The Pink Tax: Why Do Women Pay More?[*]

Please click here for the most recent version.

Kayleigh Barnes[†] (Job Market Paper)   Jakob Brounstein[‡]

October 31, 2022

Abstract

We study the question of whether women, on average, pay a price premium — a so-called "pink tax" — for the products they buy. A particular concern facing policy makers is whether such differences are a form of gender based price discrimination. Using scanner data, we find that averaged across the entire retail grocery consumption basket, women pay 4% more per unit for goods in the same product-by-location market as do men. This price differential is generated by a 15% higher average per unit price paid by women on explicitly gendered products, like personal care items, as well as a 3.8% higher average per unit price paid by women on ungendered products, like packaged food items. Higher prices paid by women could be the result of differences in demand elasticity, competitive structure, or sorting into goods with differing marginal costs. To disentangle these mechanisms, we estimate demand differences between men and women and structurally decompose price differences into markups and marginal costs. We find that women are, on average, more price elastic consumers than men, suggesting that as a consumer base women are not likely to be charged higher markups under price discrimination. Overall, we find that the pink tax is not sustained by higher markups charged to women, but by women sorting into goods with higher marginal costs and lower markups.

[*]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, Leticia Juarez, Kei Kawai, Sarah Moshary, Mathilde Muñoz, Nina Rousille, and Damian Vergara, as well as participants in the UC Berkeley IO and Trade lunch seminars and University of Washington PHEnOM seminar for helpful comments. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. While the categories of "men" and "women" do not cover the entirety of the gender spectrum, our work focuses on these two categories. Unfortunately, data indicating non-binary and alternate gender identities is scarce, and for the purposes of studying the Pink Tax, producers and retailers have historically marketed goods and services toward either men or women.

[†]University of California, Berkeley email: kayleighnb@berkeley.edu

[‡]University of California, Berkeley email: jakob.brounstein@berkeley.edu

1 Introduction

Is it more expensive to be a woman? Economic and societal forces have shaped preferences and product offerings to create disparities in the way men and women consume goods. 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 New York and California. This public discourse on the pink tax often attributes it to gender based price discrimination, where goods that are marketed to women have higher markups resulting from less elastic demand or less competitive markets. Existing studies of the pink tax find mixed evidence of its scope and magnitude and either focus on a narrow set of goods or do not delve into its underlying economic mechanisms (Moshary, Tuchman, and Bhatia 2021; Guittar et al. 2022; NYCDCA 2015; Duesterhaus et al. 2011; Manzano-Antón, Martinez-Navarro, and Gavilan-Bouzas 2018). Moshary, Tuchman, and Bhatia (2021) evaluate the existence of the pink tax for personal care items and find no evidence of higher markups on women's products when controlling for proxies of marginal costs. Controlling for marginal costs restricts comparisons to between goods with the same inputs and tests for third degree price discrimination. However, this type of comparison abstracts away from men's and women's purchase decisions and does not capture the role of differential sorting by men and women.

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, and then decomposing observed price differences into markups and marginal costs. Our paper considers a broad definition of the pink tax[1], considering any channel through which women may face higher markups in the retail consumer packaged goods (CPG) space. This definition allows us to capture the role of differential sorting between men and women and second degree price discrimination, or versioning, in generating the pink tax. We find that, averaged across the entire grocery consumption basket, women pay 4% higher per unit prices than men do for products in the same product-by-location market. We find that this price difference is sustained not just by purchases of gendered products, like men's and women's razors, but also by differences in purchasing habits between men and women for food and household items. This finding could be driven by three economic mechanisms that determine pricing: (i) women could have less elastic demand than men, (ii) women could consume products with more market power or from less competitive markets than men, or (iii) women could consume products with higher marginal costs. Pricing disparities due to markups based on demand differences or competitive structure impact consumer surplus directly, potentially driving welfare differences by gender. Price differences based on underlying production costs across the consumption baskets, on the other hand, do not reduce consumer surplus and are not perceived as an issue for "fairness" (Kahneman, Knetsch, and Thaler 1986). Disentangling the mechanisms driving the observed price premium on women's products is, thus, important to inform economic understanding and policy alternatives.

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

[1]We identify three scenarios through which the pink tax may operate: 1) different prices for goods with the 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.

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 price and composition. We find that women spend about 6% more than men do on retail CPG consumption and that their consumption bundles are larger and more diverse. The products that women purchase are on average 4% more expensive per unit than those purchased by men in the same product-by-location market. In the product space, we document a significant share of products that are exclusively bought by one gender, with the majority of these products gendered towards women. These products are particularly common in markets with explicit gender differentiation in marketing and product design, such as in beauty and personal care goods. We categorize products bought at least 90% of the time by one gender as "gendered" products, categorizing all other products as "ungendered". We then decompose the average 4% price premium paid by women into a contribution from differential sorting into ungendered products and from purchases of explicitly gendered products, finding that women pay an average of 3.9% higher prices on ungendered products relative to men and that women pay an average of 15% higher prices on gendered products relative to men. While gendered items have large price premiums, they make up a small share of actual purchases; the bulk of the price premium is being 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. Less elastic own-price elasticities put upward pressure on prices as firms can raise prices without losing much of their consumer base. Higher prices paid by women could then be consistent with women being less elastic and firms price discriminating off of the gender composition of their consumer base. Alternatively, differences in competition and market structure can also contribute to higher markups if women's markets are more concentrated, meaning that their products have higher market shares, or if women's products are more likely to be owned by multi-product firms, as substitution to products with the same parent company puts upward pressure on prices. Both the demand elasticity and competition narratives would contribute to higher prices through women paying higher markups, which has potential welfare effects for women. Finally, women could face higher prices if the products that they prefer have higher marginal costs than the products that men prefer, that is, if women differentially sort into products with higher costs of production.

To assess these possibilities, we model demand and supply, attributing differences in pricing and product choice to markups and marginal costs. We begin by estimating demand elasticity differences between men and women across the entire retail grocery consumption basket. We develop a simple, tractable model assuming constant elasticity of substitution that allows us to estimate demand by gender in the aggregate population. We aggregate individual-level purchase data to the a gender-by-product module-by-location market level and we find that, on average, women consume products more elastically than do men. This finding is consistent with women being the consumer group that is charged lower markups rather than higher markups under price discrimination.

We have shown that, on average, women are more elastic consumers than men, but in order to better understand product specific demand and decompose prices into markups and marginal costs we estimate a differentiated products demand and supply model. This model incorporates market structure and allows for flexible substitution patterns based on how "gendered" a product is. We focus on five markets: yogurt, protein bars, disposable razors, deodorant and shampoo. We selected yogurt because its pricing patterns mirror the descriptive analysis of the full consumption basket, it is representative of the most commonly bought item in the data, a packaged food item, and it has a moderate amount of differential sorting by gender. The other four markets were selected because gender is an explicit component of marketing and product design, allowing for identification of demand for explicitly gendered products which might not be well captured in the CES model.

To estimate the model, we use store-level data on quantities and prices. With this data, we gain improved inference on markets where purchase frequencies make individual-level data sparse, like personal care items. We allow for heterogeneity in preferences for the gender of a product and instrument for prices with Hausman instruments and retail chain-level leave-out mean prices following evidence from DellaVigna and Gentzkow (2019). To map our results back to consumer demographics, we analyze how our results vary with the product's woman purchase share observed in the individual-level purchase data. We find that women's products are either more elastic or have no significant differences in elasticities from ungendered or men's products and that women's products have lower markups and higher marginal costs. These results, while allowing for more granular identification and flexible demand, are largely consistent with the CES demand analysis. Overall, our findings imply that observed price differences between men and women are primarily driven by women sorting into higher cost products.

Our findings suggest that the pink tax is not a form of systemic price discrimination against women but that, if anything, women pay lower markups on average than men. Current legislation is largely focused on banning price differences for products that differ only in gender. Our paper suggests that these laws are likely to be ineffective at addressing price disparities between men and women, as the majority of our observed pink tax can be explained by men and women sorting into products that differ by more than just gender.[2] Our findings have important implications for other policy relevant issues, like potential disparities in the incidence of inflation between men and women. Finally, our findings motivate future research to study how men's and women's preference differences are formed as well as the role of preference differences in generating product differentiation through product entry and exit.

We proceed as follows: Section 2 discusses the background and history of the pink tax as well as relevant literature. Section 3 describes our data. Section 4 presents our descriptive analysis. Section 5 describes and estimates a constant elasticity of substitution demand model of men and women's consumption. Section 6 describes and estimates a differentiated

[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.

products demand model. Section 7 discusses the implications of our results and concludes.

2 Background and Literature Review

The term "The Pink Tax" was first coined in the 1990's in California, when concerns about gendered price discrimination of services, such as dry cleaning and in hair salons led to the explicit anti-price discrimination law, The Gender Tax Repeal Act. Soon after similar legislation was passed in New York City and Miami. A national version of The Gender Tax Repeal Act has been introduced federally several times since 2016 but has never been passed. More recently, there has been renewed policy interest in the Pink Tax, particularly in the setting of gendered price discrimination for consumer retail products. In 2020, the state of New York passed legislation that would outlaw gender differential pricing. In 2022, California passed a similar law. The language surrounding these laws frames the Pink Tax as a price discrimination story with the underlying assumption that markups are higher for women's products.

However, in spite of its importance as a potential component of gender inequality and its wide presence in popular discussion, there are few studies that rigorously substantiate the Pink Tax. The New York law was based on evidence collected and presented in a New York Department of Consumer Affairs (NYC DCA) study in 2015. The NYC DCA compares products in thirty-five categories and five broader industries with "clear male and female versions" sold by New York City retailers, finding that women's products cost on average seven percent more than similar products for men. While it provides key preliminary suggestive evidence of a "pink tax", the NYC DCA analysis is largely incomplete: it consists of a highly limited number of goods that were gender-matched in a subjective manner; moreover, it only documents raw list price differences rather than actual prices paid. Recently, Moshary, Tuchman, and Bhatia (2021) assess the pink tax under the definition in the New York law, products that differ only in gender. They control for brands and ingredients as a proxy for marginal costs and find no evidence of a systemic pink tax. Other works similarly focus on health and beauty products, using in store surveys of products to descriptively document price premia of around 5% on women's goods. (Duesterhaus et al. 2011; Manzano-Antón, Martinez-Navarro, and Gavilan-Bouzas 2018; Manatis-Lornell et al. 2019) Taken together, list price differences suggest that women may be paying more than men for goods with similar uses, but Moshary, Tuchman, and Bhatia (2021)'s finding of no pink tax when matching on marginal costs suggests that women and men are sorting into products that differ by more than just gender. Our paper explicitly studies differences in the prices, markups and marginal costs of the entire range of retail goods that are bought by men and women, capturing this sorting component.

Within economics, there is relatively little work that focuses on gender disparities in the pricing of goods and services. The most-related work on gendered price discrimination focuses on bargaining contexts for wages or products. Most recently, Rousille (2021) attributes nearly 100% of gender pay inequality among tech industry workers to differences in wage-asks by interviewees, underscoring the potential role for differences in bargaining power to generate gender-inequality. Ayres and Siegelman (1995) provide evidence of race and gender discrimination in bargaining for new cars, finding that women and black men paid significantly higher markups for cars than white men. This setting has been further studied by Goldberg 1996 and Trégouët 2015, with Castillo et al. (2013) also documenting systematic differences in stages of taxi-price bargaining for men and women. Fitzpatrick 2017 finds evidence of gender price discrimination in the context of bargaining for anti-malarial drugs. While these studies provide evidence and precedence of price discrimination against women, they do not capture a mechanism by which price discrimination can occur of goods with simple take-it-or-leave-it list prices nor the role of differences in preferences across product offerings.

Because we investigate the pink tax across the retail consumption basket, we view this work as closely-related to research on inequality in consumption and product offerings. Jaravel 2019 finds that poorer households experience higher inflation and price indices, exacerbating income inequality in real terms. Though we do not directly calculate differences in inflation for men and women, our work on gender explores a new angle through which price index inequality may shape wealth inequality at large. Aguiar and Hurst 2005 use survey data to demonstrate that consumption remains relatively constant among individuals as they transition into retirement, simultaneously documenting differences in sources of consumption (e.g. restaurant dining, home-production, etc.) between men and women. Aguiar and Hurst 2007 also quantify objects such as the substitution elasticity between shopping and home production and and the willingness to engaging price shopping or to take advantage of deal; while not explicitly focused on gender distinctions, their findings on the price returns of time spent shopping have important implications for understanding the differences in prices paid by men and women.

The implications of the Pink Tax for gender equality are wide-reaching: 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. Estimates of the raw gender pay gap tend to around 20% today, decreasing to about 10% after including differences in qualifications (Blau and Kahn (2017)). The presence of an aggregate Pink Tax on women's consumption augments these inequalities by reducing women's purchasing power. Moreover, by accounting principles, the existence of a Pink Tax also highlights differences in overall consumption and savings between men and women. Women, facing on average higher prices for their respective consumption bundles face both lower real wages and potentially lower scope to accrue lifetime savings and consume.

Faber and Fally 2022 study how product offerings and firm sorting drive price index inequality across incomes. They find that larger, more productive firms endogenously sort into markets that cater to richer households and that this drives asymmetry in price indices across the income distribution. This study suggests that supply side factors may play an important role in differences in product offerings and marginal costs for men and women. Simultaneously, DellaVigna and Gentzkow 2019 find substantial price mis-optimization for retail chains, where stores typically implementing uniform prices throughout all US stores irrespective of local demand and cost factors—suggesting some limitations to how and to what extent firms engage in optimal price strategy. B. J. Bronnenberg et al. 2015 study how information and experience may drive inequality in product choice on the consumer side by looking at differences in choices made between experts (by profession) and non-experts in purchasing drugs and grocery items. They find that non-experts over pay for brand name products more than experts do. While expertise may not be a direct driver of differences in product choices between genders, this work highlights the potential for misinformation and incorrect product beliefs to affect choices and prices paid.

We contribute to the large literature in industrial organization on the role of product differentiation within markets. Berry, Levinsohn, and Pakes (1995) developed the method that we use to estimate a discrete choice model in the presence of product differentiation. We incorporate the gendered-ness of a product as a characteristic over which individuals may have heterogeneous tastes. We use a revealed preference approach to identifying product gender, which means we do not need existing product characteristic data which enables us to estimate demand in multiple markets. Economists have long thought about the role of heterogeneous tastes and product differentiation in welfare. (George J Stigler and Becker 1977; Spence 1976) Product differentiation and price discrimination are sometimes thought of as separate phenomenons but in a broad view of price discrimination as any markup difference between consumers groups (like that of George J. Stigler 1987) the two are linked. Shapiro (1982) discusses second degree price discrimination through versioning, but there is no clear line of where versioning ends and product differentiation begins. Our paper contributes to this literature by analyzing how demand composition for differentiated products may lead to higher markups placed on a particular demographic group.

Though we do not estimate a general equilibrium model that incorporates endogenous product entry and exit, our findings suggest a natural next step of examining firms decisions to produce gendered products. Wollmann (2018) models entry and exit decisions of truck models finding that allowing for entry and exit moderates price increases resulting from mergers. Barahona et al. 2020 finds that firms decision to reformulate after a policy that affects demand depends on expected profits that face a tradeoff between bunching product characteristics that appeal to a larger demand group but face higher competition or differentiating a product to a smaller consumer base but facing less competition. Firm's incentives to innovate and introduce new products have also been studied in the trade literature. Work on firm and product heterogeneity stemming from differences in demand and costs among firms finds implications for innovation and competition. (Hottman, Redding, and Weinstein 2016, Faber and Fally 2022, and Atkeson and Burstein 2008) Broda and Weinstein (2010) and Bernard, Redding, and Schott (2010) also substantiate the role of innovation and turnover in driving evolution in prices, finding a substantial amount of product innovation: namely that half of firms switch their products within a span of five years and that product creation is a much stronger component of net product entry than product destruction. Although these works do not focus on gender or gendered product spaces explicitly, their findings have important implications for how we understand and motivate study of men's and women's consumer goods markets.

3 Data

We combine data on from two main sources and two supplemental sources to conduct our analysis. Our main analyses rely on the NielsenIQ data including the HomeScan Panel (HMS) and the Retailer Scanner Data (RMS). The HMS data contains purchase histories of for a rotating panel of households from 2004 to 2019. The RMS data contains anonymized purchases of products aggregated to the store-week level throughout 2007 to 2017. We 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. Lastly, we incorporate data from National Promotion Reports' PRICE-TRAK database (PromoData), which features data on wholesaler prices charged to retailers for certain products from 2008-2013. While we discuss these data in turn, see B. J. Bronnenberg et al. (2015) and Allcott, Lockwood, and Taubinsky (2019) for further discussion of the NielsenIQ data.

The entire HMS features data on the shopping trips and transactions of approximately 60k households per year. Households remain in the panel for on average 54 months, with approximately 200,000 distinct households rotating through the HMS in total. The data report on 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 the 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/sale usage, and detailed nests of product identifiers.

Our primary uses for the HMS data are to document differences in the purchasing behavior of men and women and understand how product markets differ for men and women. To confidently assign product purchases to consumer gender demographics, we restrict our consumer panel to single-individual households that log at least 12 shopping trips per year, which eliminates approximately 75% of the panelists in the HMS. This leaves us with a panel of 47,012 households which we use to study differences in consumer behavior. Summary statistics for the sample can be found in 1. Our final sample is skewed women, with about 70% of our panelists identifying as a woman. In terms of balance, the men in our sample tend to have higher income and be more educated, which we will control for in the analysis. 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 that we can confidently assign a gender to. We describe our methodology in detail in Section 3.2. The NielsenIQ data covers approximately 1.8 million products and we are able to confidently assign gender to 700,000 of them. However, these 700,000 products comprise 97% of the purchases made in our singles panel.

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

The RMS data contain product-store-week level prices and volumes of products purchased

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

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

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

by consumers from 2004 to 2018. This dataset is not tied to the consumer identifiers; rather, the strength of the RMS data is 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). These markets include yogurt, health and protein bars,

[3] These figures represent the midpoint of the discrete income buckets used for the household income field

deodorant, disposable razors and shampoo. While the HMS tracks all retail purchases for a household from any store, the RMS contains a select set of stores. For our analysis, we keep only stores that are part of a larger retail chain rather than independent stores.

Both components of the NielsenIQ data feature a highly detailed product hierarchy classification that organizes all goods into smaller nests with increasing degrees of specificity. Products in the NielsenIQ are identified with their Universal Product Code (UPC) which corresponds to a unique barcode. All UPCs fit into one of ten *departments* (the broadest category, e.g. "Health and Beauty" and "Dry Grocery"). From here, products in a department are allocated to *Product Groups*—of which there are 120 total—such as "Shaving Needs". Finally, UPCs in the same Product Group are assigned to *Product Modules*—the most granular grouping of multiple products—e.g. "Disposable Razors". The Nielsen data identifies over 1300 distinct product modules. Brand description represents an alternate grouping that features the brand name for a given set of UPCs, not strictly contained in any single Product Module or Group contained, such as "Venus", for the brand of razors. We consider Product Modules as constituting a self-contained goods market; for certain reduced-form analyses, we further divide product modules into Module-Unit groups (modules composed of goods all with the same counting units: e.g. the coffee product module contains bagged coffee measured in weight (ounces) and Keurig cup coffee measured as a count (number of K-cups).

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. Summary statistics are presented in Appendix X. 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 info 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.

Lastly, the PriceTrak PromoData data allow us to validate retailer markups relative to wholesaler price. A critical component of this work consists of assigning the source of the observed differences in prices of goods consumed by women and by men to differences in marginal costs and differences in markups. While this data does not feature information on production costs, it does provide on the intermediary costs to retailers. The PriceTrak PromoData ultimately serves to facilitate a cross-validation against our structural demand estimation that uses solely price and consumption information from NielsenIQ. The PriceTrack data features retailer cost-data of individual UPCs for a variety of time- and geographic-denominations from 2006 and 2012, with the geographic disaggregations covering 48 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—with only about 18% of the 430,000 distinct UPCs in the RMS data matching to PromoData. 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 following B. J. Bronnenberg et al. (2015).

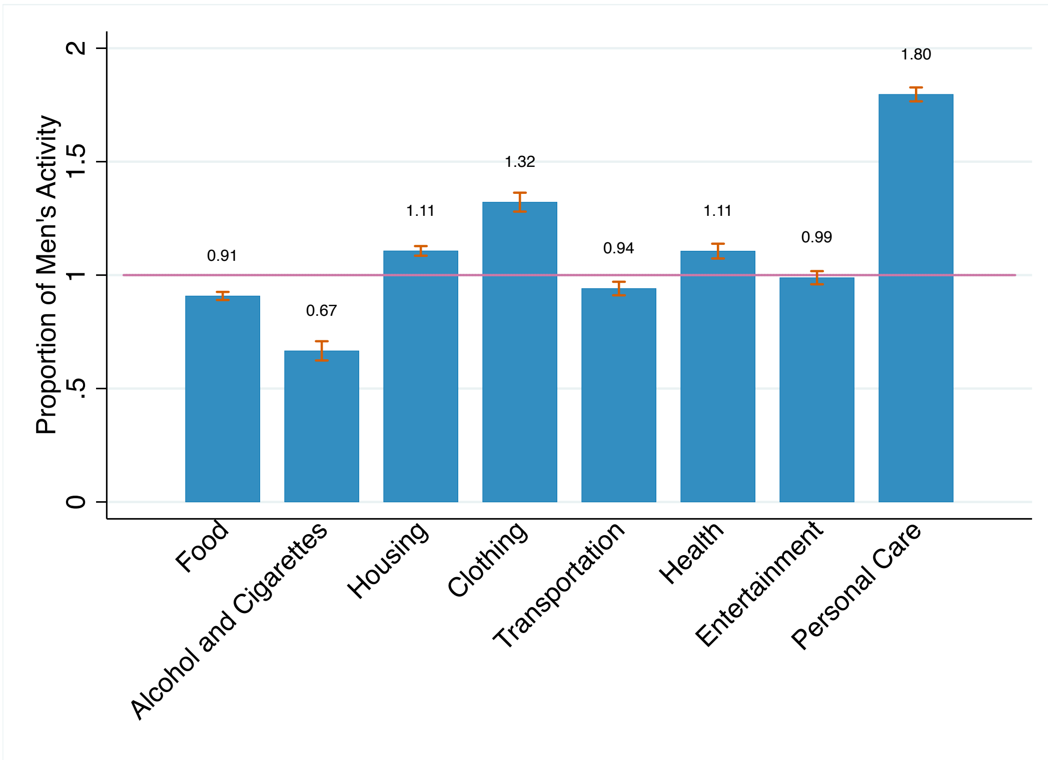
4 Price Disparities Across the Consumption Bundle

We present evidence of gender differences in both consumer behavior and the product space. We begin by examining overall differences in consumption basket composition, finding significant differences in how men and women choose to allocate their income. We document that, within grocery and big box retail purchases, women spend more than men—both overall and per item—and that products primarily bought by women are priced higher than those bought equally by men and women or primarily by men. These consumer behavior differences and product space differences indicate that gender disparities in consumption are driven by both demand and supply side forces. Women spend more per item, and there exists a larger product space of goods marketed more exclusively toward women than toward men. In line with these findings, we demonstrate the existence of a women's price premium of approximately 4% on average.

4.1 Consumer Behavior by Gender

First, we document that women's consumption bundles are different from those of men in terms of composition. Using the CE PUMD we find that women and men do not have significant differences in total yearly spending, but how they choose to allocate their spending highlights important differences in preferences across all types of spending. Figure 1 plots women's yearly spending as a percentage of men's. Each bar plots the coefficient from a regression of log spending for a category on an indicator for the individual being a woman controlling for age, income and race. 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. These findings roughly correspond with markets that are often discussed in discourse on the pink tax and gendered marketing more broadly. The focus of this paper is on differences in men and women's behavior and product space for retail markets like grocery and big box stores. These purchases largely fall under the categories of food, alcohol and cigarettes, and personal care but they do not map perfectly. A key descriptive result of our paper is that women spend more on retail purchases than do men; in the context of figure 1 this would imply that women spend more on food as groceries while men spend more on food out. Similar overall levels of spending with differing allocation patterns highlights the important role that preferences, substitution patterns and societal expectations play in evaluating the pink tax.

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

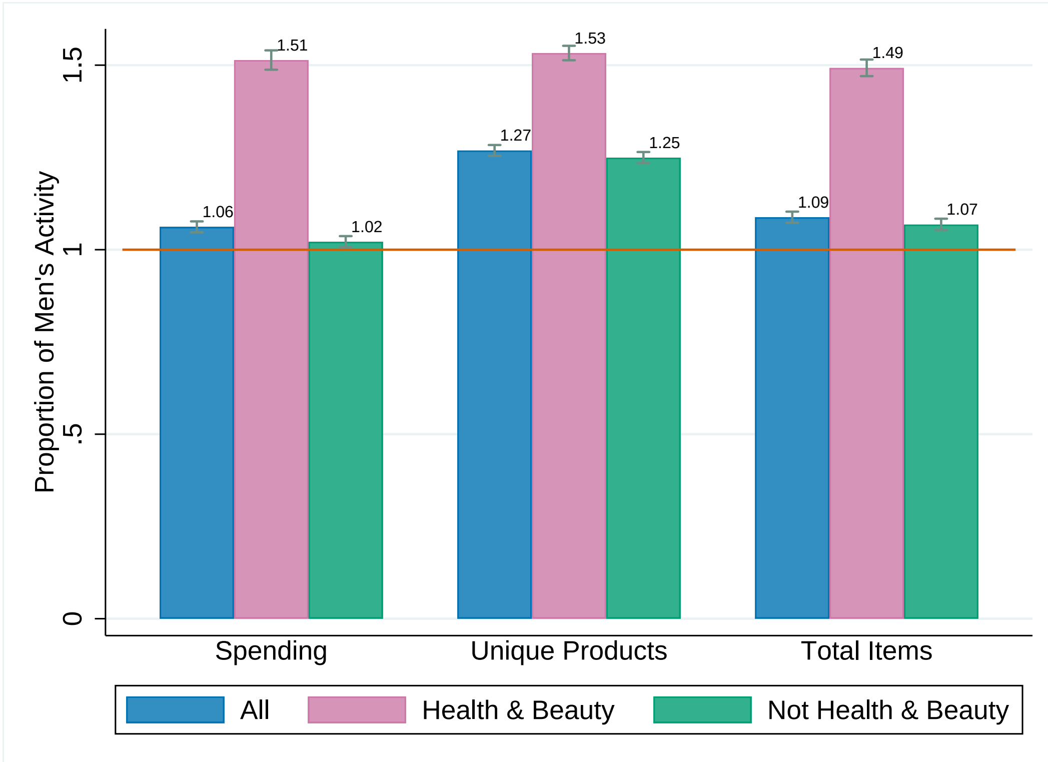


Note: This figure plots the coefficients estimated from a regression of log expenditure on an indicator for the individual being a woman and demographic controls: $\log y_{it} = \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_{i,t} + \varepsilon_{i,t}$, for spending categories food, alcohol and cigarettes,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

While figure 1 speaks to full consumption basket differences between men and women, we now turn our focus to retail spending consumption baskets and how they vary by gender. We find that women's retail consumption baskets are larger, more expensive, and filled with a greater number of unique UPCs. Figure 2 plots levels of female activity as a proportion of male activity for annual spending, unique product purchases, and overall product purchases. We find that women's yearly spending is greater than that of men by about 6%, their product diversity is greater than men's by about 27% and their consumption baskets are larger than men's in terms of items purchased by about 9%. 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.

Figure 2 compares men and women that otherwise look demographically similar in terms of location, age, race, income and education. Our panel of singles is weighted to be representative of all single men and women in the United States, and thus we can also make comparisons of how men and women's spending differs in the aggregate, by location, etc. Table 2 reports yearly spending differences between men and women subsequently adding in these demographic controls. Column (1) reports aggregate differences in spending between men and women, including only controls for year. We find small differences in yearly spending without demographic controls of about 1.6% higher yearly spending by women. Columns 2, 3, 4, 5, and 6 add in fixed effects for county of residence, income, age, race and education respectively. Column (2) compares yearly spending between men and women that live in the same county, finding 2.5% higher spending by women. We can interpret the increase in the magnitude of the coefficient across columns (1) and (2) as the contribution of geographic sorting of single men and single women to overall spending differences and is consistent with single women more often living in lower cost areas than do men. This interpretation continues as we move to columns (3) and (4) which add in controls for income and age which raise the

Figure 2: 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: $\log y_{it} = \alpha + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_{i,t} + \varepsilon_{i,t}$, 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 $X_{i,t}$ 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.

spending gap to 4.4% and 6.2% respectively. Because single women tend to skew lower income and older in age than single men, we can see the attenuating effect that lower spending among older and poorer women has in the aggregate. Columns (4) and (5) add in controls for race and education; while racial composition differences between single men and single women do contribute to the spending gap somewhat, the magnitude of the change is much smaller than the contribution from geography, age and income. Controlling for education has no contribution that cannot be accounted for by geography or the other demographic variables. While yearly spending differences vary significantly across different comparisons of interest, the same analysis on the number of unique products consumed or total number of items purchased in a year shows little variation (see appendix X). These findings suggest that while there are many factors that contribute to yearly spending differences between men and women, gender differences in consumption basket size and composition are fairly constant.

Table 2: Yearly spending differences between men and women

	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.0162**	0.0248***	0.0444***	0.0616***	0.0677***	0.0677***
	(0.0080)	(0.0077)	(0.0076)	(0.0076)	(0.0075)	(0.0075)
Observations	216890	216743	216743	216742	216742	216742
Adjusted R^2	0.018	0.097	0.105	0.125	0.133	0.133
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

Individual level clustered standard errors in parentheses

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

Note: This table presents estimates of the percent difference in yearly spending between men and women using the following regression: $\log y_{it} = \phi_t + \beta \cdot \mathbb{1}\{woman_i = 1\} + \Gamma X_i + \varepsilon_{i,t}$, where $_{it}$ 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 2 and 2 document that women's consumption bundles differ from those of men in important ways, but does not fully inform the way through which a pink tax may take form. Aggregate spending differences can arise from differences in prices paid for similar goods or from differences in quantities purchased. As a clarifying example, consider consumption habits for shampoo. Women, on average, have longer hair than men which may lead them to buy more bottles of shampoo over the course of a year, we refer to this as driving up total spending on the extensive margin, that is, buying more product. It is also possible that women have preferences for higher priced shampoos, we refer to this as the intensive margin, where women are paying higher per unit prices. Figure 2 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).

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

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

Where i denotes the individual, j denotes the product purchased and t denotes the market. Table 3 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. Similar to Table 2, we can think of the 2.3% result as the raw difference in prices paid between single men and women, not accounting for other demographic factors or location and retail chain sorting. Column (2) runs the same specification but adds in controls for age, income, and race. The large increase in the coefficient, from 2.3% to 4.67% highlights the important role 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. We can think of Column (3) as the contribution of women sorting into more or less expensive locations, because the coefficient change is small, the contribution is minimal. Similarly, column (4) can be thought of as the contribution of sorting into more or less expensive retail chains, i.e. Whole Foods vs. Walmart. Controlling for the retail chain lowers our price premium estimate to 4.19%, suggesting that retail chain sorting plays a small but significant role. Finally, in Column (5) we add in fixed effects for month rather than year. The results indicate that women spend more than 4.02% more than do men per unit of goods in the same product market, bought in the same retail chain, county, and month. We consider this our preferred specification because it attempts to control, as much as possible, for all potential differences that could arise between the two groups other than gender.

We refer to this 4% finding as our observed pink tax on the intensive margin. This price premium could be driven by many different factors. First, it can be driven by women buying products that are made specifically for and marketed to women, this would be in line with how the pink tax is traditionally thought about. Alternatively, it could be driven by differences in preferences between men and women for products that are otherwise ungendered. That is, if women happen to like organic products or name brand products more than men then we would likely observe that women pay higher per unit prices than do men. Once we understand which types of products are contributing to our observed pink tax, we want to know whether these price premiums are being driven by markups or marginal costs. The underlying implication of popular discourse on the pink tax is that it is a price discrimination story: two products differ only in their color but the pink product is priced higher, because their costs of production must be the same the women's product faces a higher markup. This price discrimination story would require that women either consume products less elastically or the competitive structure of the market is such that the products women buy face higher markups. The alternative explanation is that the products that women buy have higher marginal costs of production. This would be consistent with women having preferences for higher "quality" goods.[4] The rest of the paper strives to understand what generates our 4% pink tax by analyzing how product markets vary for men and women and estimating differences in demand between men and women.

Table 4 estimates the same equation as Table 3 while including product level fixed effects instead of module level fixed effects:

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

The interpretation of the coefficient becomes the difference in prices paid between men and women for the same exact product. Differences in prices paid for the same good can be attributed to differences in price shopping behavior, like coupon usage and sale shopping, consistent with being a more elastic consumer. We sequentially add in fixed effects in the same manner as Table 3, 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. Just like Table 3 we find that demographic differences

[4] We cannot directly attribute higher marginal costs to higher quality as quality is likely not fully innate but perceived by the individual.

Table 3: Unit prices in same product module

	(1)	(2)	(3)	(4)	(5)
Woman	0.0230***	0.0467***	0.0512***	0.0419***	0.0402***
	(0.0035)	(0.0033)	(0.0028)	(0.0020)	(0.0018)
Observations	153,333,409	153,333,409	150,059,493	143,532,160	139,739,839
Adjusted R^2	0.829	0.831	0.868	0.889	0.877
ModuleXUnits FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No
County FE	No	No	Yes	Yes	Yes
Retailer FE	No	No	No	Yes	Yes
Month FE	No	No	No	No	Yes
Demographic FE	No	Yes	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

Note: This table presents estimates from the regression: $log(P_{ijt}) = \phi_{t(j)} + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a market-time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. 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.

and differential sorting into retail chains and locations contribute to the price shopping gap. While we find that women, on average, buy more expensive products than do men, we find that they spend 0.8% less than men on the *same* product. Column (5) captures differences in prices paid for the same product by people that differ only in gender over a month, which we attribute to differences in price shopping behavior. Combining this with our result from Table 3 suggests that women are buying higher priced goods while also exhibiting behaviors associated with being more elastic consumers. Hendel and Nevo (2013) study promotional sales as a form of intertemporal price discrimination, our results would indicate that women are likely to comprise a larger share of the consumer base that benefits from this type of price discrimination.

Table 5 estimates the preferred specifications from tables 3 and 4, stratifying by department. We can see that our results hold generally across most departments with the exceptions being Alcohol and General Merchandise.[5] Among all other departments we find that women buy higher priced products relative to men while displaying acutely more price shopping behavior.

[5] However, we identify these Nielsen departments as possibly underestimating consumption, as there are many purchases of these types of products made at stores not included in the Nielsen panel.

Table 4: Unit prices for the same product

	(1)	(2)	(3)	(4)	(5)
Women	-0.0089***	-0.0055***	-0.0060***	-0.0075***	-0.0080***
	(0.0017)	(0.0017)	(0.0015)	(0.0010)	(0.0010)
Observations	151,188,750	151,191,277	139,671,522	138,165,657	135,154,990
Adjusted R^2	0.952	0.840	0.860	0.878	0.879
UPC FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No
County FE	No	No	Yes	Yes	Yes
Retailer FE	No	No	No	Yes	Yes
Month FE	No	No	No	No	Yes
Demographic FE	No	Yes	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

This table estimates reduced forms specified as $log(P_{ijt}) = \phi_{jt} + \beta \mathbf{1}_{w(i)} + \gamma X_i + \epsilon_{ijt}$ where P_{ijt} is the per-unit price of a UPC. $\mathbb{1}\{woman_i = 1\}$ is an indicator for whether the individual is a woman, ϕ_t is a UPC-market-time fixed effect and X_i is a vector of demographic controls including income, county, age, race and education which we add in sequentially. 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.

Table 5: Prices paid across departments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	H&B	Dry Groc.	Frozen	Dairy	Deli	Pack. Meat	Produce	Non-food Groc.	Alcohol	Gen. Merch.
Panel A: Per unit prices within product module										
Female	0.0534***	0.0546***	0.0489***	0.0374***	0.0348***	0.0485***	0.0182***	0.0348***	-0.1450***	-0.0484***
	(0.0029)	(0.0019)	(0.0025)	(0.0022)	(0.0063)	(0.0031)	(0.0037)	(0.0021)	(0.0204)	(0.0039)
Observations	10504032	55328879	14261546	16609254	5270903	4094787	10769045	12467554	1991861	6719882
Adjusted R^2	0.836	0.860	0.903	0.934	0.900	0.798	0.779	0.870	0.637	0.780
ModXUnitXRetXLocXMonth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Per unit price for same UPC										
Woman	-0.0212***	-0.0034***	-0.0050***	-0.0012	-0.0258***	-0.0070***	-0.0160***	-0.0133***	-0.0031	-0.0016
	(0.0022)	(0.0009)	(0.0014)	(0.0010)	(0.0055)	(0.0015)	(0.0030)	(0.0010)	(0.0019)	(0.0049)
Observations	9668836	61456951	12812841	15406630	5381334	3922662	10758430	10671760	1891549	6259655
Adjusted R^2	0.817	0.891	0.842	0.875	0.637	0.813	0.654	0.933	0.930	0.854
UPCXRetXLocXYear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

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.

Our findings are particularly strong for Health and Beauty products in Column (1), women buy products that cost on average 5.34% more than those bought by men, but when buying the same exact product women typically spend about 2.15% less than men do. Given the types of products focused on in the media when discussing the pink tax, one may expect that any results would be driven by Health and Beauty products where market segmentation by gender is particularly apparent. However, while our Health and Beauty results are relatively larger in magnitude, the pattern of our finding holds across all departments, including ones where the product space is less intuitively stratified by gender. This consistent pattern suggests that the pink tax is not just about goods marketed to men versus women but also about systematically different preferences for otherwise ungendered items.

4.2 Gender in the Product Space

We now shift our focus from consumer behavior to understanding how the the product space varies by gender. Media portrayal and public discussion depicts the pink tax as phenomenon associated with specifically gendered products. The generally suggested mechanism is that by creating products that are bought exclusively by one gender, firms can segment the market and price discriminate accordingly. Our descriptive evidence above has shown 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. To fully characterize the pink tax, we document the existence of goods that are gendered, that is they are only ever bought by one gender, and decompose our observed pink tax of 4% into into its respective contributions from gendered products and differential purchasing of ungendered products.

First, we assign values of gender-stratification to each good. We begin by calculating a woman purchase share for each UPC in our data as the projection-weighted fraction of purchases by women. We define the observed time-invariant woman purchase share of UPC j

as

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

This fraction assigns $\hat{w}_j \in [0, 1]$ where 0 denotes a good that is only bought by men and 1 denotes a good that is exclusively bought by women.[6] We sometimes use $\hat{w}_j$ as a continuous measure, but for simplicity use it to categorize goods as either gendered or ungendered. We define explicitly gendered products as those that are purchased at least 90% of the time by a single gender. That means 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. For robustness, we repeat analyses with a cutoff of .25 and .75 and include them in the Appendix.

Approximately two-thirds of the UPCs purchased by Nielsen panelists are only ever observed to be purchased once, these UPCs would always be assigned to having an explicit gender of 0 or 1. To reduce measurement error, we only assign an observed women purchase share to products that are observed to be bought with enough frequency. In theory, each UPC in our data has a true woman purchase in the population, w_j, that we do not observe. We choose a cutoff number of unique observations, n_j^*, needed to assign a UPC gender such that we are 95% sure that the true woman purchase share lies within a ten percentile bin centered around the observed value. We observe a UPC's woman purchase share, $\hat{w}_j$, and the number of unique individuals that purchase it, n_j. Our observed values represent a draw from a binomial distribution.

$$P(w_j \notin [\hat{w}_j - .05, \hat{w}_j + .05]) = \int\limits_{x < \hat{w}_j - .05, x > \hat{w}_j + .05} \binom{n}{\lceil \hat{w}_j n \rceil} x^{\lceil \hat{w}_j n \rceil} (1-x)^{n - \lceil \hat{w}_j n \rceil} f(x) dx$$

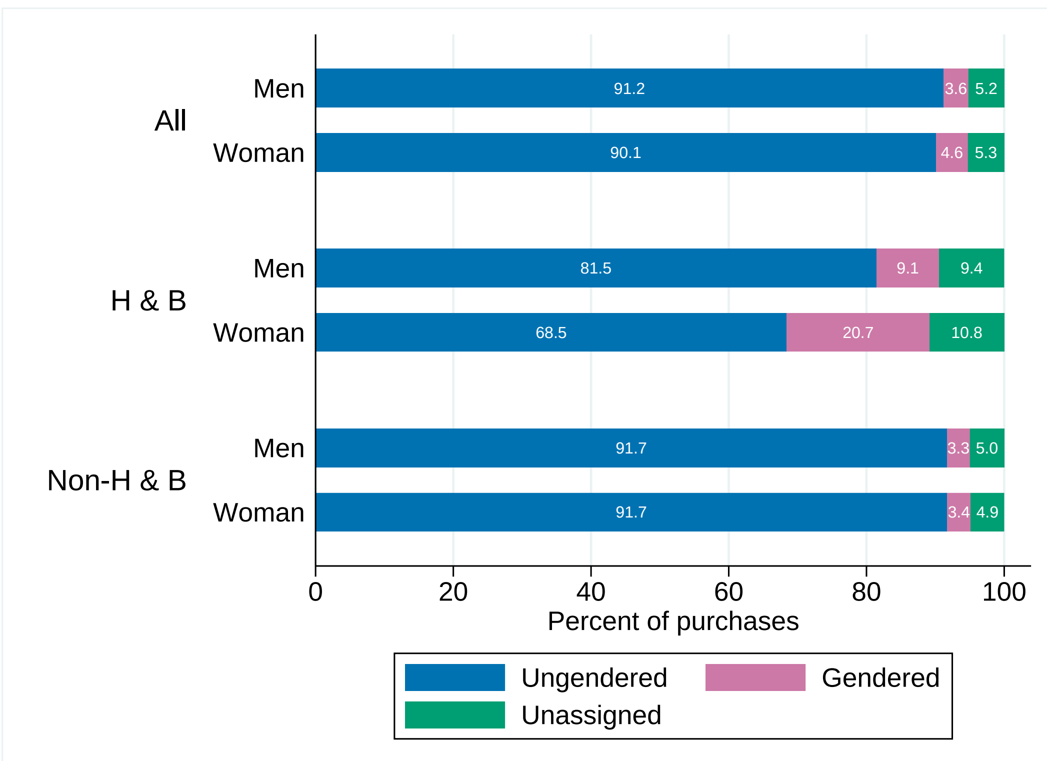
Where $f(x)$ is the empirical pdf of woman purchase share and $\lceil \hat{w}_j n \rceil$ is the closest integer to generating woman purchase share. We calculate the threshold, n_j^*, such that the probability, $P(w_j \notin [\hat{w}_j - .05, \hat{w}_j + .05])$, is .05.

[6]The HMS sample features a woman-man gender-split of approximately 70-30. We scale purchases using the proprietary Nielsen projection weights, which yields a gender-composition of 53-47. This means that the average good will be skewed slightly towards women but accurately representative of the population.

Figure A.1 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 the inclusion 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 3 shows those we are able to assign gender to account for greater than 95% of all purchases made in the data by expense. Next, we consider goods assigned to women. We find that gendered products make up a small share of purchases, 3.6% for men and 4.6% for women. Within Health and Beauty products though, gendered products make up 20% of women's purchases and 10% of men's purchases.

We plot the distribution of woman purchase share for all products, Health and Beauty

Figure 3: Consumption Basket Composition by Product Gender



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

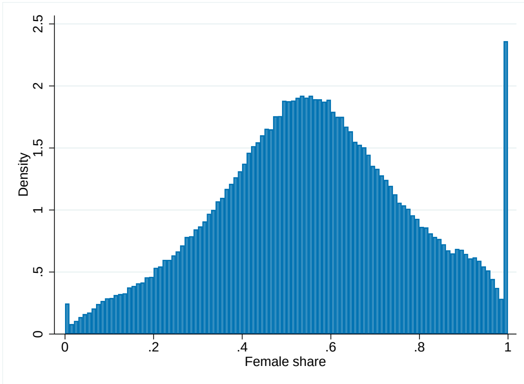
products, and all products excluding Health and Beauty in Figure 4. Panel 4a depicts the woman purchase share for all UPCs in our data after making our cutoff restriction. There is significant excess mass at the right tail of the distribution where goods are bought exclusively by women but virtually no excess mass at the left tail of the distribution where goods are bought exclusively by men. Importantly, re-weighting purchases to account for the differential gender-composition of the Nielsen panelist sample implies that this difference can be attributed to differences in consumption behaviors and preferences between men and women. Part of this difference may lead to a mechanical overstatement of gender stratification. For instance, if women purchase a greater number of unique products than do men, products will still be more frequently categorized as women's products, even after accounting for gender differential sample-composition. The general right skew of the UPC gender distribution indicates that even among goods that are not explicitly gendered, there are more goods that are bought more frequently by women. Panel 4b shows the distribution of female purchase share for Health and Beauty products, where we find that the distribution is extremely right skewed. There is a considerably large mass of goods that are purchased exclusively by women and very few that are purchased exclusively by men.

We now describe how prices vary along our measure of UPC gender. Figure 5 plots the coefficients from a regression of log unit price on ten-percentile-width bins of female purchase share and two bins for pure gender stratification at the tails, taking the 50th percentile bin as the reference point. Bin b contains goods with UPC gender $\hat{w}_j \in ((b-1)/10, b/10]$, save for the tail bins. This corresponds to estimating the following regression:

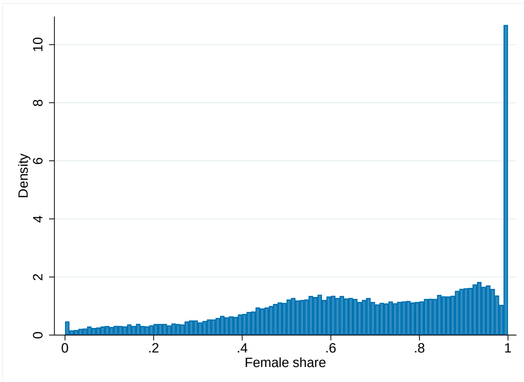
$$log(P_{jt}) = \phi_{t(j)} + \sum_{b \in \mathcal{B}} \beta_b \mathbf{1}_{d(j)=b} + \epsilon_{jt}$$

The regression contains fixed effects for the product module of the UPC, county and half-year of purchase: so 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 document significant price premiums of $\sim 10 - 40\%$ for goods purchased exclusively by either men or women relative to goods in the same market purchased by relatively equal shares of each.

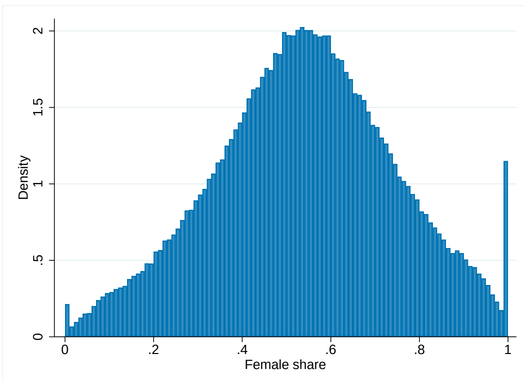
Figure 4: Distribution of Female Purchase Share Across UPCs



(a) All Departments



(b) Health and Beauty



(c) Non-Health and Beauty

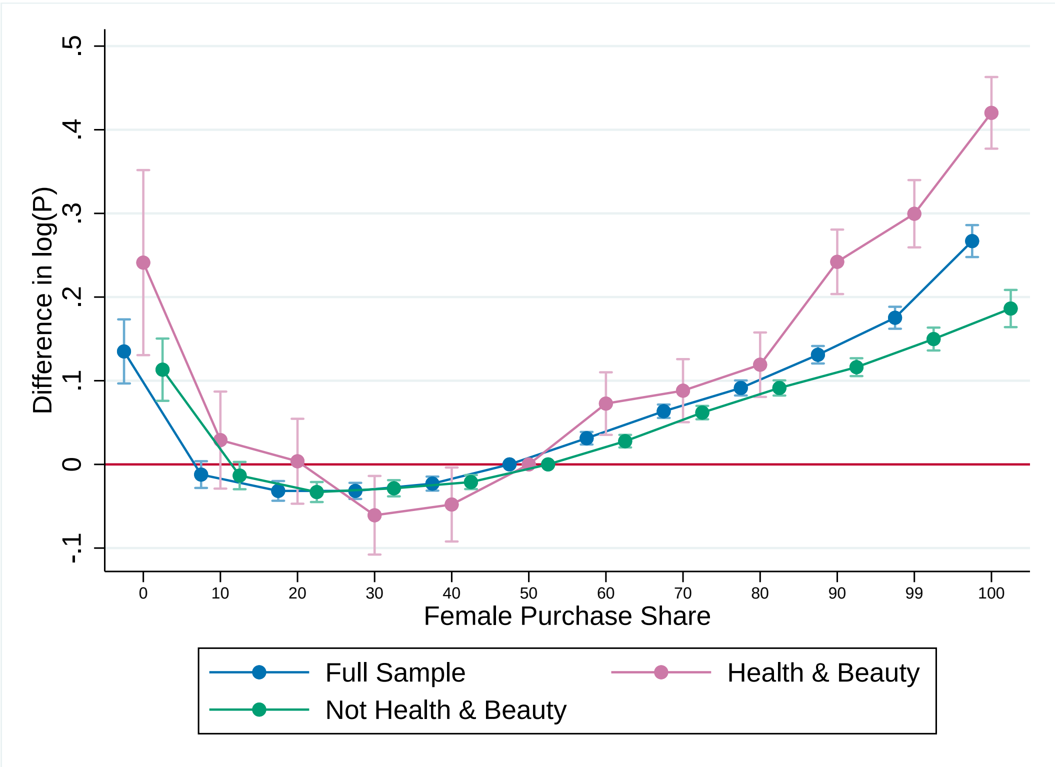
Note: This figure plots a histogram of the share of times a UPC is bought by women. We restrict to UPCs that have above a varying cutoff number of purchases by unique individuals over the panel, this cutoff number corresponds to 95% confidence that a product's true purchase share is within a 10 percentile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

We observe higher prices for women's products as compared to men's products. This difference ranges from just under 10% for non-health and beauty products to almost 20% for health and beauty products. This finding falls in line with popular depictions of the pink tax, which tend to focus on examples of price premiums for women's products relative to products that are explicitly gendered towards men. Less talked about in discourse on the pink tax is the potential for price premiums on gendered products in general which we find strong evidence of given the overall U shape of the graph. Beyond the tails of the graph, a striking pattern emerges, prices tend to monotonically increase in woman purchase share. That is, products that are bought more often, but not entirely, by men are priced lower than products that are bought more often by women. This monotonic increase in prices along woman purchase share suggests that our overall price premium of 4% from Table 3 is likely explained not just by explicitly gendered products (i.e. pink products and blue products) but also by differences in preferences for otherwise ungendered products. This finding is consistent with women having preferences for higher (perceived) quality items like, for example, organic products. This is supported by studies of differences in preferences for organic food between men and women. Ureña, Bernabéu, and Olmeda (2008) finds that women are more likely to buy and value organic food but that men are more willing to pay a higher price.

Using our definitions of product gender, we can decompose our overall 4% price premium into a contribution from differential sorting between men and women into ungendered products and their purchases of gendered products. We define gendered products as those that are bought at least 90% of the time by one gender. Because among these products it is mechanically unlikely that individuals buy a product of the other gender, we do not further divide products into men's and women's. We run the same regression specification as Table 3 column (5) but now include an indicator for whether a good is gendered and an interaction between the woman indactor and the product gender indicator:

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

Figure 5: Prices of UPCs by Female Purchase Share



Note: This figure presents plots of the results of the regression $\log P_{u,t} = \alpha + \sum_{b \in \mathcal{B}} \gamma_b \mathbb{1}\{g_u \in Bin_b\} + \theta_{m,l,t} + \varepsilon_{u,m,c,l,t}$. Bins $b \in \mathcal{B}$ include ten-percentile-width bins 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.

The results are presented in Table 6. β_1 captures the average difference in prices that women pay for ungendered products relative to men, β_2 captures the the differences in average prices that men pay for gendered products compared to the ungendered products they buy and β_3 captures the difference in average prices that women pay for gendered products relative to the ungendered products they buy. The average difference in prices that women pay for gendered products relative to gendered products bought by men is given by a linear combination of the coefficients, $\beta_1 + \beta_3 - \beta_2$. The coefficient on the woman consumer indicator in column (1) indicates that women pay a price premium of 3.83% on ungendered products relative to ungendered products bought by men. The coefficient on the gendered product indicator in column (1) shows that, across all departments, men pay lower prices on gendered products than they do ungendered products by about 1%, though the result is marginally significant. Finally, the coefficient on the interaction between the woman consumer indicator and the gendered product indicator in column (1) shows that women buying gendered products pay about 11.39% more relative to the ungendered products that they purchase. Overall, we find that women pay approximately 16.26% higher prices on gendered products than do men. While the magnitude of coefficient on women buying gendered products is large, it's contribution economically to the overall price premium is small. The 4% price premium from Table 3 is approximately the purchase weighted average of the 3.83% price premium that women pay on ungendered items and the 16.26% price premium they pay on gendered products. From Figure 3, we know that gendered products make up an overall small share of a consumption bundle. So while we find evidence of woman gendered products having significantly higher prices, the vast majority of our observed pink tax is being driven by differential sorting between men and women on ungendered products.

Columns (2) and (3) of Table 6 separately estimate the results for Health and Beauty and all other product categories. The results largely follow the same pattern as the aggregated estimation with a key deviation being prices of men's gendered products relative to men purchasing ungendered products. While in non-health and beauty products we find that man gendered products are priced lower, within health and beauty we find that men buying man gendered products spend about 3.63% more than similar products they buy that are ungendered. This finding lines up with Figure 5, man gendered items correspond to products in the 0 bucket and the 10 bucket in the graph (products bought up to 10% of the time by a woman). While we find upticks in prices for goods that are only ever bought by men in all departments, products that are bought between 0% of the time and 10% of the time (non-inclusive) by women are priced lower for non-health and beauty products. All in all, we take this as evidence that gender price premiums may exist for both men and women for health and beauty products, though premiums on women's products tend to be larger. This opens an avenue for potential price discrimination on gendered niche-ness rather than on a specific gender. However, across all departments, women pay higher prices than men regardless of if a product is gendered or not.

Table 6: Unit prices by gender of product and consumer

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
Woman consumer	0.0383***	0.0407***	0.0328***
	(0.0019)	(0.0030)	(0.0017)
Gendered Product	-0.0104*	0.0363***	-0.0158**
	(0.0054)	(0.0062)	(0.0072)
Woman Consumer & Gendered Product	0.1139***	0.1014***	0.0879***
	(0.0059)	(0.0065)	(0.0081)
Observations	131609099	9306618	120577845
Adjusted R^2	0.884	0.844	0.888
ModXUnitXRetXLocXMonth FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

Note: 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.

Together, our descriptive analysis of consumer behavior and the product space suggest that women and men make significantly different consumption choices from product markets that differ in composition. Women buy products that are more expensive and there exist significantly more products marketed to women than men. These two findings contribute to the observation that women's retail consumption baskets are more expensive, larger and more diverse than men's are. We find that women pay about 4% more than do men for products in the same market. However, this price premium is not driven solely by the existence of products that are only marketed to and bought by women but also by differential sorting into purchasing products that are otherwise ungendered. While these analyses document observable differences between the consumption habits of men and women, they do not speak to the mechanisms that give arise to them. We now turn our attention to estimating the demand side forces that could yield such a market equilibrium.

5 Men's and Women's Demand Elasticities

Following our descriptive analysis, we want to decompose the 4% price premium paid by women into markups and marginal costs. By formally estimating demand and imposing supply side competitive structure, we can back out markups and marginal costs from estimated price elasticities with a modified Lerner Rule of the form $\frac{P-MC}{P} = -\frac{1}{\varepsilon}$ where P is the price of a product, MC is the marginal cost of production and ε is the residual demand elasticity for the product incorporating cross price elasticities to other products owned by the firm. Because our descriptive results indicate that women pay higher prices across the entire consumption basket, we would like our demand analysis to speak to the entire retail consumption basket as well. However, as we broaden the set of markets that we focus on, we face a trade off between increasing the representativeness of our results and allowing for flexible substitution patterns and market structure. We strike a balance by employing two types of demand analysis: a constant elasticity of substitution model that is common in the trade literature (Faber and Fally 2019; Hottman, Redding, and Weinstein 2016) and a differentiated products model that is common in the industrial organization literature (Berry, Levinsohn, and Pakes 1995.

We use our constant elasticity of substitution (CES) model to estimate differences in price elasticities of demand between men and women using our panel of single individual households in the HMS. A benefit of this aggregated approach is that not only can we estimate elasticities for all markets in our sample, but that we can also attribute price responses to a specific gender. However, this comes at a cost of the data being relatively sparse. We have about 15,000 individuals in any given year of our sample but over one million UPCs that we observe to be purchased. This means that our estimates are largely based off of goods that are bought frequently, which tend to be ungendered products in food grocery markets. These products comprise the bulk of the consumption basket, but we won't have good identification on infrequent purchases. Because of this, our CES model captures the role of demand composition and sorting across ungendered products and the relative value of women and men as consumer bases to price discriminate against.

We structurally estimate markups and marginal costs with a differentiated products market demand model, the model validates our findings from the CES model while also incorporating market structure, flexible substitution patterns across a product's gender, as well as identification in markets where purchases are infrequent. To estimate the model we use store level weekly sales data from the RMS, this data is subject to considerably less sparsity than the individual purchases which allows us identification in markets with products are purchased relatively infrequently, namely Health and Beauty products. However, using store level data comes at the cost of not being able to attribute purchases to a specific gender. To overcome this we study how elasticities, marginal costs and markups vary with woman purchase share within a product market. Allowing for flexibility in substitution patterns requires significant computational power, therefore we restrict our analysis to five product markets that have significant dispersion of woman purchase share across the product space.

5.1 CES Model and Estimation

To estimate demand elasticity differences between men and women, we augment the constant elasticity of substitution (CES) model used in Faber and Fally (2019). This approach allows us to aggregate elasticities and make comparisons of the purchasing habits of men and women across a wide range of products. The model characterizes a representative consumer for each location and period, l, that varies in gender, g. The consumer allocates their income between retail goods, G, and consumption of the outside option:

$$U(g) = U(U_G(g), C(g))$$

We assume that the basket of goods that comprise the outside option, C, is consumed normally.

The model aggregates products in two tiers: the consumer allocates consumption across product modules with Cobb-Douglas elasticity and substitutes between goods with module-specific constant elasticity of substitution. We denote product modules with n and refer to a market, t as a product module within a location and time period. The consumer maximizes their utility subject to their budget constraint by choosing a vector of quantities, q, that represents their consumption bundle across all goods:

$$U_G(g,l) = \max_{q} \prod_{n \in \mathcal{N}_l} \left[\sum_{j' \in G_t} \left(q_j \varphi_j(g) \right)^{\frac{\sigma_t(g)-1}{\sigma_t(g)}} \right]^{\alpha_t(g) \frac{\sigma_t(g)}{\sigma_t(g)-1}} \tag{1}$$

$\mathcal{N}_l$ refers to the set of product modules that the representative consumer in location and time l consumes from, and j refers to a specific UPC (product) within a product module. Some studies that estimate demand elasticities with the Nielsen data study products at the brand-level, whereas we consider the UPC-level due to inconsistencies in the gender-marketing of products within brands. To illustrate, in the disposable razors market, all product brands produced by Gillette map to the gender of the product (e.g. Gillette Venus marketed toward women versus Gillette Fusion marketed toward men), but other razor brands like Bic do not always have brand names that map to one gender (e.g. Bic Plus razors have both female- and male-marketed UPCs under the same brand). φ_j refers to the perceived product quality of a product j in module n. σ_t represents the elasticity of substitution within a market, and α_t denotes the share of expenditures allocated to a market $n \in \mathcal{N}_l$[7]. σ_t is our main parameter of interest, as it captures differences in price responsiveness between men and women.

Specifying the upper tier as Cobb Douglas implies that comparisons of consumption amounts

[7] We assume that $\sum_{n \in \mathcal{N}_l} \alpha_t(g) = 1$ for set of modules $\mathcal{N}_l$

between products within the same module depend on their relative quality-adjusted prices:

$$\frac{b_{jt}(g)}{b_{kt}(g)} = \left(\frac{p_j/\varphi_j(g)}{p_k/\varphi_k(g)}\right)^{1-\sigma_t(g)}, \tag{2}$$

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

$$\Delta log(b_{gjt}) = (1 - \sigma_t(g))\Delta log(\bar{P}_{gjt}) + \eta_{gt} + \varepsilon_{gjt}. \tag{3}$$

Where differences are taken from one time period to the next and η_{gt} captures the change in the price index. Though we derive this estimating equation from a CES demand model, it has the additional benefit of being interpreted in other useful ways. Deviating from constant elasticity of substitution, this estimating equation can be interpreted in a reduced form way as an average of heterogeneous price responses within a market. In our estimation we define markets, t, as a product module-county-half year or product module-county-retail chain-half year. This ensures that our estimated results are based off of true changes in behavior rather than changes in composition of out overall sample over time. We estimate our model at the half-year level because many product categories are prone to stockpiling, which in shorter time intervals would bias our demand estimates towards greater elasticity. To address auto-correlation in the error term, we cluster standard errors at the UPC-county level.

We face the standard issues of simultaneity in demand estimation where price changes may be correlated with demand shocks. To address this issue, we rely on two identifying assumptions typically employed in empirical works. First, we assume that local demand shocks are uncorrelated and idiosyncratic across localities while supply shocks are correlated across space and retailers 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 $(1 - \sigma_t(g))$ using two instruments. The first are Hausman instruments, which we construct as national leave-out means in price changes at the county level, $\frac{1}{N-1}\sum_{c \neq c'} \Delta \log(P_{gjt})$. The second are instruments that follow DellaVigna and Gentzkow (2019) developed by Allcott, Lockwood, and Taubinsky (2019) which are constructed as national leave-out means of price changes at the county-retailer chain level, $\frac{1}{N-1}\sum_{r,c\neq r',c'} \Delta\log(P_{gjt})$. Much of the variation in the DellaVigna Gentzkow instrument is driven by variation in how often a product is placed on a promotional sale. The timing of these sales is driven by a bargaining process between the retailer and the manufacturer and typically only one manufacturer is put on promotional sale at a time. If competition among manufacturers is strong enough, then promotional sale decisions are largely independent of demand shocks as well.

Equation (3) estimates the elasticity of substitution across products within the same 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, σ_{gt}, and market share, $s_{jt}(g)$. Solving Equation (1) yields:

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

Where $P_t(g)$ is a price index, $P_t(g) = \left[\sum_{i\in G_t} p_{jt}^{(1-\sigma_t(g))}\varphi_{jt}(g)^{(\sigma_t(g)-1)}\right]^{\frac{1}{1-\sigma_t(g)}}$. From here we can directly derive the own-price elasticity of demand as:

$$\varepsilon_{jt}(g) = \sigma_t(g) - (\sigma_t(g)-1)\cdot s_{jt}(g) \tag{4}$$

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

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

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

5.2 CES Model Results

We begin by estimating differences in the elasticity of substitution, $\sigma_t(g)$, between men and women. Table 7 presents results of estimating Equation (3) and pooling the elasticities across all departments. The main coefficient of interest is $\sigma_m - \sigma_w$, the difference in elasticity of substitution between men and women. In column (1) We include a UPC-market fixed effect and estimate differences in demand elasticities between men and women for the same price change for the same product. If we assume that demand shocks affect men and women in the same way, this regression does not need to be instrumented since the endogenous portion is differenced out. We find that for the same UPC in the same market, women are about 4.45 percentage points (pp) more elastic than men. Column (1) restricts only to UPCs purchased by both men and women in the same market, columns (2-4) include a market level fixed effect and the results correspond to our full CES model, incorporating differing product choices between men and women. Column (2) includes a market fixed effect at the module, county, half year level and instruments with Hausman instruments only. We find that women are 11.6 pp more elastic than men. Columns (3) and (4) define markets at the module, county, retail chain, half year level. Column (3) Instruments for price with the DellaVigna Gentzkow instruments and finds similar results that women are 11 pp more elastic consumers than men. Finally, column (4) includes both instruments and finds that women are 6.9 pp more elastic consumers than men.

To test for robustness, we estimate the model letting the representative consumer vary in income or age in addition to gender and present the results in Appendix X. Overall, we find that the single individuals in our sample are relatively inelastic consumers, with estimates of men's elaticities of substitution around -.7 and women's elasticities of substitution of around -.8. In a similar analysis, Faber and Fally (2022) estimate σ for all households in the Nielsen data of around -2. When we run our specification on all households rather than our panel of single individuals we find similar levels of elasticity, indicating that our singles panel is significantly more inelastic than non-single households. Taken together, the results suggest that women substitute more elastically than men.

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

Table 7: Elasticities of Substitution

	(1) County	(2) County	(3) County-Retailer	(4) County-Retailer
$1-\sigma_m$		0.3055***	0.2777***	0.2548***
		(0.0193)	(0.0219)	(0.0181)
$\sigma_m-\sigma_w$	-0.0445***	-0.1161***	-0.1097***	-0.0686***
	(0.0091)	(0.0221)	(0.0257)	(0.0209)
Observations	1,054,187	18,271,669	11,007,333	12,431,472
F-Stat		12,764	8,184	5,397
UPCXTimeXCountyXGen FE	Yes	No	No	No
ModXTimeXCountyXGen FE	No	Yes	No	No
ModXTimeXCountyXRetXGen FE	No	No	Yes	Yes
Hausman IV	No	Yes	No	Yes
DellaVigna Gentzkow IV	No	No	Yes	Yes

UPC-County level clustered standard errors in parentheses

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

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

and find that women are either more elastic than men are or are not significantly different than men in terms of elasticity. Table 8 presents elasticity of substitution results pooled to the department level. We present results defining markets at the retail chain, designated marketing area (DMA), half year level. DMAs are more aggregated geographic areas than counties but less aggregated than states. Using DMAs does not change our results in terms of magnitude but improves power by reducing the amount of sparsity in the data. We find that across almost all food products women are significantly more elastic consumers than are men, with $\sigma_m - \sigma_w \in [-0.15, -0.46]$. Among non-food retail products we find no significant differences in the elasticities of substitution between men and women and the magnitude of the coefficient for Health and Beauty products suggest the possibility that women are less elastic in that market space.[8] 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 take this as evidence that women appear to be more elastic across markets with little explicit gendering, but we cannot refute that women are less elastic in markets with significant gendering.

[8] We find that Health and Beauty and General merchandise products tend to be less elastic than other departments. The finding that Health and Beauty products are more inelastic than other types of products is consistent with the findings in Faber and Fally (2022) and our findings in Section 6.2. General Merchandise contains many products which can either be purchased or have substitutes sold at retailers not included in the Nielsen data and thus many of the purchase habits from this department cannot be considered complete. Examples include tools, automotive, household appliances, photographic supplies and stationary.

Table 8: 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.
$1-\sigma_m$	0.4347***	0.2619***	0.4946***	0.1788***	0.1447	0.1667*	0.0001	0.2238***	-0.5720	0.4893***
	(0.0867)	(0.0315)	(0.0717)	(0.0357)	(0.1310)	(0.0953)	(0.0974)	(0.0672)	(0.5679)	(0.1156)
$\sigma_m-\sigma_w$	0.1037	-0.2682***	-0.4578***	-0.2709***	-0.1488	-0.2688**	-0.2145***	0.0161	0.7583	-0.0384
	(0.0974)	(0.0369)	(0.0907)	(0.0456)	(0.1650)	(0.1204)	(0.0798)	(0.0744)	(0.6057)	(0.1251)
Observations	718302	5335802	1314605	1680282	401229	467441	1084136	1144523	63143	265534
Adjusted R^2	-0.256	-0.287	-0.280	-0.192	-0.278	-0.212	-46.792	-0.275	-0.337	-0.333
ModuleXTimeXDMAXRetXGender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DMA IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Retailer IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

UPC-DMA level clustered standard errors in parentheses

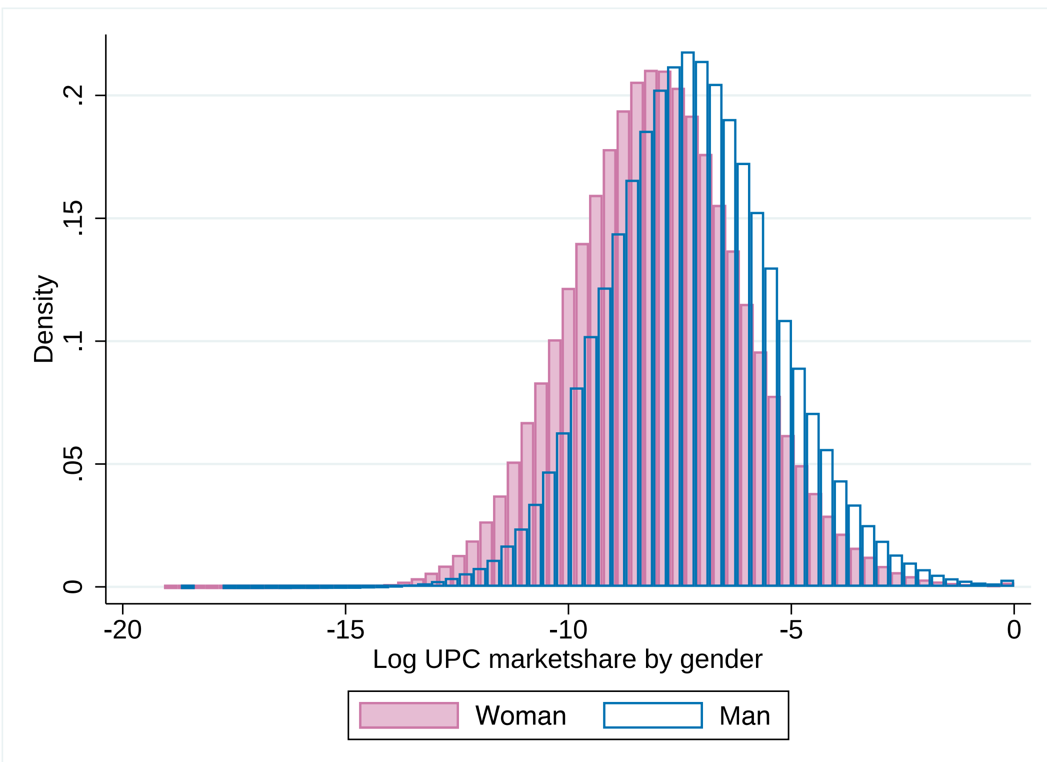
* $p < .10$, ** $p < .05$, *** $p < .01$

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

So far we have estimated the elasticity of substitution, $\sigma_t(g)$, while actual price elasticities of demand are given by Equation (4) and are a function of the elasticity of substitution and market shares. This means that price elasticities of demand will range from $\sigma_t(g)$, under monopolistic competition where market shares are approximately 0, to 1, under monopoly where the market share of the single good is 1. Because we have found that women generally substitute more elastically than men, the only remaining channel for them to be less elastic consumers is through market competition being significantly less competitive in women's markets than in men's. From Figure 2 in the descriptive analysis, we know that women buy more unique products than do men by about 27%. This suggests that women's markets are more diverse than men's and are also likely more competitive.

In Figure 6 we show the histogram of log market shares for the men and women in our sample.

Figure 6: Market Competition by Gender



Note: This figure presents histograms of log market share of products for men and women separately.

The entire distribution of market shares for women is shifted to the left, indicating that their markets are more competitive. Further, market shares in our data are very small, on the order of 0.05% for the median UPC. This means that, in our setting, elasticities of substitution are close approximations of price elasticities of demand. On average, we can conclude that women are more price elastic consumers than men are. This finding is consistent with women being less likely to be the consumer group to pay higher prices under price discrimination. Abstracting from the role of multiproduct firms, this finding is also consistent with products having relatively more women as their consumer base being associated with higher marginal costs.

6 Markups and Marginal Costs

Our constant elasticity of substitution demand model speaks to differences in demand elasticities between men and women across their retail consumption baskets. To do this, we leveraged individual level purchase data aggregated to the by-gender market level. This method allowed us to capture consumer level average demand differences across a broad scope of products, but at the cost of model complexity in terms of flexible substitution patterns and market structure. Additionally, individual level purchase data faces sparsity issues in markets where purchases are relatively infrequent, like Health and Beauty products. To structurally decompose prices into markups and marginal costs, we allow for significantly more model complexity at the cost of narrowing our focus to less markets. To do this, we use weekly store level data that does not face the same sparsity issue that the aggregated individual level data does. This lack of sparsity comes at the cost of no longer being able to attribute purchases to a specific gender. To overcome this, 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.

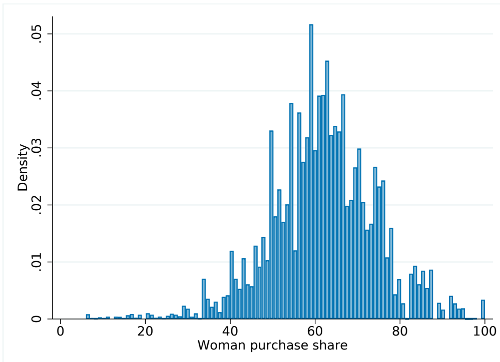
We model demand in five product markets: yogurt, protein bars, deodorant, disposable razors, and shampoo. We focus on these markets because they have a high level of dispersion of $\hat{w}_j$ across their product spaces. Specifically, we select yogurt because prices and consumer behavior look similar to descriptive results of the entire grocery consumption basket. We think of the yogurt market as representative of grocery food markets generally. While yogurt seems to have significant heterogeneity in preferences across gender, marketing and advertising is less explicitly gendered than the other markets we focus on. The four other markets were selected because they contained a large amount of explicitly gendered products while also matching the price and consumption trends that we see in descriptive section. Figure 7 plots histograms of woman purchase share in each of the selected markets. Yogurt follows a similar normal distribution to what we see across the data at large, but the other four markets are either bimodal (deodorant and disposable razors) or somewhat uniform (protein bars and shampoo). In the descriptive analysis, we found evidence that products explicitly gendered to men or women had higher prices than ungendered products in the same market (see Figure 5). Suggesting the potential for price discrimination on gendered products as a whole, rather than just women's products. Because of this we select two markets that have very few ungendered products, deodorant and razors, and two markets that have a relatively high amount of both gendered and ungendered products, protein bars and shampoo. Deodorant and razors allow us to test for price discrimination on women's products versus men's products while shampoo and protein bars are additionally able to test for price discrimination relative to ungendered products. Finally, three of the markets we analyze, deodorant, shampoo and razors are discussed in concurrent work on gender price discrimination by Moshary, Tuchman, and Bhatia (2021). Appendix xx contains a replication of the descriptive analysis for each market chosen.

6.1 Differentiated Products Demand Model and Estimation

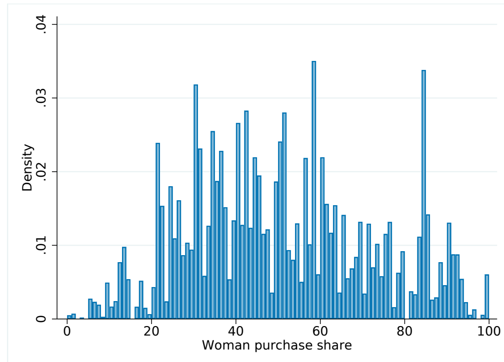
We follow the standard differentiated products market demand model presented in Berry, Levinsohn, and Pakes (1995) (BLP). Our main departure is that instead of typical product characteristics, we include our measure of the woman purchase share of a product, $\hat{w}_j$ and allow for heterogeneity in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

$$u_{ijt} = \alpha p_{jt} + \beta_{\mathbf{i}} \mathbf{x_j} + \xi_{jt} + \epsilon_{ijt}, \tag{5}$$

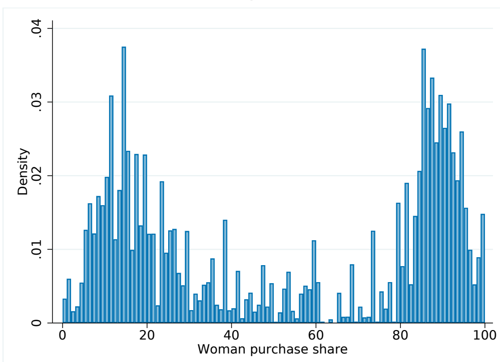
Figure 7: Woman purchase share distribution



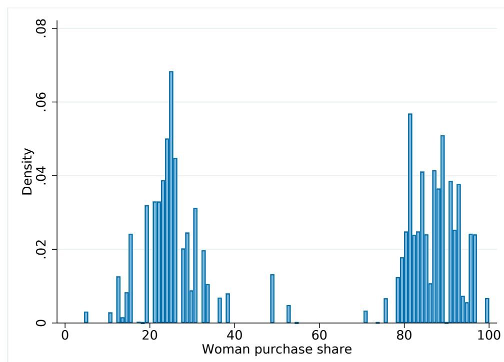
Yogurt



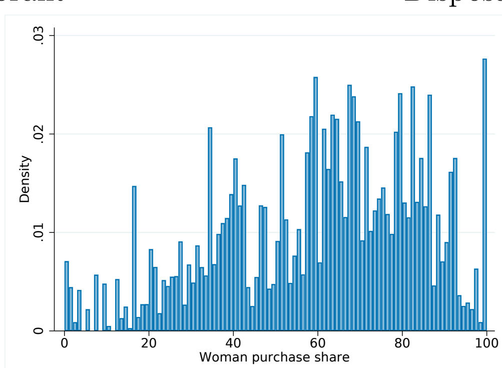
Protein Bars



Deodorant



Disposable Razors



Shampoo

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

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

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

Heterogeneity in preferences for product gender will generate more reasonable substitution patterns than our CES demand model does. Under CES demand, price increases on a woman's razor will lead to equal levels of substitution from the women's razor into other women's razors and men's razors. Now, the random coefficient on women purchase will generate substitution patterns that have men's razors substituting to men's razors and women's razors substituting to women's razors. Allowing for heterogeneity in preferences for the outside option is important as the value of the outside option is likely different 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. 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.

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

$$s_{jt} = \int \frac{\exp(\alpha p_{jt} + \beta_i \mathbf{x_j} + \xi_{jt})}{1 + \sum_k (\exp(\alpha p_{kt} + \beta_i \mathbf{x_k} + \xi_{kt}))} d\beta_i \tag{6}$$

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

The methods used for estimation of this class of models is standard and well documented in the industrial organization literature. The indirect utility that an individual receives from consuming product j can be written as a linear component and a non-linear component:

$$u_{ijt} = \delta_{jt} + \Sigma \cdot v_i + \epsilon_{ijt},$$

where $\delta_{jt} = \alpha p_{jt} + \beta \mathbf{x_j} + \xi_{jt}$ is the fixed component of utility from product j. Given a guess of the variances of the taste parameters for the woman purchase share and outside option, $\hat{\Sigma}$, we can construct estimates of the market shares:

$$\hat{s}_{jt} = \int \frac{\exp(\delta_{jt} + \hat{\Sigma} v_i)}{1 + \sum_k (\exp(\delta_{kt} + \hat{\Sigma} v_i))} dv_i \tag{7}$$

Using our estimated market shares and observed market shares, we iteratively solve for δ_{jt} using the contraction mapping:

$$\delta'_{jt} = \delta_{jt} + \log(s_{jt}) - \log(\hat{s}_{jt}).$$

From the converged estimates of δ_{jt}, we can recover the price parameter, α, and the mean taste parameters, β from a regression of δ_{jt} on prices. In practice, we include product-retail chain fixed effects as well as time fixed effects in our specification which allows the mean taste parameters to vary at the product-retail chain level. This regression also provides estimates of ξ_{jt}, which we use to estimate the variance of taste parameters, Σ with the following moment condition:

$$\mathbb{E}[\xi_{jt} Z_{jt}] = 0,$$

where ξ_{jt} and Z_{jt} are the residuals of the unobserved product characteristics and demand side instruments after all fixed effects have been partialled out.

We instrument for prices with the same instruments we use for our aggregate elasticity 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 where differences are summed over products that are true rivals, that is, products that are owned by other firms and one for products produced by the same firm. The instrument captures "closeness" in the product space in terms of woman purchase share and is rooted in the idea that substitution likely happens among products that are similar in gender.

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

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

We construct an ownership matrix, Ω, that maps each product in our data to a common owner so that element jk is 1 if product j and product k are owned by the same firm and 0 otherwise.[9] 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:

$$\frac{p^* - mc}{p^*} = -(\Omega J)^{-1} \frac{s(p^*)}{p^*} \tag{8}$$

Because price is observed, identified markups also identify marginal costs. The estimated parameters are presented in Table A.10.

[9] We construct the ownership matrix through manual search, Capital IQ, and newspaper articles.

6.2 Differentiated Products Demand Model Results

We begin by plotting prices by decile of woman purchase share for each market. These prices are observed in the data and we normalize them relative to the size of the good.[10] We plot the median price of a product within a woman purchase share decile along with the interquartile range. Figure 8 presents the data. Generally, we find that prices are increasing in woman purchase share. The average men's razor in our data priced at about $1.2, while the average women's razor is priced at about $1.5. We find that women's yogurt is generally priced about 5 cents higher per ounce than ungendered yogurt, women's protein bars are priced about 5 cents higher per ounce as well. Women's deodorant is priced about 20 cents more per ounce than men's.[11] Finally, women's shampoo can cost 20-25 cents more per ounce than men's or ungendered shampoo.

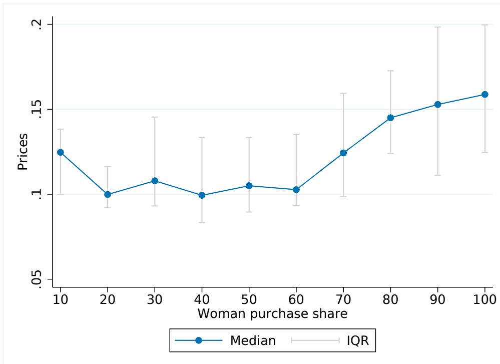
We plot median estimated own price elasticities and interquartile range by woman purchase share in Figure 9. Own price demand elasticities in our model are given by $\varepsilon_{jt} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}}$. Generally we find that women's products are either more elastic or no differently elastic than men's or ungendered products. Most of the markets exhibit a downward trend in elasticities along woman purchase share. These findings are generally inconsistent with a price discrimination story, where we would expect to find that women's products or gendered products in general have less elastic demand. Instead we find that women's products are much more elastic and men's products are either slightly more elastic (yogurt and shampoo) or are no differently elastic than ungendered products (deodorant and razors). Our results are generally consistent with our CES demand estimation and suggest that women as a consumer base seem to be generally more elastic consumers than men across both gendered and ungendered products.

Firm's pricing decisions and markups are made based on the own price elasticity of the product, the cross elasticities with other products owned by the firm and marginal cost. Even

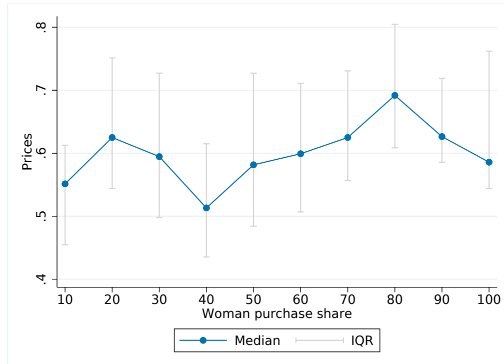
[10] Yogurt, protein bar, deodorant and shampoo prices are all presented as price per ounce while razors are presented as price per count.

[11] An interesting finding about the price discrepancy in deodorant is that it is mostly generated by women's deodorant in slightly smaller amounts but having the same list price. The outlier in the 70th decile for deodorant is primarily driven by the brand Tom's of Maine, it is a natural health product that is generally priced higher than other deodorants.

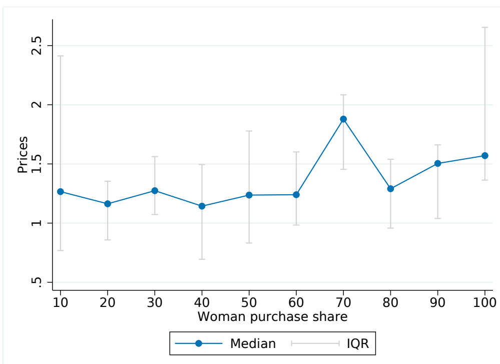
Figure 8: Observed Prices



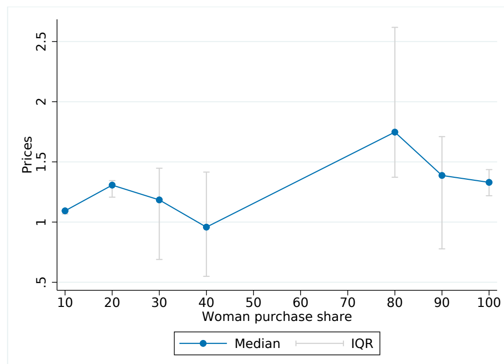
Yogurt



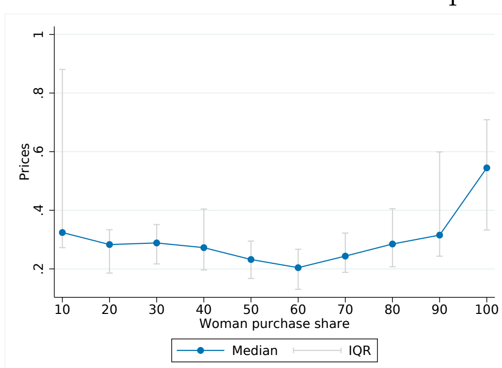
Protein Bars



Deodorant



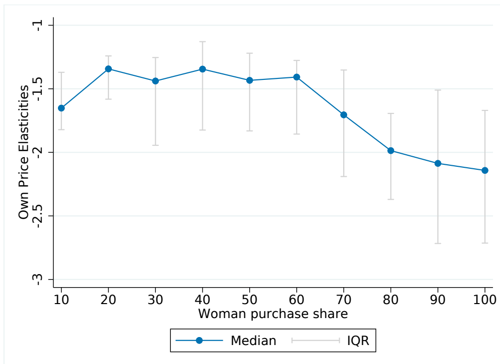
Disposable Razors



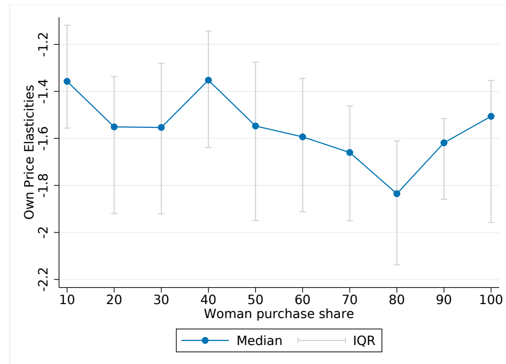
Shampoo

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

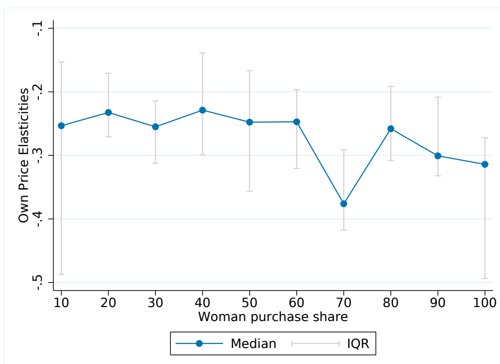
Figure 9: Own Price Elasticities



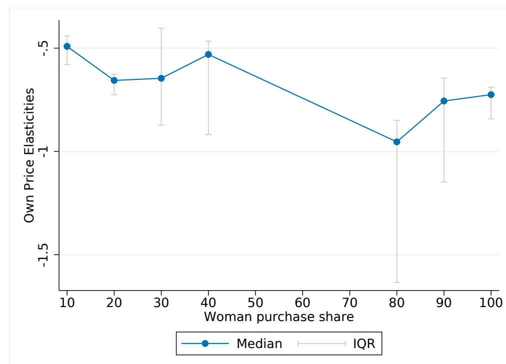
Yogurt



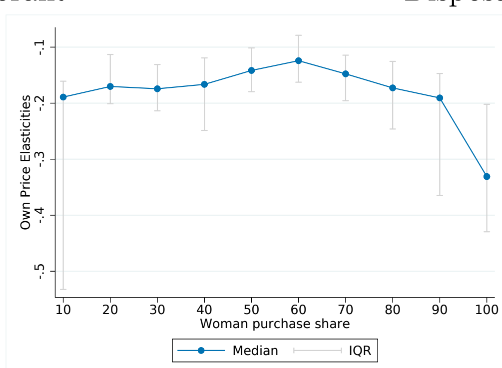
Protein Bars



Deodorant



Disposable Razors



Shampoo

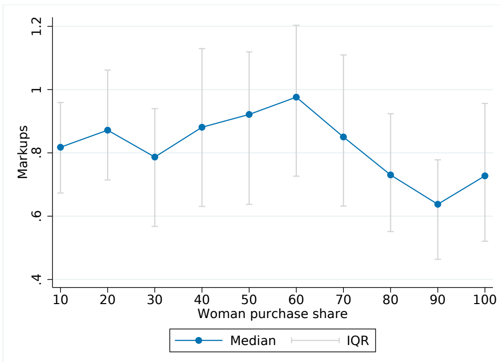
Note: This figure presents median estimated own-price elasticities of products by decile of woman purchase share. Grey bars represent the inter quartile range of the estimates.

though women's products have more elastic demand, they could face higher markups through substitution patterns or the competitive structure of the market. Multiproduct firms have incentives to price higher because some of the lost demand is funneled into other products that they own. Women's products could still face higher markups if they are more likely to be owned by large multiproduct firms and consumers strongly substitute to other products owned by the firm.

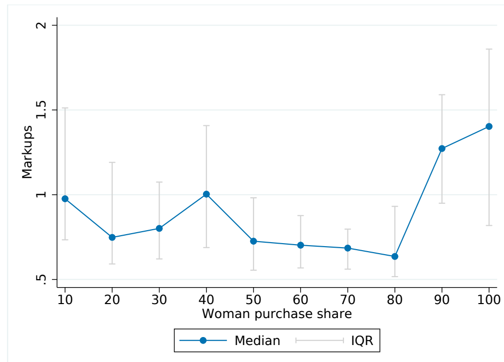
We plot median estimated markups along with interquartile range by woman purchase share decile in Figure 10. We find that markups are generally decreasing in woman purchase share in all product markets except for protein bars. Protein bars are the only product market where we find that women pay significantly higher markups than men. Looking at the elasticities in Figure 9, women's own price elasticities are slightly more elastic than men's. That means that this result is generated by substitution patterns and the competitive structure of market. When we look into this, we find that this result is driven entirely by substitution of Luna bars into Clif bars, Luna is Clif's woman oriented protein bar brand and both Clif command's a significantly large share of the market. Because of this, we do not take this as evidence of price discrimination but rather the result of firm's market power. Overall, the results suggest that women's products are associated with lower markups.

From the markups we directly calculate marginal costs and present them in Figure 11. Among yogurt, razors, and shampoo we find that marginal costs are increasing in woman purchase share, that is the products that women sort into are more expensive to produce. We find weakly higher marginal costs for women's deodorant but this is dwarfed by higher marginal costs in ungendered deodorants. Ungendered deodorants make up a small share of the market, and tend to be either natural products, like Tom's of Maine, or clinical strength products, like Certain Dri, that seem reasonable to have higher marginal costs. Again, the only market where women are not sorting into higher marginal cost products is protein bars, but this is primarily driven by competitive structure and does not seem to be consistent with the narrative across the consumption basket.

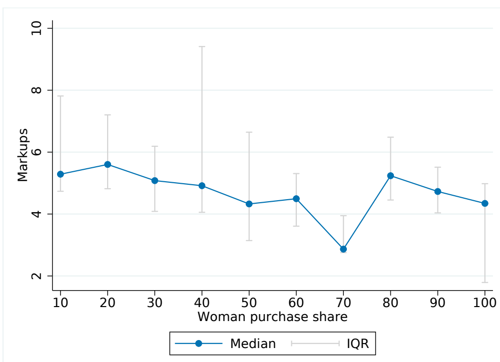
Figure 10: Markups



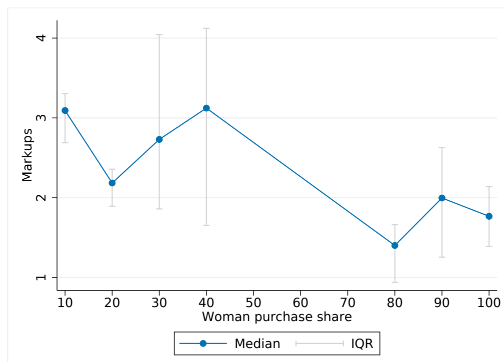
Yogurt



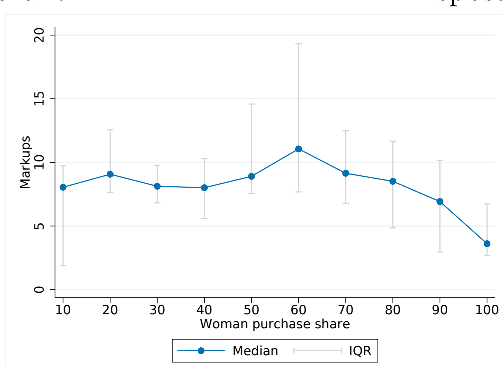
Protein Bars



Deodorant



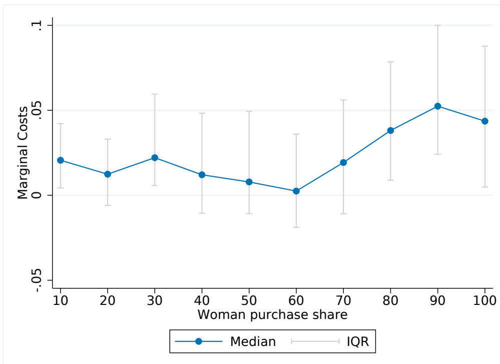
Disposable Razors



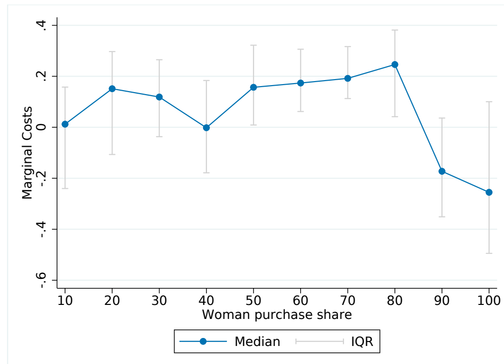
Shampoo

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

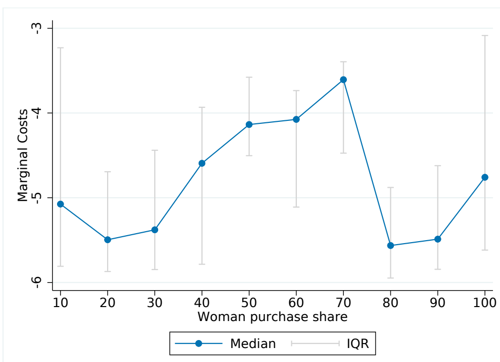
Figure 11: Marginal Costs



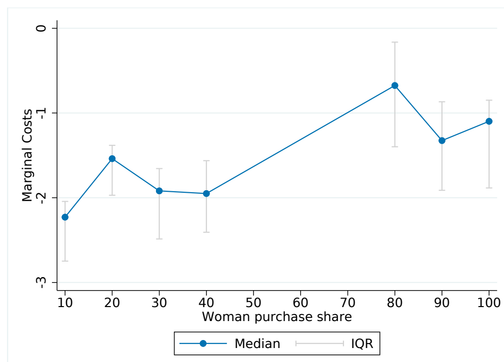
Yogurt



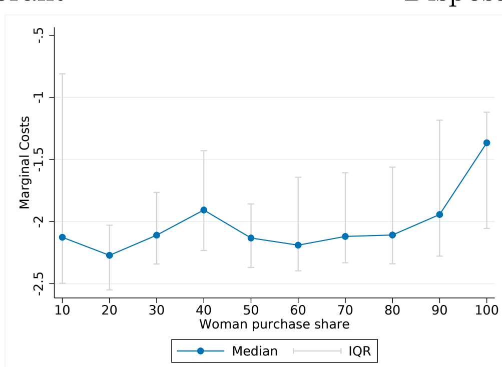
Protein Bars



Deodorant



Disposable Razors



Shampoo

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

7 Conclusion

Our paper documents retail spending differences between men and women and decomposes observed spending differences into demand and supply side mechanisms. Our work was motivated by the hypothesized “Pink Tax”, the idea that women’s products are priced higher due to price discrimination. Three economic mechanisms factor into firms’ pricing decisions: price elasticity of demand, competitive structure, and marginal costs. We document price premiums paid by women but when we decompose these price premiums into their economic mechanisms we find they are primarily driven by differences in marginal costs. Our paper suggests that public discourse on the pink tax, which often cites cherry picked examples of price differences for gendered products, fails to capture differences in actual consumption choices between men and women that result from differential sorting. Our work also suggests that current legislation in New York City and proposed legislation in California, which place bans on price differences for products that differ only in gender, are likely to be ineffective as the majority of price disparities between men and women can be explained by sorting into products that likely differ in more than just gender.

We allow for a broad definition of the pink tax, considering any and all consumption differences that may lead to women’s consumption bundles being more expensive or women paying more in markups. We show that women do buy higher priced products, paying an average price premium of 4% relative to similar goods that are bought buy men. We decompose this observed price premium into differential sorting between men and women into otherwise ungendered products and price premiums on explicitly gendered products. We document the existence of gendered products, products that are almost exclusively purchased by one gender, and show that there are significantly more women’s products than men’s products. We document that price disparities for gendered products are about 15%. However, purchases of gendered products make up a small fraction of the overall consumption bundle and we can attribute the majority of the 4% price premium paid by women to sorting in ungendered products. These descriptive findings provide a more nuanced view of the pink tax, or more broadly, gender differences in consumption habits and product markets.

To understand if observed price premiums are driven by markups (price discrimination) or preferences for goods with higher marginals costs, we formally model demand in two ways, incorporating methods from the international trade and industrial organization literature. Our first model allows us to attribute demand elasticities to the gender of the consumer and speaks to the entire consumption bundle. Our second model allows us to identify elasticities for explicitly gendered products that are not well captures in the first model and more carefully models demand by allowing for more flexible substitution patterns and competitive structure. Our aggregate demand model finds that women are generally more elastic consumers than men are, this finding suggests that women are sorting into products that have higher marginal costs, at least among the most frequently bought ungendered products. Our differentiated products demand model allows for heterogeneity in preferences for the gender of a product and finds that women's products are either more elastic or not significantly differently elastic than men's products or ungendered products. Taken together, our demand estimations show that price premiums paid by women are generated by women sorting into higher marginal cost goods.

The policy implications of our findings are nuanced. We can confidently state that existing and proposed legislation that bans pricing differences based on gender is likely to have little to no effect on the observed pink tax paid by women, a finding supported by related work by Moshary, Tuchman, and Bhatia (2021). It is harder to prescribe optimal policy or welfare improving policy when the underlying mechanism is differences in preferences for quality or marginal costs. Classical economic theory that assumes rational consumers and that revealed preferences are utility maximizing would suggest that policy that interferes with markets here would be welfare decreasing. However, growing literature on biased beliefs in retail consumption suggest that consumer preferences do not necessarily map to utility, its possible that men and women may be biased to different degrees and this could affect optimal policy.(B. J. Bronnenberg et al. 2015; Allcott, Lockwood, and Taubinsky 2019) Ultimately, our work describes gender differences in consumption behavior and the product space but does not address how these differences come to be.

Preference formation has long been a topic of interest in economics, since George J Stigler and Becker (1977) first put forth their theory of accumulated consumption capital. More recently this theory has been applied to study generational differences in preferences (B. Bronnenberg, Dubé, and Joo (2022)). Given that men and women are socialized to consume and value goods in very different ways, one would expect that a woman's accumulated consumption capital would be very different from a man's. The relevant policy, welfare and research questions then become how are preferences shaped, can preferences be changed, and can changing preferences increase utility? Finally, we estimate a partial equilibrium model where we take the set of products produced as given. A natural question to arise is how do systematically different preferences between men and women shape product entry and exit along with innovation (and vice versa)?

References

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

– (2007). "Life-Cycle Prices and Production". In: *American Economic Review* 97.5, pp. 1533–1559.

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.

Atkeson, Andrew and Ariel Burstein (2008). "Pricing-to-Market, Trade Costs, and International Relative Prices". In: *The American Economic Review* 98.5, pp. 1998–2031. ISSN: 00028282. URL: `http://www.jstor.org/stable/29730160`.

Ayres, Ian and Peter Siegelman (1995). "Race and gender discrimination in bargaining for a new car". In: *The American Economic Review*, pp. 304–321.

Barahona, Nano et al. (2020). "Equilibrium effects of food labeling policies". In: *Available at SSRN 3698473*.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott (Mar. 2010). "Multiple-Product Firms and Product Switching". In: *American Economic Review* 100.1, pp. 70–97. DOI: `10.1257/aer.100.1.70`. URL: `https://www.aeaweb.org/articles?id=10.1257/aer.100.1.70`.

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. DOI: `10.1257/jel.20160995`. URL: `https://www.aeaweb.org/articles?id=10.1257/jel.20160995`.

Broda, Christian and David E. Weinstein (June 2010). "Product Creation and Destruction: Evidence and Price Implications". In: *American Economic Review* 100.3, pp. 691–723. DOI: `10.1257/aer.100.3.691`. URL: `https://www.aeaweb.org/articles?id=10.1257/aer.100.3.691`.

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.

Bronnenberg, Bart, Jean-Pierre Dubé, and Joonhwi Joo (2022). "Millennials and the Takeoff of Craft Brands: Preference Formation in the US Beer Industry". In: *Marketing Science*.

Castillo, Marco et al. (2013). "Gender differences in bargaining outcomes: A field experiment on discrimination". In: *Journal of Public Economics* 99, pp. 35–48.

Conlon, Christopher and Jeff Gortmaker (2020). "Best practices for differentiated products demand estimation with pyblp". In: *The RAND Journal of Economics* 51.4, pp. 1108–1161.

DellaVigna, Stefano and Matthew Gentzkow (2019). "Uniform pricing in us retail chains". In: *The Quarterly Journal of Economics* 134.4, pp. 2011–2084.

Dubé, Jean-Pierre, Günter J Hitsch, and Peter E Rossi (2009). "Do switching costs make markets less competitive?" In: *Journal of Marketing research* 46.4, pp. 435–445.

Duesterhaus, Megan et al. (2011). "The cost of doing femininity: Gendered disparities in pricing of personal care products and services". In: *Gender Issues* 28.4, pp. 175–191.

Faber, Benjamin and Thibault Fally (2019). *Firm heterogeneity in consumption baskets: Evidence from home and store scanner data*. Tech. rep. National Bureau of Economic Research.

– (2022). "Firm heterogeneity in consumption baskets: Evidence from home and store scanner data". In: *The Review of Economic Studies* 89.3, pp. 1420–1459.

Fitzpatrick, Anne (2017). "Shopping While Female: Who Pays Higher Prices and Why?" In: *American Economic Review* 107.5, pp. 146–49.

Flagg, Lee A et al. (2014). "The influence of gender, age, education and household size on meal preparation and food shopping responsibilities". In: *Public health nutrition* 17.9, pp. 2061–2070.

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.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein (Mar. 2016). " Quantifying the Sources of Firm Heterogeneity *". In: *The Quarterly Journal of Economics* 131.3, pp. 1291–1364. ISSN: 0033-5533. DOI: 10.1093/qje/qjw012. eprint: https://academic.oup.com/qje/article-pdf/131/3/1291/30636473/qjw012.pdf. URL: https://doi.org/10.1093/qje/qjw012.

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.

Kahneman, Daniel, Jack L Knetsch, and Richard Thaler (1986). “Fairness as a constraint on profit seeking: Entitlements in the market”. In: *The American economic review*, pp. 728–741.

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 Bhatia (2021). “Investigating the Pink Tax: Evidence Against a Systematic Price Premium for Women in CPG”. In: *Available at SSRN 3882214*.

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.

Rousille, Nina (2021). “The Central Role of the Ask Gap in Gender Pay Inequality”. In: *Working paper*.

Shapiro, Carl (1982). “Consumer information, product quality, and seller reputation”. In: *The Bell Journal of Economics*, pp. 20–35.

Spence, Michael (1976). “Product differentiation and welfare”. In: *The American Economic Review* 66.2, pp. 407–414.

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

Stigler, George J and Gary S Becker (1977). “De gustibus non est disputandum”. In: *The american economic review* 67.2, pp. 76–90.

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.

Wollmann, Thomas G (2018). “Trucks without bailouts: Equilibrium product characteristics for commercial vehicles”. In: *American Economic Review* 108.6, pp. 1364–1406.

Appendix A Additional figures and tables

Table A.1: Nielsen panelist behavior per month

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

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

Table A.2: 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.3: Price paid per good unit by department

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

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

Table A.4: Demographics of CE PUMD single-member households

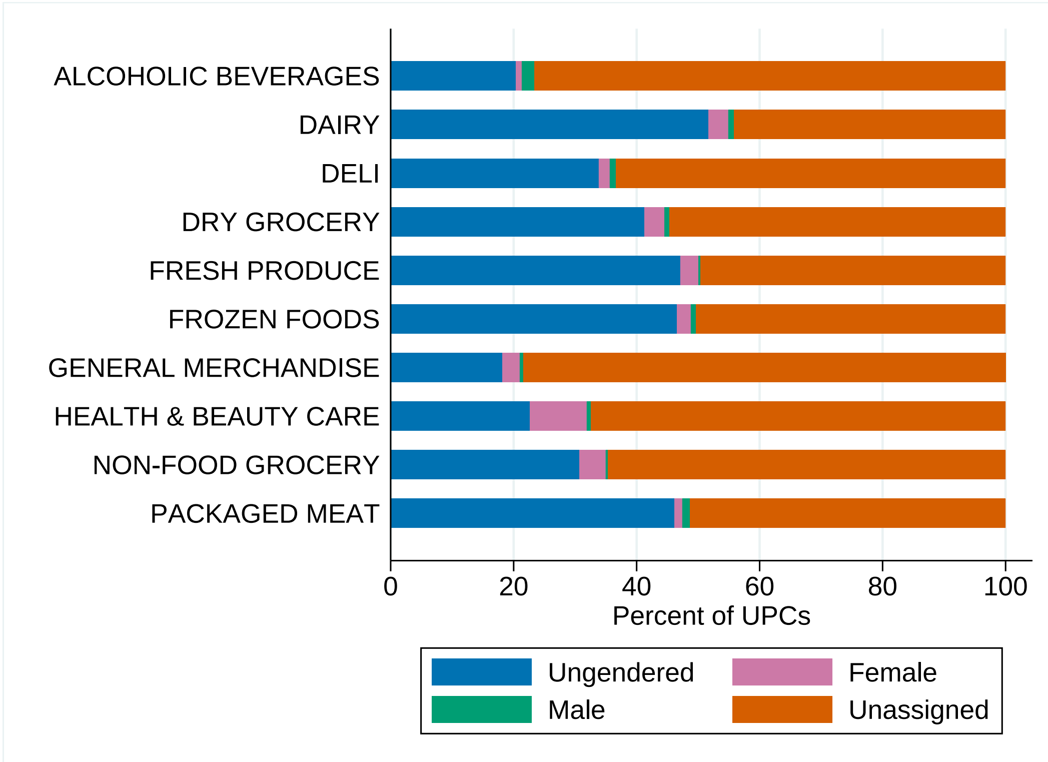
–	Total	Women	Men	Difference
Income	30530	26950	34665	-7715.418**
	(42896.3)	(36923.05)	(48568.25)	(335.0263)
Age	54.72	58.93	49.86	9.071**
	(20.2861)	(20.2295)	(19.2376)	(.1516)
High school	0.482	0.478	0.486	-.008*
	(.4997)	(.4995)	(.4998)	(.0038)
College	0.284	0.278	0.291	-.013**
	(.4508)	(.448)	(.4541)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

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

Table A.5: Most Popular Brands by Product Gender - Deodorant

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

Figure A.1: 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.6: Unit prices in same product module by UPC gender

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
female=1	0.0322***	0.0182***	0.0286***
	(0.0019)	(0.0032)	(0.0017)
gendered2=1	0.0084***	0.0817***	0.0066***
	(0.0022)	(0.0038)	(0.0024)
female=1 × gendered2=1	0.0848***	0.1002***	0.0648***
	(0.0024)	(0.0042)	(0.0026)
Observations	131501221	9299678	120478978
Adjusted R^2	0.884	0.844	0.888
ModXUnitXRetXLocXMonth FE	Yes	Yes	Yes
Demographic FE	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

Figure A.2: Consumption basket composition as share of purchases, 75-25 Cutoff

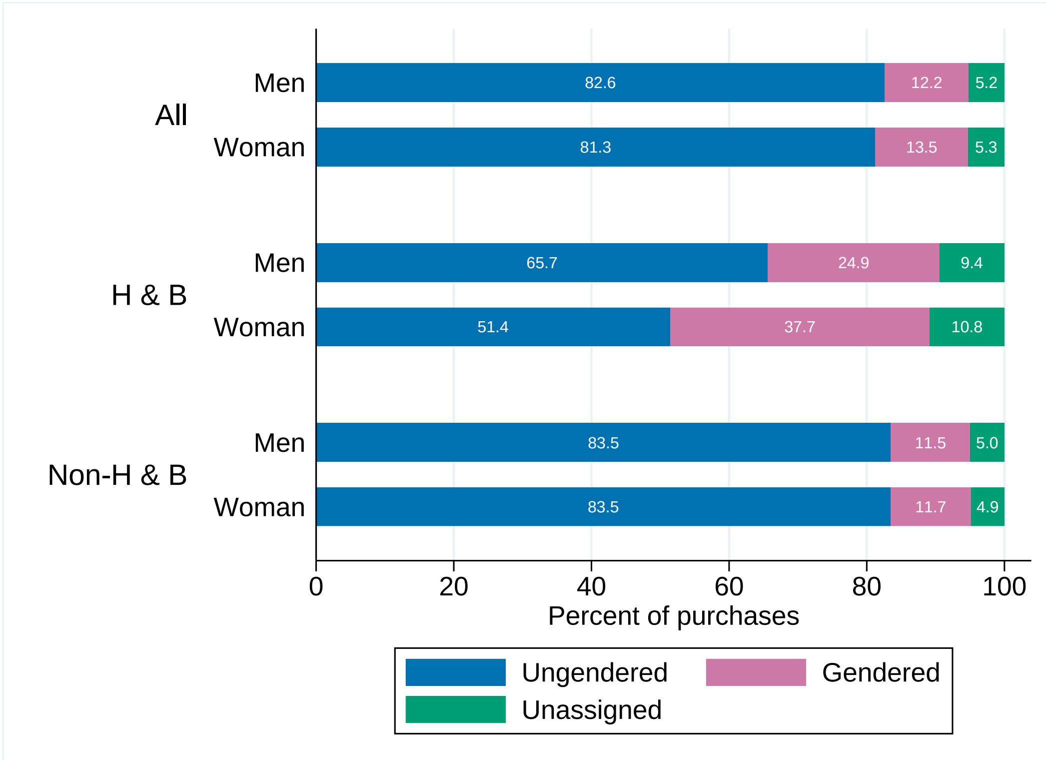


Table A.7: OLS Elasticities (No Instruments)

	(1) COunty-Half Year	(2) County-Retailer-Half Year
$1 - \sigma_m$	0.6886***	0.7784***
	(0.0170)	(0.0075)
$\sigma_m - \sigma_w$	-0.0181***	-0.0073*
	(0.0065)	(0.0044)
Observations	17,010,404	14,939,386
Adjusted R^2	0.016	0.000
ModuleXTimeXCountyXRetXGender FE	Yes	Yes
County IV	No	No
Retailer IV	No	No

UPC-County level clustered standard errors in parentheses

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

Table A.8: First Stage

	(1) Hausman	(2) Dellavigna-Gentzkow
Hausman	0.3280***	
	(0.0044)	
Dellavigna-Gentzkow		0.2215***
		(0.0036)
Observations	16,351,076	11,018,742
Adjusted R^2	0.008	0.006
ModuleXTimeXCountyXRetXGender FE	Yes	Yes
County IV	Yes	No
Retailer IV	No	Yes

UPC-County level clustered standard errors in parentheses

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

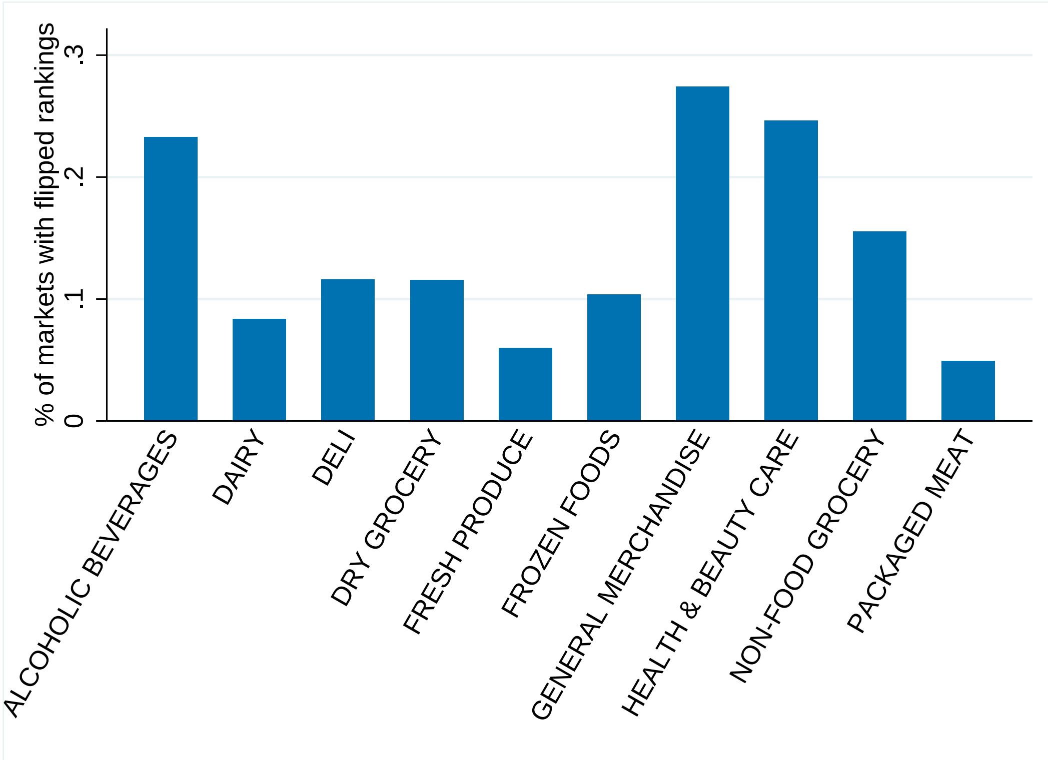
Table A.9: Reduced Form Results

	(1) Hausman	(2) Dellavigna Gentzkow
Hausman	0.0650***	
	(0.0077)	
WomanXHausman	0.0102	
	(0.0076)	
Dellavigna-Gentzkow		0.0568***
		(0.0052)
WomanXDellavigna-Gentzkow		-0.0198***
		(0.0060)
Observations	16,336,260	11,007,333
Adjusted R^2	-0.022	-0.052
ModuleXTimeXCountyXRetXGender FE	Yes	Yes
County IV	Yes	No
Retailer IV	No	Yes

UPC-County level clustered standard errors in parentheses

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

Figure A.3: Markets with Flipped Product Rankings between Men and Women



This figure displays the percentage of markets with flipped product rankings across departments. We define flipped product rankings as markets where items that are in the top 25% of products by market share for women are in the bottom 25% of products for men and vice versa.

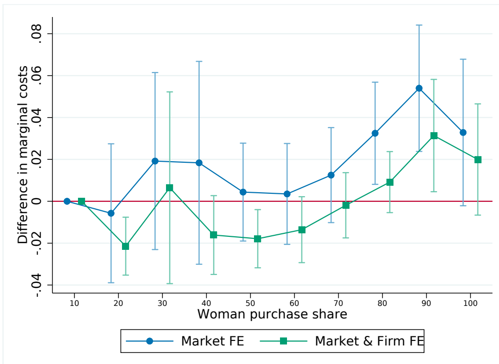
Table A.10: Differentiated Products Model Results

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

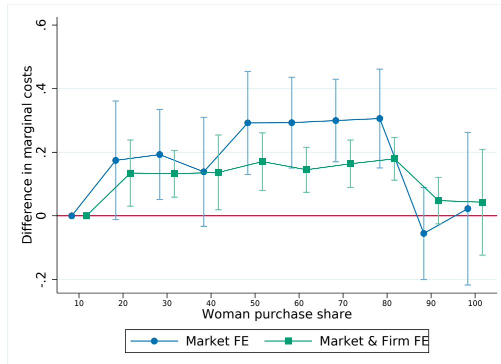
Market level clustered standard errors in parentheses

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

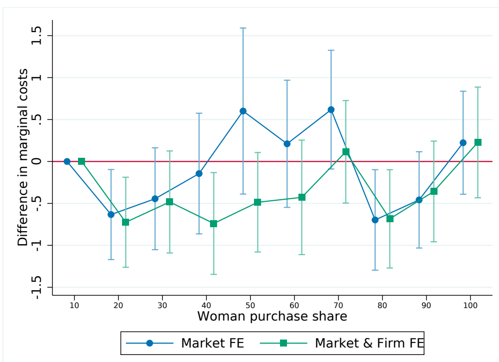
Figure A.4: Marginal Costs with Market and Firm FE



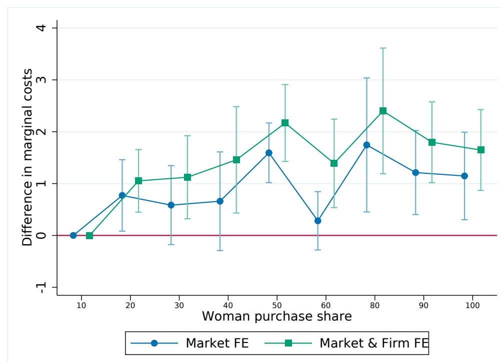
Yogurt



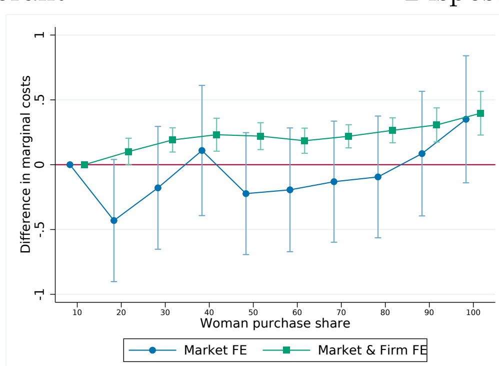
Protein Bars



Deodorant



Disposable Razors



Shampoo

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

Appendix B Markups under CES Demand and Oligopolistic Competition

Our aggregated demand analysis rules out gender based price discrimination as the primary driver of our observed 5% price premium on products bought by women. Instead, it suggests that women sort into products that are higher marginal cost. Our demand analysis thus far has been focused on assessing differences in demand behavior between men and women, which required focusing on single individuals who are an incomplete portion of the market. Firm's pricing decisions will be based on the average price elasticity that they face, a large portion of which will be household purchases made by families. To close our study of aggregate demand for largely ungendered products, we compute price elasticities for all households in the Nielsen data and study how the elasticities of the products that women sort into compare to the products men sort into. Attributing household purchases to a specific gender is difficult; women often take the role of primary shopper in the household (Flagg et al. 2014) but men increasingly play a role in role in shopping and all decisions are likely some aggregation of the preferences of the shopper, their partner and (or) their children. From our analysis on single individuals we have ruled out that the price premiums paid by women are from systemic price discrimination from women being more inelastic consumers. Still, the demand behavior of non-single households along with differences in consumption basket composition between men and women could lead to women spending more of their income in markups than men. We calculate price elasticities of demand using our model for the entire population of consumers in the Nielsen data, and then compare the difference in price elasticity of demand for the average purchase made by women to the average purchase made by men.

The results are presented in Table B.1. Column (1) aggregates across all product departments, while columns (2) and (3) separate Health and Beauty products from non-Health and Beauty Products respectively. Though we find that women are more elastic than are men for the same type of good, overall the goods that comprise a woman's consumption basket are slightly more inelastic than men's. Our results are significant but economically small. The average product in our sample has a price elasticity of about -2, the average difference in markups paid by women as opposed to men is then less than a half of a percent. This number is significantly smaller than the overall difference in spending between between men and women of 6% and the similar good price premium of 4%. Therefore, we can confidently say that the main driver of the observed price premium paid by women is due to marginal costs.

Table B.1: Purchase-weighted differences in price elasticities between men and women

	(1) All	(2) Health & Beauty	(3) Non-Health & Beauty
$\Delta\varepsilon$	0.0137***	0.0175***	0.0064**
	(0.0025)	(0.0031)	(0.0026)
Observations	146718945	8907946	137810996
Adjusted R^2	0.006	0.003	0.007
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Income FE	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Race FE	Yes	Yes	Yes

Individual level clustered standard errors in parentheses

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

script here

Appendix C Brand Loyalty and Forward Looking Firms

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

$$u_{ijt} = \alpha p_{jt} + \beta_{\mathbf{i}} \mathbf{x_j} + \xi_{jt} + \gamma \mathbb{1}(state_{it} \neq j) + \epsilon_{ijt}, \tag{9}$$

where γ is a negative number that represents the utility cost of switching to a product not consumed in the previous period. The individual's probability of choosing product j is given by:

$$s_{ijt} = \frac{\exp(\alpha p_{jt} + \beta_i \mathbf{x_j} + \xi_{jt} + \gamma \mathbb{1}(state_{it} \neq j))}{1 + \sum_k (\exp(\alpha p_{kt} + \beta_i \mathbf{x_k} + \xi_{kt} + \gamma \mathbb{1}(state_{it} \neq k)))} \tag{10}$$

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

$$s_{jt} = \int \int \frac{\exp(\alpha p_{jt} + \beta_i \mathbf{x_j} + \xi_{jt} + \gamma \mathbb{1}(state_{it} \neq j))}{1 + \sum_k (\exp(\alpha p_{kt} + \beta_i \mathbf{x_k} + \xi_{kt} + \gamma \mathbb{1}(state_{it} \neq k)))} d\beta_i df(i), \tag{11}$$

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

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

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

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

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

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

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