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

Kayleigh Barnes† Jakob Brounstein‡

November 3, 2025

Abstract

We evaluate the existence of the pink tax: the hypothesized price premium on women's consumer goods. Using retail scanner data, we find women pay 5.1% more for consumer packaged goods (CPGs) in the same product-by-location market. When we observe wholesale prices, we find women's goods feature lower markups. Estimating a constant elasticity of substitution model of demand on the near universe of CPGs and a differentiated products demand model within the disposable razors market yields similar conclusions: the pink tax is not sustained by higher markups, but by women sorting into goods with higher marginal costs and implied quality valuations.

JEL codes: J16, D12, L11

Keywords: Gender inequality, Pink tax

*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. Our own analyses and calculations are based in part on NielsenIQ Retail Measurement Service and Consumer Panel Service data for all categories for the 16-year period ending in 2019, for the total US. The conclusions drawn from the NielsenIQ data are those of our own 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 or in developing, reviewing or confirming the research approaches used in connection with this report. The views expressed in this work are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

†Corresponding Author. Federal Reserve Board of Governors, 20th Street and Constitution Avenue N.W., Washington, D.C. 20551; 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 that explicitly feature gender-based marketing (Moshary, Tuchman, and Vajravelu 2023; Guittar et al. 2022; NYCDCA 2015; Duesterhaus et al. 2011; Manzano-Antón, Martinez-Navarro, and Gavilan-Bouzas 2018). However, by considering the widest array of consumer goods yet studied in a more generalized demand framework, we show that the pink tax actually encompasses broad gender differences in preferences and consumption that contribute to overall differences in cost of living. We seek to not only evaluate the existence of the pink tax, but also speak more generally to differences in the nature of men's and women's demand for goods.

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]based on any channel through which women may face higher prices 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 5.1% 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. By definition, this finding could possibly be driven by several distinct economic mechanisms that determine pricing: for instance (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 our understanding of this phenomenon as well as the implications of policies that aim to address the pink tax.

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 16-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 argue that this step is crucial in correctly assigning product gender characteristics. Additionally, our main specifications control for demographic differences that might otherwise present compositional concerns in comparing single men and single women. 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. Lastly, we incorporate data from PriceTrak to inform the wholesale costs for a subset of goods in our data, from which we can directly compute retailer markups.

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 6.7% 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 5.1% more

[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.

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 (after applying the Nielsen probability weights to reweight purchases) as "gendered" products, categorizing all other products as "ungendered" (with alternate cutoffs demonstrating the robustness of our results). We then decompose the average 5.1% price premium paid by women into a contribution from differential sorting into ungendered products and from purchases of highly gender-segregated products, finding that women pay an average of 4.8% higher prices on *ungendered* products relative to men and that women pay an average of 21.2% higher prices on gendered products relative to those 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 and estimate demand and supply under a variety of assumptions, attributing differences in pricing and product choice to markups and marginal costs.

We estimate 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 exhibit a price-elasticity of demand between 11% and 26% greater in magnitude than do men. This finding implies that women are charged lower markups rather than higher markups, on average, under price discrimination. This elasticity difference varies across CPG departments (e.g. dry grocery, health and beauty goods, etc.), but we find no department in which women exhibit significantly lower price elasticity than men. We further adapt the model of Faber and Fally (2022) to our setting, allowing us to estimate by-gender subjective product quality, intrinsic quality, and quality-adjusted price across goods. We find that women's consumer goods are assigned 25% greater intrinsic quality than goods consumed by both genders, whereas those consumed by men are assigned only 10% greater quality. Individuals with the greatest female-goods concentration of their consumption baskets feature an average quality-consumption level about 10% higher than mean-product quality, where those with the greatest male-goods concentration of their consumption baskets see a 5% lower average quality-consumption level. We also document that women's goods do not operate in less-competitive markets than do those of men's, implying that a competition-channel is unlikely to generate greater markups for women.

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. Studying the goods themselves, we find slightly higher markups placed on men's goods than on women's goods: within market, the most-male goods face a 10pp higher markup than the most-female goods; when weighting by expense, this difference expands to 30pp.

Finally, we estimate overall markups and marginal costs of production for disposable razors using a differentiated products demand model (Berry, Levinsohn, and Pakes (1995)). Our work up to this point speaks very broadly to gender differences in consumption. For this reason, we view that incorporating additional model complexity while focusing in on a specific market that serves as a canonical pink tax example complements our analysis. In addition to prices, we incorporate product gender as a characteristic over which consumers can have heterogeneous taste. This procedure produces product-market level estimates of elasticities and marginal costs; combining these estimates with observed prices, we also compute markups relative to manufacturing costs. We find that razors disproportionately consumed by women have higher marginal costs of production. We estimate that women's razors see 14pp higher markup unconditionally across razor brands (weighting by overall units sold); however, including firm fixed effects generates a negative markup difference, with women's razors exhibiting an 11pp lower markup than men's razors produced by the same firm.

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 and quality. Our findings carry key implications for policy. Recently ratified and proposed legislation across different countries and subnational jurisdictions aim to ban differential pricing of products that differ only in gendered marketing. Our results suggest 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. Moreover, because we attribute gendered price differences to difference in marginal cost, such legislation if enforced, would likely lead to product exit and welfare loss for women consumers.[2]

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 controlling for manufacturer 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

[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.

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.

1.1 Related literature

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 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 phenomena but Stigler (1987) defines price discrimination as any markup difference between consumers groups. Relatedly, we speak to the body work studying 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 pricing of differentiated products could be a form of second degree price discrimination with certain consumer groups being charged systematically higher markups, which we conclude is *not* the case along the gender dimension.

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 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 Universal Product Code (UPC) $\times$ Store $\times$ 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.

[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 Section A.2.

HomeScan Panel ("HMS"). The entire HMS features data on the shopping trips and transactions of approximately 60,000 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 household-years, which eliminates 74.6% of panelist-years in the HMS. We also restrict to households-years with at least 12 shopping trips, which eliminates only 15 out of nearly 162 million panelist-years for both family and single-person panelist households. Lastly, we drop single-person households that report multiple genders compositions (31 individuals in the singles sample) for the reason that we cannot distinguish coding errors from panelists that undergo gender transition. This leaves us with a panel of 49,256 households which we use to study differences in consumer behavior. We report summary statistics for the sample in Table A.1.

Additionally, our main specifications control for demographic differences that might otherwise present compositional concerns in comparing single men and single women.

Table A.1 shows that the single men and women in our sample appear similar in many respects, but do exhibit relatively small but statistically significant differences in observable characteristics. For instance, women in our sample are 0.6 years younger than are men. Additionally, the men in our sample report on average 20% greater income[4] and slightly higher levels of education, which we control for in the analysis. We frequently control for these demographic characteristics in our specifications in order to avoid comparing men and

[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, following Allcott, Lockwood, and Taubinsky (2019).

women that differ significantly along these characteristics.

Women are overrepresented in our sample as they are in the general HMS, comprising about 70% of our panelists. However, applying the HMS projection weights yields a near 50-50 gender balance. We employ these weights for the entirety of our analysis that uses the HMS. 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. Purchases in our main sample comprise approximately 1.87 million UPCs. We are able to confidently assign gender to 1.17 million UPCs that comprise 98.7% of the purchases made in our singles panel.

Retailer Scanner Data ("RMS"). 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).[5] 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 which is important for constructing our price instruments described in Section 4. The Nielsen data has limited coverage of independent stores, so this constraint has limited impact on differential selection *within* the Nielsen data.

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. We define products at the UPC level. 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 most granular grouping of multiple products—e.g. "Disposable

[5]The RMS is also used to generated the prices paid as observed in the HMS.

Razors". The Nielsen data identifies over 1,300 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; however for certain reduced-form analyses, we further divide product modules into module-units to compare quantities and quantity-weighted prices across purchases. For example, the coffee product module contains bagged coffee measured in weight (ounces) and Keurig cup coffee measured as a count (number of K-cups).

We construct transaction-level UPC unit prices as the total amount paid for a UPC divided by the observed quantity. We construct observed quantity as product of 1) the quantity of units contained in a single purchase-unit (e.g. 5 ounces) 2) "multipack" (a variable indicating the number of units contained in the package; for non-multipacks, this variable is equal to one), and 3) the number of overall units purchased. For example, we calculate the unit price corresponding to a \$10 purchase of a 5 count multipack of razors, where each multipack contains 2 units, and each unit contains 4 razors as
$\$10/(5\ multipacks\ \cdot\ 2\ units\ per\ multipack\ \cdot\ 4\ razors\ per\ unit) = \$0.25\ per\ razor.$[6]

Pricetrak PromoData ("Pricetrak"). Lastly, the PriceTrak PromoData data allow us to validate retailer markups relative to wholesaler prices. While this data does not feature information on manufacturing 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 match to PromoData (largely because the PriceTrak data report on a smaller subset of UPCs);

[6] We find that women consume multipack items about 9% less frequently than do men and that multipack purchases constitute around 5% of transactions. Table A.9 estimates a set of regressions analogous to the module-level results in our main specification of Table 1 Panel (a) while including the log number of units contained in a purchase (equal to $\log(1) = 0$ for non-multi-packs). The table shows that our descriptive price differences are unaffected by multi-pack purchases.

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 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 characteristics, such as year, county, income, age, race and education that may be related to other purchase habits.[8] We find that women's yearly spending is greater than that of men by about 6.7%, their product diversity is greater than men's by about 31.3% and their consumption baskets are larger than men's in terms of items purchased by about 9.3%.[9] 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.

[7]We use the CE PUMD to analyze differences in full consumption baskets in Figure A.7 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 finding that men spend more on food suggests that men are more likely to substitute food expenditure to eating out than women.

[8]Note that applying the Nielsen proprietary probability weights alleviates demographic differences between men and women due to sampling variability.

[9]We compare differences in yearly spending subsequently adding in controls in Table A.4. We convert coefficient estimates $\hat{\beta}$ throughout the paper from log price regressions to percentage differences as $e^{\hat{\beta}} - 1$. We find that the raw annual spending gap is about 1.5%, while the gap between demographically similar men and women is about 6.8%.

Our documented total spending differences in Figure 1 could arise from differences in prices paid for similar goods, differences in quantities purchased, or differences in the goods purchased themselves. 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 an *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 which are typically contained in their own modules).

Popular discussion of the pink tax is often focused on differences in prices paid between men and women for similar goods in the same market. We compare per unit prices paid by men and women for products in the same market with the following specification:

$$log(P_{ijct}) = \phi_{m(j)ct} + \beta\mathbb{1}\{Woman_i = 1\} + \gamma X_i + \epsilon_{ijct}, \tag{1}$$

where i denotes the individual, j denotes the product purchased, $m(j)$ is the module of product j, and c denotes geography, and t denotes time. 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; Column (2) estimates this same regression on the most-restrictive sample specification from column (7), for the purposes of comparison across a constant sample. We interpret the 2.9% estimate 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. Columns (3)-(7) sequentially add permutations of fixed effects and additional covariates. Note that sample size changes across columns reflect singleton purchases that cannot be estimated under different fixed effect specifications.

Column (3) runs the same specification but adds in controls for age, income, and race. The increase in the coefficient estimate from 2.9% to 5.1% 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 (4) 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 (5) 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 4.2%, suggesting that retail chain sorting plays a small but significant role. Finally, in Column (7) we add in fixed effects for month rather than year. The results indicate that women spend about 5% 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.[10]

Table 1 Panel (b) estimates Equation (1) while including UPC-level fixed effects instead of module-level fixed effects. In this specification, the interpretation of the coefficient becomes the difference in prices paid between men and women for the *exact same* 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 (3) 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

[10]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.

gap, which is consistent with women sorting into higher price chains. Column (7) shows that women pay 1.4% 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. This finding is further corroborated by Table A.2, which also shows that women make greater usage of deals and coupons. 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 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 develop a continuous measure of gender on the product level, and decompose observed descriptive pink tax of 5% 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

$$\hat{w}_j = \frac{\sum_{i \in \mathcal{I}} Purchase_{ij} \mathbb{1}\{woman_i = 1\} \cdot NielsenWeight_i}{\sum_{i \in \mathcal{I}} Purchase_{ij} \cdot NielsenWeight_i}. \tag{2}$$

This fraction assigns to each product a value $\hat{w}_j \in [0, 1]$ where 0 denotes a good that is only bought by men and 1 denotes a good that is exclusively bought by women. $NielsenWeight_i$ refers to the Nielsen projection weight assigned to individual i in the HMS data. 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.[11] Because employing these weights yield a sample representative of US consumers, we define UPC gender parity as $\hat{w}_j = 0.5$, as opposed to the initial gender balance of sample when unweighted. We use the continuous measure of UPC gender for most of our analysis, but we also employ these binary categories to more intuitively capture comparisons involving "highly gendered goods".

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. 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 observing a single purchase, these UPCs would always be assigned to having an explicit gender of 0 or 1. 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. We can express the probability that the value $\hat{w}_j$ lies more than 5 percentage points away from its true value w_j as

$$P(w_j \notin (\hat{w}_j - .05, \hat{w}_j + .05)) = \int\limits_{x \geq |\hat{w}_j \pm .05|} \binom{n_j}{\lceil \hat{w}_j n_j \rceil} x^{\lceil \hat{w}_j n_j \rceil} (1-x)^{n_j - \lceil \hat{w}_j n_j \rceil} f(x) dx, \tag{3}$$

where $f(x)$ is the empirical probability density function of woman purchase share. We use Equation (3) to calculate thresholds for discrete bins of UPC-gender of radius 0.00025 from 0 to 1, $n^*_{jb(\hat{w}_j)}$, such that the probability 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

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

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] Table A.8 replicates the results from Table 1 using only the set of goods with identifiable UPC gender, yielding quantitatively and qualitatively similar results.

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:

$$\log P_{jct} = \theta_{m_jct} + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{w_j \in Bin_b\} + \varepsilon_{jct}, \tag{4}$$

where $\mathcal{B}$ is a set of ten-percentile bins that partition the interval $[0, 100]$, allowing for different price estimates along the UPC gender distribution. θ_{n_jct} is a set of indicators for module m_j, county c, and half-year t. 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 significant price premiums of between $10 - 40\%$ for goods purchased exclusively by women relative to similar goods purchased at gender parity. We observe no price

[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.1 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 however, gendered products make up 20% of women's purchases and 10% of men's purchases.

premium for goods purchased exclusively by men. However, beyond the tails of the graph, a striking pattern emerges: prices tend to monotonically increase in woman purchase share beginning around goods with 30% female purchase share. This increase in prices along woman purchase share suggests that our overall price premium 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:

$$\begin{aligned} log(P_{ijct}) = \phi_{m(j)ct} + \beta_1 \mathbb{1}\{Woman_i = 1\} + \beta_2 \mathbb{1}\{GenderedProduct_j = 1\} \\ + \beta_3 \mathbb{1}\{Woman_i = 1\} \cdot \mathbb{1}\{GenderedProduct_j = 1\} + \gamma X_i + \epsilon_{ijct}. \end{aligned} \tag{5}$$

This specification has the desirable quality that by interacting consumer gender with a gendered product indicator, we allow for differential price estimates for male-gendered and female-gendered goods. Table 2 presents the results of these regressions. We find that women pay a price premium of 4.8% on ungendered products relative to ungendered products bought by men. Across all departments, men pay lower prices on their gendered products than they do for ungendered products by about 1.5%. The interaction coefficient shows that women pay about 14.5% more for gendered goods relative to ungendered goods. Overall, we find that women pay approximately 21.2% higher prices on gendered products than do men (calculated by exponentiating the difference of sums of coefficients for gendered product purchases between women and 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.1 indicates that gendered products make up an overall small share of a consumption bundle. Indeed, the

5% 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

To estimate demand differences between men and women, we augment the constant elasticity of substitution (CES) model used in Faber and Fally (2022). A CES model offers several desirable characteristics in our setting. We allow men and women to differ both in their price elasticity and in their subjective product quality evaluations. Additionally, the nesting structure of the model also features flexibility in allowing for heterogeneous price sensitivity across product modules or departments. Furthermore, beyond the model structure, the estimating equation we derive also features the intuitive interpretation of capturing the average difference in price elasticity of women versus men over product modules.

This approach ultimately allows us to make aggregated comparisons of the purchasing habits of men and women across a wide range of products to make inference on their average price elasticities and average product quality consumption. If the per-unit price premium observed on women's goods and on goods purchased by women more broadly is attributable to differences in markups, we should observe that women exhibit lower price-sensitivity on average as a consumer demographic. However, if the observed price-premium on good purchased by women is attributable to differences in marginal costs, we should observe that women consumer higher-quality goods than do men.

The model characterizes a representative consumer of gender g. The consumer allocates their income in an additively separable manner between a vector of retail goods G and consumption of the outside option:

$$U = U(V_G(g), C_g).$$

We assume that the basket of goods comprising the outside option C_g is consumed as a normal good.

The model aggregates products in two tiers. The consumer allocates consumption *across product modules* with Cobb-Douglas elasticity and *substitutes between goods within modules* with module-specific constant elasticity of substitution. We index product modules as $n \in \mathcal{N}$, which perfectly partition the set of UPCs. Each module n contains goods G_n. 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:

$$V_G(g) = \prod_{n \in \mathcal{N}} \left[\sum_{j \in G_n} \Big(q_j(g) \varphi_j(g) \Big)^{\frac{\sigma_n(g)-1}{\sigma_n(g)}} \right]^{\alpha_n(g) \frac{\sigma_n(g)}{\sigma_n(g)-1}}. \tag{6}$$

Here, q_j refers to quantity consumed of UPC j, $\varphi_j(g)$ refers to the subjective product quality of a product j as evaluated by a consumer of gender g. $\sigma_n(g)$ represents the elasticity of substitution for gender g within module n, and $\alpha_n(g)$ denotes the share of expenditures by gender g allocated to a module $n \in \mathcal{N}$.[13]

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:

$$\frac{b_j(g)}{b_k(g)} = \left(\frac{p_j / \varphi_j(g)}{p_k / \varphi_k(g)} \right)^{1-\sigma_n(g)}, \tag{7}$$

where $b_j(g)$ is the budget share spent on product j. From Equation (7), expressing gender in subscripts for notational ease, we derive our estimating equation while also allowing for consumption and prices to vary over geography and time:

[13]Under a Cobb-Douglas upper nest it is the case that $\sum_{n \in \mathcal{N}.} \alpha_n(g) = 1$.

$$\Delta log(b_{jgct}) = (1 - \sigma_{ng})\Delta log(p_{jct}) + \eta_{ngct} + \varepsilon_{jgct}, \tag{8}$$

where differences are taken between two consecutive periods and η_{ngct} represents module $\times$ gender $\times$ county $\times$ half-year fixed effects to capture the local CES price index. We derive this estimating equation from a CES demand model, but this estimating equation can also be interpreted as an average price response within a market. Additionally, pooling all module markets together allows us to estimate a module-market average price responsiveness $1 - \sigma_g$.

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 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 $(1 - \sigma_g)$, the average substitution elasticity, using two instruments for prices. The first are Hausman instruments, which we construct as national leave-out means in price changes at the county or DMA level, $\Delta H_{jct} := \frac{1}{N^c_{jt}-1}\sum_{c'\neq c}\Delta\log(p_{jc't})$ where N^c_{jt} refers to the number of observations of UPC/Brand j across counties in half-year t.[14] 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, $\Delta DV_{jcry} := \frac{1}{N^{cr}_{jy}-1}\sum_{c'\neq c, r'\neq r}\Delta\log(p_{jc'r't})$ where N^{cr}_{jy} refers to the number of observations of UPC/Brand j across retailer-counties in the half-year.[15][16] Both instruments may see exclusion restriction violations from national demand shocks, for example from national advertising campaigns or a national holidays that drive certain purchases (e.g. hot dogs on

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

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

[16] 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.

the fourth of July). To help control for this possibility, we include market-by-time fixed effects in our estimation.

In our estimation procedure, we aggregate observations to the Brand $\times$ Designated Market Area (DMA) $\times$ consumer gender $\times$ semester level. We aggregate UPCs to the brand-level in order to overcome sparsity issues where individual UPCs are not necessarily observed in consecutive semesters within a specific market.[17] This decision also allows us to compare our estimates with those of Faber and Fally (2022) to gauge the validity of our approach. We also include an additional distinction for the retail chain r depending whether we use instruments from DellaVigna and Gentzkow (2019). We estimate our model at the half-year level because 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 $\sigma_{nf} - \sigma_{nm}$ if men and women exhibit differential stockpiling behavior. To address auto-correlation in the error term, we cluster standard errors at the brand-DMA level.

Section B.1 gives additional detail to the model. We derive own-price elasticity of demand:

$$\varepsilon_{jt}(g) = \sigma_g - (\sigma_g - 1) \cdot s_{jg} \tag{9}$$

Where s_{jg} is the module-level market share of product j for gender g. Thus, we can calculate ε_{jg} as a function of known and estimated parameters. In the case of monopolistic competition, all market shares are approximately zero and ε_{jg} coincides with the elasticity of substitution, σ_g. To map elasticities to markups, we assume firms compete on prices and maximize firm profits given the demand that they face.[18] We focus primarily on σ_g, as our setting is better described by monopolistic competition rather than monopoly, but we investigate gender-differential competition channels explicitly in Section 4.2.

[17]We aggregate prices from UPCs to the brand level using a Tørnquist price index. We use the procedure outlined in Section 3.2 to confidently assign continuous values of brand genders to 70.6% of 240,000 brands that appear in our data, comprising 99.85% of aggregate purchase volume.

[18]We do not assume single-product firms in computing markups. Section 5 calculates markups using retailer costs and sale prices directly and Section 6 assumes multi-product firms and constructs an ownership matrix to compute markups.

4.1 CES Model Results

We begin by estimating differences in the elasticity of substitution, $\sigma_n(g)$, between men and women. Table 3 presents results of estimating Equation (8) 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. The second row displays estimates of $1-\hat{\sigma}_m$. In odd-numbered columns we include gender-module-market-semester fixed effect and even-numbers columns include gender-module-market-retailer-semester fixed effects. These specifications aim to capture the differences in demand elasticities between men and women for the same price change for the same product in the same market and purchasing setting, for which reason our preferred specification includes the retailer fixed effects.

If it is the case 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.[19] The OLS specification estimates that women exhibit 7.6 percentage points lower price elasticity than do men, with no difference when not controlling for retailer. However, as expected, the OLS estimates for the level of elasticity of substitution, $(1-\hat{\sigma}_m)$, exhibit substantial bias, differing in sign from the IV estimates.

Across instrumental variables specifications, we find that women consistently demonstrate significantly *greater* price sensitivity than do men. We find that for within the same market, women are between about 11 and 26 percent more price-elastic than men. Our preferred specification, column (8), uses both Hausman and DellaVigna-Gentzkow retailer instruments as well as retailer fixed effects and gives an elasticity of substitution for men of 1.51 and for women of 1.77. Interestingly, we estimate elasticities of substitution slightly lower in magnitude than estimates from Faber and Fally (2022) ($\hat{\sigma} = 1.66$ in column (6) versus $\hat{\sigma} = 2.20$ in the analogous specification in Faber and Fally (2022)). However, our sample differ from Faber and Fally (2022), as our subsample of the HMS data only consists of single individuals, whereas Faber and Fally (2022) do not make this restriction. Nonetheless, our estimates are consistent with monopolistic competition.

We now turn our focus to how elasticities of substitution vary across product depart-

[19]Table B.1 presents the first stage and reduced form results respectively.

ments. Table 4 presents elasticities of substitution results pooled to the department level in all of these specifications, we use both instruments as well as retailer fixed effects (in addition to our other market definition fixed effects).[20] We find that women are either more elastic than men are or are not significantly different than men in terms of elasticity across all departments. However, for some departments, namely Health & Beauty, Alcohol, and General Merchandise, we estimate less elastic demand with substantial noise. These departments have substantially less coverage in our data, as many purchases of these products can be made at stores that are not included in the Nielsen data. Further, these products tend to be bought less frequently, which can exacerbate sparsity issues. In contrast, for departments that see reliable coverage in the Nielsen data and do not face sparsity issues due to frequent purchases, like Dry Grocery, we estimate reasonable demand elasticities with precision.

The table shows that across almost all food products and general non-food merchandise women are significantly more elastic consumers than are men. We find that $\hat{\sigma}_w - \hat{\sigma}_m \in [-1.23, -0.49]$ among precisely estimated specifications. For half of the ten departments, including Health and Beauty, dry groceries, deli products, non-food groceries, and alcohol, we find no significant difference in elasticity of substitution between men and women.[21] 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.

[20] Table B.2 displays this disaggregation without retailer fixed effects, yielding similar results.

[21] We find that Health and Beauty, Alcohol and General merchandise products tend to exhibit less price elastic demand 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.

4.2 The competition channel

Up to this point, we have estimated elasticities of substitution, $\sigma(g)$. However, price elasticities of demand are given by Equation (9) and are a function of the elasticity of substitution and market shares. Under this derivation, price elasticities of demand will range from σ_g (in case of monopolistic competition where market shares approach 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 purchase 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.

To further assess the competition channel, we study how brand and UPC market shares vary through the product-gender distribution as well as in men's and women's consumption baskets. First, we construct the market share of brand/UPC j in semester t as the aggregate expense observed in the HMS on j as a share of the overall national expense on the module of brand/UPC j in the same semester (weighted using Nielsen's projection weights):[22]

$$s_{jt} = \frac{\sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{J}_{ij}} Exp_l \cdot NielsenWeight_i}{\sum_{i\in\mathcal{I}} \sum_{k\in G_j} \sum_{l\in\mathcal{J}_{ik}} Exp_l \cdot NielsenWeight_i}, \tag{10}$$

where $\mathcal{J}_{ik}$ is the set of transactions of individual i purchasing product k. We correlate this object with measures of product and consumer measures of gender. We assign each individual in the HMS a gender-consumption rank based on their time-invariant percentile of expense-weighted UPC-gender average.[23]

To speak to the market share composition of individuals' consumption baskets, we assign

[22] We construct these measures in the HMS. Constructing these shares in the RMS could yield greater precision, but 1) our results demonstrate that the HMS weights yield considerable precision and 2) computing national-level market shares for all products using the RMS would be computationally infeasible due to the size of the RMS.

[23] We construct the expense-weighted average UPC-gender consumed by individual i as $\frac{\sum_{j\in\mathcal{J}_i} \hat{w}_j \cdot x_j}{\sum_{j\in\mathcal{J}_i} \hat{x}_j}$ for all transactions $\mathcal{J}_i$ of household i.

each individual $\times$ half-year an expense-weighted average market share and assign each individual the mean of these values over half-years. Figure 4 plots the conditional distribution of expense-weighted average market share of goods consumed along gendered-consumption rank. On both panels, the left-most values correspond with individuals with the highest consumption share of "male" goods, whereas those on the right correspond with individuals with the highest consumption share of "female".

First, Panel (a) shows an inverse U-shape of raw expense-weighted average market share consumed along gendered consumption rank, which is expected since ungendered goods mechanically exhibit consumption from both men and women and would exhibit greater market shares. Second, observe that the female-end of the gendered consumption distribution features slightly lower average market share than the male end, suggesting the presence of greater levels competition among female-gendered goods than ungendered goods and, to a lesser extent, male goods, indicating that female-male price differentials are not augmented by differences in competition environments. Panel (b) elaborates on this result, plotting average market shares demeaned against semester fixed effects and module fixed effects, which controls for gendered sorting across modules as well as potential changes in competition over time. Interestingly, the graph demonstrates a negative relationship between market share and gendered consumption rank, albeit of significantly smaller magnitude than Panel (a). We interpret this graph to indicate that when controlling for the competitive environment across modules (and over time), individuals consuming female goods (who are increasingly likely to be women by definitions) consume products with lower market shares.

Figure 5 displays a similar set of results, however using brand gender as the variable on the x-axis. As expected, we observe greater market share of brands at near-gender parity. However, in both Panel (a) that plots raw average market shares and Panel (b) that residualizes brand market shares against half-year and module fixed effects, we observe little difference between female and male brands in terms of their average market shares. Panel (a) shows slightly lower market shares among brands consumed at least 80% by women compared to those consumed at least 80% by men, but this difference is largely eliminated after controlling for half-year and module fixed effects. However, we note a drop-off of half a percentage point market share on average in both specifications occurring among brands consumed nearly entirely by women.[24]

We observe that women tend to consume goods with *lower* market shares and that female-gendered goods themselves also exhibit *lower* market shares. Abstracting from the role of pricing in the context of multiproduct firms, these results rule out a differential competition channel in driving differences in price paid by men and women. We can conclude that women are more price elastic consumers than are men and 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.

4.3 Differences in product quality along the gender distribution

Our model also prescribes a method for estimating subjective and intrinsic quality. Applying a logarithm to and rearranging Equation (7) yields an expression for subjective product quality of good j to good k in the same module n:

$$\log \frac{\varphi_j(g)}{\varphi_k(g)} = \left(\frac{1}{\sigma(g) - 1} \right) \cdot \log \frac{x_j(g)}{x_k(g)} - \log \frac{p_j(g)}{p_k(g)}. \tag{11}$$

As in Faber and Fally (2022), we let household quality evaluations depend on an intrinsic quality term $\log \phi_j$ and multiplicative term $\gamma(g)$ such that $\log \varphi_j(g) = \gamma(g) \log \phi_j$. Using a normalization such that that the average multiplicative term over genders averages to one, the intrinsic quality term corresponds with the democratic average quality evaluation across households:

$$\log \phi_j = s_{female} \cdot \log \varphi_j(female) + (1 - s_{female}) \cdot \log \varphi_j(male) \approx \bar{\varphi}_j \tag{12}$$

for s_{female} corresponding with the share of female consumers in the entire market (typically around 0.5; i.e. not only consumers of the good).[25] In the case that a good is only purchased

[24] Figure A.3 display analogous results for UPC gender and competition; Figure A.4 replicates Figure 5, weighting by log aggregate module expense. Both sets of results yield consistent conclusions.

[25] Note that in this setup, subjective quality-adjusted price can be written as $\log \frac{\tilde{p}_i(g)}{\tilde{p}_j(g)} := \log \frac{p_i}{p_j} - \log \frac{\varphi_i(g)}{\varphi_j(g)} =$

by a single gender, the intrinsic quality equals the subjective quality of the purchasing gender, as implicitly, goods that are never purchased a given consumer gender correspond with a subjective quality valuation of $-\infty$ for that consumer gender. Following Faber and Fally (2022), we view the consumer gender that does not purchase a good at all as *not* comprising part of the consumer base that constitutes the democratic average.

We follow this structure to estimate brand-level intrinsic quality and quality-adjusted price for brands relative to their module-averages in the half year, as well subjective quality for men and women separately. This process features several steps. First, we construct budget shares for each individual $\times$ brand $\times$ half-year, which we express relative to their own module-average budget shares and apply a logarithm (per Equation (11)). We then average these values within each gender $\times$ brand $\times$ half-year (applying the appropriate Nielsen proprietary weights). Next, we assign each gender-department an elasticity of substitution from Table 4. Our main specification uses only departments with the department codes with elasticities that are significantly greater than one in absolute value.[26] We measure brand- and module-level unit prices as the ratio of aggregate expense to aggregate amounts purchased in a half-year. From here, we use Equation (11) to compute gender $\times$ brand $\times$ half-year levels of subjective quality and related objects and Equation (12) to compute brand $\times$ half-year measures of "intrinsic" quality and related measures. We residualize these measures relative to half-year and module averages.[27]

This process allows us to speak to subjective quality, intrinsic quality, and quality-adjusted prices perspectives of consumers and from the perspectives of goods themselves.

$\frac{1}{1-\sigma(g)} \cdot \log \frac{x_i(g)}{x_j(g)}$ and intrinsic quality-adjusted price as $\log \frac{\tilde{p}_i}{\tilde{p}_j} := \log \frac{p_i}{p_j} - \log \frac{\phi_i}{\phi_j}$.

[26] To demonstrate the robustness of our approach, we also include results using all departments with elasticities of substitution bounded at least one standard error above one (i.e. also including the Meat department) as well as results using all departments with point estimates of elasticities of substitution greater than one (i.e also including Dairy and Non-good Groceries in addition to meat). See Figure B.2 and Figure B.3.

[27] We validate our quality measures by replicating Figure 2 in Faber and Fally (2022) in our setting. Figure B.1 shows a monotonically increasing relationship between average quality consumed and total individual expense (proxying for income, which is imprecisely measured in the Nielsen survey) and a decreasing relationship between quality-adjusted price and total individual expense. The magnitudes of our estimates are slightly smaller than those documented in Faber and Fally (2022), which we attribute to differences in our sample construction and our methodological approach (namely, we study individuals where Faber and Fally (2022) studies households, and we leverage the continuity of the HMS panel over time whereas Faber and Fally (2022) effectively treat their data as half-year cross sections.

To speak to the quality-composition of individuals' consumption bundles, we compute for each individual $\times$ half-year their residualized expense-weighted quality consumed and quality-adjusted price consumed and assign each individual their mean over half-years (as in Section 4.2). We then plot values of expense-weighted average intrinsic quality and quality-adjusted price consumed along household gendered-consumption rank. Figure 6 plots the evolution of average quality consumed and quality-adjusted unit price faced along ranks of gendered consumption. The first panel shows a flat relationship between quality and gendered-consumption rank until around the 55th percentile of the gendered-consumption rank, at around 5% lower quality than module $\times$ half-year mean quality. Starting at this point, the relationship broadly monotonically increases to around 10% higher quality at the highest levels of gendered-consumption. We take this result to corroborate our finding that the pink tax price differential is driven by marginal costs and quality rather than markups. Panel (b) elaborates on this result, plotting quality-adjusted price. The figure shows that, after accounting for product quality, individuals with the highest expense-concentration of female goods actually pay around 1% lower per unit relative to the average quality-unit $\times$ unit price within the module $\times$ half-year. We also observe slightly lower quality-adjusted prices paid by individuals with high expense-concentration of male goods.

Figure 8 disaggregates this result by plotting female- and male-subjective quality valuations of goods along brand-gender. The result demonstrates two interesting asymmetries which likely drive our intrinsic quality results. First, women assess the quality of female goods higher than ungendered goods and male goods, both of which are assigned approximately equal quality valuations. However, men assess male goods higher valuations than ungendered goods and female goods—but with female goods valued substantially higher than ungendered goods.[28] Second, the magnitude of the subjective quality valuations differ between men and women. Namely, Panel (a) shows that women assess the quality of women's goods at around 30% higher than ungendered and male goods, but men assess the quality of men's goods around 20% higher than ungendered goods and around 10% higher than female goods. These asymmetries allocate relatively greater quality-mass to female goods than do

[28]We interpret men and women as most-highly valuing their respective goods as further validating our approach.

men's goods.[29]

We take these results to demonstrate that women's goods are assessed as of higher quality than men's goods. To the extent that quality serves as a proxy for marginal cost (e.g. as in the supply-side model of Faber and Fally (2022)), we take these results to demonstrate that women's consumer goods exhibit higher marginal cost. Overall, this section demonstrates that within this model environment, women 1) demonstrate greater price-sensitivity than do men, 2) do not consume bundles that exist in less-competitive environments, and 3) consume goods of higher assessed quality than do men. These results jointly imply that the Pink Tax is not driven by markups and demand factors, but rather due to quality, preferences, and gender-differential sorting into goods along these dimensions.

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.[30]

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.[31] 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

[29] Figure 7 replicates Figure 6 using brand-gender along the x-axis, demonstrating similar results.

[30] 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).

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

inference on gender differences in *overall* markups $\mu_f - \mu_m$ under the following condition:

$$\mathbb{E}[\log(\mu_f) - \log(\mu_m) \mid \log(\mu_f^r) - \log(\mu_m^r)] \approx \mathbb{E}\Big[\Delta\%c + \Delta\%\mu^m + \Delta\%\mu^d \mid \Delta\%\mu^r\Big] = 0 \quad (13)$$

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 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.[32] This matched subsample accounts for 37% of purchase volume we observe in the HMS panel during this timeframe.[33] 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

[32] We report the following match rate by department: 1) Health and beauty (5.5%), 2) Dry grocery (12.5%), 3) Frozen foods (14.2%), 4) Dairy (12.8%), 5) Deli (8.3%), 6) Packaged meats (16.5%), 7) Produce (3.3%), 8) Non-food grocery (10.3%), 9) Alcohol (0.2%), 10) General merchandise (2.7%). We exclude alcohol from this part of the analysis due to the low match rate.

[33] Table C.1 elaborates on the characteristics of balance based on whether UPC$\times$market$\times$year observations merge to the PriceTrak data. Around 14% of transactions map to the PriceTrak Data. Women are about 3% more likely than men to make a purchase that maps to the PriceTrak data (+0.4pp relative to a 14.2% baseline). Restricting to purchases made within the same year $\times$ module $\times$ county $\times$ retailer combination, observations that merge to the PriceTrak data are between 8 and 9.4% more expensive per unit than are those that do not merge. These relationships do not threaten the internal validity of our approach to using the PriceTrak data, but may have important external validity implications.

matches to a wholesale price, around 8% of transactions exhibit a negative markup (i.e. observed wholesale prices greater than unit prices). We discard these goods with observed negative markups in this part of the analysis because we cannot apply the logarithm to these negative values; moreover, we cannot easily distinguish whether such observations reflect measurement error or loss-leader pricing.[34] 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.[35]

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.[36]

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, indicating 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 77% reduction in the gender-differential unit prices paid relative to as reported in column (5)—after conditioning

[34] 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).

[35] 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.

[36] Section C presents additional conceptual and empirical evidence on markups as inferrable from PriceTrak data. See Section C.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.

on location, age, race, retailer, and income demographic.

Figure 9 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.[37]

6 Differentiated Products: Markups and Marginal Costs

Having demonstrated that women tend to consume more price-elastically and that they tend to purchase products with higher wholesale prices, we now turn our attention to estimat-

[37]Table C.2 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.

ing 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 complement the prior analyses (which we view as largely broader in scope) and address these limitations, we estimate markups and marginal costs while allowing for significantly more model complexity at the cost of narrowing our focus to fewer markets. We use data on the store-week level do not face the same sparsity issues as the aggregated individual-level data. 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 the market for disposable razors. We choose disposable razors because they have a high level of dispersion of $\hat{w}_j$ across their product space and because they are commonly referred to as the canonical pink tax item. Disposable razors exhibit near complete gender segregation, and they also feature observable product characteristics. Figure D.1 plots the histogram of woman purchase share for the sales of razors included in our sample for demand analysis. The distribution is bi-modal with the vast majority of razors either bought mostly by women or mostly by men.

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:

$$u_{ijt} = \beta_i \mathbf{x}_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{14}$$

where x_{jt} is a three dimensional vector of a constant term, the woman purchase share of product j and the price of product j in market t. $\xi_{jt} = \xi_{jr(t)} + \xi_{m(t)} + \Delta\xi_{jt}$ 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 system is that the three dimensional vector β_i of coefficients on prices, woman purchase share, and the outside option, are individual-specific coefficients.[38] 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:

$$\beta_i = \mathbf{\Sigma} \cdot \mathbf{v}_i, \qquad \mathbf{v}_i \sim N(0, \mathbf{I_3})$$

Heterogeneity in preferences for product gender will generate more flexible substitution patterns than 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 woman purchase share 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

[38] We define the outside option to include no purchases of any razor or shaving product as well as purchases of non-disposable razors and depilatories. In our data, men and women that have ever purchased a shaving product are similarly likely to have purchased a disposable razor (53% of men vs 61% of women). Men and women are equally likely to purchase a non-disposable razor (18%) while women are more likely to purchase a depilatory (20% vs 5%). Despite this observation, the men and women in our sample appear equally likely to substitute or switch product type purchases. The average man in our sample that purchases a disposable razor purchases from 1.85 product modules that are related to shaving and the average women purchases from 1.8 shaving-related product modules.

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.[39]

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

$$s_{jt} = \int \frac{\exp(\beta_i \mathbf{x}_{jt} + \xi_{jt})}{1 + \sum_k \exp(\beta_i \mathbf{x}_{kt} + \xi_{kt})} d\beta_i \tag{15}$$

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 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 of moments. We estimate the model using the store level RMS data, which contains weekly observations of prices and quantities sold of each product offered. Because disposable razors constitute infrequent and storable purchases, we aggregate the data to the quarter level. We make a variety of restrictions to our sample to aid in estimation. We use data from

[39] 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.

2013 and 2014[40] and keep only larger retail chains: those that have above median count of store-quarter observations and require that each store has observed razor purchases in each quarter or week of the data. We drop stores that offer relatively few products: those less than the 25th percentile. We also drop products that are in the bottom 25th percentile of number of stores they are offered at. We make an additional sample restriction to ensure that our instruments have enough power for identification: we calculate the coefficient of variation of prices within the retail chain level and keep chains that have an above median coefficient of variation. These retail chains are those that put items on sale at higher frequency, which give our price instruments greater power.

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 a square ownership matrix of J products, Ω, that maps each product in our data to a common owner so that element jk is equal to 1 if product j and product k are owned by the same firm and 0 otherwise.[41] 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:

$$\frac{p^* - mc}{p^*} = -(\Omega J)^{-1} \frac{s(p^*)}{p^*}, \tag{16}$$

Because we observe price, identified markups also identify marginal costs.[42]

6.2 Differentiated Products Demand Model Results

Table 6 shows average differences in prices, own-price elasticities, markups, and marginal costs between men's and women's disposable razors across all products in our sample in

[40] We restrict the data to two post-recession years for computational tractability purposes.

[41] We construct the ownership matrix through manual search, Capital IQ, and newspaper articles. Although the Nielsen data contain information about brands, firms in the data frequently own multiple brands.

[42] Table D.1 presents the estimated parameters.

Panel (a) and within firms in Panel (b). We define a razor as a women's razor if its woman purchase share is above 0.6, otherwise we classify the razor as a men's (or non-gendered) razor; this definition fits the bimodal distribution of razor female purchase share according to Figure D.1. We present results from a series of regressions on the product-store-quarter level of the outcome of interest on an indicator for whether the razor is a women's razor or a men's razor and include fixed effects for the store-quarter; these regressions are weighted by the total sales volume for each product-store-quarter observation. We interpret the coefficient as the difference between the average men's razor and a woman's razor purchased at the same store in the same time period.

Panel (a) shows results without including firm fixed effects. Across razors, women's razors exhibit both higher marginal costs and markups, but statistically identical price elasticities. The first column shows the difference in prices, which we observe in the data. Women's razors are, on average, priced 9 cents higher per unit than men's in the data. The second column presents the results for our estimated own-price elasticities which are given by $\varepsilon_{jt} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}}$. We find that women's razor purchases are not significantly differently elastic from men's on average, with a difference close to zero. In the third column we present the difference in estimated marginal costs for men's and women's razors. We find that women's razors are associated with 9 cent higher marginal cost than men's on average. However, in column (4) we present our estimated markups, which are given $\frac{p-mc}{p}$. We find that women's razors have higher markups by about 14 percentage points. This result may seem counterintuitive, as on average the difference in marginal costs fully offsets the difference in price. The finding that women's razors have higher markups is due to the non-linear relationship between markups and price along with the correlation structure of prices and marginal costs in the men's razors versus women's. Overall, we take our results as evidence that although women's razors have higher markups on average, gendered price differences in razors do not seem to be driven by systemic price discrimination, as women appear to be no differently elastic than men are.

Panel (b) includes firm fixed effects. The majority of our work thus far has looked at differences in prices, elasticities, and markups between men and women across a product market. However, price discrimination is often thought of as a way for a firm to segment their market and profit maximize. In our CES analysis, we could not explore multi-product firm offerings, as the Nielsen data does not include parent company data for products. Restricting our analysis to a smaller product market, we can more feasibly construct the product ownership matrix for disposable razors. This parent company data is contained within the ownership matrix, Ω, that captures multi-product firm's pricing incentives.[43]

Table 6 Panel (b) includes firm level fixed effects to test for price discrimination on men's and women's razors by capturing average differences for men's and women's products within-firm. In column (1), we show that, within firm, women's razors are priced significantly higher than men's by about 35 cents per razor. Column (2) shows that, within firm, women's razors exhibit significantly greater price elasticity than those of men by 35 percentage points. The average price difference between men's and women's razors is almost completely explained by differences in average marginal costs, as shown in column 3. Finally, column 4 shows that within firm, women's razors have lower markups than men's by about 11 percentage points on average. These results directly refute the common narrative around the pink tax in the media, that women are charged unfair markups on nearly identical products. The differences in our results between Panels (a) and (b) that vary firm fixed effects also highlight the important role of sorting between product markets. Our results suggest that men and women differentially sort into different brands so as to mitigate the observed within-firm differences in razor supply and demand attributes.

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 5% 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

[43] 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.2.

for goods purchased at gender-parity. Not only do we observe a consistent women's price premium of 21.2% on overtly gendered goods, but we also observe that women sort into purchasing less-overtly-gendered products with higher prices 4.8%, such as organic foods.

We proceed by studying the causal components of this pink tax. We distinguish three 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 approximately 15% more price sensitive than are men. We further adapt the model of Faber and Fally (2022) to our setting allowing us to estimate by-gender subjective product quality, intrinsic quality, and quality-adjusted price across goods. We find that women's consumer goods are assigned 25% greater intrinsic quality than goods consumed by both genders, whereas those consumed by men are assigned only 10% greater quality. Individuals with the greatest female-goods concentration of their consumption baskets feature an average quality-consumption level about 10% higher than mean-product quality, where those with the greatest male-goods concentration of their consumption baskets see 5% lower average quality-consumption level about 10% higher than mean-product quality. We also document that women's goods do not operate in less-competitive markets than those of men's, implying that a competition-channel is unlikely to generate greater markups for women.

We corroborate this central finding by implementing several additional designs that leverage complementary data and identification techniques. We combine our scanner data with data from PriceTrak for a subset of our goods on wholesaler prices paid by retailers. Wholesale prices represent the cost of the product charged to the retailer. We demonstrate that coverage 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. Studying the goods themselves, we find slightly higher markups placed on men's goods than on women's goods: within market, the most-male goods face a 10pp higher markup than the most-female goods; when weighting by expense, this difference expands to 30pp.

Finally, we estimate overall markups and marginal costs of production for disposable razors using a differentiated products demand model (Berry, Levinsohn, and Pakes (1995)). We view this part of our analysis as complementary insofar as we can focus on one of the canonical examples of "pink tax goods" and incorporate additional model complexity. We use this model to estimate product-market level demand elasticities, marginal costs, and markups. We find that products disproportionately consumed by women have higher marginal costs of production. We estimate that women's razors see 14pp greater markups unconditionally across razor brands (weighting by overall product expense); however, including firm fixed effects generates a negative markup difference, with women's razors exhibiting an 11pp lower markup than men's razors produced by the same firm.

We conclude from our analysis a novel set of facts to frame the discussion of the pink tax: women pay around 5% more per unit for similar goods than do men; when we study overtly gendered goods, this price difference rises to 21.2%. 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 is 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 recent and ongoing 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.

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 (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.

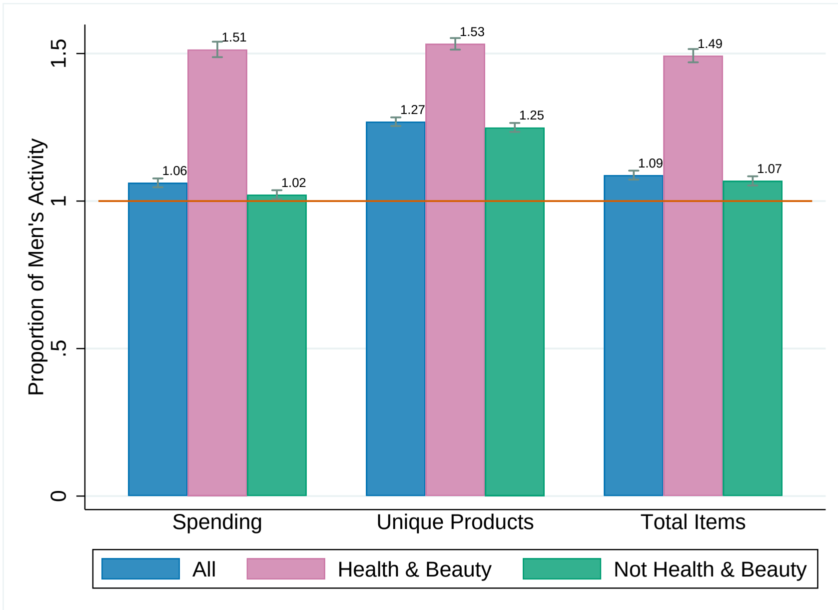
Stigler, George J. (1987). *The Theory of Price*. Macmillan Publishing Co., Inc.

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.

Main exhibits

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 log unit prices

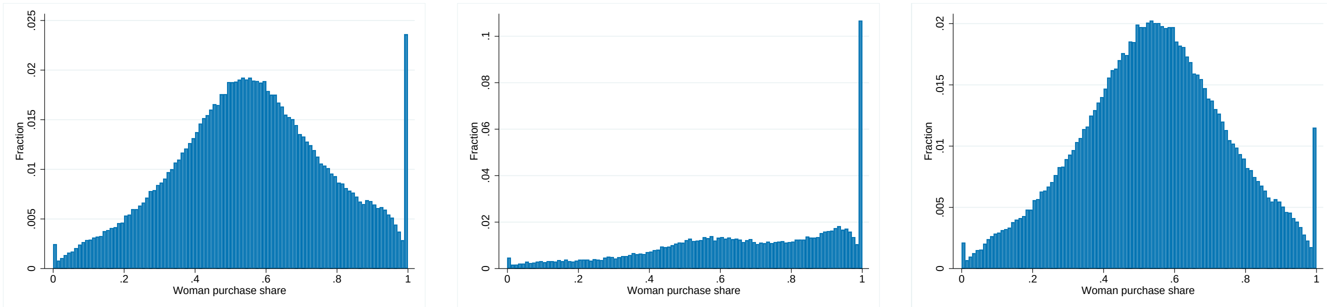
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Unit prices in same product module							
Women	0.0230***	0.0299***	0.0514***	0.0545***	0.0489***	0.0381***	0.0501***
	(0.0035)	(0.0038)	(0.0037)	(0.0031)	(0.0025)	(0.0027)	(0.0026)
Men's Average	$1.329	$1.276	$1.276	$1.276	$1.276	$1.276	$1.276
Mod. X Units FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.83	0.84	0.84	0.88	0.91	0.92	0.92
N	153,333,383	977,157,26	97,715,726	97,715,726	97,715,726	97,715,726	97,715,726
Number of clusters	49,256	49,248	49,248	49,248	49,248	49,248	49,248
Panel (b): Unit prices for same product							
Women	-0.0089***	-0.0104***	-0.0030	-0.0060	-0.0120***	-0.0190***	-0.0139***
	(0.0017)	(0.0031)	(0.0031)	(0.0045)	(0.0034)	(0.0040)	(0.0039)
Men's Average	$1.293	$1.709	$1.709	$1.709	$1.709	$1.709	$1.709
UPC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.95	0.95	0.95	0.96	0.97	0.97	0.97
N	151,184,396	441,147,94	44,114,794	44,114,794	44,114,794	44,114,794	44,114,794
Number of clusters	49,256	49,202	49,202	49,202	49,202	49,202	49,202
Common fixed effects sample	No	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	No	No
Month × Year FE	No	No	No	No	No	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes	Yes
Retailer FE	No	No	No	No	Yes	Yes	Yes
Demographic FE	No	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: $log(P_{ijt}) = \phi_{t(j)} + \beta Woman_i + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $Woman_i$ indicates whether individual i 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: $log(P_{ijt}) = \phi_{jt} + \beta Woman_i + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $Woman_i$ 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. Column 1 can be thought of as a raw gap between single men and single women within the same year, and each subsequent column demonstrates the contribution of controlling for an additional market or demographic factor. Note that the sample size in column (1) is larger than in the other columns, which use the sample estimated under the most restrictive fixed effect specification in column (7), where many singletons are dropped. All standard errors are clustered at the individual-level.

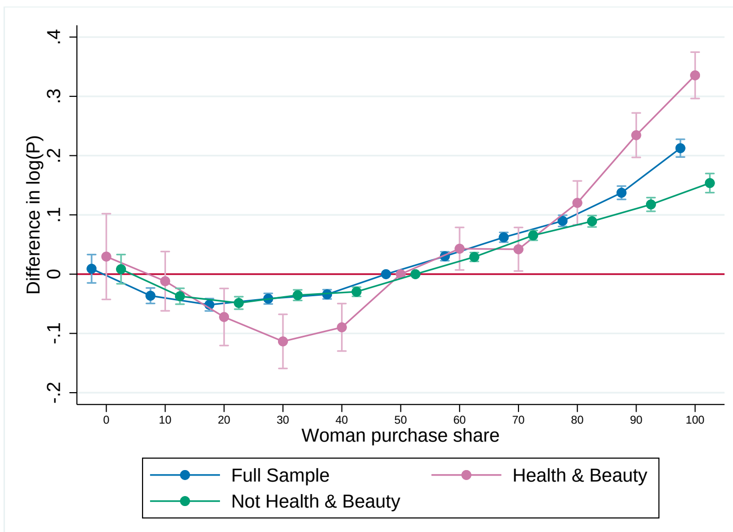
Figure 2: Distribution of Woman Purchase Share (UPC-gender) Across UPCs



(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.

Figure 3: Prices of UPCs by Woman Purchase Share



Note: This figure presents plots of the results of the regression $\log P_{jct} = \theta_{n_j ct} + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{w_j \in Bin_b\} + \varepsilon_{jct}$. Bins $b \in \mathcal{B}$ include five-percentile-radius bins centered at values $5 + 10k$ and two bins for pure gender stratification at the tails partitioning the interval $[0, 1]$. The regression includes fixed effects for product module m_j, county c, and half-year t. Results are presented for the whole sample, also separating out Health and Beauty and Dry Grocery. Standard errors are clustered at the UPC-county level.

Table 2: Log unit prices by gender of product and consumer

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
Women	0.0473***	0.0654***	0.0468***
	(0.0026)	(0.0078)	(0.0026)
Gendered Product	-0.0523***	-0.0027	-0.0571***
	(0.0065)	(0.0195)	(0.0068)
Women × Gendered Product	0.1452***	0.1220***	0.1424***
	(0.0076)	(0.0201)	(0.0085)
Men's Ungendered Average	\$1.242	\$4.012	\$1.153
Module × unit × retailer × county × month FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes
Adj. R-squared	0.92	0.88	0.92
N	92,29,1891	3,867,947	88,423,944
Number of clusters	49,247	47,492	49,243

Note: This table presents estimates from the regression: $log(P_{ijct}) = \phi_{jct} + \beta_1 Woman_i + \beta_2 GenderedProduct_j + \beta_3 Woman_i \cdot GenderedProduct_j + \gamma X_i + \epsilon_{ijct}$. ϕ_{jct} is a vector of fixed effects for the interaction of product module, units denomination, retailer chain, county, and month defined by product-location-time tuple (j, c, t). 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. Standard errors are clustered on the individual-level.

Table 3: Elasticities of Substitution

	OLS		Hausman IV		Retailer IV		Both IVs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_m - \sigma_w$	-0.00679	-0.0760***	-0.143***	-0.205**	-0.236***	-0.332***	-0.159***	-0.221***
	(0.00529)	(0.00941)	(0.0455)	(0.0922)	(0.0707)	(0.112)	(0.0417)	(0.0792)
$1 - \sigma_m$	0.289***	0.680***	-0.272***	-0.623***	-0.162**	-0.275***	-0.243***	-0.511***
	(0.00638)	(0.00828)	(0.0381)	(0.0814)	(0.0635)	(0.101)	(0.0356)	(0.0709)
% Elasticity dif. (|women| v. |men|)	0.95	23.75	11.24	12.63	20.31	26.04	12.79	14.63
1st Stage F-Stat	-	-	1121	839.6	788.2	552.2	3173	2560
Module× gender×DMA× half-year FE	Yes	No	Yes	No	Yes	No	Yes	No
Mod.× gen. × retailer× DMA× HY FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	13,398,189	7,928,011	9,578,130	5,429,742	13,398,189	7,928,011	9,578,130	5,429,742
Number of clusters	1.02E+06	756,534	763,079	573,630	1.02E+06	756,534	763,079	573,630

Brand-DMA level clustered standard errors in parentheses

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

This table presents the results OLS and IV regressions on the product-consumer gender-retailer-DMA-semester level of estimating average log budget share on log price changes, controlling for different fixed effects: $\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$.

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.
$\sigma_m - \sigma_w$	0.0752	-0.105	-0.637**	-0.489**	0.638	-0.862**	-0.274*	-0.164	-0.612	-1.232**
	(0.458)	(0.126)	(0.284)	(0.190)	(0.749)	(0.370)	(0.145)	(0.470)	(2.796)	(0.598)
$1 - \sigma_m$	0.189	-0.661***	-0.901***	-0.0828	-1.766**	-0.418	-0.361**	-0.0412	1.668	0.795
	(0.395)	(0.109)	(0.237)	(0.169)	(0.693)	(0.312)	(0.152)	(0.446)	(2.763)	(0.568)
% Elasticity dif. (|women| v. |men|)	-9.27	6.32	33.51	45.16	-23.07	60.79	20.13	15.75	-91.62	600.98
Observations	326,867	2,737,720	588,628	538,573	103,133	204,639	283,830	521,080	37,210	87,925
Number of clusters	58,072	268,531	62,751	38,296	11,819	17,077	27,068	62,713	8,527	18,661
MGDRT FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hausman DMA IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Retailer IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

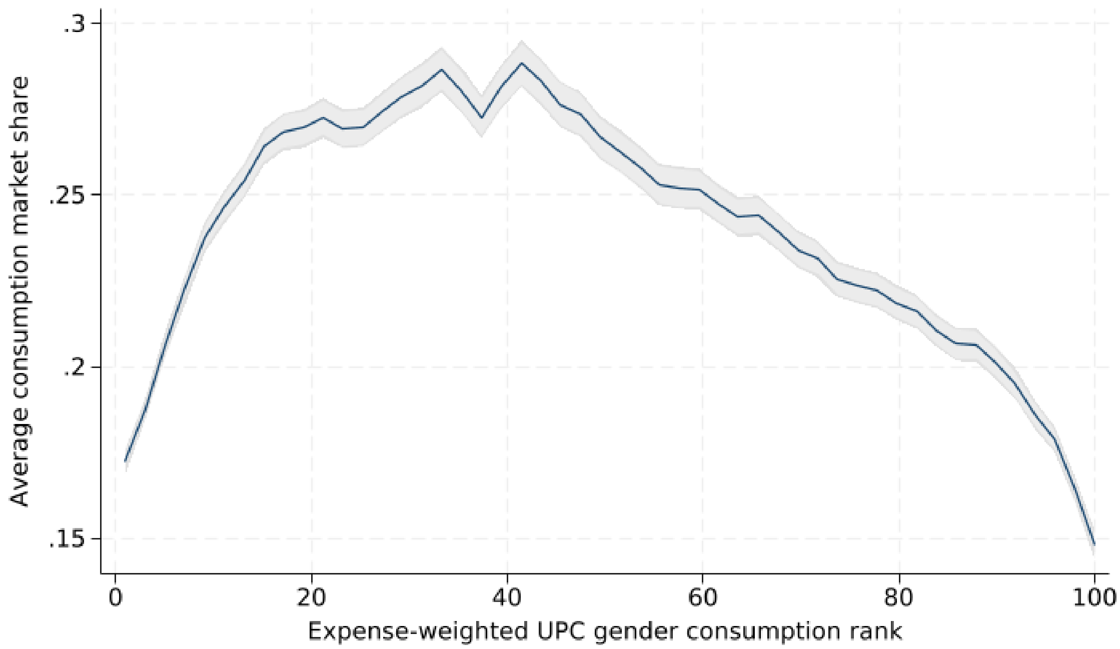
Brand-DMA level clustered standard errors in parentheses

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

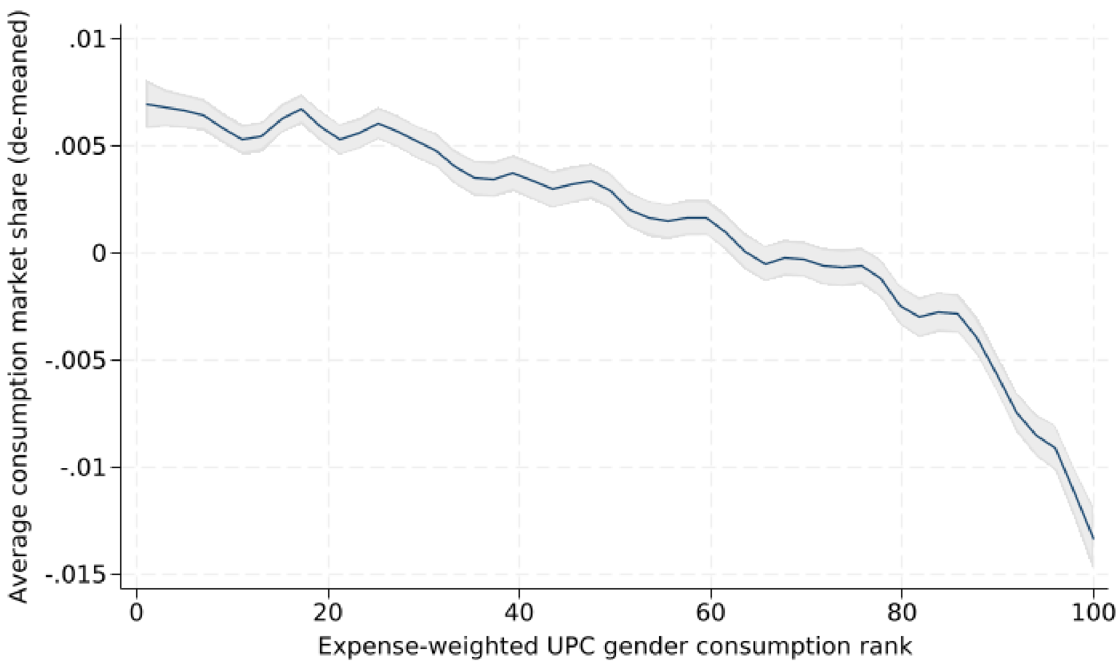
This table presents the results OLS and IV regressions on the product-consumer gender-retailer-DMA-semester level of estimating average log budget share on log price changes, controlling for different fixed effects: $\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$. Each column stratifies the data by department; every column uses both Hausman and Retailer leave-out instruments in an instrumental variables regression. "MGDRT FE" refers to module×gender×DMA×retailer×half-year fixed effects.

Figure 4: Average market shares consumed by households ranked by gendered consumption

(a) Raw average market share



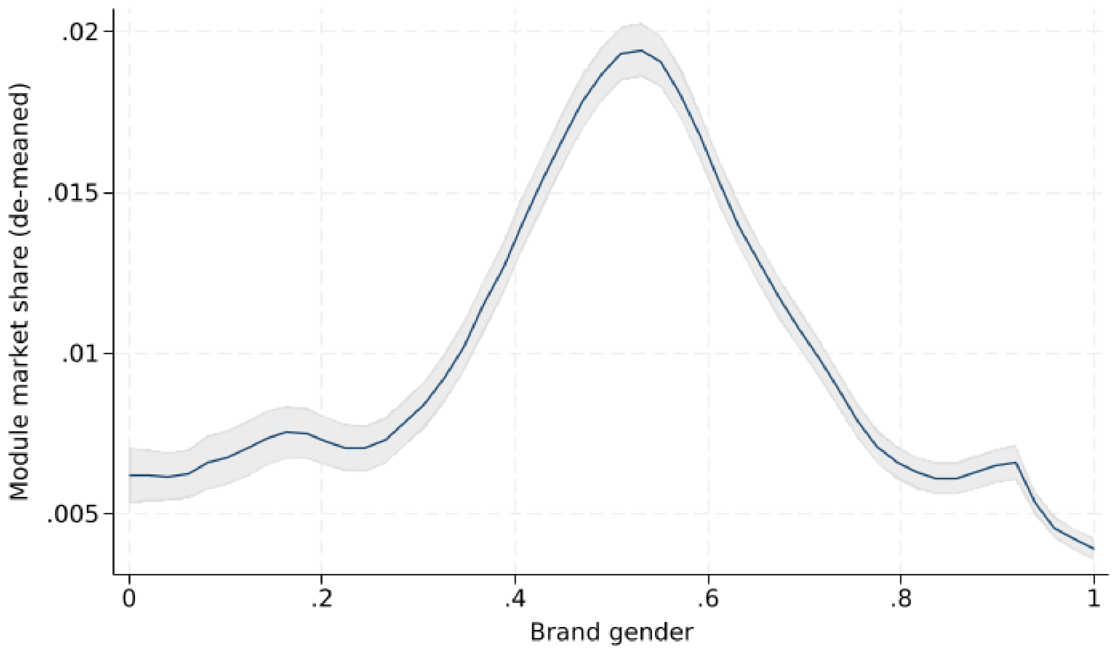
(b) Market shares residualized against module-average and half-year



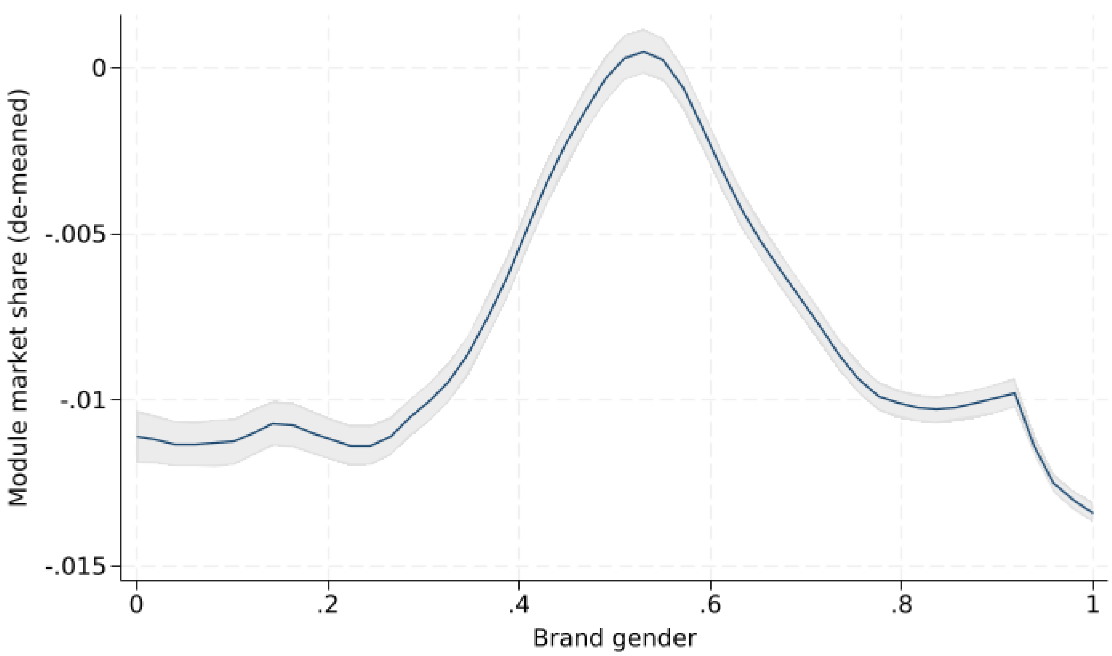
Note: These figures display average market shares consumed along percentile ranks of the distribution of gendered consumption. On the x-axis in both panels, the left-most values corresponds with individuals consuming the greatest share of most "male products" and the right-most values corresponds with individuals consuming the greatest share of most "female products". Panel (a) plots the average expense-weighted market share consumed as the dependent variable; Panel (b) plots the average of these market shares residualized against half-year and module fixed effects. Each observation is a household. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

Figure 5: Market shares by brand gender

(a) Raw average market share



(b) Market shares residualized against module-average and half-year

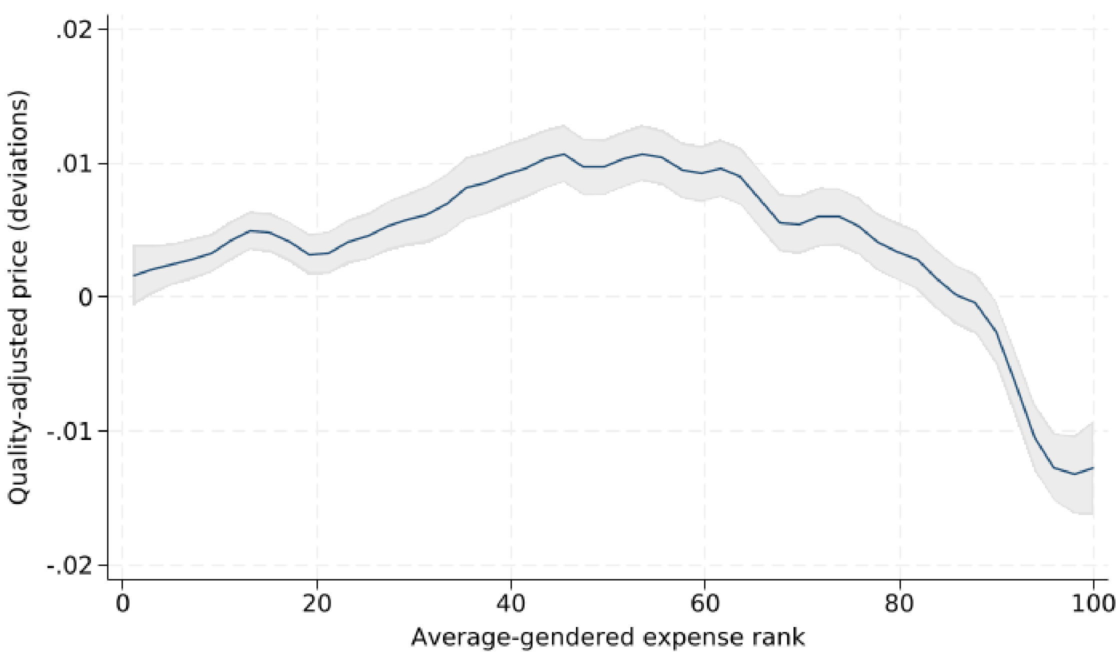


Note: these graphs display the average national market shares of brands within modules based on their observed brand gender. Each observation is a brand-semester between 2004 and 2019. Panel (a) shows raw average market shares; Panel (b) plots average market shares residualized within each half-year and module. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

Figure 6: Average intrinsic quality and quality-adjusted price
Households ranked by gendered consumption (male-to-female)

(a) Intrinsic quality

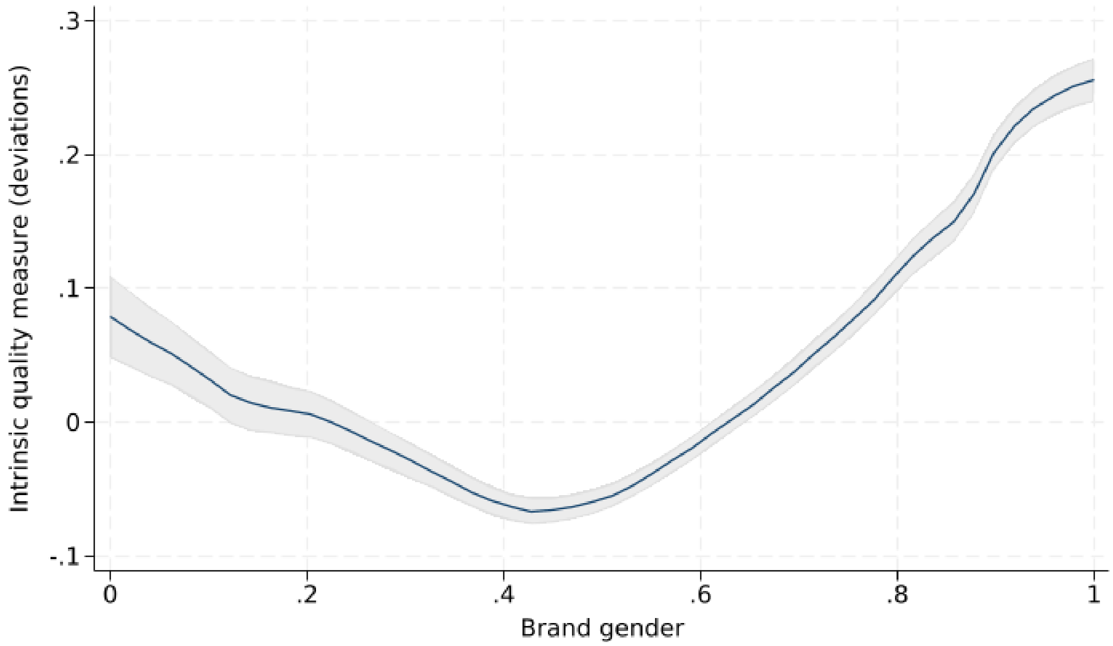
(b) Quality-adjusted price



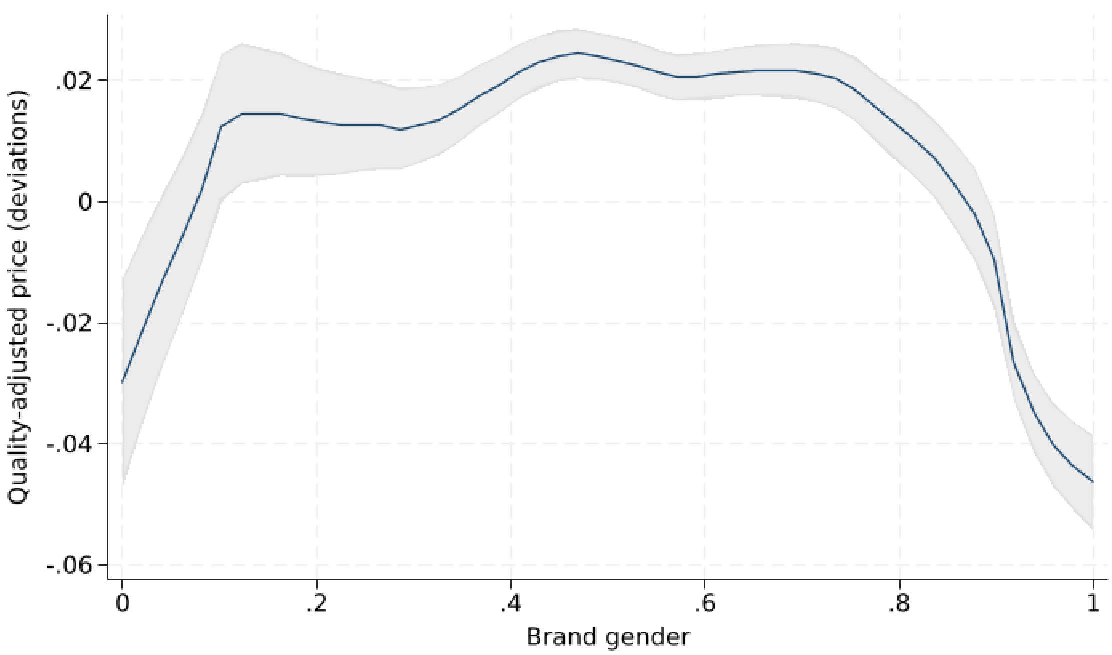
Note: These graphs plot intrinsic quality (Panel (a)) and quality-adjusted price (Panel (b)) along individual-ranks of gendered share of consumption. Each observation is a unique individual in the Nielsen HomeScan Panel. Individuals on the left have the highest expense-concentration of "male goods" in their consumption baskets; individuals on the right have the highest expense-concentration of "female goods" in their consumption baskets. Individual observations are weighted using Nielsen proprietary probability weights. Quality and quality-adjusted price are measured in units percent deviation relative to the mean good in each module $\times$ half-year. The fitted relationships correspond to local polynomial regressions, and the displayed confidence intervals are at the 95% level.

Figure 7: Average intrinsic quality and quality-adjusted price by brand-gender

(a) Intrinsic quality



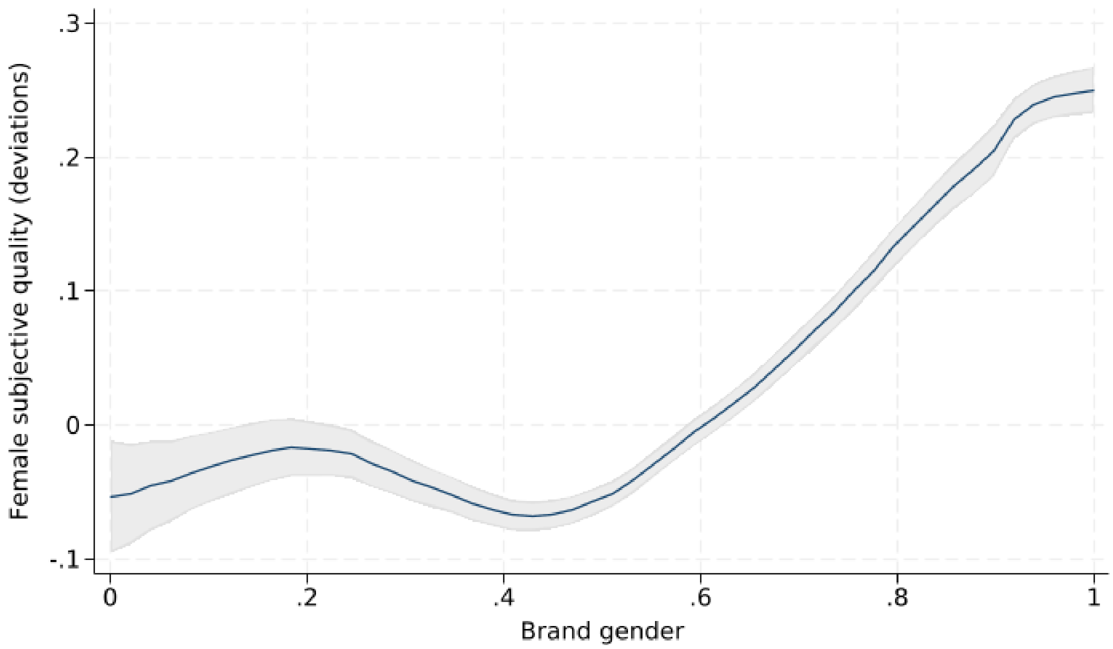
(b) Intrinsic quality-adjusted price



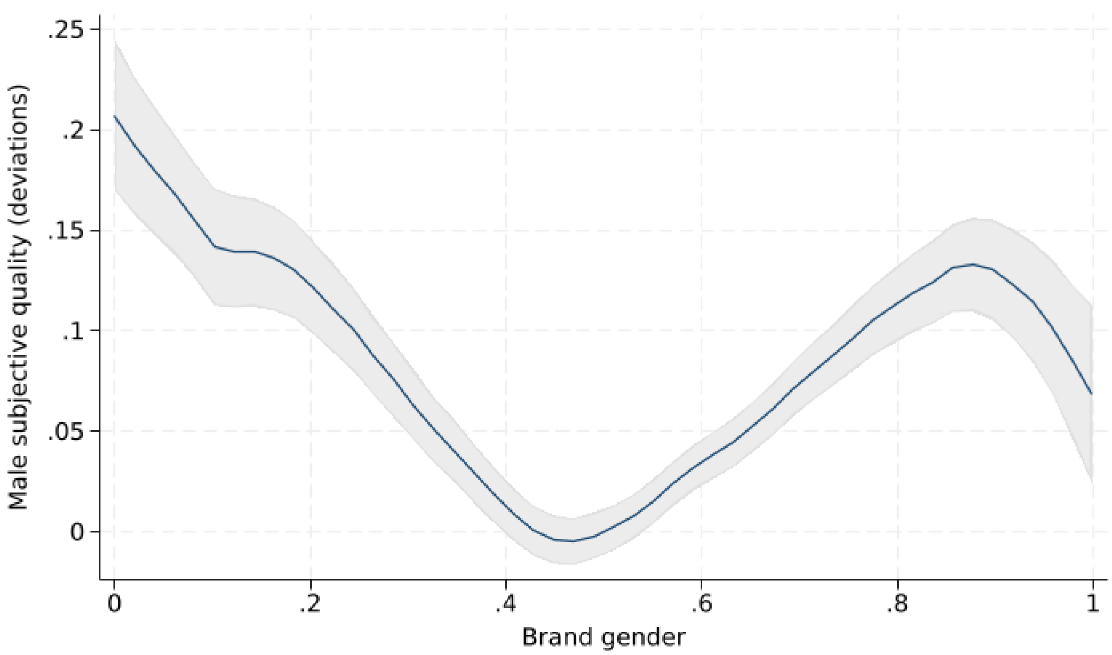
Note: These graphs plot intrinsic quality (Panel (a)) and quality-adjusted price (Panel (b)) along brand-gender. Each observation is a brand-semester. Brand gender is constructed as the expense-weighted mean UPC-gender of UPCs contained within the brand. The x-axis can be interpreted thus as the share of the consumption purchased by women (i.e. increasing in "female-ness" from left to right"). Quality and quality-adjusted price are measured in units percent deviation relative to the mean good in each module × half-year. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

Figure 8: Subjective quality valuations by brand-gender

(a) Female subjective quality valuations



(b) Male subjective quality valuations



Note: These graphs plot subjective quality valuations by women and men (Panel (a) and Panel (b), respectively) along brand-gender. Brand gender is constructed as the expense-weighted mean UPC-gender of UPCs contained within the brand. The x-axis can be interpreted thus as the share of the consumption purchased by women (i.e. increasing in "female-ness" from left to right"). Subjective quality is measured in units percent deviation relative to the mean good in each module × half-year, and constructed for men and women separately. The fitted relationships correspond to local polynomial regressions, and the displayed confidence intervals are at the 95% level.

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**	0.0501***	0.0394***	0.0092***
	(0.0035)	(0.0045)	(0.0034)	(0.0026)	(0.0039)	(0.0033)
Log wholesale cost			0.7600***			0.7041***
			(0.0018)			(0.0036)
Men's mean (levels USD)	$1.32	$0.43	$0.43	$1.28	$0.37	$0.37
PriceTrak sample	No	Yes	Yes	No	Yes	Yes
Module × Units	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	No	No	No
Month	No	No	No	Yes	Yes	Yes
Adj. R-squared	0.83	0.75	0.86	0.92	0.88	0.92
N	153,333,383	17,901,305	17,901,305	97,715,726	9,690,298	9,690,298
Number of clusters	49,256	28,412	28,412	49,248	28,356	28,356

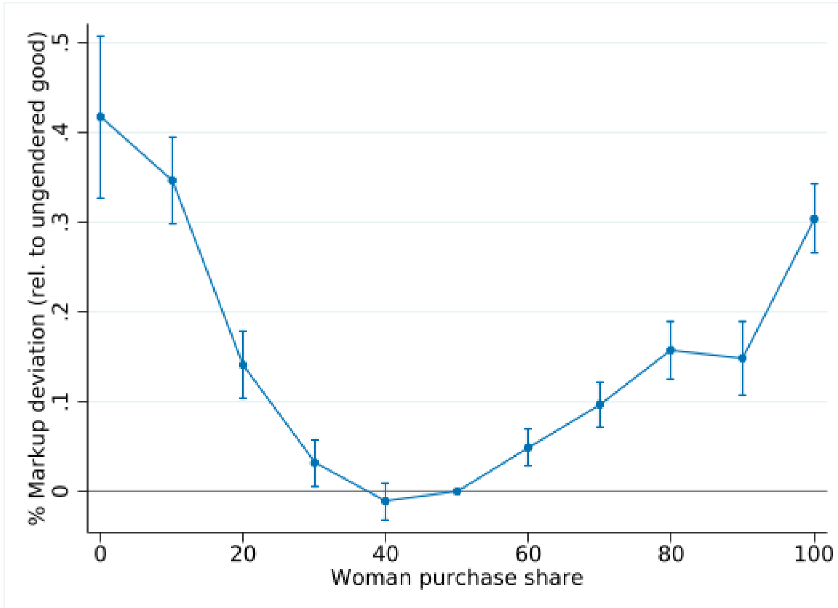
Panel B: Log markups, controlling for retailer cost

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0086	0.0035	0.0090	0.0060*	0.0098***	0.0044
	(0.0073)	(0.0061)	(0.0061)	(0.0032)	(0.0029)	(0.0041)
Men's mean (percent markup)	89%	89%	89%	90%	91%	93%
Demographics	No	Yes	Yes	Yes	Yes	Yes
Module × Units	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	No
Month × Year FE	No	No	No	No	No	Yes
Adj. R-squared	0.00	0.47	0.47	0.63	0.71	0.75
N	18,076,261	18,076,147	18,076,147	17,258,171	15,310,301	9,834,314
Number of clusters	28,412	28,412	28,412	28,406	28,393	28,357

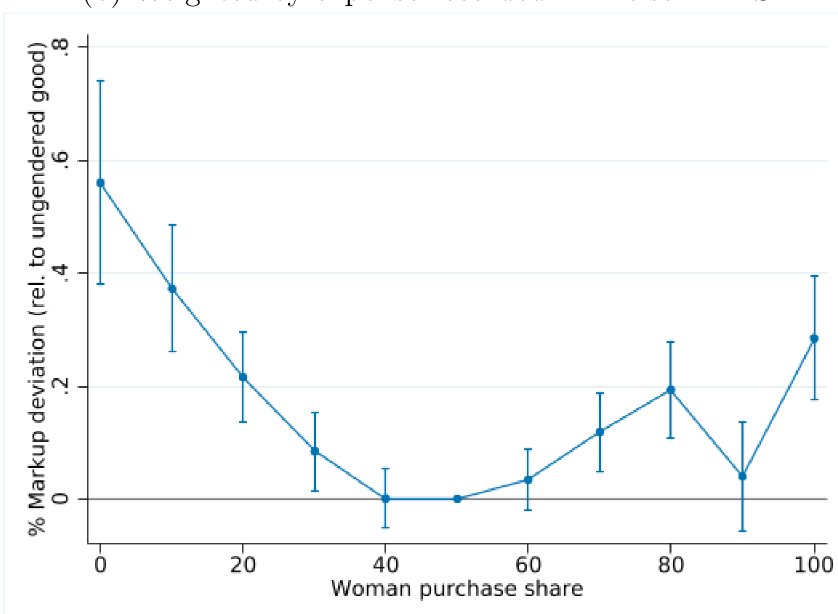
This table presents estimates from transaction-level regressions uniquely identified by an individual i purchasing product j at time t. Panel (a) estimates the form: $log(P_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \gamma C_{jt} + \Gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price 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 9: Log 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.

Table 6: Differentiated Products Model Estimates

	(1) Prices	(2) Own Elasticities	(3) Marginal Costs	(4) Markups
Panel A: Across all products				
Female razor indicator	0.0940***	-0.0033	0.0995***	0.1433***
	(0.0028)	(0.0035)	(0.0095)	(0.0026)
Observations	1,071,486	1,071,486	1,071,486	1,071,486
Adjusted R^2	-0.005	0.030	-0.040	-0.036
Quarter FE	Yes	Yes	Yes	Yes
Zip FE	No	No	No	No
Store FE	No	No	No	No
StoreXQuarter FE	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No

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

	(1) Prices	(2) Own Elasticities	(3) Marginal Costs	(4) Markups
Panel B: Within Firm				
Female razor indicator	0.3536***	-0.3526***	0.3510***	-0.1159***
	(0.0026)	(0.0027)	(0.0183)	(0.0033)
Observations	1,071,486	1,071,486	1,071,486	1,071,486
Adjusted R^2	0.094	0.195	-0.040	-0.028
Quarter FE	No	No	No	No
Zip FE	No	No	No	No
Store FE	No	No	No	No
Store× Quarter FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes

Heteroskedasticity-robust standard errors in parentheses

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

Note: This table displays regressions of prices and estimated price elasticities, marginal costs, and markups of disposable razors on a female-razor indicator. Female razors are defined as razors purchased at least 60% of the time by women, although 99% of razors see female purchase share above 80% or below 42%. Prices are directly observed in out data. Price elasticities, marginal costs, and markups are estimated from our differentiated products demand model described in Section 6. Standard

Appendix A Additional metadata and descriptive results

A.1 Additional descriptive evidence on the pink tax and product gender from the Nielsen data

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

	Total	Women	Men	Difference
Income	44249	39178	50231	-11053.8**
	(36755.08)	(33639.89)	(39288.29)	(509.3439)
Age	53.39	53.12	53.71	-.593**
	(16.5003)	(17.2805)	(15.5241)	(.2221)
High school	0.600	0.636	0.558	.078**
	(.4899)	(.4812)	(.4966)	(.0063)
College	0.240	0.209	0.276	-.068**
	(.4269)	(.4064)	(.4472)	(.0054)
Post-grad	0.120	0.113	0.129	-.017**
	(.3253)	(.3163)	(.3355)	(.0041)
White	0.781	0.762	0.802	-.04**
	(.4138)	(.4258)	(.3982)	(.0055)
Black	0.135	0.158	0.107	.051**
	(.3417)	(.3651)	(.3097)	(.0044)
Asian	0.0260	0.0230	0.0280	-.005*
	(.1577)	(.1499)	(.1663)	(.0023)
Hispanic	0.0690	0.0700	0.0670	0.00200
	(.2531)	(.2549)	(.2508)	(.0038)
No. households	49,256	35,428	13,828	21,600

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: Nielsen panelist behavior per month

	Total	Women	Men	Difference
Months in Panel	53.35	50.85	56.26	-5.407**
	(48.378)	(46.675)	(50.1261)	(.4468)
Trips	9.395	9.018	9.833	-.815**
	(6.5983)	(6.0547)	(7.1526)	(.0609)
Spending	258.8	259.6	257.9	1.644
	(177.0685)	(175.8798)	(178.4388)	(1.6378)
Spending inc. share	0.0120	0.0140	0.0100	.004**
	(.0208)	(.0235)	(.017)	(.0002)
Purchases	53.95	55.78	51.84	3.941**
	(32.122)	(32.2948)	(31.7906)	(.2966)
Unique products	25.67	28.44	22.45	5.985**
	(14.7973)	(15.2127)	(13.6116)	(.1341)
Unique modules	6.597	7.516	5.531	1.986**
	(15.3426)	(16.422)	(13.9114)	(.1416)
Unique groups	3.500	3.955	2.973	.982**
	(7.0203)	(7.3166)	(6.6215)	(.0648)
Coupon value	11.65	12.80	10.31	2.487**
	(15.3496)	(15.6305)	(14.9068)	(.1415)
Coupon use	8.229	9.159	7.150	2.009**
	(5.4355)	(5.6248)	(4.995)	(.0494)
Deal use	2.972	3.223	2.682	.541**
	(2.1307)	(2.2144)	(1.9902)	(.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.3: 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.
$*p < .05$, $**p < .01$.

Table A.4: Yearly spending differences between men and women

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0148*	0.0230***	0.0425***	0.0597***	0.0656***	0.0656***
	(0.0079)	(0.0075)	(0.0075)	(0.0074)	(0.0074)	(0.0074)
Men's Average (log)	7.789	7.789	7.789	7.789	7.789	7.789
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	0.09	0.10	0.12	0.13	0.13
N	232,205	232,061	232,061	232,060	232,060	232,060
Number of clusters	49,256	49,142	49,142	49,141	49,141	49,141

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.5: Yearly differences in log total items purchased by consumer gender

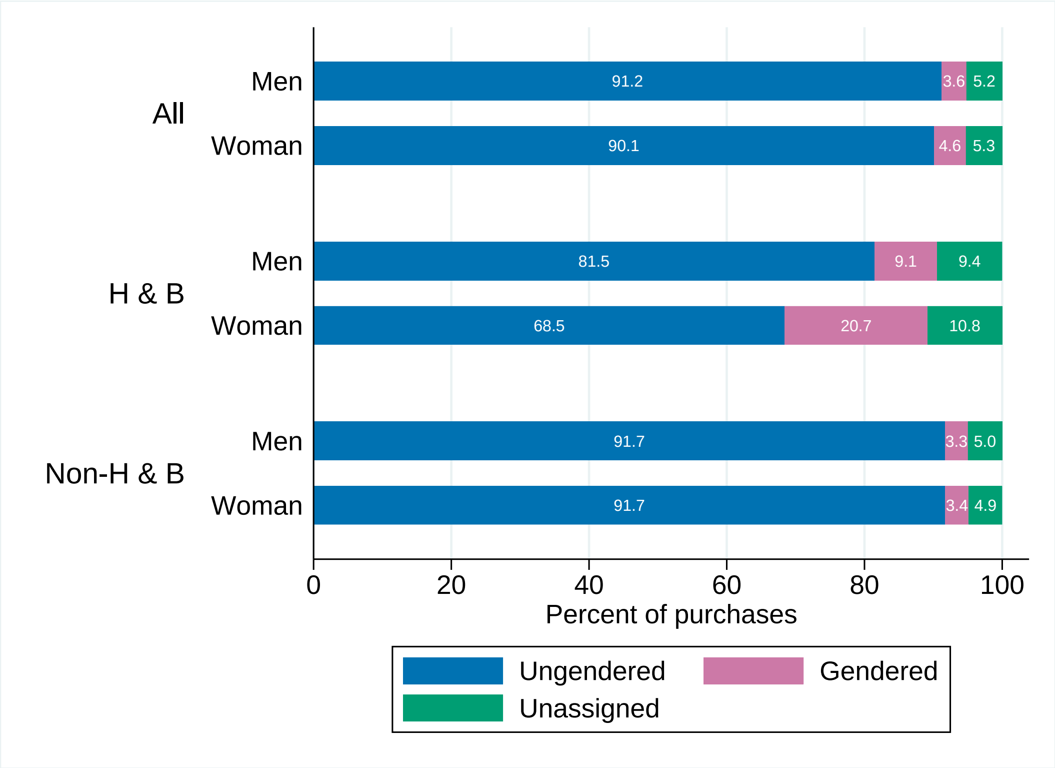
	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0799***	0.0804***	0.0694***	0.0880***	0.0921***	0.0921***
	(0.0079)	(0.0076)	(0.0076)	(0.0075)	(0.0075)	(0.0075)
Men's Average (log)	6.602	6.602	6.602	6.602	6.602	6.602
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Income FE	No	No	Yes	Yes	Yes	Yes
Age FE	No	No	No	Yes	Yes	Yes
Race FE	No	No	No	No	Yes	Yes
Education FE	No	No	No	No	No	Yes
Adj. R-squared	0.01	0.09	0.10	0.12	0.13	0.13
N	232,205	232,061	232,061	232,060	232,060	232,060
Number of clusters	49,256	49,142	49,142	49,141	49,141	49,141

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.

Figure A.1: Consumption Basket Composition by Product Gender



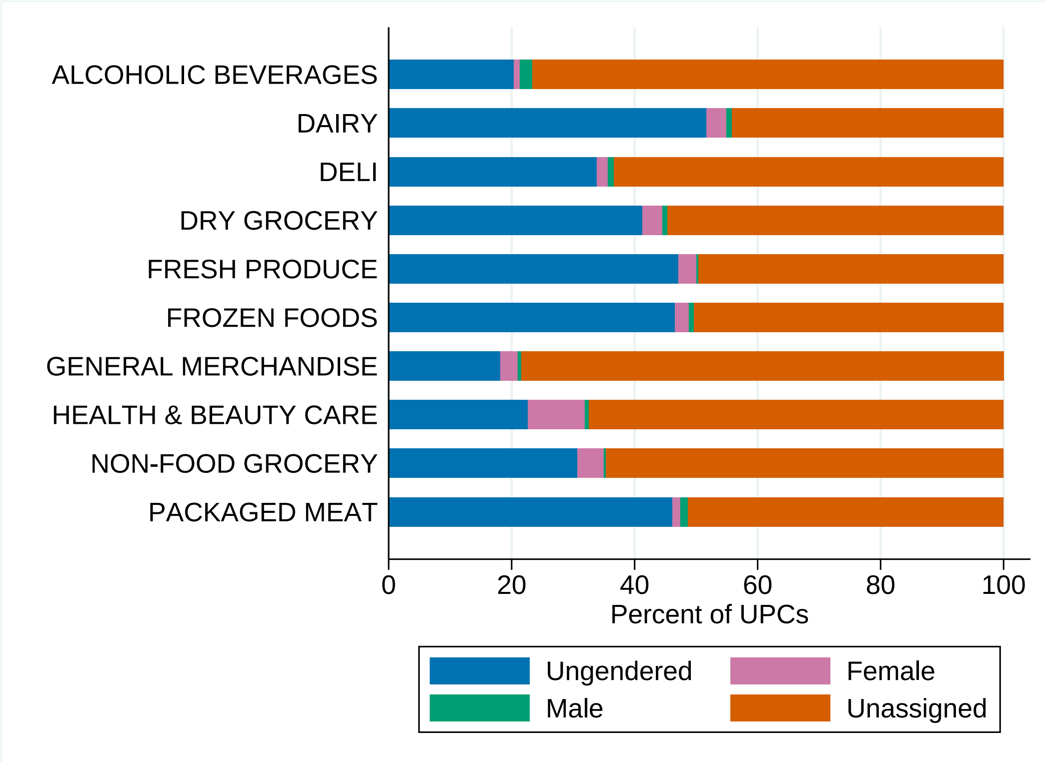
Note: This figure presents plots the decomposition of purchases made by men and women into gendered, ungendered and unassigned products. The first rows show this for all product departments while the next two separate out health and beauty products.

Table A.6: Examples of Popular Brands by Product Gender

Ungendered	**Woman-gendered**	**Man-gendered**
	Deodorant	
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
	Granola and Protein Bars	
Clif	Think Thin! Pink	Oh Yeah! Victory
Quaker Chewy	Luna	MLO Xtreme
Power Bar	Kashi Go Lean Roll!	Musclepharm Combat
	Disposable Razors	
Personna	Venus Embrace	Gillette Good News
Bic Comfort Twin	Bic Simply Soleil	Gillette Sensor 5
Super Max	Schick Quattro for Women	Bic Metal

Note: this table illustrates the validity of our measure of UPC gender. Within the product module of deodorant, ganola and protein bars, and disposable razors, we list popular gendered and ungendered brands which align with our categorization. Woman-gendered is defined as having a purchase-weighted woman purchase share above 90%. Male-gendered is defined as having a purchase-weighted woman purchase share lower than 10%. Ungendered products are those that lie within the 10% to 90% range.

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.

Table A.7: Yearly differences in number of unique products by consumer gender

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.2639***	0.2616***	0.2615***	0.2715***	0.2721***	0.2721***
	(0.0075)	(0.0072)	(0.0073)	(0.0073)	(0.0073)	(0.0073)
Men's Average (log)	5.606	5.606	5.606	5.606	5.606	5.606
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	0.07	0.16	0.16	0.16	0.17	0.17
N	232,205	232,061	232,061	232,060	232,060	232,060
Number of clusters	49,256	49,142	49,142	49,141	49,141	49,141

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.8: Differences in prices paid between men and women for products with identifiable UPC gender

	(1)	(2)
	Price per Unit	Same UPC
Female	0.0487***	-0.0139***
	(0.0026)	(0.0039)
Men's Average	$1.24	$1.74
ModuleXUnits FE	Yes	No
UPC FE	No	Yes
Month FE	Yes	Yes
County FE	Yes	Yes
Retailer FE	Yes	Yes
Demographic FE	Yes	Yes
Adj. R-squared	0.92	0.97
N	92,291,891	42,702,383
Number of clusters	49,247	49,194

Individual level clustered standard errors in parentheses

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

Note: The table replicates the results from Table 1 Column (6) of Panels (a) and (b). all products for which we can identify UPC gender using our technique developed in Section 3.2. This table presents estimates from the regression $log(P_{ijt}) = \phi_{t(j)} + \beta Woman_i + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $Woman_i$ indicates whether individual i 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: $log(P_{ijt}) = \phi_{jt} + \beta Woman_i + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $Woman_i$ 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.

Table A.9: Differences in log unit prices paid within module
Including multi-pack controls

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0234***	0.0473***	0.0515***	0.0469***	0.0383***	0.0503***
	(0.0034)	(0.0033)	(0.0028)	(0.0022)	(0.0027)	(0.0026)
Log multipack count	-0.0892***	-0.0920***	-0.0899***	-0.0598***	-0.0481***	-0.0487***
	(0.0018)	(0.0018)	(0.0014)	(0.0015)	(0.0021)	(0.0021)
Men's Average	$1.33	$1.33	$1.29	$1.23	$1.28	$1.28
Module × Units FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Month X Year 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
Adj. R-squared	0.83	0.83	0.87	0.91	0.92	0.92
N	153,333,383	153,333,383	150,043,380	137,137,029	97,715,726	97,715,726
Number of clusters	49,256	49,256	49,256	49,252	49,248	49,248

Individual level clustered standard errors in parentheses

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

Note: This table presents estimates from the regression: $log(P_{ijkt}) = \phi_{t(j)} + \beta Woman_i + \delta \log Multi_k + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $Woman_i$ indicates whether individual i is a woman, $Multi_k$ counts the number of units included (where non-multipacks are assigned a value of log(1)), 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 clusted on the household-level

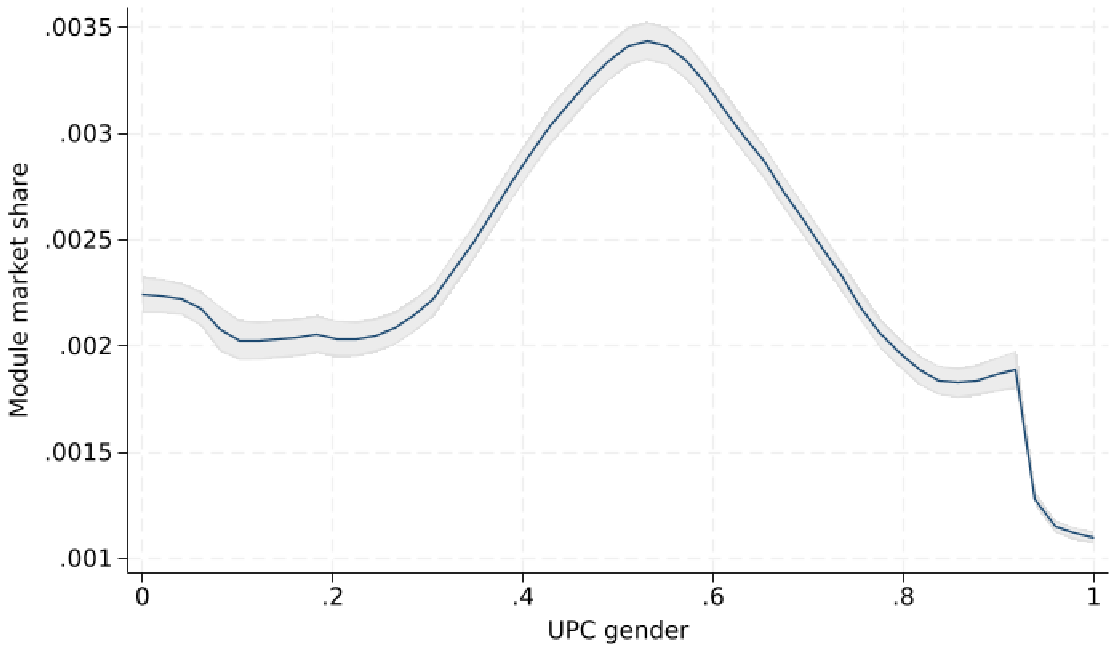
Table A.10: Prices paid across departments

	(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	0.0764***	0.0649***	0.0494***	0.0494***	0.0311***	0.0534***	0.0168***	0.0414***	0.0302***	-0.0094
	(0.0073)	(0.0029)	(0.0041)	(0.0033)	(0.0087)	(0.0052)	(0.0055)	(0.0044)	(0.0103)	(0.0171)
Men's Average	\$3.86	\$0.28	\$1.12	\$0.36	\$3.56	\$0.71	\$1.57	\$0.96	\$1.93	\$12.99
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	0.88	0.87	0.92	0.94	0.90	0.85	0.78	0.91	0.99	0.86
N	4,443,135	44,889,271	10,084,043	12,196,328	4402975	2,742,710	8,416,509	5,688,481	1,299,644	3,546,033
Number of clusters	47,790	49,221	48,443	48,765	45,445	45,626	45,803	48,212	26,642	46,953
Panel B: Per unit price for same UPC										
Women	-0.0201	-0.0016	-0.0101***	0.0009	-0.0350***	-0.0156***	-0.0229***	-0.0110***	0.0053*	0.0300
	(0.0140)	(0.0016)	(0.0029)	(0.0015)	(0.0103)	(0.0038)	(0.0044)	(0.0021)	(0.0030)	(0.0207)
Men's Average	\$3.95	\$0.28	\$1.21	\$0.38	\$3.79	\$0.74	\$1.45	\$0.76	\$2.36	\$15.18
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	0.95	0.97	0.97	0.98	0.92	0.95	0.90	0.99	1.00	0.94
N	4,253,347	40,792,894	9,247,164	12,428,193	4,412,234	2,623,437	8,947,522	5,968,132	1,261,528	2,657,264
Number of clusters	46,854	49,208	48,089	48,718	44,770	44,809	46,444	47,919	23,278	44,421

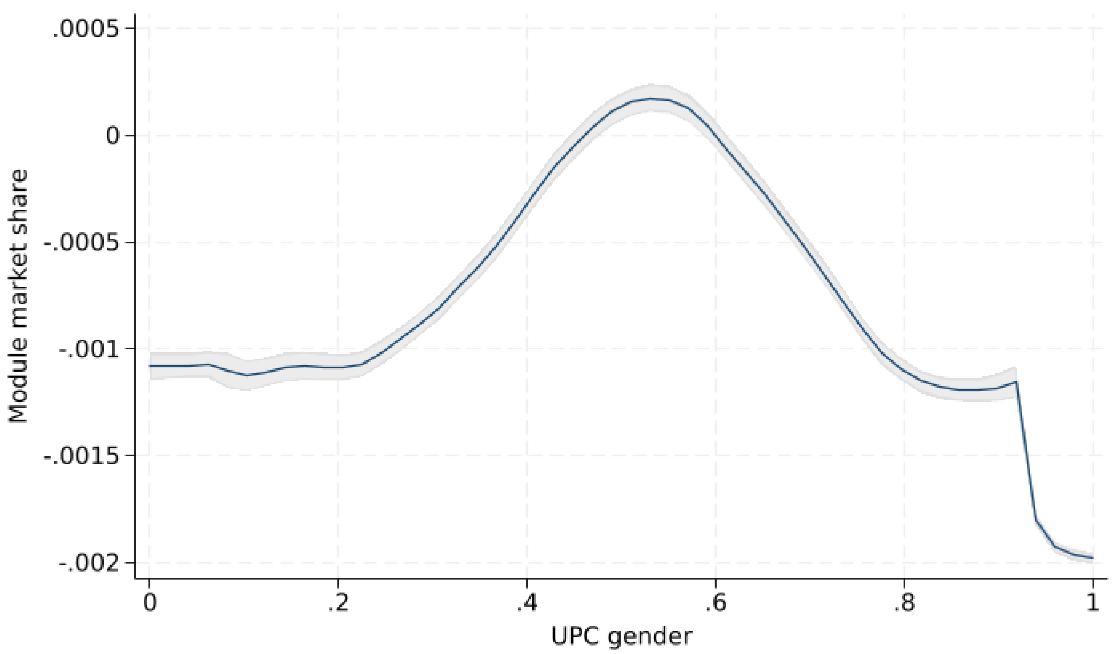
This table estimates $log(P_{ijt}) = \phi_t + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$, stratifying by department across columns. P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{Woman_i = 1\}$ is an indicator for whether the individual is a woman and and X_i is a vector of demographic 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.

Figure A.3: Market shares by UPC gender

(a) Raw average market share



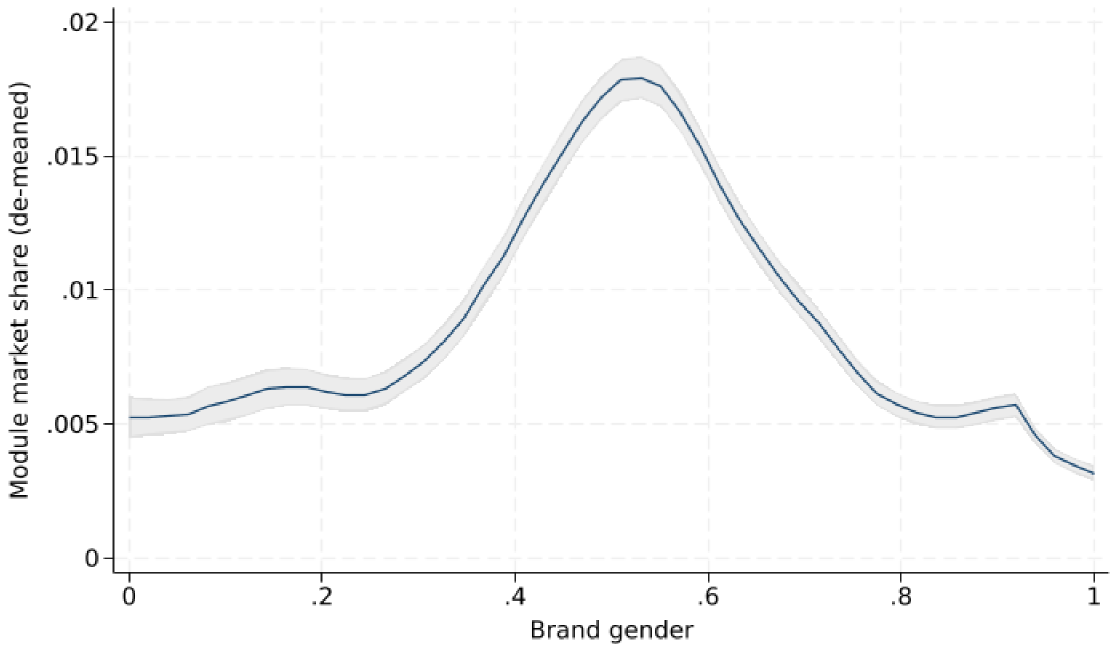
(b) Market shares residualized against module average and half-year



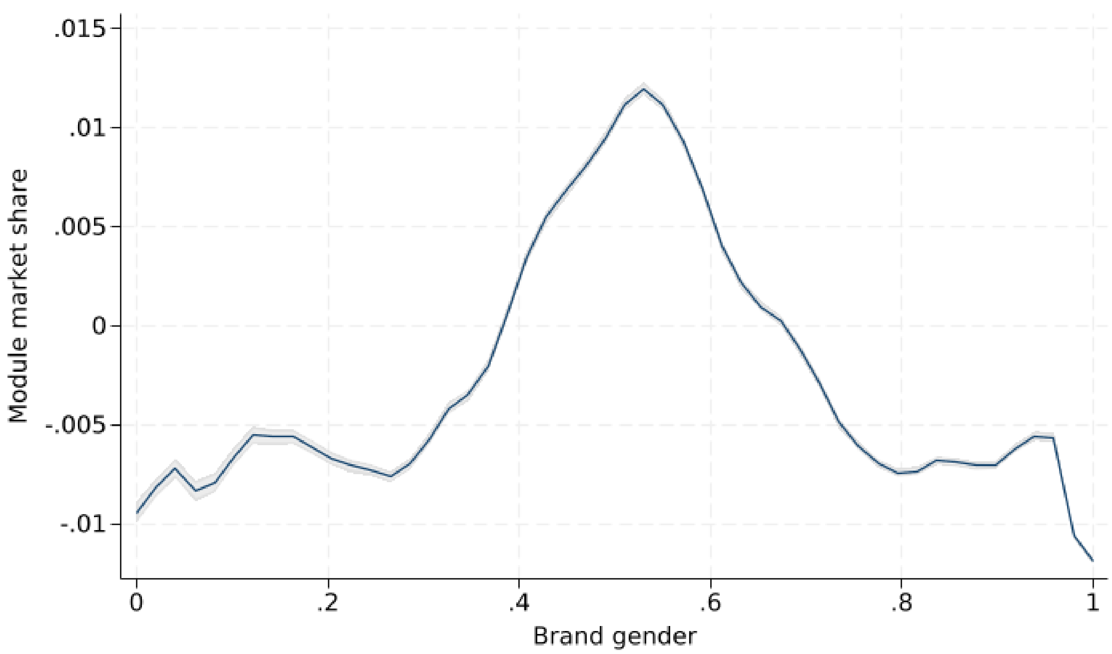
Note: these graphs display the average national market shares of UPCss within modules based on their observed brand gender. Panel (a) shows raw average market shares; Panel (b) plots average market shares residualized within each half-year and module. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

Figure A.4: Market shares by brand gender
Weighted by log module expense

(a) Raw average market share



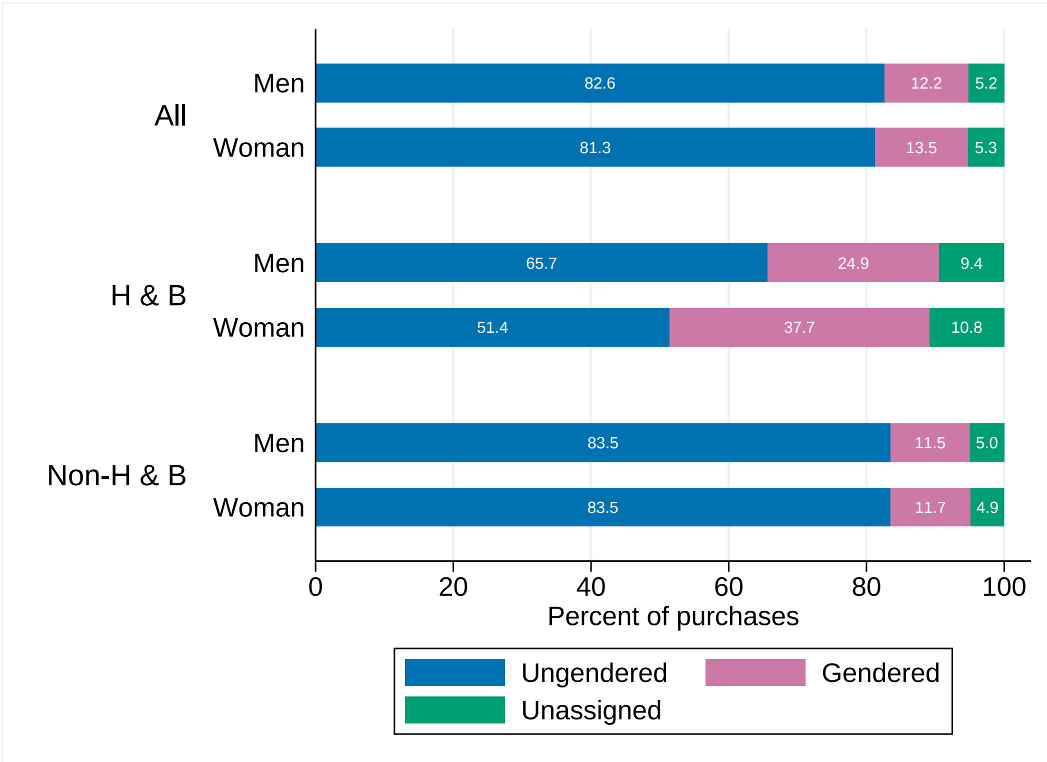
(b) Market shares residualized against module average and half-year



Note: these graphs display the average national market shares of brands within modules based on their observed brand gender. Panel (a) shows average market shares weighted by log aggregate module expense; Panel (b) plots average market shares residualized within each half-year and module weighted by log aggregate module expense. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

A.1.1 Results on UPC gender using less-restrictive cutoff

Figure A.5: 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.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	0.0408***	0.0407***	0.0408***
	(0.0026)	(0.0083)	(0.0026)
Gendered UPC	-0.0186***	0.0455***	-0.0212***
	(0.0029)	(0.0090)	(0.0031)
Female × Gendered UPC	0.1177***	0.1402***	0.1144***
	(0.0034)	(0.0104)	(0.0036)
Men's Ungendered Average	$1.24	$4.01	$1.15
MURLM FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes
Adj. R-squared	0.92	0.88	0.92
N	92,291,891	3,867,947	88,423,944
Number of clusters	49,247	47,492	49,243

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 evidence from the Consumer Expenditure Survey Public Use Microdata

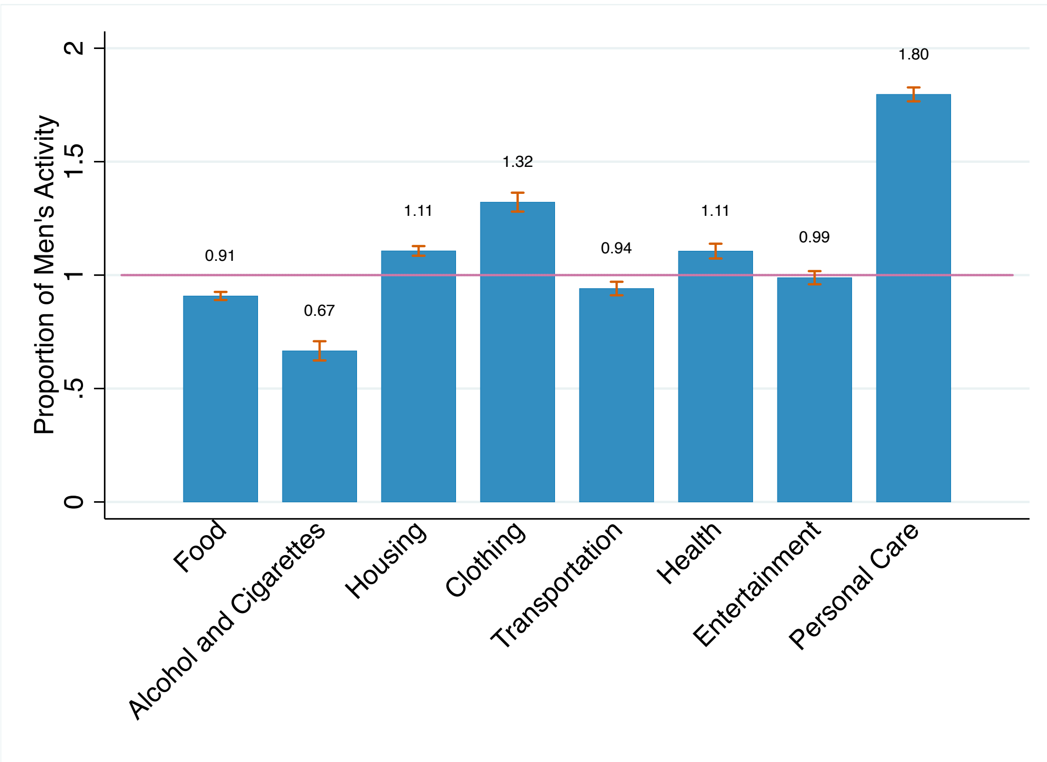
The Consumer Expenditure Survey Public Use Microdata (CE PUMD) is publicly available from the Bureau of Labor Statistics and provides information on a household's expenditures and income. The CE PUMD is comprised of a quarterly interview survey of 6,000 households that tracks overall spending and large purchases and a diary survey of 3,000 households that tracks all purchases over a two week period. We utilize only the quarterly interview surveys to inform aggregate consumption basket price and composition differences. Similar to the Nielsen HMS data, we restrict our analysis to individuals that live alone which allows us to attribute spending to one gender. We use data from years 2010 to 2017 which comprise 67,950 person-quarter observations. We present summary statistics in Table A.12. Similarly to our HMS single household panel, our CE PUMD single household panel shows that women tend to be older and poorer than the men in the sample, but otherwise are roughly similar demographically. The CE PUMD interview survey contains quarterly spending information for several categories; we focus on the eight categories that comprise the vast majority of spending: food, housing, clothing, transportation, health, entertainment, personal care, and alcohol and cigarettes. Each category aggregates all of the spending made by the individual in the quarter before their interview. Thus the food category contains all spending related to food: groceries, restaurants, convenience stores, etc. The housing category includes both rental and mortgage spending, health includes health insurance, payments to health care providers and prescriptions, and personal care includes hygiene, well being and beauty spending.

Table A.12: Demographics of CE PUMD single-member households

_	Total	Women	Men	Difference
Income	30,530	26,950	34,665	-7,715.418**
	(42,896.3)	(36,923.05)	(48,568.25)	(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	67,950	36,417	31,533	4,884

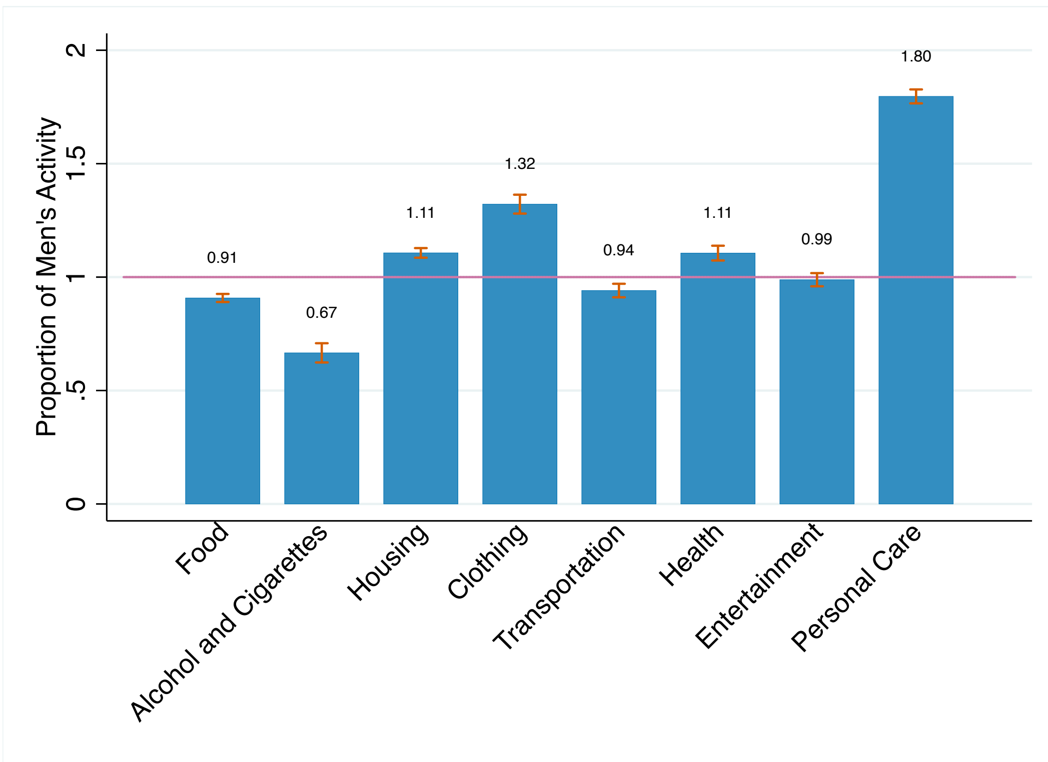
This table displays demographic data of men and women constituting single-member households as well as their differences. Dollar amounts are expressed in USD 2016.
$*p < .05$, $**\,p < .01$

Figure A.6: 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. $\mathbb{1}\{Woman_i = 1\}$ 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, age, race and education. Standard errors are clustered at the individual-level.

Figure A.7: 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. $\mathbb{1}\{Woman_i = 1\}$ 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, age, race and education. Standard errors are clustered at the individual-level.

Appendix B Additional CES model and results

B.1 Additional CES model setup

Equation (8) 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, $\sigma_n(g)$, and market share, $s_j(g)$. Solving the first order condition for Equation (6) yields:

$$q_j(g) = \left(P_n(g)\frac{\varphi_j(g)}{p_j}\right)^{\sigma_n(g)-1} \frac{\alpha_n(g)E(g)}{p_j}$$

Where $P_n(g)$ is a price index, $P_n(g) = \left[\sum_{j\in G_n} p_j^{(1-\sigma_n(g))}\varphi_j(g)^{(\sigma_n(g)-1)}\right]^{\frac{1}{1-\sigma_t(g)}}$ (for which we allow p_j to vary by transaction) and $E(g)$ represents the total expenditure of gender g.

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

$$\mu_j = \frac{p_j - c_j}{p_j} = \frac{\sum_g x_j(g)}{\sum_g \varepsilon_j(g)x_j(g)}.$$

Where $x_j(g)$ is the sales of product j to gender g. 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.

B.2 Additional CES results

Table B.1: CES First stage and reduced form results of price change instruments

Panel A: First stage

	(1)	(2)	(3)	(4)	(5)	(6)
Hausman IV	0.130***	0.139***			0.112***	0.113***
	(0.00261)	(0.00219)			(0.00253)	(0.00203)
Retailer IV			0.233***	0.237***	0.252***	0.244***
			(0.00350)	(0.00425)	(0.00506)	(0.00617)
MGDT FE	Yes	No	Yes	No	Yes	No
MGDRT FE	No	Yes	No	Yes	No	Yes
Observations	9,578,130	5,429,742	13,398,189	7,928,011	9,578,130	5,429,742
F-Statistic	2,469	4,019	4,452	3,108	3,173	2,560
Adjusted R^2	0.272	0.473	0.24	0.447	0.274	0.476
Number of clusters	76,3079	573,630	1.02E+06	756,534	763,079	573,630

Panel B: Reduced form

,

	(1)	(2)	(3)	(4)	(5)	(6)
Hausman IV	-0.0326***	-0.0782***			-0.0304***	-0.0740***
	(0.00451)	(0.0100)			(0.00455)	(0.0103)
Female×Hausman IV	-0.0242***	-0.0437***			-0.0190***	-0.0340***
	(0.00565)	(0.0120)			(0.00573)	(0.0124)
Retailer IV			-0.0328**	-0.0555***	-0.0337*	-0.0436
			(0.0128)	(0.0203)	(0.0176)	(0.0290)
Female×Retailer IV			-0.0692***	-0.102***	-0.0636***	-0.0841**
			(0.0152)	(0.0238)	(0.0211)	(0.0338)
MGDT FE	Yes	No	Yes	No	Yes	No
MGDRT FE	No	Yes	No	Yes	No	Yes
Observations	9,578,130	5,429,742	13,398,189	7,928,011	9,578,130	5,429,742
Adjusted R^2	0.182	0.314	0.153	0.299	0.182	0.314
Number of clusters	763,079	573,630	1,015,000	756,534	763,079	573,630
F-Statistic	120.1	136.4	55.09	59.06	67.53	74.58

Brand-DMA level clustered standard errors in parentheses

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

This table presents the first stage and reduced form results of the instrumental variables design using Hausman and retailer leave-out instruments to substitution elasticities of men and women. The first stage displayed in Panel A regresses within-market log price changes between half years on within-market changes in the values of the instruments between half years. "MGDT FE" refers to module×gender×DMA×half-year fixed effects. "MGDRT" refers to module×gender×DMA×retailer×half-year.

Table B.2: Elasticities of Substitution by Department
No retailer fixed effects

	(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.
$\sigma_m - \sigma_w$	0.233	-0.141**	-0.446***	-0.185**	0.0148	-0.401*	-0.187*	0.0484	-0.0823	-0.682*
	(0.203)	(0.0715)	(0.168)	(0.0771)	(0.248)	(0.231)	(0.0953)	(0.162)	(0.67)	(0.389)
$1 - \sigma_m$	-0.11	-0.303***	-0.497***	-0.0766	-0.575***	-0.429**	-0.293***	-0.0982	0.565	0.471
	(0.127)	(0.0592)	(0.141)	(0.0649)	(0.218)	(0.207)	(0.0934)	(0.145)	(0.566)	(0.382)
% Elasticity dif. (women v. men)	-20.99	10.82	29.79	17.18	-0.94	28.06	14.46	-4.41	18.92	128.92
Observations	633,365	4,626,095	964,377	947,026	255,957	317,418	574,360	961,507	60,436	236,745
Number of clusters	82,060	351,177	78,704	48,601	15,561	21,475	33,362	90,969	11,303	29,481
MGDR FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hausman DMA IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Retailer IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

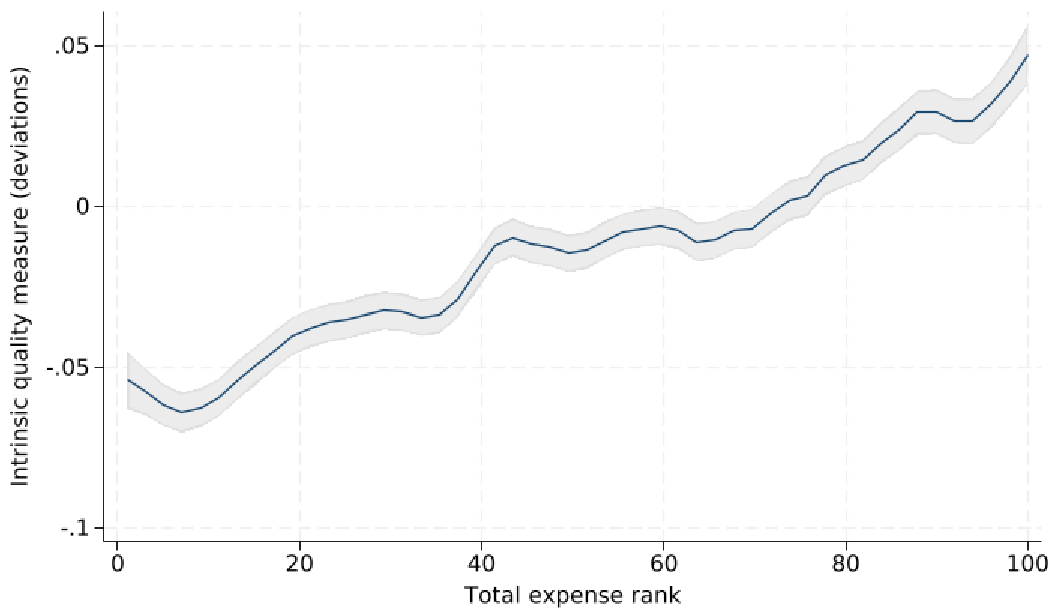
Brand-DMA level clustered standard errors in parentheses

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

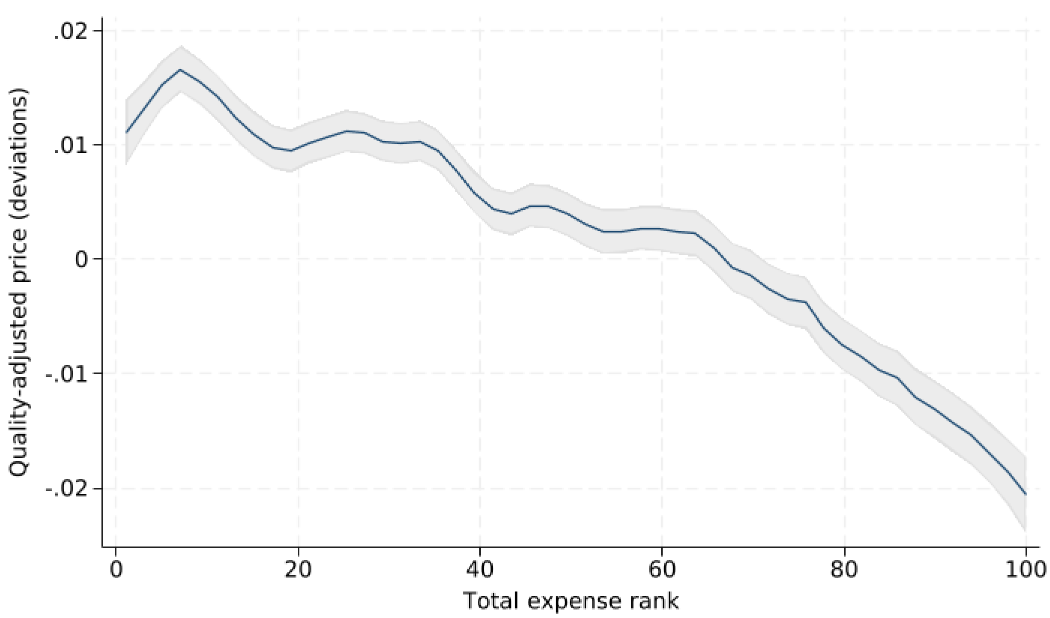
This table presents the results of estimating elasticities of substitution by regressing changes in the log budget share of a product on changes in log price for men and women controlling for the location, retail chain, and half-year corresponding to the following regression: $\Delta log(b_{gjt}) = (1-\sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}$. Each column stratifies the data by department; every column uses both Hausman and Retailer leave-out instruments in an instrumental variables regression. "MGDR FE" refers to module×gender×DMA×half-year fixed effects.

Figure B.1: Average intrinsic quality and quality-adjusted price
By total expense rank of individuals
Faber and Fally (2022) replication - validation of our approach

(a) Intrinsic quality

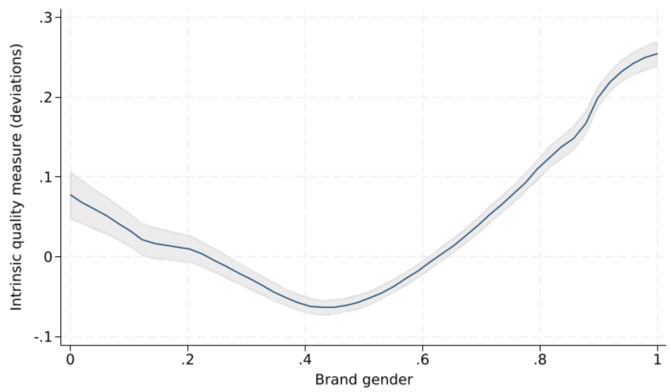


(b) Intrinsic quality-adjusted price

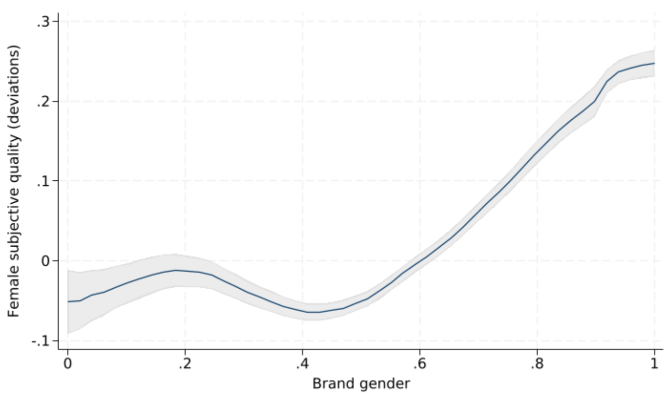


Note: These graphs replicate Figure 2 from Faber and Fally (2022) in our setting, validating the construction of our quality-measures. The graphs plot intrinsic quality (Panel (a)) and quality-adjusted price (Panel (b)) along household-expenditure ranks (weighted by Nielsen proprietary probability weights). Brand gender is constructed as the expense-weighted mean UPC-gender of UPCs contained within the brand. Quality and quality-adjusted price are measured in units percent deviation relative to the mean good in each module × half-year. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

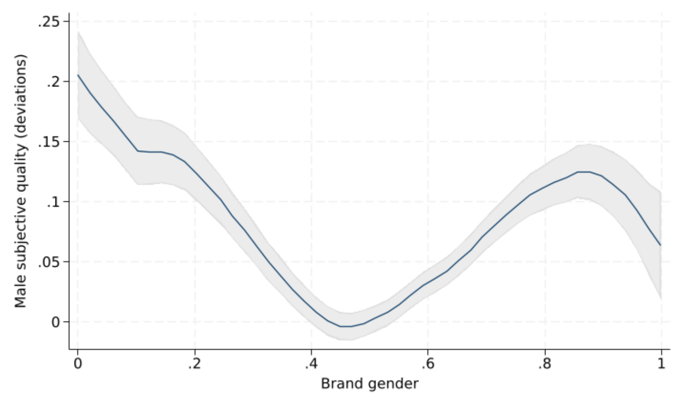
Figure B.2: Intrinsic and subjective quality by brand-gender
Robustness: Departments with elasticities of substitution bounded above one by at least one standard error



Panel (a) Intrinsic quality



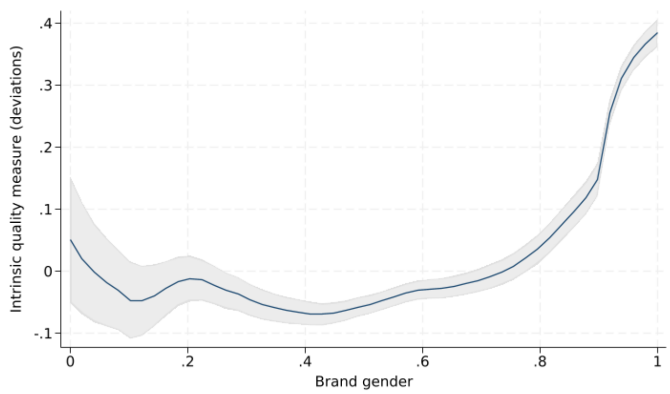
Panel (b) Female subjective quality valuations



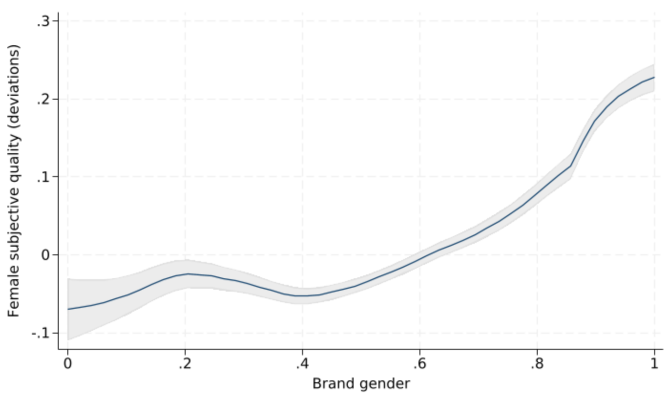
Panel (c) Male subjective quality valuations

Note: These graphs plot intrinsic quality (Panel (a)), and subjective quality valuations by women and men (Panel (b) and Panel (c), respectively) along brand-gender. Brand gender is constructed as the expense-weighted mean UPC-gender of UPCs contained within the brand. This graph replicates results from Section 4.3 however using all departments with point estimates for elasticity of substitution at least one standard error above one (see Table 4). The x-axis can be interpreted thus as the share of the consumption purchased by women (i.e. increasing in "female-ness" from left to right"). Quality is measured in units percent deviation relative to the mean good in each module $\times$ half-year. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

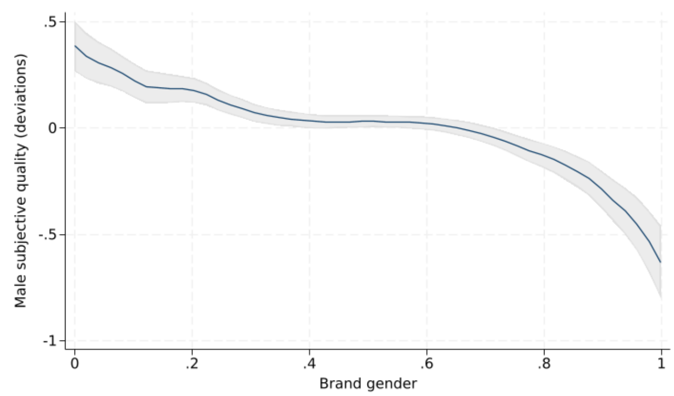
Figure B.3: Intrinsic and subjective quality by brand-gender
Robustness: Departments with elasticities of substitution greater than one



Panel (a) Intrinsic quality



Panel (b) Female subjective quality valuations



Panel (c) Male subjective quality valuations

Note: These graphs plot intrinsic quality (Panel (a)), and subjective quality valuations by women and men (Panel (b) and Panel (c), respectively) along brand-gender. Brand gender is constructed as the expense-weighted mean UPC-gender of UPCs contained within the brand. This graph replicates results from Section 4.3 however using all departments with point estimates for elasticity of substitution greater than one (see Table 4). The x-axis can be interpreted thus as the share of the consumption purchased by women (i.e. increasing in "female-ness" from left to right"). Quality is measured in units percent deviation relative to the mean good in each module × half-year. The fitted relationships correspond to local polynomial regressions, and the shading represents 95% confidence intervals.

Appendix C Additional evidence on retailer 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 (13), but this condition is less restrictive and amounts to a bounding condition: that on 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 C.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 $\mathbb{E}[\Delta\%c] > 0$.

C.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:[44]

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

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-

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

ously:

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

and

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

I.e. given our observation that 1) female prices at retail exceed male prices at retail, 2) female retail costs (i.e. distributor prices) 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, lower retailer markup and greater elasticity on part of women likely imply a lower overall markup between manufacturing to final sale to consumer, which determines overall demand.

C.2 Additional figures and tables on retailer costs and markups

Table C.1: Balance on PriceTrak merge

Panel (a): Regressing consumer gender on the merge indicator

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0068***	0.0045***	0.0044***	0.0065***	0.0053***	0.0034***
	(0.0007)	(0.0003)	(0.0003)	(0.0015)	(0.0005)	(0.0003)
Constant	0.1402***	0.1414***	0.1421***	0.1112***	0.1118***	0.1154***
	(0.0026)	(0.0008)	(0.0004)	(0.0055)	(0.0007)	(0.0003)
Weight	Nielsen	Nielsen	Nielsen	Expense	Expense	Expense
Year	No	Yes	Yes	No	Yes	Yes
Module	No	Yes	Yes	No	Yes	Yes
County	No	No	Yes	No	No	Yes
Retailer	No	No	Yes	No	No	Yes
Adj. R-squared	0.000	0.503	0.721	0.000	0.515	0.777
N	1.55e+08	1.55e+08	1.38e+08	1.53e+08	1.53e+08	1.37e+08
Number of clusters	1,869,584	1,869,373	1,661,846	1,858,083	1,857,874	1,650,113

Panel (b): Regressing the merge indicator on transaction log unit price

	(1)	(2)	(3)	(4)	(5)	(6)
Merges to PriceTrak	-0.2121***	0.1352***	0.0894***	-0.7880***	0.1054***	0.0768***
	(0.0589)	(0.0050)	(0.0037)	(0.2029)	(0.0056)	(0.0035)
Constant	-1.4507***	-1.5006***	-1.5151***	-0.6812***	-0.7838***	-0.8536***
	(0.0581)	(0.0040)	(0.0034)	(0.2027)	(0.0030)	(0.0016)
Weight	Nielsen	Nielsen	Nielsen	Expense	Expense	Expense
Year	No	Yes	Yes	No	Yes	Yes
Module	No	Yes	Yes	No	Yes	Yes
County	No	No	Yes	No	No	Yes
Retailer	No	No	Yes	No	No	Yes
Adj. R-squared	0.002	0.829	0.907	0.014	0.897	0.959
N	1.53e+08	1.53e+08	1.37e+08	1.53e+08	1.53e+08	1.37e+08
Number of clusters	1,858,045	1,857,874	1,650,113	1,858,045	1,857,874	1,650,113

UPC-clustered standard errors in parentheses

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

Note: This table displays regressions pertaining to the balance of the PriceTrak merge. Panel (a) regresses an indicator for whether a UPC-market-year from the PriceTrak data merges onto the Nielsen data onto a female-consumer indicator. Panel (b) regresses log unit transaction price on the PriceTrak merge indicator. Each column varies the fixed effect and weighting specification.

Table C.2: Log markup by purchaser gender, budgetshare-weighted

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0294**	-0.0175	-0.0182	-0.0054	0.0014	-0.0012
	(0.0142)	(0.0121)	(0.0118)	(0.0045)	(0.0044)	(0.0057)
Male mean (levels)	89.3%	89.3%	89.3%	90.0%	91.5%	93.9%
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	0.77	0.82
N	17,901,420	17,901,305	17,901,305	17,084,000	15,142,190	9,690,298
Number of clusters	28,412	28,412	28,412	28,406	28,393	28,356

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 C.3: 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	-0.2652***	-0.0875***	-0.0054	0.0176***	0.0482***	0.0070	0.0087	-0.0202	0.0293	-0.0305***
	(0.0146)	(0.0178)	(0.0096)	(0.0052)	(0.0134)	(0.0095)	(0.0192)	(0.0155)	(0.0198)	(0.0111)
Male mean (levels)	73.9%	67.3%	71.35%	77.2%	31.7%	30.7%	352.1%	358.1%	44.5%	43.6%
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	0.75	0.00	0.75	0.00	0.58	0.00	0.82	0.01	0.68
N	755,352	225,292	9,405,824	5,310,685	2,380,817	1,491,274	1,788,708	1,059,547	612,284	439,534
Number of clusters	27,572	21,102	28,395	28,257	27,794	26,300	27,933	26,398	25,155	21,780

Individual level clustered standard errors in parentheses

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

	(11) Pack. Meat	(12) Pack. Meat	(13) Produce	(14) Produce	(15) Non-food Groc.	(16) Non-food Groc.	(17) Gen. Merch.	(18) Gen. Merch.
Female	0.0203*	0.0051	0.0237	0.0239	-0.0437***	-0.0065	0.0493***	-0.0498**
	(0.0117)	(0.0106)	(0.0250)	(0.0342)	(0.0110)	(0.0117)	(0.0128)	(0.0242)
Male mean (levels)	39.1%	37.4%	272.4%	287.4%	47.1%	43.2%	102.8%	102.3%
Module	No	Yes	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes	No	Yes
County	No	Yes	No	Yes	No	Yes	No	Yes
Retailer	No	Yes	No	Yes	No	Yes	No	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R-squared	0.01	0.50	0.02	0.83	0.00	0.68	0.01	0.63
N	752,874	453,095	259,734	108,374	1,759,695	635,798 351,650	109,466	
Number of clusters	26,190	23,200	22,113	14,437	28,147	25,031	26,403	17,796

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. Alcoholic products are largely not covered in the Pricetrak Data and so are excluded from this table. All transaction-observations are given equal weight. Standard errors are clustered at the individual-level.

Table C.4: Log retailer 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)	\$0.29	\$0.29	\$0.29	\$0.28	\$0.27	\$0.26
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	18,076,261	18,076,147	18,076,147	17,258,171	15,310,301	9,834,314
Number of clusters	28,412	28,412	28,412	28,406	28,393	28,357

Panel (b): Budgetshare-weighted

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0361***	0.0274***	0.0344***	0.0461***	0.0433***	0.0471***
	(0.0081)	(0.0040)	(0.0041)	(0.0026)	(0.0022)	(0.0027)
Male mean (levels)	\$0.43	\$0.43	\$0.43	\$0.38	\$0.34	\$0.31
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	0.91	0.93
N	17,901,420	17,901,305	17,901,305	1,7084,000	15,142,190	9,690,298
Number of clusters	28,412	28,412	28,412	28,406	28,393	28,356

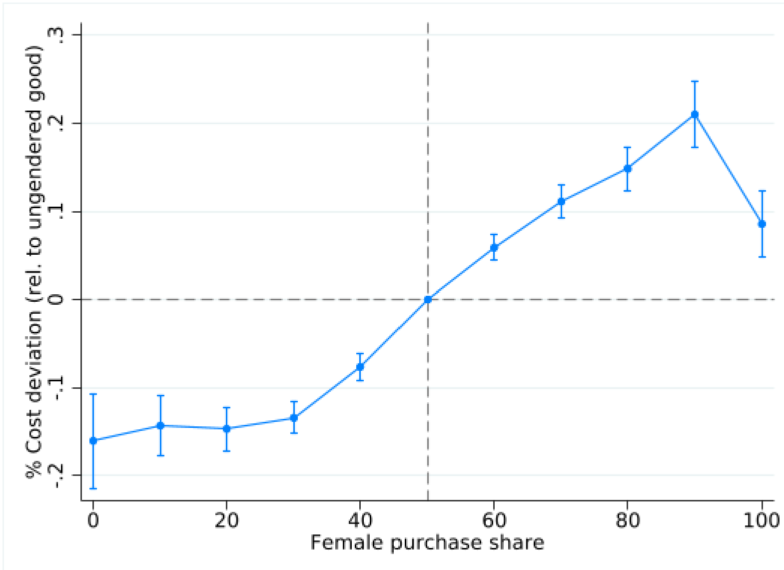
Individual-level clustered standard errors in parentheses

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

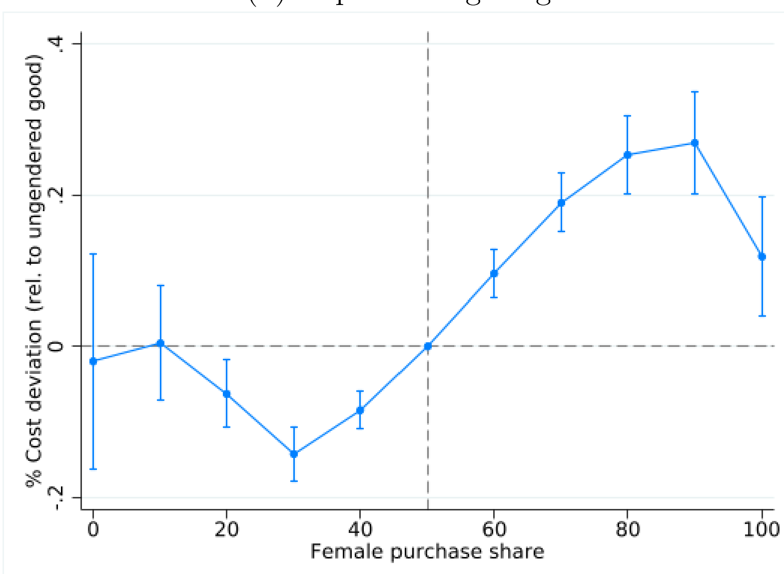
This table presents estimates from the transaction-level regression: $log(C_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \Gamma X_i + \epsilon_{ijt}$ where C_{jt} is the wholesale price of UPC j in year t as observed in PriceTrak. $\mathbb{1}\{Woman_i = 1\}$ is an indicator for whether the individual 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 C.1: Average retailer costs, within market, by female purchase intensity

(a) Equal weighting



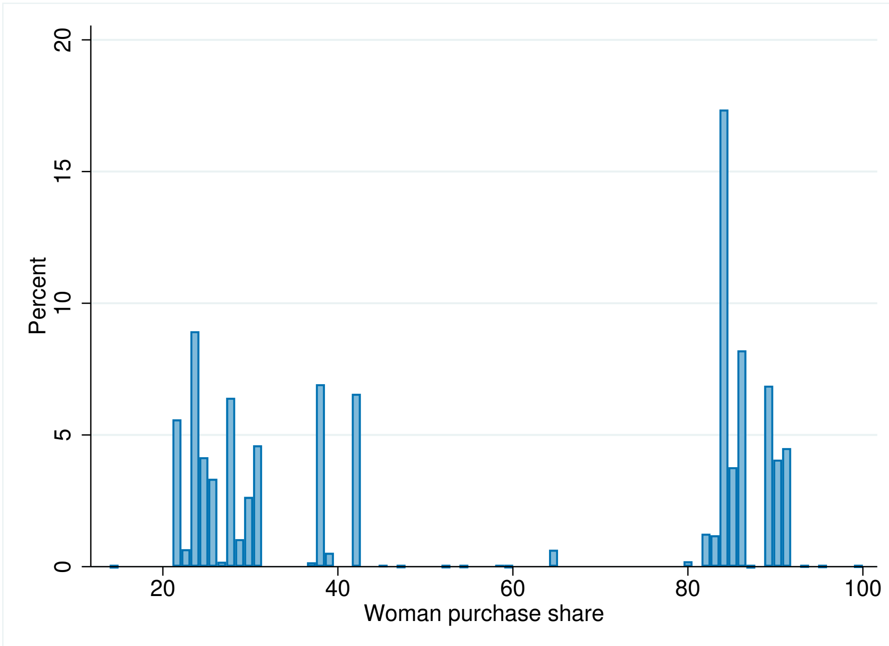
(b) Expense-weighting



Note: These figures display the coefficients estimated from the following regression on the UPC-year level: $y_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Bins $b \in \mathcal{B}$ 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 uses log markup as the dependent variable; Panel (b) uses directly observed log retailer cost (distributor price) as the dependent variable. Both specifications feature equal weighting for all observations. Standard errors are clustered at the UPC level.

Appendix D Additional differentiated products demand results

Figure D.1: Distribution of brand-gender for disposable razors



Note: This figure presents the distribution of brand gender, or woman purchase share, $\hat{w}_j$, for disposable razors. Brand gender is defined as the time-invariant expense-weighted average share of a product brand purchased by women.

Table D.1: Differentiated Products Model Parameter Estimates

	Disposable Razors
Linear Coefficients	
Price (α)	-2.29***
	(0.22)
Nonlinear Coefficients	
Price (Σ_p)	0.81***
	(0.08)
Outside Option (Σ_1)	0.13
	(10.27)
Woman Purchase Share (Σ_w)	7.54E-05
	(70.22)

Market level clustered standard errors in parentheses

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

This table presents estimated parameters from our differentiated products model using PyBLP. We allow for heterogeneity in taste for the outside option, price and the woman purchase share of the razor. We include product, store and week fixed effects.

D.1 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.2. 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 pricing and production of products.

Appendix E 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 E.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 E.1.

Table D.2: 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.

Table E.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.

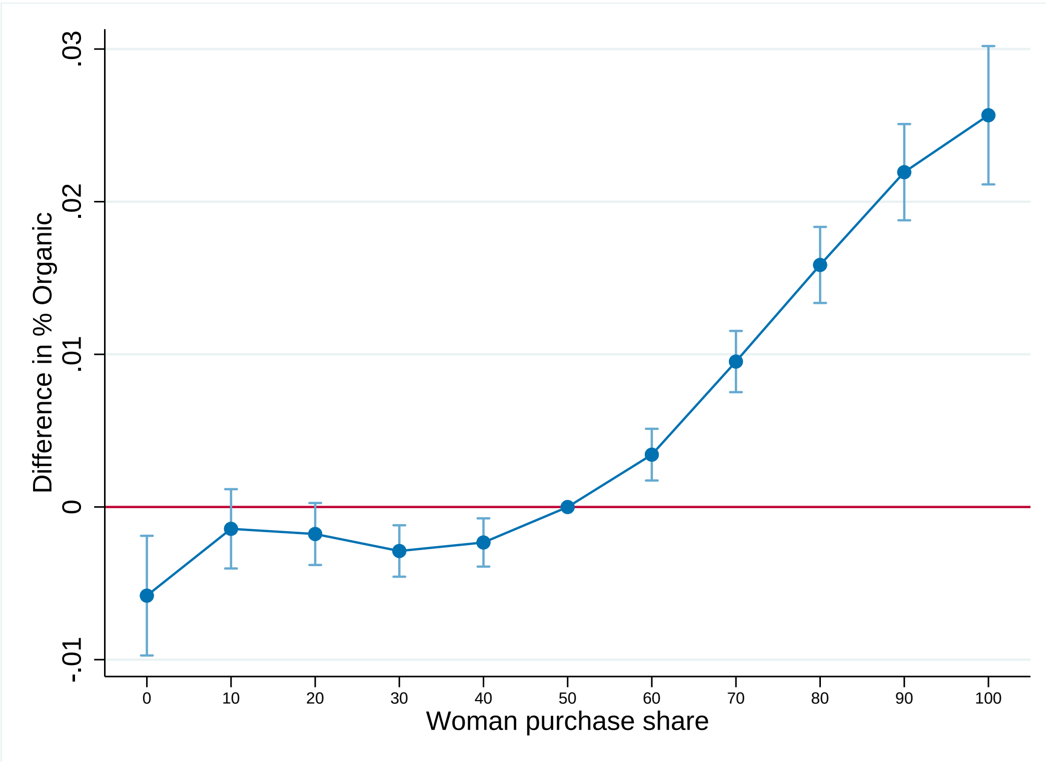
Figure E.1 plots the coefficients from a regression of organic status of a product on decile of woman purchase share normalized to products that are bought equally by men and women. The graph shows that products primarily bought by men are slightly less likely to be organic while products more often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

Table E.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.

Figure E.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.