47,619	47,945	49,166	36,103	48,979
Panel B: Per unit price for same UPC										
Women	-0.0211***	-0.0035***	-0.0049***	-0.0014	-0.0258***	-0.0063***	-0.0158***	-0.0134***	-0.0020	-0.0032
	(0.0022)	(0.0009)	(0.0014)	(0.0010)	(0.0055)	(0.0016)	(0.0030)	(0.0010)	(0.0019)	(0.0049)
Men's Average	$4.428	$1.874	$2.904	$2.098	$3.597	$3.010	$1.779	$3.370	$8.819	$6.567
URLY FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.82	0.89	0.84	0.88	0.64	0.81	0.65	0.93	0.93	0.85
N	9,654,273	61,381,504	12,796,632	15,383,639	5,371,930	3,918,400	10,743,008	10,662,589	1,889,857	6252419
Number of clusters	49,126	49,246	48,920	49,043	47,661	47,501	47,855	49,127	35,122	48,917

This table estimates $log(P_{ijt}) = \phi_t + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$, stratifying by department across columns. P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman and and X_i is a vector of demographic controls including income, county, age, race. In panel A, ϕ_t is a vector of fixed effects for the interaction of product module, units, retailer chain, county, and half-year. In Panel B ϕ_t is a vector of fixed effects for the interaction of product (UPC), retailer chain, county, and half-year. "MURLM FE" refers to Module × Unit × Retailer × County × Month fixed effects; "URLY" refers to UPC × Retailer × County × Year. Standard errors are clustered at the household-level.

Table A.11: Unit prices in same product module by UPC and consumer gender, 75-25 Cutoff

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
Female	0.0322***	0.0182***	0.0286***
	(0.0019)	(0.0032)	(0.0017)
Gendered UPC	0.0084***	0.0817***	0.0066***
	(0.0022)	(0.0038)	(0.0024)
Female × Gendered UPC	0.0848***	0.1002***	0.0648***
	(0.0024)	(0.0042)	(0.0026)
Men's Ungendered Average	-1.529	-0.999	-1.551
Male SD	1.571	1.904	1.556
MURLM FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes
Adj. R-squared	0.88	0.84	0.89
N	131501221	9299678	120478978
Number of clusters	49256	49127	49256

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: This table presents estimates from the regression: $log(P_{ijt}) = \phi_{t(j)} + \beta_1 \mathbf{1}_{w(i)} + \beta_2 \mathbf{1}_{g(j)} + \beta_3 \mathbf{1}_{w(i)} \cdot \mathbf{1}_{g(j)} + \gamma X_i + \epsilon_{ijt}$. $\phi_{t(j)}$ is a vector of fixed effects for the interaction of product module, units denomination, retailer chain, county, and half-year. X_i includes with demographic controls for income, age, race and education. Columns 2 and 3 separate out Health and Beauty products. This table corresponds to table 6 in the paper but with the gendered product cutoff at 25-75 rather than 10-90. "MURLM FE" refers to module×unit×retailer×county×month fixed effects.

A.2 Additional CES model and results

A.2.1 Additional CES model setup

Equation (4) estimates the elasticity of substitution across products within the same module-market but does not explicitly estimate the price elasticity of demand. We now derive overall price elasticities associated with our model in terms of the elasticity of substitution, $\sigma_t(g)$, and market share, $s_{jt}(g)$. Solving Equation (1) yields:

$$q_{jt}(g) = \left(P_t(g) \frac{\varphi_{jt}(g)}{p_{jt}} \right)^{\sigma_t(g)-1} \frac{\alpha_t(g) E(g)}{p_{jt}}$$

Where $P_t(g)$ is a price index, $P_t(g) = \left[\sum_{i \in G_t} p_{jt}^{(1-\sigma_t(g))} \varphi_{jt}(g)^{(\sigma_t(g)-1)} \right]^{\frac{1}{1-\sigma_t(g)}}$.

Firms price their products in response to the sales weighted average demand elasticity that they face across the population:

$$\mu_{jt} = \frac{p_{jt} - c_{jt}}{p_{jt}} = \frac{\sum_g x_{jt}(g)}{\sum_g \varepsilon_{jt}(g) x_{jt}(g)}.$$

Where $x_{jt}(g)$ is the sales of product j to gender g in market t. Because we can only attribute purchases to a gender for single individuals, we are limited to extrapolating the results from our singles to the whole population.

A.2.2 Additional CES results

Table A.12: OLS Elasticities

	(1) County-Half Year	(2) County-Retailer-Half Year
$\sigma_m - \sigma_w$	-0.0181***	-0.0073*
	(0.0065)	(0.0044)
$1 - \sigma_m$	0.6886***	0.7784***
	(0.0170)	(0.0075)
Observations	17,010,404	14,939,386
Adjusted R^2	0.016	0.000
MTCRG FE	Yes	Yes

UPC-County level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents the results of estimating elasticities of substitution by regressing changes in the log budget share of a product on changes in log price: $\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$. "MTCRG FE" refers to module×half-year×county×retailer×gender fixed effects.

Table A.13: First stage results of price change instruments

	(1) Hausman	(2) DellaVigna-Gentzkow
Hausman	0.3280***	
	(0.0044)	
DellaVigna-Gentzkow		0.2215***
		(0.0036)
Observations	16,351,076	11,018,742
Adjusted R^2	0.008	0.006
F-statistic	24634	20712
MTCRG FE	Yes	Yes
County IV	Yes	No
Retailer IV	No	Yes

UPC-County level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents the results of of regressing average prices changes on the UPC-locality-half-year level on Hausman and DellaVigna-Gentzkow instruments "MTCRG FE" refers to module×half-year×county×retailer×gender fixed effects.

Table A.14: Reduced form results

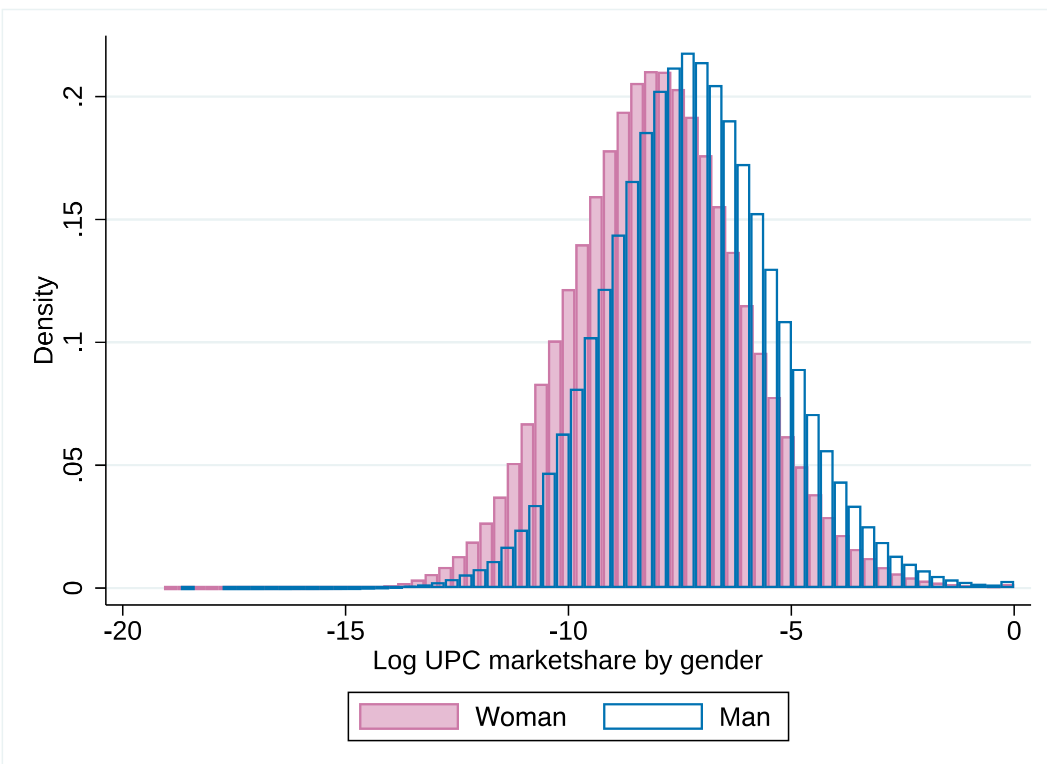
	(1) Hausman	(2) DellaVigna-Gentzkow
$\Delta \log P_{IV}$	0.0650***	0.0568***
	(0.0077)	(0.0052)
Woman $\times \Delta \log P_{IV}$	0.0102	-0.0198***
	(0.0076)	(0.0060)
Observations	16,336,260	11,007,333
Adjusted R^2	-0.022	-0.052
MTCRG FE	Yes	Yes
County IV	Yes	No
Retailer IV	No	Yes

UPC-County level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

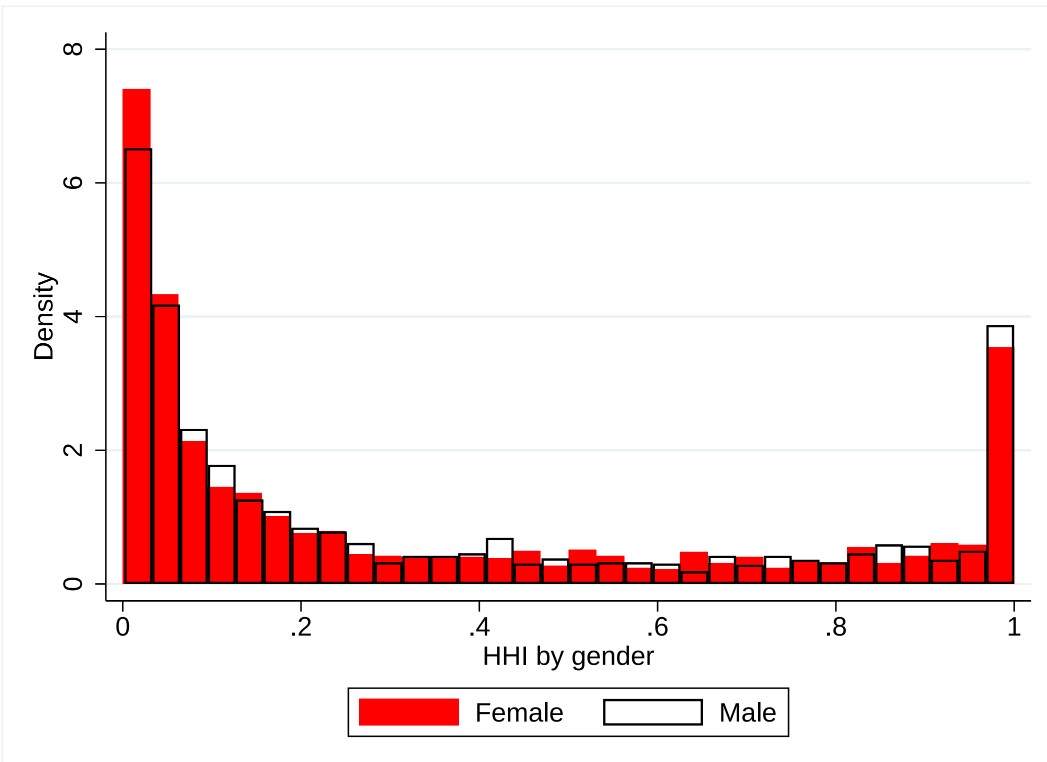
This table presents the reduced-form of regressing changes in consumption of goods on the UPC-locality-half-year level on Hausman and DellaVigna-Gentzkow instruments "MTCRG FE" refers to module×half-year×county×retailer×gender fixed effects.

Figure A.5: Market Competition by Gender (UPC-level)



Note: This figure presents histograms of log market share of UPC for men and women separately. Market-shares are computed on the UPC×consumer-gender×half-year level.

Figure A.6: Market Competition by Gender (Module-level)



Note: This figure presents histograms of Hirschfield-Herfindahl Index (HHI) observations of product markets for men and women separately. Individual HHI observations are computed on the module×consumer-gender×half-year level and constructed from UPC×consumer-gender×half-year level data.

A.3 Additional evidence on retail markups

How do our results on retailer markups inform the mechanisms underlying the pink tax? We find that gendered differences in unit prices nearly disappear and possibly become negative after conditioning on retailer costs. Additionally, we find no average difference in retailer markups paid by men and women and even greater retail markups on male-gendered goods than on female-gendered goods. Setting aside concerns on external validity to goods that do not match to PriceTrak and goods under alternative vertical integration settings, under what conditions would these results imply that all of the pink tax is attributable to differences manufacturing cost?

Consider a decomposition of average difference in prices paid for a female and male good:

$$
\begin{aligned}
&\mathbb{E}[\Delta\%p] = \mathbb{E}[\Delta\%c + \Delta\%\mu^m + \Delta\%\mu^d + \Delta\%\mu^r] > 0 \\
&\implies \mathbb{E}[\Delta\%c] + \mathbb{E}[\Delta\%\mu^m] + \mathbb{E}[\Delta\%\mu^d] + \underbrace{\mathbb{E}[\Delta\%\mu^r]}_{\leq 0} > 0 \\
&\implies \mathbb{E}[\Delta\%c] > -\Big(\underbrace{\mathbb{E}[\Delta\%\mu^m] + \mathbb{E}[\Delta\%\mu^d]}_{\text{Not observed}} + \underbrace{\mathbb{E}[\Delta\%\mu^r]}_{\leq 0} \Big).
\end{aligned}
$$

In rationalizing the observed pink tax in this setting, it is necessarily the case that $\mathbb{E}[\Delta\%c] > 0$ if the expected sum of $\mathbb{E}[\Delta\%\mu^m] + \mathbb{E}[\Delta\%\mu^d] < 0$. Therefore, a natural question to ask is: given our observation that $\mathbb{E}[\Delta\%\mu^r] \leq 0$, is it reasonable to suppose also that $\mathbb{E}[\Delta\%\mu^m] + \mathbb{E}[\Delta\%\mu^d] < 0$? This condition is implied by the sufficient but not necessary condition in Equation (6), but this condition is less restrictive and amounts to a bounding condition: that on average, average differences in markups by UPC-gender are of identical sign along the vertical integration chain. We explore this question in a simplified section in Section A.3.1. We find that absent gender-differential competitive environments between layers, markups at each production/supply layer are set according to ultimate consumer demand, so that is in fact likely that $\mathbb{E}[\Delta\%\mu^r]$ and $\mathbb{E}[\Delta\%\mu^m] \leq 0$ given our observation that $\mathbb{E}[\Delta\%\mu^d] < 0$; i.e. it is likely the case that $\mathbb{E}[\Delta\%c] > 0$.

A.3.1 Markups in a vertically integrated setting

Consider the environment of vertical integration as in Section 5. To explore the bounding problem of the average gender difference in manufacturer and distributor markups, we want to explore under what conditions they differ in sign from the average gender difference in retailer markups.

We suppose a manufacturer that manufacturers a female and male good, a wholesaler/distributor, and a retailer that sells to a final consumer. We maintain this structure of vertical integration in order to align with the PriceTrak data as well as our discussion in Section 5, but this discussion generalizes to other vertical integration structures as well, such as one with a distinct wholesaler and distributor.

A single manufacture produces two goods to respective gender demand bases $h \in \{f, m\}$ separately at marginal costs c_m and c_f. The final consumer demand bases for these products exhibit iso-elastic price-sensitivity ε_m, and ε_f respectively in a manner independent of consumption of the other good.

The manufacturer sells both products to a single wholesaler/distributor, the wholesaler/distributor sells these products a single retailer, and the retailer resells these products as final goods to the ultimate consumers.

The manufacturer's problem is

$$\max_{p_m^m, p_f^m, Q_m, Q_f} (p_m^m - c_m)Q_m + (p_f^m - c_f)Q_f,$$

consisting of the price and quantity combination that maximizes rents from the retailer. Let superscripts refer to stages of the production process ($k \in m$ for manufacturer, d for distributor, and r for retailer) and subscripts refer to UPC-gender.

The distributor takes marginal costs as exogenously given with $c_m^d = p_m^m$ and $c_f^d = p_f^m$.

The distributor's problem follows a similar structure in selling the goods to a retailer:

$$\max_{p_m^d, p_f^d} (p_m^d - c_m^d)Q_m + (p_f^d - c_f^d)Q_f.$$

Finally, the retailer sets prices in selling to the final consumer with $c_m^r = p_m^d$ and $c_f^r = p_f^d$:

$$\max_{p_m, p_f, Q_m, Q_f} (p_m - c_m^r)Q_m(p_m) + (p_f - c_f^r)Q_f(p_f),$$

facing their differentiated iso-elastic consumer bases and taking wholesaler prices as exogenous. We define the final price as the retail price $p_h := p_h^r$.

The setup yields a standard multi-marginalization problem with the each stage setting prices according to a standard Lerner markup rule:[38]

$$p_h^k = c_h^k (1 - \frac{1}{|\varepsilon_h|})^{-1},$$

for each stage of the production process $k \in \{m, d, r\}$. This price-setting process results in a final price to consumers of

$$p_h = c_h (1 - \frac{1}{|\varepsilon_h|})^{-3}.$$

We are interested in knowing whether it is possible to observe the following simultane-

[38]Quantities are ultimately set by the consumer. Because firms linearly maximize profit, each stage internalizes consumers' down-the-line demand response to prices.

ously:

$$
\begin{aligned}
p_f^r := p_f > p_h =: p_h^r,\\
c_f^r = p_f^d > c_m^r = p_m^d = c_m^r,\\
\frac{p_f^r}{c_f^r} - 1 = \mu_f^r < \mu_m^r = \frac{p_m^r}{c_m^r} - 1,\\
|\varepsilon_f| > |\varepsilon_m|,
\end{aligned}
$$

and

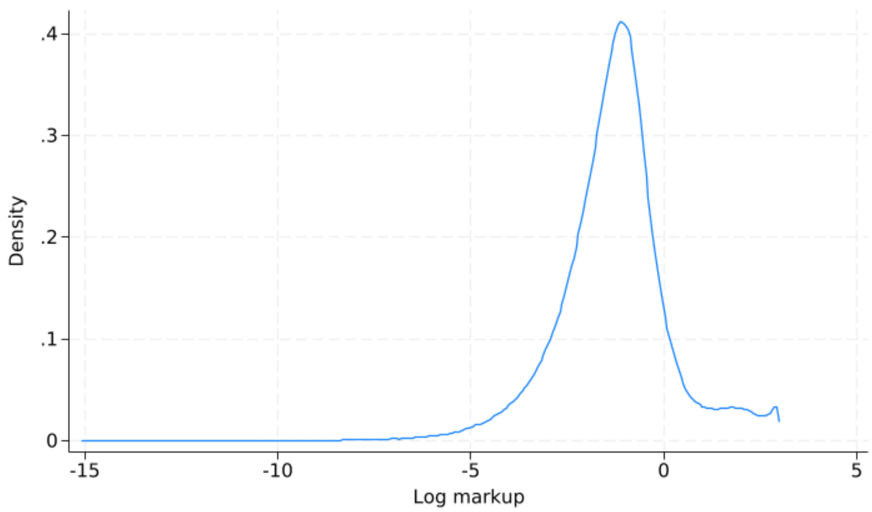
$$
\mu_f^m + \mu_f^d < \mu_m^m + \mu_m^d.
$$

I.e. given our observation that 1) female prices exceed male prices, 2) female retail costs exceed male retail costs, 3) male retail markups exceed female retail markups, and 4) elasticity of demand on the female goods exceeds that of the male good in absolute value, can it be the case that the sum of manufacturer and distributor markup for female goods exceeds that of male goods? In this simplified environment, the answer is no. Without alternate assumptions on the structure of competition within and between layers, greater retailer markup and low elasticity on part of women implies a

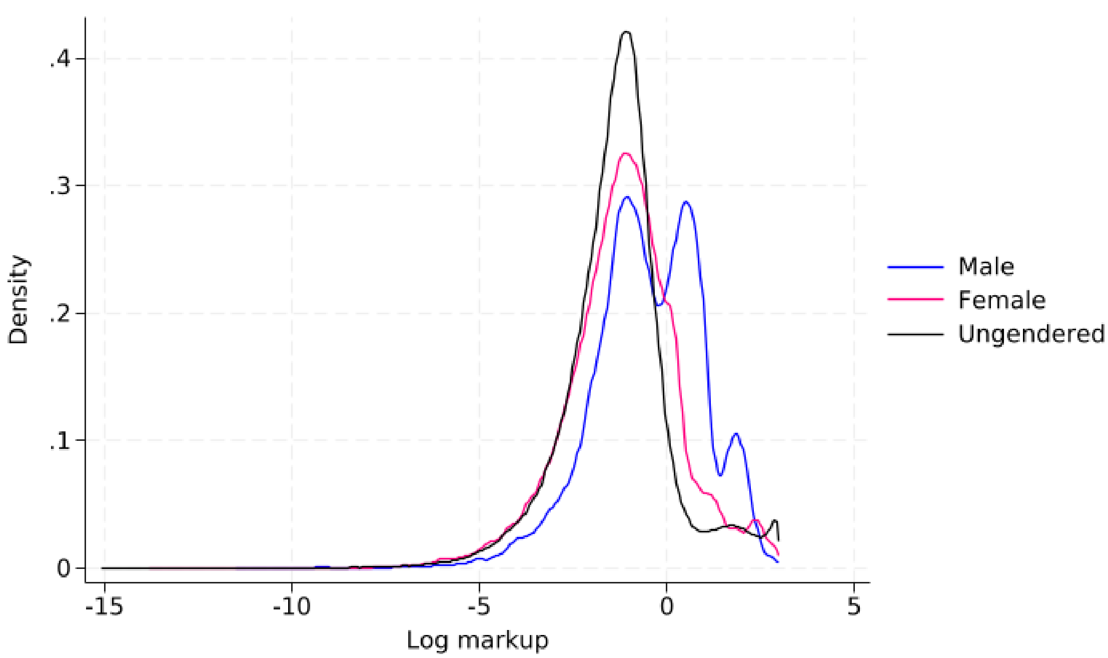
A.3.2 Additional figures and tables on retailer costs and markups

Figure A.7: Distribution of markups

(a) Markups



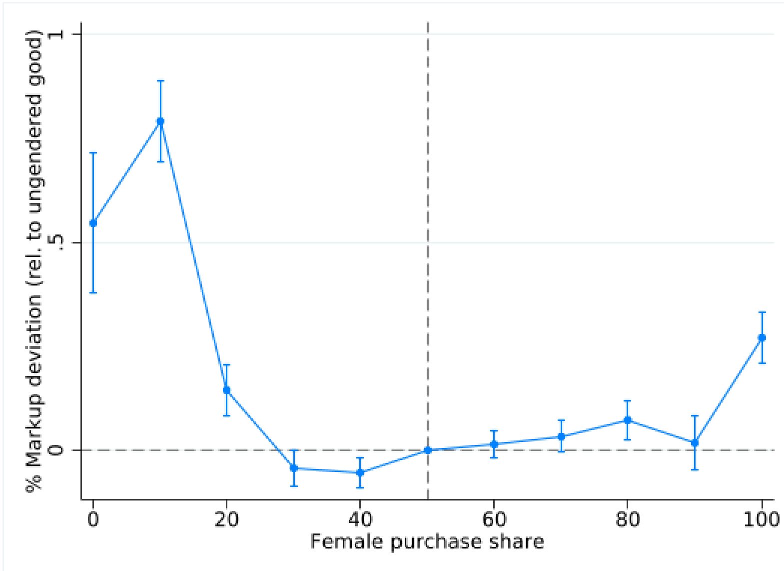
(b) Markups by UPC-gender



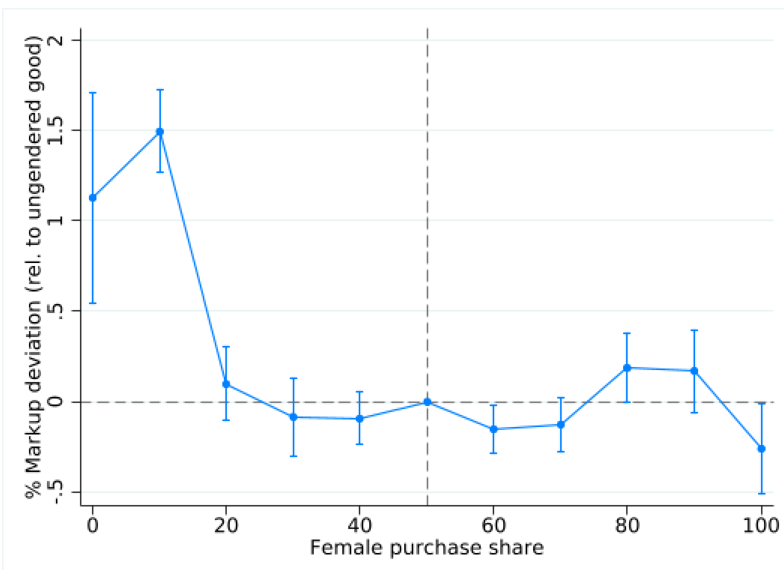
Note: These figures display the estimated probability density functions of log markups estimated using an Epanetchnikov kernel. Panel (a) plots the distribution of log markups; Panel (b) plots the distribution of log markups by UPC-gender. "Female" and "male" goods refer to UPCs purchased 90% more by women or men respectively. "Ungendered" goods are purchased between 40 and 60 percent by men/women.

Figure A.8: Unconditional markup by female purchase intensity

(a) Unweighted



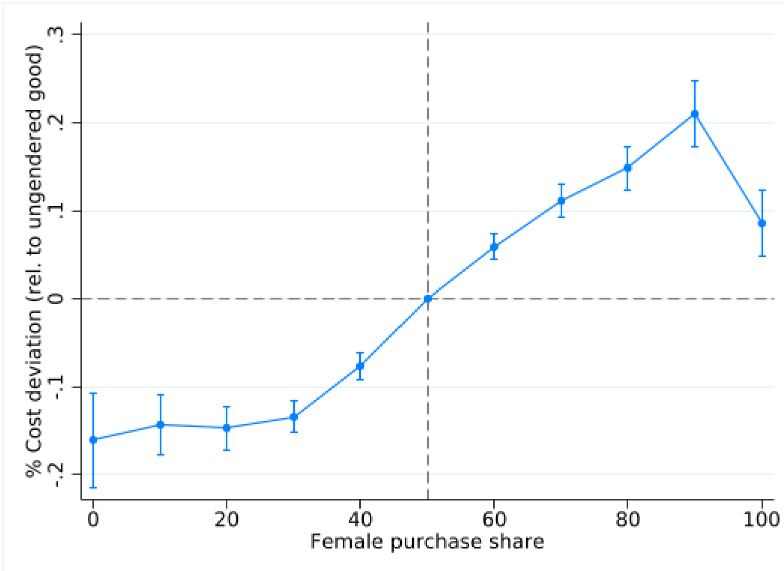
(b) Weighted by expense recorded in Nielsen HMS



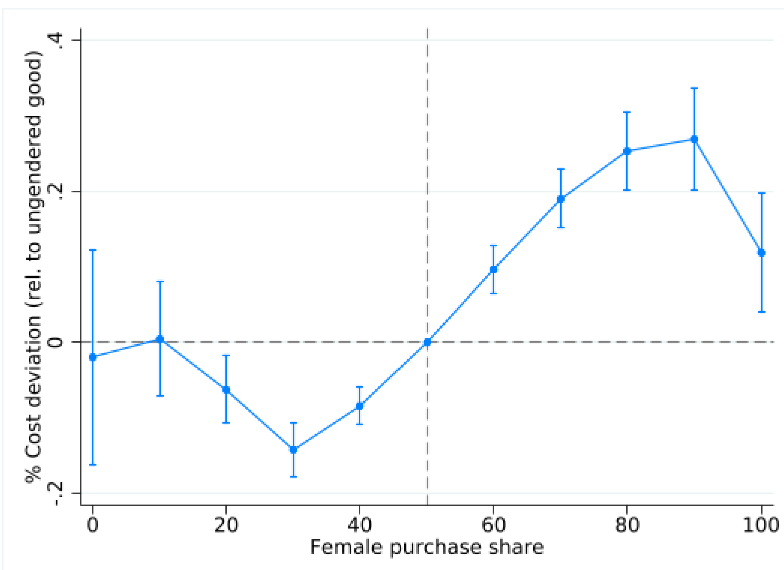
: These figures display the coefficients estimated from the following regression on the UPC-year level: $\log \mu_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Markup μ is constructed using PriceTrak data on wholesale prices and Nielsen final sale prices. Bins $b \in \mathcal{B}$ represent ten-percentile-width bins centered at multiples of 10 (truncated at 0 and 100) partitioning the interval $[0, 1]$; these bins reflect the aggregate amount of a UPC purchased by single women (as opposed to single men). The regression includes year fixed effects. Coefficients γ_b are estimated relative to goods in the same product module purchased at approximate gender-parity (between 45 and 55%). Panel (a) estimates this regression with equal weighting for all observations. Panel (b) presents the coefficients estimated from an analogous regression with analytic weights on UPC-year expenditure as recorded in the Nielsen HMS data. Standard errors are clustered at the UPC level.

Figure A.9: Distribution of retailer costs

(a) Unweighted



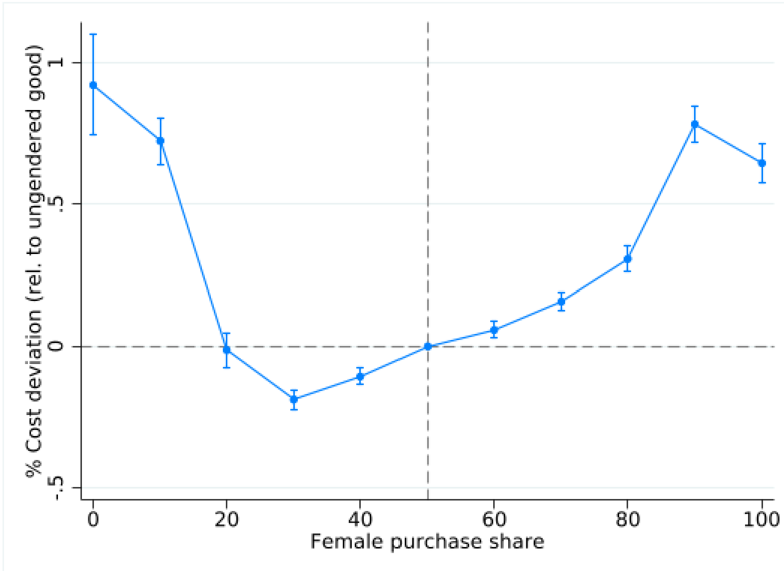
(b) Weighted by expense recorded in Nielsen HMS



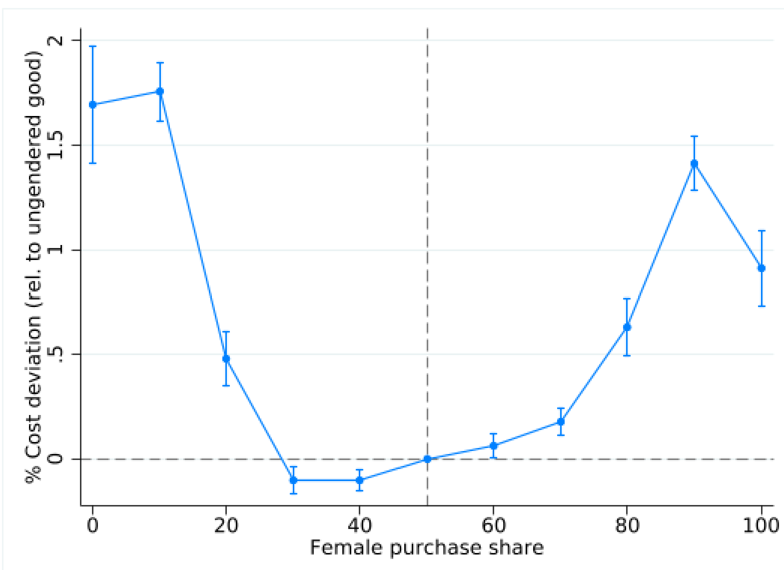
: These figures display the coefficients estimated from the following regression on the UPC-year level: $\log C_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Retailer cost C_{ut} μ observed from PriceTrak data. Bins $b \in \mathcal{B}$ represent ten-percentile-width bins centered at multiples of 10 (truncated at 0 and 100) partitioning the interval $[0, 1]$; these bins reflect the aggregate amount of a UPC purchased by single women (as opposed to single men). The regression includes year fixed effects. Coefficients γ_b are estimated relative to goods in the same product module purchased at approximate gender-parity (between 45 and 55%). Panel (a) estimates this regression with equal weighting for all observations. Panel (b) presents the coefficients estimated from an analogous regression with analytic weights on UPC-year expenditure as recorded in the Nielsen HMS data. Standard errors are clustered at the UPC level.

Figure A.10: Unconditional distribution of retailer costs

(a) Unweighted



(b) Weighted by expense recorded in Nielsen HMS



: These figures display the coefficients estimated from the following regression on the UPC-year level: $\log C_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Retailer cost C_{ut} μ observed from PriceTrak data. Bins $b \in \mathcal{B}$ represent ten-percentile-width bins centered at multiples of 10 (truncated at 0 and 100) partitioning the interval $[0, 1]$; these bins reflect the aggregate amount of a UPC purchased by single women (as opposed to single men). The regression includes fixed effects for product module, county and half-year. Coefficients γ_b are estimated relative to goods in the same product module purchased at approximate gender-parity (between 45 and 55%). Panel (a) estimates this regression with equal weighting for all observations. Panel (b) presents the coefficients estimated from an analogous regression with analytic weights on UPC-year expenditure as recorded in the Nielsen HMS data. Standard errors are clustered at the UPC level.

Table A.15: Log markup by purchaser gender, budgetshare-weighted

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0294**	-0.0174	-0.0181	-0.0049	-0.0011	-0.0048
	(0.0142)	(0.0121)	(0.0117)	(0.0045)	(0.0044)	(0.0043)
Male mean (percent markup)	89.3%	89.3%	89.3%	90.0%	91.6%	94.5%
Demographics	No	Yes	Yes	Yes	Yes	Yes
Module	No	No	Yes	Yes	Yes	Yes
County	No	No	No	Yes	Yes	Yes
Retailer	No	No	No	No	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	No	No	No	No	Yes
Adj. R-squared	0.00	0.46	0.46	0.69	0.75	0.78
N	17901420	17901327	17901327	17088442	15570832	14342604
Number of clusters	28412	28412	28412	28406	28400	28403

Individual-level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents estimates from the transaction-level regression: $log(\mu_{ijt}) = \phi_{t(j)} + \beta\mathbf{1}_{w(i)} + \Gamma X_i + \epsilon_{ijt}$ where μ_{ijt} is the retailer markup UPC inferred from the PriceTrak and Nielsen data. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect, and X_i is a vector of demographic controls including income, county, age, race and education. Observations are weighted by the transaction expense as a share of the individual's annual income. Standard errors are clustered at the individual-level.

Table A.16: Log markup by purchaser gender × department

	(1) H&B	(2) H&B	(3) Dry Groc.	(4) Dry Groc.	(5) Frozen	(6) Frozen	(7) Dairy	(8) Dairy	(9) Deli	(10) Deli
Female	-0.2651***	-0.0626***	-0.0054	0.0183***	0.0482***	0.0117*	0.0087	-0.0102	0.0294	-0.0239***
	(0.0146)	(0.0081)	(0.0096)	(0.0041)	(0.0134)	(0.0063)	(0.0192)	(0.0095)	(0.0198)	(0.0082)
Male mean (levels)	.73849409	.74317808	.71253358	.78596188	.31703158	.31037346	3.5213589	3.5524422	.4452701	.44226254
Module	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
County	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Retailer	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R-squared	0.01	0.69	0.00	0.73	0.00	0.53	0.00	0.81	0.01	0.63
N	755352	647325	9405824	6663221	2380817	2223118	1788708	1649061	612284	583801
Number of clusters	27572	27060	28395	28355	27794	27572	27933	27714	25155	24467

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

	(1) Pack. Meat	(2) Pack. Meat	(3) Produce	(4) Produce	(5) Non-food Groc.	(6) Non-food Groc.	(7) Alcohol	(8) Alcohol	(9) Gen. Merch.	(10) Gen. Merch.
Female	0.0203*	0.0067	0.0239	0.0137	-0.0437***	-0.0011	-0.3454	0.0000	0.0492***	-0.0312***
	(0.0117)	(0.0078)	(0.0249)	(0.0214)	(0.0110)	(0.0056)	(0.2787)	(.)	(0.0128)	(0.0106)
Male mean (levels)	.39090748	.37807859	2.7235146	2.8282553	.47068611	.45971075	.73357241	.84023755	1.0281722	.97306338
Module	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
County	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Retailer	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R-squared	0.01	0.43	0.02	0.74	0.00	0.59	0.43	0.97	0.01	0.51
N	752874	706334	259734	209314	1759695	1548076	171	87	351650	280888
Number of clusters	26190	25734	22113	19721	28147	27951	81	29	26403	25244

Individual level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents estimates from the transaction-level regression: $log(\mu_{ijt}) = \phi_{t(j)} + \beta\mathbf{1}_{w(i)} + \Gamma X_i + \epsilon_{ijt}$ where μ_{ijt} is the retailer markup UPC inferred from the PriceTrak and Nielsen data. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect, and X_i is a vector of demographic controls including income, county, age, race and education. Regressions are stratified by product department. All transaction-observations are given equal weight. Standard errors are clustered at the individual-level.

A.4 Additional BLP results

Table A.17: Log cost by purchaser gender

Panel (a): By transaction

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0262***	0.0283***	0.0434***	0.0484***	0.0404***	0.0429***
	(0.0048)	(0.0028)	(0.0028)	(0.0021)	(0.0017)	(0.0022)
Male mean (levels)	.29	.29	.29	.28	.27	.26
Module	No	No	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
County	No	No	No	Yes	Yes	Yes
Retailer	No	No	No	No	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	No	No	No	No	Yes
Adj. R-squared	0.00	0.76	0.76	0.84	0.88	0.89
N	18076261	18076169	18076169	17262606	15320158	9834314
Number of clusters	28412	28412	28412	28406	28394	28357

Panel (b): Budgetshare-weighted

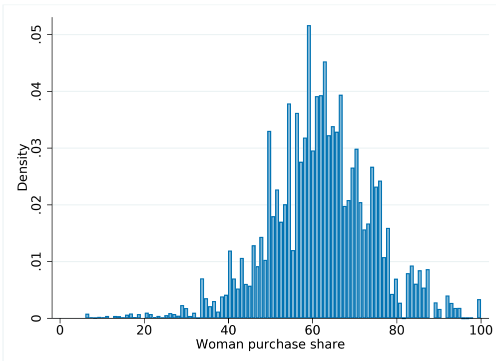
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0361***	0.0274***	0.0344***	0.0461***	0.0431***	0.0471***
	(0.0081)	(0.0040)	(0.0041)	(0.0026)	(0.0022)	(0.0027)
Male mean (levels)	.43	.43	.43	.38	.34	.31
Module	No	No	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
County	No	No	No	Yes	Yes	Yes
Retailer	No	No	No	No	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	No	No	No	No	Yes
Adj. R-squared	0.00	0.79	0.79	0.88	0.92	0.93
N	17901420	17901327	17901327	17088442	15151947	9690298
Number of clusters	28412	28412	28412	28406	28394	28356

Individual-level clustered standard errors in parentheses

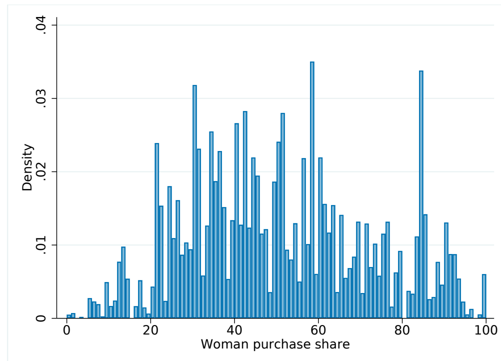
* $p < .10$, ** $p < .05$, *** $p < .01$

This table presents estimates from the transaction-level regression: $log(C_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \Gamma X_i + \epsilon_{ijt}$ where C_{jt} is the wholesale price of UPC j in year t as observed in PriceTrak. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect, and X_i is a vector of demographic controls including income, county, age, race and education. Panel (a) estimates this regression with equal weighting for all transaction-observations; Panel (b) weights each transaction-observations by the transaction expense as a share of the individual's annual income. Standard errors are clustered at the individual-level.

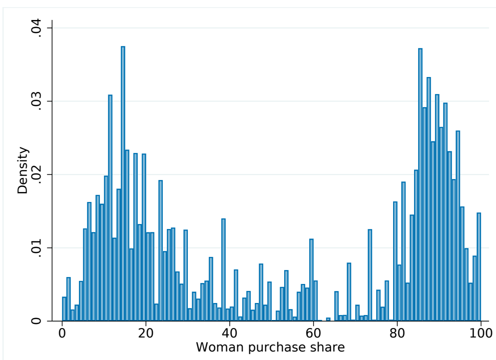
Figure A.11: Woman purchase share distribution



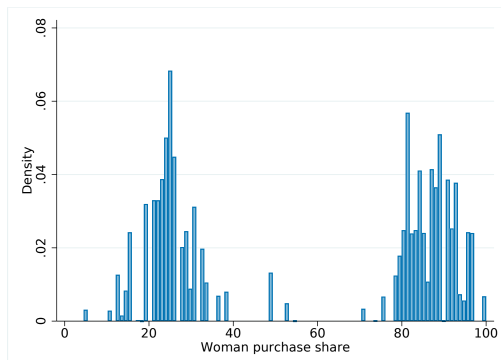
Yogurt



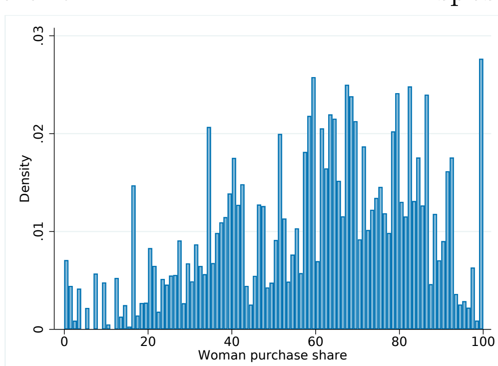
Protein Bars



Deodorant



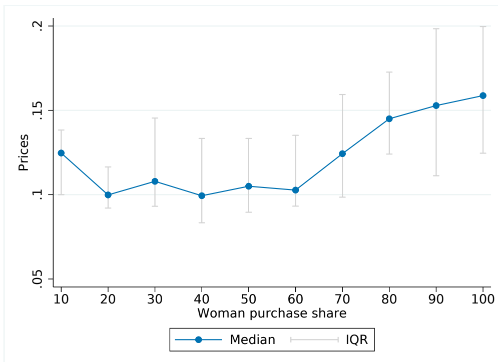
Disposable Razors



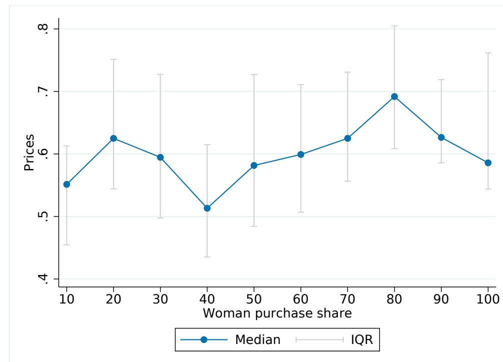
Shampoo

Note: This figure presents the distribution of products across woman purchase share for the five selected product markets.

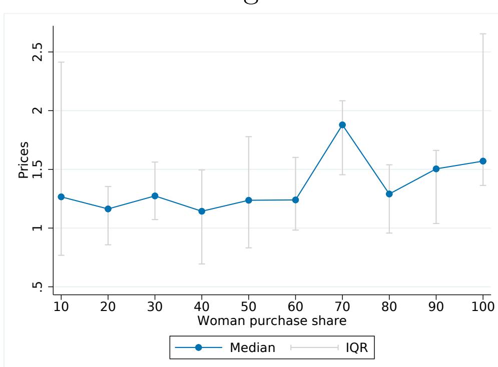
Figure A.12: Observed Prices



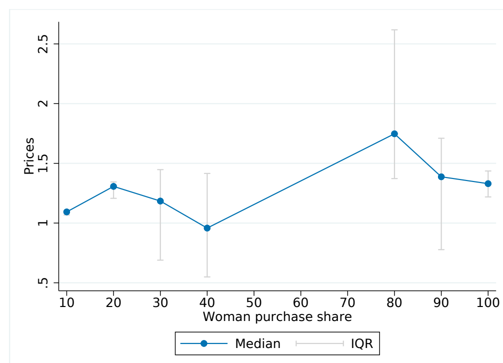
Yogurt



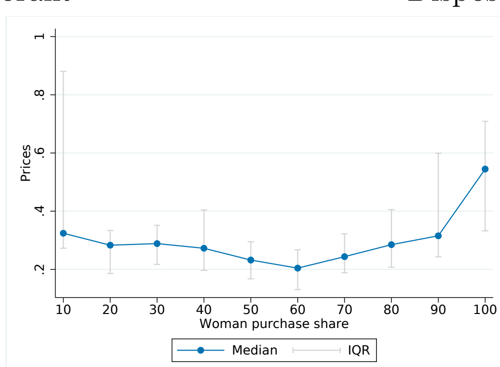
Protein Bars



Deodorant



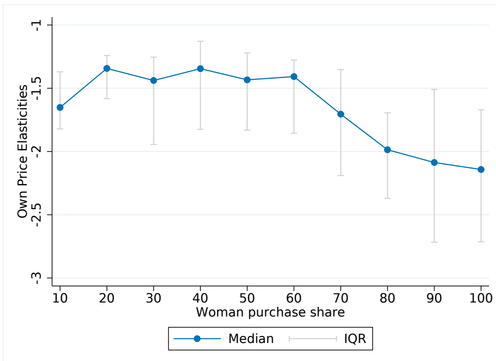
Disposable Razors



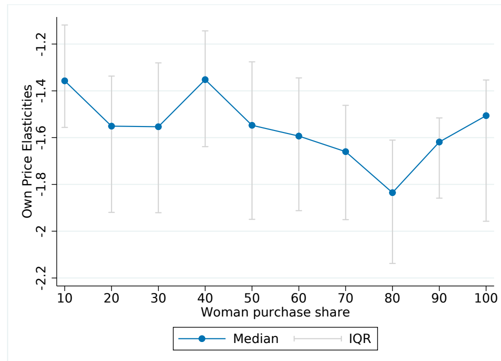
Shampoo

Note: This figure presents median prices of products by decile of woman purchase share. Prices are observed in the data and are not estimated. Grey bars represent the inter quartile range.

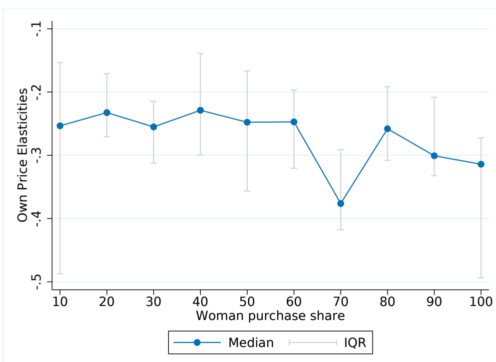
Figure A.13: Own Price Elasticities



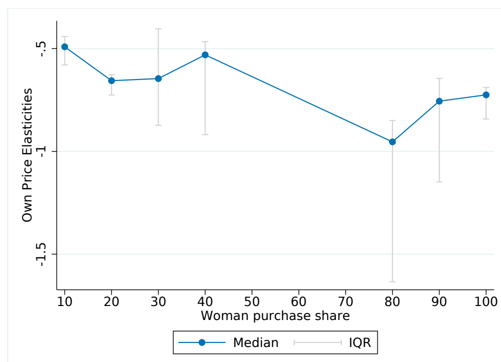
Yogurt



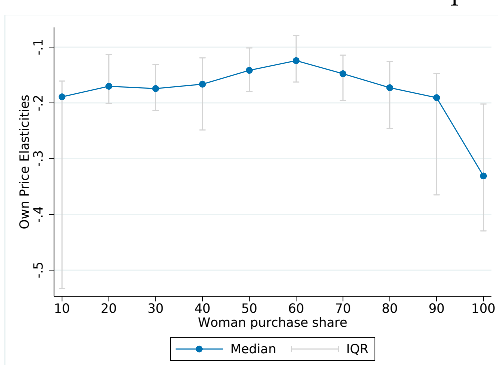
Protein Bars



Deodorant



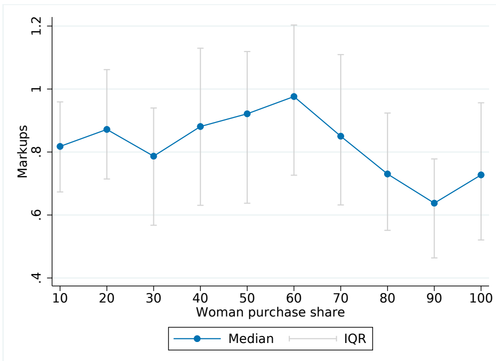
Disposable Razors



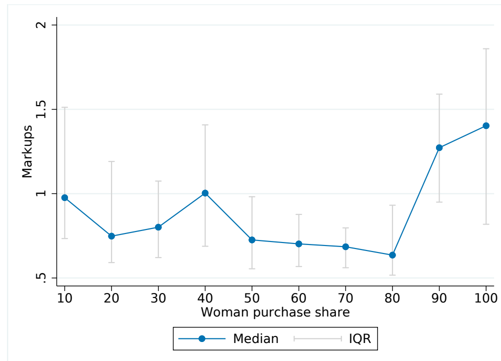
Shampoo

Note: This figure presents median estimated own-price elasticities of products by decile of woman purchase share. Grey bars represent the inter quartile range of the estimates.

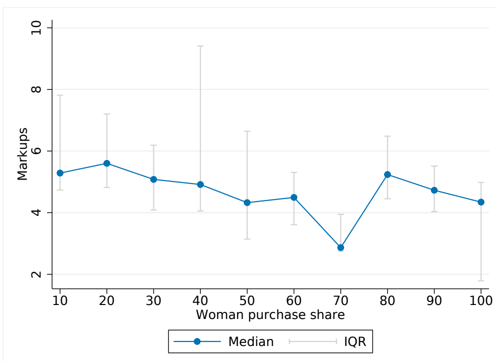
Figure A.14: Markups



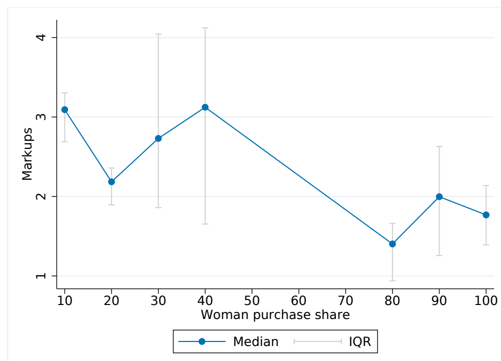
Yogurt



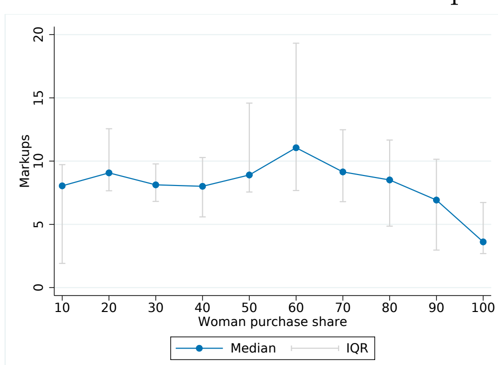
Protein Bars



Deodorant



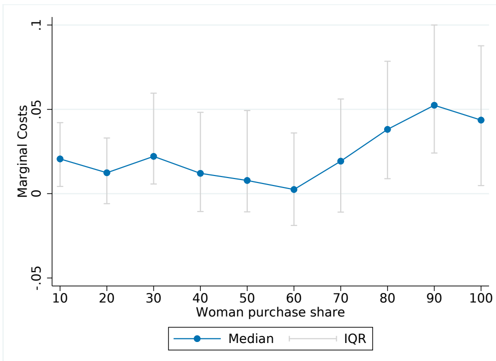
Disposable Razors



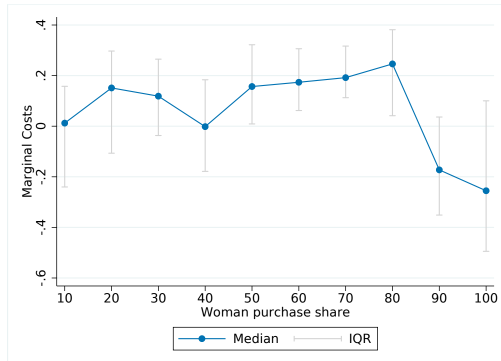
Shampoo

Note: This figure presents median estimated markups of products by decile of woman purchase share. Grey bars represent the inter quartile range of the estimates.

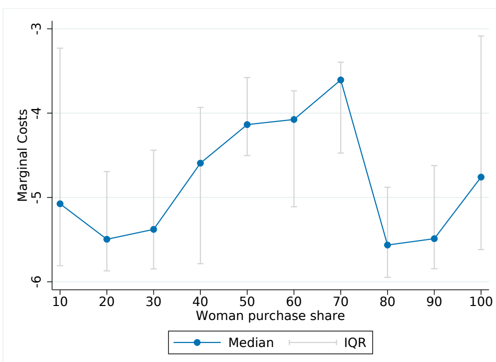
Figure A.15: Marginal Costs



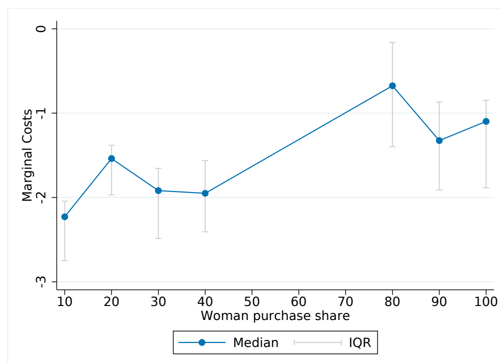
Yogurt



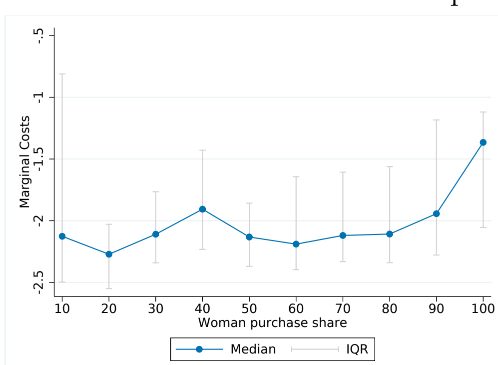
Protein Bars



Deodorant



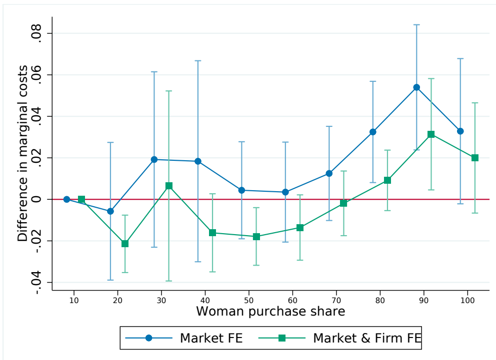
Disposable Razors



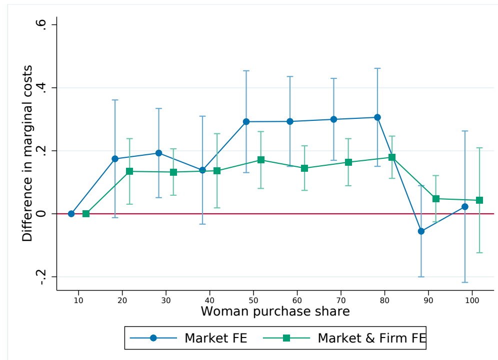
Shampoo

Note: This figure presents median estimated marginal costs of products by decile of woman purchase share. Grey bars represent the inter quartile range of the estimates.

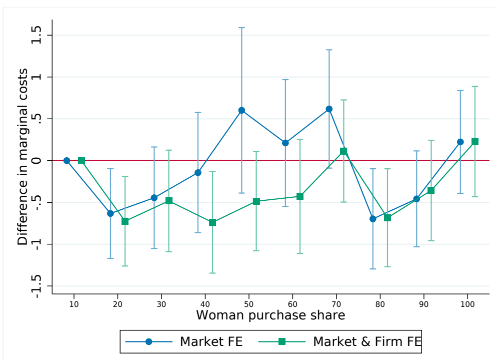
Figure A.16: Marginal Costs with Market and Firm FE



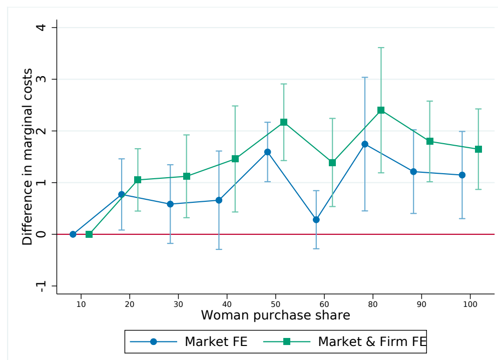
Yogurt



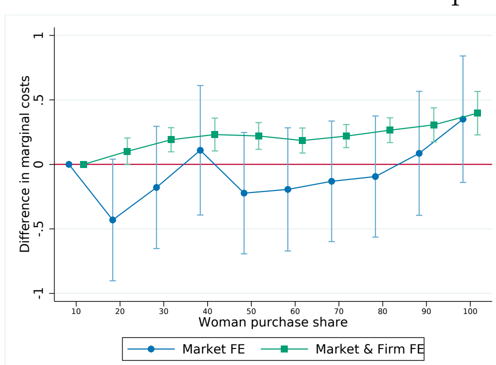
Protein Bars



Deodorant



Disposable Razors



Shampoo

Note: This figure plots average average markups for each decile of woman purchase share relative to goods that are bought up to 10% of the time by men within a market and within a market and firm. Standard errors bars were computed taking estimated values as truth.

Table A.18: Differentiated Products Model Results

	(1) Yogurt	(2) Deodorant	(3) Protein Bars	(4) Razors	(5) Shampoo
Price (α)	-13.778***	-0.201***	-2.710**	-0.563***	-0.612
	(0.198)	(0.0045)	(0.171)	(0.125)	(0.392)
σ_1	10.187***	9.386***	7.262	62.911***	4.417
	(1.377)	(1.849)	(15.661)	(25.780)	(6.575)
σ_W	15.509*	1.906	23.738	19.833	17.670
	(8.611)	(5.603)	(17.414)	(13.808)	(68.242)
Observations	728,428	3,425,548	1,443,840	466,059	694,939
$\bar{\varepsilon}$	-1.875	-0.329	-1.716	-0.766	-0.215
$\bar{\mu}$	0.879	5.278	0.925	3.377	11.109

Market level clustered standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix B Brand Loyalty and Forward Looking Firms

In our differentiated products demand model, we consistently estimate demand elasticities for Health and Beauty products that yield negative marginal costs under static competition over prices. While many alternate models of firm conduct can could rationalize the pricing decisions of firms and produce positive marginal costs, in this appendix we explore how brand loyalty and forward looking firms could lead to less elastic demand and lower equilibrium prices. We build on the model presented in Dubé, Hitsch, and Rossi (2009), where brand loyalty is incorporated as a psychological switching cost and firms maximize their present discounted stream of profits. The individual's indirect utility from consuming product j in market t is now:

$$u_{ijt} = \alpha p_{jt} + \beta_{\mathbf{i}}\mathbf{x_j} + \xi_{jt} + \gamma \mathbb{1}(state_{it} \neq j) + \epsilon_{ijt}, \tag{10}$$

where $\gamma leq 0$ represents the utility cost of switching to a product not consumed in the previous period. The individual's probability of choosing product j is given by:

$$\mathrm{s}_{ijt} = \frac{\exp(\alpha p_{jt}+\beta_i \mathbf{x_j}+\xi_{jt}+\gamma \mathbb{1}(state_{it}\neq j))}{1+\sum_k(\exp(\alpha p_{kt}+\beta_i \mathbf{x_k}+\xi_{kt}+\gamma \mathbb{1}(state_{it}\neq k)))}$$

(11)

To arrive at population-level choice probabilities, or market shares, we integrate over the distribution of random taste shocks as well as the distribution of the state space.

$$s_{jt} = \int\int \frac{\exp(\alpha p_{jt} + \beta_i \mathbf{x_j} + \xi_{jt} + \gamma \mathbb{1}(state_{it} \neq j))}{1 + \sum_k(\exp(\alpha p_{kt} + \beta_i \mathbf{x_k} + \xi_{kt} + \gamma \mathbb{1}(state_{it} \neq k)))} d\beta_i df(i), \tag{12}$$

where $f(i)$ is the state space distribution and maps to the previous period's market share. Incorporating brand loyalty provides firm's with an additional dimension over which they can increase market shares. In the standard BLP model, firm's can increase their market shares by adjusting prices in that time period. Now, firm's market shares are not only dependent on current period prices, but also indirectly by previous periods' prices through the previous period's market share. Note that the existence of brand loyalty means we will observe consumers being less elastic, as it would take a larger price change to incentive a consumer to switch products than without switching costs.

If we kept the static model of competition that is standard in BLP, the existence of brand loyalty and inertia should always lead to higher equilibrium prices. This is because in a one shot game, there is a benefit of cannibalizing on existing inertial customers. The effect on prices for forward looking firms, however, is ambiguous. We now assume that firms maximize the present discounted stream of future profits, making supply dynamic rather than static. The firm's problem is given by:

$$V(\pi_{ft}) = \sum_{j \in \mathcal{J}_f} \sum_{l} \beta^l (p_{jt} - mc_{jt}) s_{jt},$$

Firms compete over prices and the solution is defined by a set of strategies, $\sigma(f)$, that satisfy Markov perfect equilibrium. Because the supply side is now dynamic, the game does not have a closed form solution and must be solved with computational methods. However, we can build intuition for how strategies change. In a static supply model, firms maximize profit in a single period and face a trade off between prices and market shares. If a firm raises prices, it makes more money on the marginal consumer that stays, but loses out on the consumers that leave. Firms set prices such that the marginal benefit of raising prices is exactly offset by the marginal loss of losing customers. When consumers are brand loyal and firms are forward looking, prices in the current period have an enduring effect on market shares in the future. That is, lower prices today not only increases today's market shares but tomorrow's as well.

This additional incentive expands the range of potential equilibrium price outcomes relative to the static model. That is because there is now an additional trade off decision being made: firms may have incentive to cannibalize on their inertial consumer base with higher prices, but they also may have incentive to lower prices in order to gain and maintain a larger consumer base in future periods. Dubé, Hitsch, and Rossi (2009) simulate equilibrium prices for consumers with a standard logit utility function and assuming single product firms and find that at very high levels of brand loyalty equilibrium prices are higher than in static equilibrium, but at lower levels of brand loyalty equilibrium prices are lower than they would be in static competition.They find that equilibrium prices are initially decreasing in brand loyalty then the trend inverts and prices begin increasing in brand loyalty. Empirically, they find that the level of brand loyalty observed in orange juice and margarine markets is consistent with lower equilibrium prices.

These results are consistent with our finding that prices for Health and Beauty products are low given their observed demand elasticities. Estimating this model is ongoing work and will be included in future iterations of this paper. We now discuss how our results in the main body of the paper can be interpreted in the context of brand loyalty and a dynamic supply side. Our paper finds that marginal costs tend to be increasing in woman purchase share, that is products that are more often bought by women have higher marginal costs. The introduction of brand loyalty has the potential to change this relationship if women and men are heterogeneously brand loyal.

Holding the level of brand loyalty constant between men and women would likely lead to a level shift up of our marginal cost estimates, as the pricing incentives for men's products and women's products would change in the same way. In order for our results to be flipped, women would need to have significantly different brand loyalty levels than men. Specifically, men would need to have moderate brand loyalty levels with women either having close to no brand loyalty or fairly high levels of brand loyalty.

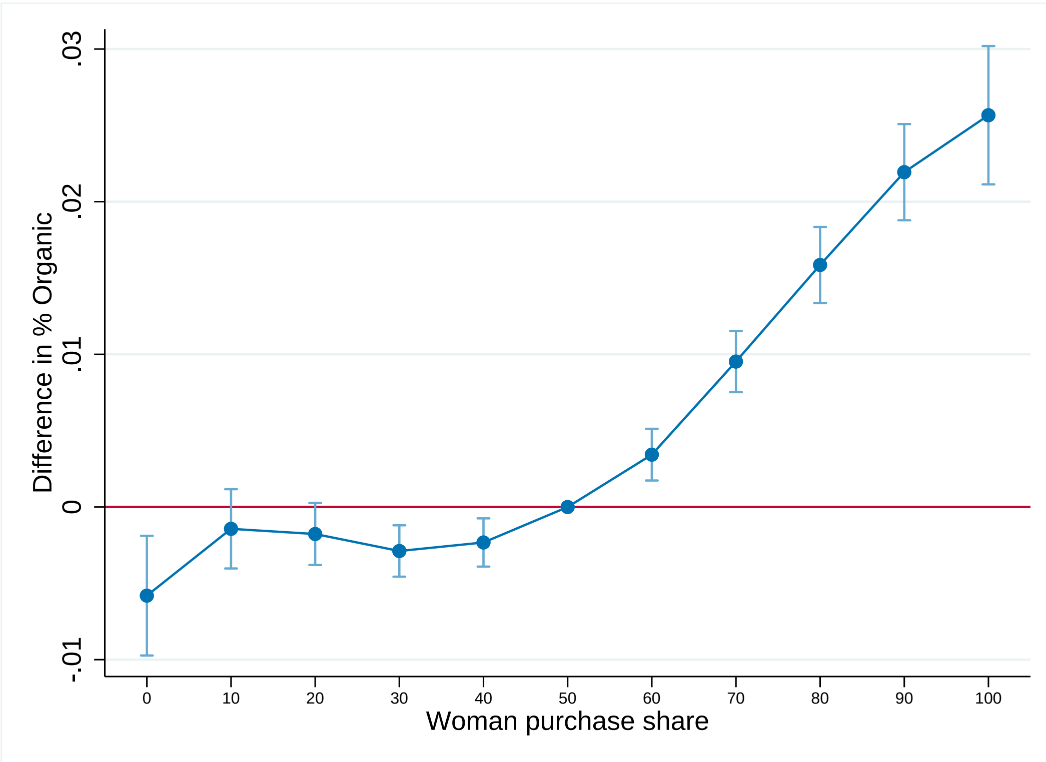
Appendix C Purchases of Organic Products

Our analysis of prices and marginal costs paid by women suggests that women are sorting into products that have higher prices, higher marginal costs and lower markups. In this section we look at how one product attribute associated with higher costs of production, organic products, vary with gender. Organic products typically have higher costs of production because they cannot be grown with lower cost pesticides and require certification with the US government. For this analysis we restrict to consider only food products, as these departments have more reliable information on organic status. We present two main analyses that capture the difference between men and women in the purchasing habits of organic products. First, we estimate the difference in the share of organic purchases within a given market for men and women in Table C.1. Second, we plot the difference in share of products that are organic by woman purchase share relative to products bought equally by men and women in the same market in Figure C.1.

Table C.1 shows that women are about 0.2pp more likely to purchase an organic product than men are for products in the same product market. While this number is small, it is highly significant and reflects the overall low level of organic purchases. On average, the men in our sample have an average organic purchase rate of about 0.8% meaning that we estimate that women are between 25% and 32% more likely to buy organic products, a notable difference in propensity.

Figure C.1 plots the coefficients from a regression of organic status of a product on decile of woman purchase share normalized to products that are bought equally by men and women. The graph shows that products primarily bought by men are slightly less likely to be organic while products more often bought by women are significantly more likely to be organic. The orders of magnitude are comparable to those found in Table C.1, with women's products being about 0.2 pp more likely to be organic.

Figure C.1: Share of Organic UPCs by Women's Purchase Share



Note: This figure presents plots of the results of the regression $\mathbb{1}_{O(jt)} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}_{g(j)} \in Bin_b\} + \theta_t + \varepsilon_{jt}$. Bins $b \in \mathcal{B}$ include ten-percentile-width bins centered at and two bins for pure gender stratification at the tails partitioning the interval $[0, 1]$. The regression includes fixed effects for product module, county and half-year. Results are presented for the whole sample and also separating out Health and Beauty and Dry Grocery. Standard errors are clustered at the UPC-county level.

Table C.1: Purchases of Organic Products

	(1)	(2)	(3)	(4)
Women	0.0023***	0.0026***	0.0022***	0.0021***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Men's average	0.008	0.008	0.008	0.008
Module X Units FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No
Month FE	No	No	No	Yes
County FE	No	Yes	Yes	Yes
Retailer FE	No	No	Yes	Yes
Demographic FE	Yes	Yes	Yes	Yes
Adj. R-squared	0	0	0	0
N	122,260,336	120,476,974	114,028,782	108,583,777
Number of clusters	49,252	49,250	49,245	49,249

Note: This table presents estimates from the regression: $\mathbb{1}_{O(ijt)} = \phi_{t(j)} + \beta \mathbb{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$ where $\mathbb{1}_{O(ijt)}$ is an indicator turned on if the purchase is an organic product. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. Standard errors are clustered at the individual-level. Column 1 can be thought of as a raw gap between single men and single women, each subsequent column demonstrates the contribution of controlling for an additional market or demographic factor.

Appendix D Marginal Cost Validation: Razors Case Study

We validate our marginal cost findings for disposable razors with information about product characteristics that are likely correlated with the costs of production. Specifically, we scrape information on a razor's number of blades, the existence of a moisture strip, and the shape an contents of the handle. We create an indicator for whether the handle is ergonomic based on it having a shape that requires more plastic in comparison to a straight handle or whether it has additional rubber grip in the handle. We are able to gather information on product attributes for 90 out of the 176 razor product lines in our data (226 UPCs), however we capture those products that have the largest market share and are able to capture information for 73% of purchases that are made on private label disposable razors.

We present purchase weighted comparisons of the product characteristics of the average women's razor to the average men's razor in Table D.1. In Panel A, we find that women's razor purchases have 0.3 more blades than men's, with the average razor purchase having between two and three blades. Women's razor purchases are slightly less likely to have a moisture strip, by about 5pp. Finally, women's razor purchases are about 17pp more likely to have a ergonomic handle. We take this as evidence that the razors that women purchase have characteristics associated with having higher cost of production as they require more materials to produce than men's razor purchases. Panel B presents UPC level results. We do not find significant differences in the average number of blades or moisture strips between men's and women's product offerings but do find that women's product offerings are significantly more likely to have ergonomic handles. The difference in findings between Panels A and B highlights the important role that sorting plays when considering differences in how men and women consume products.

Overall, we find that the product attribute data support our finding that women's razors have higher marginal costs of production. This should give confidence that while our marginal cost estimates may be biased downwards due to using a static model or other competitive factors, the trend lines and elasticity estimates are capturing meaningful differences in firm's

Table D.1: Women's and Men's Razor Attributes

	(1) Blades	(2) Moisture Strip	(3) Ergonomic Handle
Panel A: Purchase Level			
Women's Razor	0.3050***	-0.0477***	0.1663***
	(0.0001)	(0.0000)	(0.0000)
Men's Razor Average	2.239	0.758	0.310
Adj. R-squared	0.04	0.00	0.03
N	664,347,126	661,511,494	661,511,494
Panel B: UPC Level			
Women's Razor	0.0829	0.0504	0.2127***
	(0.1108)	(0.0521)	(0.0718)
Men's Razor Average	2.484	0.815	0.369
Adj. R-squared	-0.00	-0.00	0.03
N	226	224	224

* $p < .10$, ** $p < .05$, *** $p < .01$

This table plots coefficients from regressions of a given product characteristic on whether or not the product is a women's razor. Data on gender and characteristics were created by searching product and brand descriptions. Panel A presents results weighting each razor by the number of purchases observed in the RMS data. Panel B does not weight by number of purchases. Robust Standard errors are presented in parentheses.

pricing and production of products.