Who Pays for Protectionism?

The Welfare and Substitution Effects of Tariffs

Torsten Søchting Jaccard

University of Toronto

April 22, 2022

CLICK HERE FOR THE LATEST VERSION

Abstract

This paper studies the distributional costs to US consumers of country-specific tariffs. By linking detailed household purchase records with a barcode-specific country-of-origin, I estimate a demand model with both detailed consumer heterogeneity and rich import substitution patterns in the face of country-specific tariff changes. Simulations using this model show that tariffs placed on low-income countries are regressive and anti-rural whereas tariffs on high-income countries are progressive and anti-urban. I provide novel evidence that the urban/rural disparity in exposure to tariff policy is driven by the extent to which retail market characteristics differ across urban and rural counties in the US. When modeling import substitution, I combine descriptive text on the packaging of each barcode with unsupervised clustering algorithms to place barcodes into market segments of observable similarity. In general, I estimate lower tariff costs when compared to a model in which varieties are segmented based on their production location, as is common in the trade literature. These findings caution against the practice of estimating consumption gains from trade in the absence of (1) detailed variety attribute data and (2) information regarding the domestic alternatives available to consumers.

This paper forms a chapter of my thesis, and I am indebted to my supervisor - Daniel Trefler - and committee members Victor Aguirregabiria and Peter Morrow for their support and guidance on this project. I have benefitted from discussions with Nina Pavcnik, Thomas Sampson, Kevin Lim, Pamela Medina-Quispe, and my peers at the University of Toronto. I thank Sebastian Stumpner for generously providing data, and the Rotman China Initiative for financial support. Adam Morrison provided excellent research assistance. The analysis in this paper is my own, calculated 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 my own and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All remaining errors are my own.

1 Introduction

This paper studies how households differ in their consumption of imported goods and therefore the distribution of costs associated with changes in the relative price of imports, such as a change in trade policy. Two empirical relationships are central to understanding the consumption consequences of such events: (1) the extent to which households differ in their relative expenditure of imported goods and (2) the extent to which households differ in their ability to substitute consumption away from these same goods. Despite the importance of these two relationships in shaping the distributional costs of trade policy, our current understanding of both suffers from a number of limitations.

Consider the extent to which households differ in their relative expenditure of imported goods. A substantial body of literature has documented that low-income households spend a greater share of their income on tradeable sectors – such as manufactures and agricultural goods – when compared to high-income households which would suggest that low-income households are relatively more exposed to trade policy (Fajgelbaum and Khandelwal [2016]; Carroll and Hur [2020]). Yet this finding is also entirely compatible with a world in which low-income households do not spend a single dollar of expenditure on imported goods. If low-income households *only* consume domestically produced varieties *within* manufactures and agricultural goods, then sector-level expenditure shares offer little information about a household's true exposure to trade policy[1]. What is needed are data linking detailed household characteristics to expenditure on imported goods, rather than expenditure on tradeable sectors[2].

In order to provide credible estimates of the welfare effects associated with any shock to relative prices, one must also propose a model of substitution. Almost all current approaches relevant to the study of import price shocks assume that the substitutability of varieties is heavily influenced by the country in which those varieties were produced[3]. Yet this assumption is problematic given the primacy of multinational firms in observed trade flows (Bernard et al. [2012]; Bernard et al. [2018]). Consider an American multinational relocating production from a plant in Texas to a

[1] Throughout this paper I will refer to varieties as unique barcodes within a given product category. I follow Faber and Fally [2021] and define a firm as a brand.

[2] A number of papers study plausibly exogenous trade shocks to infer distributional differences in import consumption, such as: Cravino and Levchenko [2017]; Amiti et al. [2020]; Bai and Stumpner [2019]; Cravino and Levchenko [2018]; Li [2019]; Hottman and Monarch [2021]; and Auer et al. [2021]. Borusyak and Jaravel [2021] provide a detailed study concerning the role of within- and between-sector expenditure shares in shaping the consumption gains from trade. With respect to mapping household exposure to changes in trade policy, this paper uses detailed data to both confirm their findings and identify additional relevant dimensions of household heterogeneity.

[3] Recent examples include Feenstra et al. [2018] and Auer et al. [2021].

maquiladora across the Mexican border. Traditional models of import substitution – estimated using customs data – are susceptible to overstating the gains to US consumers associated with this shift in production for two reasons. First, while shifting production to Mexico might lead to lower prices, such a model would erroneously assume that these newly imported varieties constitute an expansion in the set of varieties offered US consumers. Second, it is often assumed *a priori* that imported and domestic goods are poor substitutes even though in this case the goods in question had previously belonged to some domestic market segment in which they competed with other domestic, and presumably also imported, alternatives. Both of these concerns can again be viewed as a shortcoming in data: in the absence of detailed attribute data for each imported and domestic variety, one cannot construct a more intuitive model of import substitution.

In this paper, I address both of these concerns by providing a dataset linking individual barcodes to their country of origin. These data constitute all text on the packaging of approximately 65,000 unique barcodes within the personal care/cosmetics category. Packaged consumer goods sold within the US are required to provide some statement equivalent to "Made in . . .", which allows for the recovery of each barcode's production location[4]. I link these data to detailed household purchasing records and socio-economic characteristics for over 42,000 US households. I use these data to both inform and estimate a model of demand with detailed consumer heterogeneity and rich patterns of import substitution in response to country-specific relative price shocks. The detailed nature of these data naturally come at the expense of scope: the categories studied in this paper constitute approximately $240 of expenditure per household-year. The goal of this paper is therefore not to provide aggregate welfare statements concerning the entire consumption basket but rather to guide future research by providing an empirical study rich in detail and novel in flexibility.

Within narrow product categories, wealthier households purchase varieties from wealthier origin countries. Conditional on income and a battery of socio-economic characteristics, I find that urban households consume a greater share of imported varieties than rural households, but that this pattern is completely reversed for Chinese imports. This reversal is driven by differences in retail environment: a household shopping for shampoo at a dollar store is seven times more likely to purchase a Chinese variety than an otherwise identical household shopping at a grocery/drug

[4]Barcodes generally do not contain information as to their country of origin. Auer et al. [2021] link barcodes to an import/domestic classification for 8,689 products and 3,000 household purchase records using Swiss data, and Borusyak and Jaravel [2021] map US scanner data to brand-specific import propensities. By linking attribute data for 60k barcodes directly to a country-of-origin, this paper estimates a more flexible demand model in the face of *country-specific* tariff changes. Faber and Fally [2021] also use scanner data to study distributional outcomes of changes in trade costs, but do not link household-specific expenditure directly to a country of origin. Ghai and Hottman [2019] use a similar dataset to this paper to estimate Armington elasticities.

store. When combined with the prevalence of dollar stores in rural ZIP codes, this analysis provides evidence that ZIP code population density is strongly correlated with household exposure to country-specific tariffs. The notion that retail markets may play a role in shaping the heterogeneous costs of trade policy has received little attention so far in the literature, and as such it is important to emphasize that this relationship is estimated conditional on household characteristics more commonly studied in this context such as income, education, race, and gender.

I therefore incorporate two dimensions of household heterogeneity when estimating my demand system: income and ZIP code population density. In order to do so, I first cluster households into types based on income and population density and then estimate separate demand parameters for each type. This approach to modeling heterogeneity is generally referred to as a two-stage grouped fixed effects (GFE) estimator (Bonhomme and Manresa [2015]; Bonhomme et al. [2017]). When applied to a discrete choice setting, however, the fact that different household types consume different sets of goods could introduce bias as this leads to a substantial proportion of varieties with zero market share for any given type. I therefore augment the GFE estimator with a bias correction for sparse data as outlined in Dubé et al. [2021] which allows for a flexible incorporation of multi-dimensional heterogeneity within a nested logit model.

In constructing a model of import substitution, I leverage both the detailed text data available for each barcode as well as intuition from the discrete choice demand estimation literature. The dataset used in this paper provides all text on the packaging of each barcode, which allows for a uniquely granular approach to understanding product differentiation. However it is not immediately obvious how to make use of descriptive text within the context of a random coefficients logit model of demand. Instead, I make use of text similarity measures and unsupervised clustering algorithms to place varieties into market segments based on the similarity of descriptive text on their packaging as well as their similarity across other dimensions, such as price and size. These clusters are then implemented as nests in a nested logit demand system, which I refer to as an attribute-cluster model[5].

I use the estimated model to run counterfactuals in which country-specific goods experience a relative price increase. I find that an increase in the price of Chinese imports is both regressive and anti-rural while an increase in the price of European imports is progressive and anti-urban[6].

[5]While the core intuition of this model is standard in the discrete choice literature (McFadden [1974]; Berry [1994]; Berry et al. [1995]; Cardell [1997]), the application of this intuition to questions of international trade is rare. Notable exceptions include Goldberg [1995]; Head and Mayer [2019]; and Coşar et al. [2018].

[6]These same patterns hold more generally for counterfactuals targeting, respectively, low- and high-income origin countries.

Averaging across all categories studied in this paper, I find that the within-category welfare costs for the most poor and rural households are 60% greater than their urban and wealthy counterparts when considering a positive price shock to Chinese goods. When targeting European goods, I find that the most urban and wealthy households have welfare costs which are 44% greater than poor and rural households.

I estimate a negative relationship across all categories between the price sensitivity of demand and both household income and ZIP code population density. Although this result is not surprising in itself, these estimates have implications for the outcomes of tariff policy: tariffs on wealthier origin countries are effective at generating government revenue but less effective at generating substitution towards domestically produced alternatives, whereas tariffs on low-income countries exhibit the exact opposite pattern. A corollary of this statement is that the incidence of any given tariff policy borne by US consumers is generally increasing in the GDP per capita of the country being targeted.

A tariff increase on NAFTA (Canada and Mexico) imports, however, both raises revenue *and* generates domestic substitution at a lower cost than any other tariff policy studied in this paper. This result is driven by the overwhelming presence of multinationals exporting to the US from Mexico and Canada. Varieties belonging to multinationals often have a number of highly substitutable but domestically produced alternatives - often within the same firm - which allows consumers to more readily substitute away from tariff increases. I provide novel evidence that this result is of first-order importance to understanding the costs of tariff policy: two thirds of all import expenditure in my dataset accrues to American multinationals with off-shored production.

This result underscores a key conceptual contribution of this paper: imported varieties provide welfare gains to consumers insofar as they consist of attribute combinations which are scarce in the domestic economy. In placing varieties into market segments based on attribute similarity, I find that imported varieties enter into a wide array of market segments alongside domestic alternatives and that this is especially true for imported varieties belonging to large multinationals. I provide a number of specification tests which all confirm that the attribute-cluster model provides a closer fit of the underlying substitution data when compared to a model based on more typical Armington-style assumptions of import substitution and I outline how future research should exercise caution when empirically estimating the consumption gains from trade in the absence of detailed attribute data and/or data concerning the domestic alternatives available to consumers.

This paper is organized as follows: Section 2 provides an overview of the data and three stylized

facts which inform the demand system specification outlined in Section 3. Section 4 provides an overview of the estimated demand parameters while Section 5 and Section 6 provide counterfactual policy simulations discussing, respectively, the distributional costs of tariffs and the efficacy with which tariffs raise revenue and/or generate domestic substitution. Section 7 concludes and Appendix A provides additional tables, figures, and details.

2 Data and Stylized Facts

This section provides an overview of the two key datasets used in this paper as well as three stylized facts which motivate the demand system to follow.

2.1 Data

Household Scanner Data: This paper uses the NielsenIQ Homescanner dataset. These data consist of a rotating panel of ∼60,000 American households. Each household records the price, date, and store that they visited for each shopping trip and the barcode-specific purchases made within each trip. These data include a full suite of socio-demographic information for each household. I follow Faber and Fally [2021] and proxy for household income using total expenditure per household member.

Barcode Country of Origin: A particular issue for applying the NielsenIQ data to questions of international trade is that barcodes do not contain explicit information as to their country of origin. I therefore merge the NielsenIQ Homescanner data with barcode-specific country-of-origin information purchased from Label Insight, Inc. (LI), a firm that specializes in extracting and organizing information found on the labelling of consumer packaged goods[7].

Label Insight uses an AI to extract from packaging the ingredients, branding, and any other text information that may be included for thousands of barcodes sold across the majority of retail chains in the US. Since imported goods in the US are required to contain some statement equivalent to "Made in . . .", the Label Insight AI incidentally recovers a country of origin for each barcode they collect. Naturally, Label Insight can only cover a segment of total consumption and their coverage is best for personal care products and packaged food. I therefore purchased the

[7] I am aware of one other paper that has used data from Label Insight in a trade/international macroeconomics context: Ghai and Hottman [2019] use similar data to estimate an Armington elasticity using retail data for 10 consumer packaged goods categories.

origin country, ingredients list, brand, and barcode description (as read directly off package) for approximately 65,000 barcodes spanning 15 product categories within personal care and cosmetics[8].

Data Summary: The final raw dataset used in this analysis contains purchase-level observations that combine a barcode with a household, store, and date. I restrict my sample to the years 2015 - 2017 and only include households that are present in each year of that period[9].

This final dataset contains 42,074 households purchasing barcodes from 15 distinct personal care and cosmetic product categories over three years[10]. These purchases span 23,806 barcodes (1,689 brands) and 20.6 million USD in total expenditure (3.3 million units purchased). Given that observations are at the purchase level, these data comprise 2.75 million observations. Table A.1 provides an overview of the data for each category, whereas Table A.2 provides data on sales by origin country. The final merged dataset contains purchases from 44 distinct origin countries, with an overall import share of 14.4% which amounts to 3.0 million USD of expenditure on imported varieties over three years. By merging the NielsenIQ categories to the BEA's Consumer Expenditure Survey I find that US households spend, on average, $240 USD per year on the categories studied in this paper. The merged NielsenIQ-LI dataset used in this paper covers approximately 68% of all predicted expenditure within these categories.

Given the novelty of the data used in this paper to study questions of trade policy, I first establish three stylized facts which guide the demand model to follow. I turn to presenting these three stylized facts now.

2.2 Stylized Facts

(1) Wealthy households purchase varieties from wealthier countries. This section studies how *within*-category import expenditure differs across income deciles in the USA. I first estimate a logit model in which the dependent variable is a simple indicator equal to one if a given purchase was of an imported variety and zero otherwise. I then include all socio-demographic household characteristics as potential explanatory variables along with fixed effects for half-year time periods,

[8]The US Customs and Border Protection (CBP) require that the country-of-origin printed on the label corresponds to the last country in which the good underwent a "substantial transformation".

[9]The LI data provides a current snapshot of the marketplace for personal care products and given the significant year-over-year turnover in varieties offered for these products, I find that years prior to 2015 exhibit a significant drop in merged barcodes. Changes to US trade policy began in 2018 which directly affected the categories studied in this paper so I choose 2017 as the final year of my sample.

[10]These categories are: Bathing Accessories, Body Soap, Dental Care, Deodorant/Perfume, Eye Cosmetics, Face Cosmetics, Facial Care, Hair Care, Hair Styling, Hand Soap, Lip Cosmetics, Nail Cosmetics, Shaving Care, Skin Care, and Toothpaste.

Figure 1: Purchase Propensity by Income and Origin

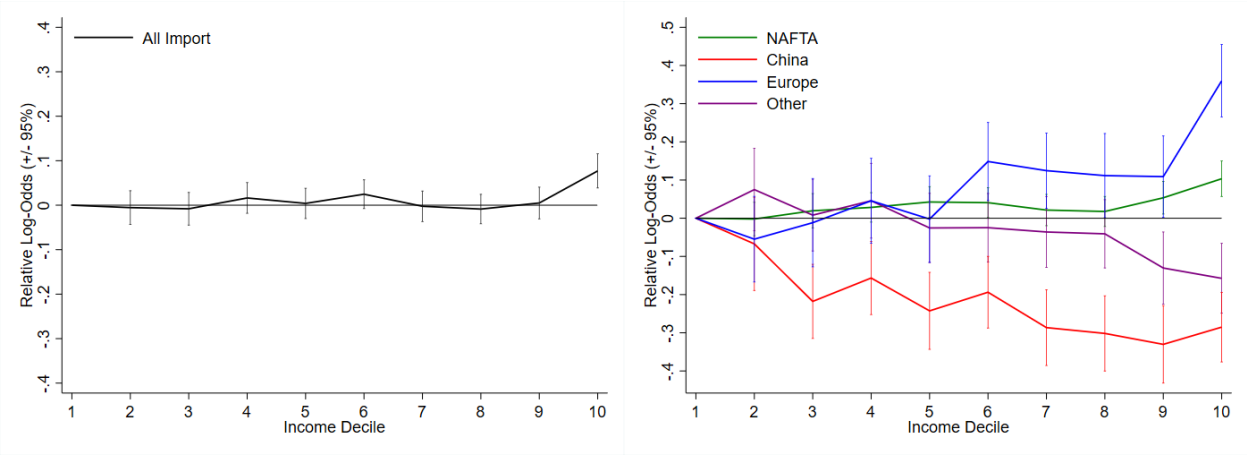

Figure 1 provides estimates of the propensity of a given income decile (x-axis) to purchase varieties from a specific origin. The left panel provides estimates with a dependent variable equal to one if the purchase is imported, and zero otherwise. The right panel provides similar estimates for specific origin countries. These estimates represent log-odds of a purchase relative to the lowest-income decile. All results include controls for education, race, age, presence of children, married household heads, and ZIP code latitude and longitude. I also include region, half-year, and product category fixed effects. 95% confidence intervals are provided, and standard errors are clustered at the product-region-period level.

geographic regions within the US, and the product category being purchased. The set of household characteristics includes income decile of each household, the population density of each household's ZIP code, household head educational attainment, household race and composition (including gender), and the longitude and latitude of each household's ZIP code centroid. This allows me to recover a relationship between the propensity to purchase imported varieties within narrow categories and household income. I then estimate identical models but with the dependent variable equal to one if a given purchase was from a specific origin country - for example, China - and zero otherwise.

Figure 1 provides purchase propensity estimates for five different origins across all income deciles. The left-hand panel provides estimates for the propensity to purchase any imported variety. The right-hand panel provides estimates for the propensity to purchase varieties from specific import origins: NAFTA, Europe, China, and all other import origins. All reported estimates are relative to the lowest income decile. Similar to other studies, I find that the propensity to purchase imports is relatively flat across the income distribution. The only exception is for the wealthiest decile of households, with an estimate of 0.085 that is significant at the 1% level. To place these estimates in context, consider a household in the poorest decile with an import share of consumption around

14.0% (this is slightly less than the average in the data). Figure 1 suggests that a household in the wealthiest decile would have an import share of consumption within the same product category of 15.1%[11].

The more relevant gradient along which income is correlated with exposure to changes in trade policy, however, lies in the origin country composition of import purchases. As shown in Figure 1, the extent to which households source consumption of the same product category from different origin countries changes dramatically as one moves along the income distribution. This is especially true when comparing the purchase of European and Chinese imports. I find that a household in the wealthiest decile is 40% more likely to purchase a European variety and 25% less likely to purchase a Chinese variety than an otherwise identical household in the poorest decile[12].

I re-cast this analysis as a linear model with the origin country GDP per capita as the dependent variable and estimate this model using OLS. Conditional on purchasing a non-NAFTA imported variety, households in the wealthiest decile source their consumption from countries with an average GDP per capita that is 3,035 USD greater than households in the poorest decile[13].

(2) Conditional on household income, retail markets play an important role in shaping import consumption. I provide novel evidence in this section relating import consumption to retail market composition. I estimate the logit model outlined earlier but with an additional set of dummy variables: fixed effects for the retail format at which each purchase took place[14]. Table 1 provides estimates of relative purchase propensities for different origins across retail formats.

A number of estimates stand out from Table 1. In terms of overall import propensities - Column (1) - I find that dollar stores have the highest propensity for import purchases, followed by grocery stores, discount stores, drug stores, and other formats. Similar to the estimates regarding household income, the more interesting variation across format types lies in the origin countries sourced by these formats, rather than the aggregate import propensity. Consider the propensity for households

[11] Other papers have found results similar to the left panel of Figure 1 (Borusyak and Jaravel [2021]; Auer et al. [2021]).

[12] Borusyak and Jaravel [2021] find a similar, albeit muted, relationship between income and the China share of expenditure by calculating average China import shares at the firm/brand level. I find a stronger relationship as I find that even within firms/brands, low-income households purchase varieties produced in lower-income countries. Bai and Stumpner [2019] and Hottman and Monarch [2021] both find evidence that the rapid expansion of Chinese exports to the US was likely "pro-poor".

[13] I estimate this relationship for non-NAFTA imports as imports from NAFTA completely reverse this trend - wealthy households purchase Mexican varieties and poorer households purchase Canadian varieties. I discuss this finding further in the third stylized fact.

[14] The NielsenIQ data provide indicators for different format types, and I aggregate the 30+ number of formats they provide into five categories: Dollar Stores, Discount Stores, Grocery Stores, Drug Stores, and all Other formats.

Table 1: Relative Purchase Propensities by Retail Format and Origin

	(1) Import	(2) NAFTA	(3) China	(4) Europe	(5) Other
Other	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)
Dollar Store	0.371	0.224	1.224	-0.032	0.456
	(0.045)	(0.058)	(0.073)	(0.124)	(0.065)
Discount Store	0.064	0.143	-0.169	-0.201	0.075
	(0.020)	(0.024)	(0.064)	(0.052)	(0.046)
Drug Store	0.045	0.234	-0.846	-0.279	-0.303
	(0.034)	(0.037)	(0.069)	(0.060)	(0.050)
Grocery Store	0.268	0.429	-0.497	0.120	-0.110
	(0.028)	(0.031)	(0.062)	(0.050)	(0.060)
N	3.3M	3.3M	3.3M	3.3M	3.3M
r^2	0.047	0.078	0.280	0.157	0.051
			Controls		
$\bar{\boldsymbol{x}}_i$	X	X	X	X	X
Income Decile	X	X	X	X	X
Category FE	X	X	X	X	X
Half-Year FE	X	X	X	X	X
Region FE	X	X	X	X	X

Table 1 provides estimates of the relative log-odds of purchasing a variety from a given origin conditional on the format visited. All regressions include controls for household income decile, education, race, age, presence of children, married household heads and ZIP code latitude and longitude. I also include region, half-year, and product category fixed effects. Estimates are relative to the "Other" retail format category, which is the reference category. Standard errors are provided in parentheses and are clustered at the product-region-period level.

to purchase varieties imported from China (Column (3)). A household purchasing a given product category at a dollar store is 7.69 *times* more likely to purchase a Chinese variety than an identical household purchasing the same product category at a drug store. They are also 5.44 and 3.93 times more likely to purchase a Chinese variety than when shopping, respectively, at a grocery store and discount store. These results are flipped when discussing propensities for purchasing European and NAFTA varieties, with grocery stores exhibiting the strongest propensity for both of these origins.

Similar to the analysis using income, I re-estimate this relationship as a linear model with origin GDP per capita as the dependent variable. I then limit my sample to non-NAFTA imports and

estimate the average origin GDP per capita of imports associated with different retail formats. I find that grocery store import purchases are, on average, from non-NAFTA countries with an average GDP per capita that is $9,673 greater than countries sourced by dollar stores. This is equivalent to the difference between South Korea and Japan, or China and the Czech Republic. Again, these findings are estimated conditional on the socio-economic characteristics of the household making each purchase, including income decile.

The results shown here suggest a novel channel through which households are exposed to trade policy: their retail environment. Naturally these findings are descriptive rather than causal. The goal of this section then is to provide evidence that even after controlling for household characteristics which a researcher might reasonably believe are associated with heterogeneity in import consumption - such as income or education - there are additional dimensions of heterogeneity across consumers which might be equally important in shaping exposure to trade policy[15].

When constructing the demand system used in this paper, I capture both the heterogeneity associated with income and retail format exposure in my definition of household "types". Of course households have the ability to shop at multiple retail formats and so I group households using a household characteristic that is strongly correlated with retail format shopping frequency: ZIP code population density. Figure A.1 in Appendix A.1 provides estimates of a logit model in which the dependent variable is an indicator associated with whether or not any given shopping trip was to a certain retail format. These estimates include all of the demographic controls mentioned earlier, as well as household income decile and a dummy variable for each decile of ZIP code population density.

I find that conditional on income and other socio-economic characteristics, population density is strongly correlated with the retail formats households frequent. Households in the least dense decile of ZIP codes are 2.32 times more likely to shop at a dollar store and 2.41 times more likely to shop at a discount store than an identical household living in the most dense decile of ZIP codes. Opposite patterns hold for shopping at drug stores and grocery stores, with households in the most dense decile of ZIP codes 2.66 times more likely to shop at a grocery store than an identical household living in the least dense decile of ZIP codes.

When combined with the estimates in Table 1, these findings suggest that conditional on standard socio-economic household characteristics, ZIP code population density is strongly correlated

[15]These findings also suggest a link between the "food desert" literature and the literature studying the role of retailers in international trade (Bernard et al. [2010]; Basker and Van [2010]; Handbury [2021]).

with household-level exposure to country-specific trade policy. In specifying my demand system, I use both income and population density as the two dimensions of household heterogeneity relevant to estimating distributional costs of tariff changes[16].

(3) 63% of import expenditure accrues to American firms, while 18% of expenditure on domestically produced varieties accrues to foreign firms. The data used in this paper allow for a unique study as to how expenditure based on production location differs from that of design location. To continue the example cited in the introduction: a variety produced in Mexico by an American firm has a production location in Mexico but a design location in America. This study therefore offers a departure from papers using customs data as such analyses are limited to only studying the production origin of goods.

In order to do so, I manually link the ∼1,700 firms in my final dataset to a country of origin. This is either the headquarter location of that firm or - when these data are not available - the country in which that brand was founded. Table 2 provides a decomposition of aggregate expenditure based on production and design origins. Each element of the table provides the share of aggregate expenditure accruing to that combination of production and design origins, with the columns representing production origin and the rows representing design origin.

I find that 85.6% of all expenditure accrues to domestically produced varieties and 14.4% accrues to foreign-produced varieties. This 14.4% constitutes the import share of expenditure one would infer using standard datasets. However by design location, these percentages change to 79.0% and 21.0%, respectively, which suggests that the penetration of foreign firms in the US market is almost 50% greater than what one would estimate using customs data. Further decomposition shows that 69.9% of all expenditure accrues to varieties designed and produced in America, whereas 9.1% of all expenditure accrues to domestically designed varieties produced off-shore. When compared to the expenditure share of varieties both designed and produced abroad (5.3%), I find that 63% of all expenditure on imported varieties in fact accrues to American firms with off-shored production. Lastly, I find that 15.7% of all expenditure accrues to varieties designed by foreign firms but produced within the USA. Even if one could decompose customs data into American and non-American firms, one would still be missing almost three quarters of all expenditure on foreign-

[16] I provide direct evidence that population density is related to origin country exposure in Figure A.2 in Appendix A.1. These estimates are identical to those provided for income except in this case the fixed effects represent ZIP code population density deciles.

Table 2: Expenditure by Production and Design Origin

	Production Origin		
Design Origin	Domestic	Foreign	
Domestic	69.9	9.1	79.0
Foreign	15.7	5.3	21.0
	85.6	14.4	

Table 2 provides aggregate expenditure shares across all categories decomposed by the production and design origin of each barcode. Production origin refers to the location in which that barcode was produced, and design origin refers to the origin country of the brand/firm associated with that barcode.

designed varieties as these are produced mainly in the USA and therefore never cross the border[17]. The difference between production-based expenditure shares and design-based expenditure shares is even starker when calculating country-specific import shares in the US market. Based on production location, Canada, Mexico and China account for 10.9% of all expenditure in my dataset. France, Germany, Italy and the UK account for only 1.2%. However when calculated using a design-based expenditure share, Canada, Mexico, and China account for only 0.2% of expenditure, whereas the expenditure share for France, Germany, Italy, and the UK increases substantially to 11.4%.

The three stylized facts introduced here motivate the demand system introduced in the following section. The first two stylized facts suggest that the demand system to follow must be able to flexibly account for the fact that households differ in their exposure to country-specific price changes along two key dimensions: income and ZIP code population density. The third stylized fact suggests that typical Armington-style models - in which substitutability and production location are often analogous - may not be useful in this setting. If consumers primarily view barcodes as belonging to a brand - and that brand's country of origin - then a model of substitution based primarily on production location may be mis-specified. I address both of these points in my demand model specification, which I turn to now.

[17] These data provide some of the first expenditure-based estimates at the intersection of multinationals and import consumption (Head and Mayer [2019]; Ramondo [2014]; and Tintelnot [2017]). Blonigen and Soderbery [2010] show that the greatest variety gains for American automobile consumers due to foreign firms in fact derives from these firms producing in America.

3 Demand Estimation

Rather than relying on product-based demand systems - such as the CES or AIDS - I leverage the detailed data of this paper to estimate a demand system in the variety attribute space[18]. I opt for the nested logit demand system for three reasons. First, in the absence of non-price continuous attributes the random coefficients logit model - as in Berry et al. [1995] - is as general as the nested logit (Grigolon and Verboven [2014]). The only non-price continuous attribute present in the data is the weight of a given variety. Second, the data presented in this paper contain descriptive text from the packaging of each barcode. These data provide a remarkably granular approach to understanding variety differentiation but do not lend themselves easily to the random coefficients logit model[19]. Third, I account for consumer heterogeneity by estimating a separate demand system for each household type with a bias correction for sparse data. This allows for more flexible consumer heterogeneity at a lower computational cost when compared to a random coefficients logit model.

3.1 Nested logit with Consumer Heterogeneity

This section provides the main estimating equations of this paper. Consider a barcode j of product category k. This barcode is of brand b_j and has a binary indicator $m_j = 1$ associated with j being imported. I define a market as a region-time (r, t) pair, where each time period consists of a half-year and each region refers to one of five non-overlapping geographic areas of the US[20]. For all equations to follow, I drop the subscript k as it is implied that each set of demand parameters is category-specific.

Assume that for each category k there exists an exhaustive set of market segments denoted by Γ_k with each individual segment indexed as $\gamma \in \Gamma_k$. Equation 1 then provides the demand curve - as derived in Cardell [1997] and Berry [1994] - for barcode j belonging to market segment γ. For ease of exposition, I define $\boldsymbol{\beta X_j}$ as the sum of utility over all observable non-price attributes that

[18] This is due partly to the fact that many of the categories studied in this paper constitute well over a thousand unique barcodes.

[19] Consider a matrix of bilateral (dis)similarity in text across all varieties, in which each element represents the text dissimilarity between varieties i and j. It is not clear how such a matrix could be transformed into a vector representing some continuous attribute for each variety that one might consider relevant for understanding substitution patterns.

[20] North-East, South-East, Mid-West, South-West, and Mountain/Pacific.

are specific to j[21].

$$\ln\left(\frac{s_{jrt}}{s_{0rt}}\right) = -\alpha \ln p_{jrt} + \boldsymbol{\beta X_j} + \sigma \ln(ns_{jrt\gamma}) + \zeta_{jrt} \tag{1}$$

s_{jrt} denotes the *within*-category quantity market share of variety j. The market share of the outside option is given by s_{0rt}, and I outline my approach to determining this quantity in the following paragraph. The within-nest market share of variety j in segment γ is given by $ns_{jrt\gamma}$. Introducing consumer heterogeneity into Equation 1 is often accomplished by allowing parameters to vary linearly with some market characteristic, such as average household income. Given that household heterogeneity in this paper consists of two dimensions, I instead estimate a completely separate demand system for each household type h. I discuss this approach in the next section.

I therefore create a separate panel for each household type and product category at the barcode level within each half-year period (t) and region of the US (r). Equation 2 then provides the main estimating equation of this paper.

$$\ln\left(\frac{s_{jhrt}}{s_{0hrt}}\right) = -\alpha_h \ln p_{jrt} + \boldsymbol{\beta_h X_j} + \sigma_h \ln(ns_{jhrt\gamma}) + \zeta_{jhrt} \tag{2}$$

In order to construct the aggregate market size, and thereby identify the market share of the outside option s_0, I make use of the consumer expenditure survey (CEX) from the BEA. For each year, I allocate the CEX total expenditure data to the 15 categories studied in this paper based on each category's sales share in the NielsenIQ data. This leaves me with an average annual household expenditure for each category and year. I then adjust this in three ways. First, I use a household-type and year-specific average price for each category to transform the average annual expenditure into an estimate of the average units purchased per household. I then adjust this average units purchased per household to account for differences in overall expenditure across household types and to adjust for seasonality in demand[22]. I then simple multiply the per household quantity purchase estimates with the number of households in each region and of each type, and define this value as the total expected market size.

In the following sections, I describe my approach to assigning household type, creating the nesting structure, implementing the bias correction for sparse data, and addressing the endogeneity of price and within-nest market share in Equation 2.

[21] Specifically: $\boldsymbol{\beta X_j} = \beta^w w_j + \phi_{m_j}$ where w_j is the weight of variety j and ϕ_{m_j} is some utility shifter associated with purchasing a foreign variety.

[22] I run a regression at the product-half-year level of total units purchased against a fixed effect for the second half of each year. I then allocate units purchased per household to each half-year based on the second-half fixed effect estimate.

3.2 Household Heterogeneity

This paper accounts for heterogeneous preferences by implementing a two-stage grouped fixed effects estimator (GFE) as outlined in Bonhomme and Manresa [2015] and Bonhomme et al. [2017]. As discussed, income and ZIP code population density are both strongly correlated with consumption exposure to origin-specific trade cost shocks. I therefore use these two dimensions of heterogeneity in order to place households into types. Consider a household i with income Y_i living in a ZIP code with population density D_i. I implement the k-medians algorithm in order to group each household into clusters which minimize the within-cluster average distance between households in the (Y_i, D_i) space. I opt for ten household types as this allows for substantial heterogeneity while maintaining sufficient variation in the data to estimate demand systems for each household type. Table A.3 in Appendix A.1 describes the median household characteristics of each type h.

The approach used in this paper to incorporate household heterogeneity differs from standard techniques often employed when estimating discrete choice models. A more common approach would be to allow the estimates of certain parameters to vary linearly with some demographic characteristic, such as income. Of course these demographic parameters could enter non-linearly, but it would be difficult to incorporate multiple dimensions of heterogeneity across parameter estimates within a single parsimonious demand system. The GFE estimator allows for the introduction of multiple dimensions of consumer heterogeneity. However in a discrete choice setting, the extent to which households of different types purchase entirely different sets of varieties could introduce bias. Discrete choice demand models do not admit zero market shares which is a cause for concern when implementing a GFE estimator in this setting. I therefore augment the GFE estimator by explicitly modeling consideration sets and thereby alleviating concerns of selection and bias[23]. I now turn to placing varieties into market segments.

3.3 Attribute-Cluster Model of Import Substitution

The nested logit demand system outlined in Equation 1 requires that the researcher construct an *a priori* nesting hierarchy across varieties within a given product category. Although there are a number of barcode characteristics within the data that could be used to place varieties into market segments - such as brand or country-of-origin - I instead rely on unsupervised machine learning clustering algorithms to determine the clusters which best capture where varieties exist in the space

[23] I outline this approach in Section 3.4.

of observable attributes. Specifically, I employ the Partitioning Around Medoids (PAM) clustering algorithm to group varieties within each category based on their relative (dis)similarity across three characteristics: price, weight, and descriptive text found on the packaging.

This process involves three steps, which I outline in greater detail in Appendix A.2. First, for a given product category, I calculate a simple distance between all varieties for price and weight. For any two varieties j and j', define these distances as $\Sigma^p_{jj'} =\mid p_j - p_{j'} \mid$ and $\Sigma^w_{jj'} =\mid w_j - w_{j'} \mid$, respectively. I then calculate a distance measure between the text that constitutes each variety's brand and packaging[24], and denote this distance as $\Sigma^t_{jj'}$. Last, I standardize all distance measures to have the same mean and standard deviation, before calculating the total bilateral Euclidean distance between each variety pair:

$$\Sigma_{jj'} = \left({\Sigma^p_{jj'}}^2 + {\Sigma^w_{jj'}}^2 + {\Sigma^t_{jj'}}^2\right)^{1/2}$$

After calculating a bilateral distance for all varieties, I implement the PAM clustering algorithm to cluster varieties based on the objective of minimizing the within-cluster average bilateral distance: $\Sigma_{jj'}$. I augment this approach using the Silhouette Method to select the "optimal" number of clusters (for details, see Appendix A.2). I refer to this model of substitution as an "attribute-cluster" model.

By clustering varieties based on price and weight, this approach to the nested logit effectively mimics the desirable properties of the random coefficients logit model - varieties that are more similar in the space of continuous characteristics will be seen as closer substitutes when compared to varieties that are more dissimilar. The use of text-based data to model product differentiation, however, forms a key contribution of this paper[25]. As an example, when clustering deodorant varieties into partitions of strong and weak substitutability, there are a number of partitions almost identical in terms of average weight and price. However these clusters are differentiated by the presence of words such as "Men", "Women", "Sport", or "Antiperspirant" on the packaging. Such a granular understanding of the attribute space would be difficult to incorporate into a random coefficients structure as it is not immediately obvious how one could map text (dis)similarity into a single cardinal measure or characteristic[26].

[24] Again, there is greater detail plus an example in Appendix A.2.

[25] Clustering has been used in the marketing literature to understand market segments since at least the 1980s (Punj and Stewart [1983]; Green et al. [1990]; Iacobucci et al. [2000]). Recently, Harding and Lovenheim [2017] use clustering to create aggregate product categories when estimating the AIDS model of demand.

[26] In concurrent and ongoing work Almagro and Manresa [2021] provide a data-driven approach to selecting the nests within a nested logit model. The approach implemented here is in response to a separate concern: when

Notice that the clustering algorithm described above categorizes varieties by observable similarity without relying on origin country information[27]. In doing so, this approach eschews any *a priori* assumption about whether imported varieties are seen as fundamentally different from domestic varieties, and instead employs data-driven methods of placing imports within market segments of observably similar varieties.

This approach marks a departure from the standard "Armington" approach to understanding substitution patterns often assumed in the trade literature. These models tend to assume that imported and domestic varieties are fundamentally poor substitutes. Once translated into the language of discrete choice models, this assumption is equivalent to stating that there exists only one attribute relevant for understanding substitution patterns: the country in which each good was produced. Even in the absence of specifying a nesting structure, any study of trade with CES preferences fundamentally assumes that all new varieties consist of novel attribute combinations regardless of whether or not a given variety has observably similar alternatives available to consumers. The explicit assumption being in this paper is that the gains to consumers are larger when new varieties offer new attribute combinations, whereas varieties with a number of identical alternatives offer lower utility gains[28].

The stylized facts introduced in this paper cast some doubt on the underlying intuition of the Armington model. I find that almost two thirds of all expenditure on imported varieties accrues to domestic brands, and that 15% of all expenditure accrues to foreign brands producing varieties in the domestic market. While these exact estimates may differ by sector and country of interest, the fundamental argument being made is that the production location of a variety does not necessarily dictate the observable attributes of that variety in the eyes of consumers. The data studied in this paper allow for a unique opportunity to test whether this is in fact the case, and I provide a specification test of the attribute-cluster model introduced in this paper versus alternatives based on production origin of varieties in Section 4.3. I find that models based on Armington-style assumptions fail to provide a meaningfully closer fit of the data when compared to the attribute-cluster model.

attributes consist of lengthy text data, it is not clear how to incorporate this information within a random coefficients logit model of demand.

[27]The algorithm does include the brand of each variety, so to the extent that consumers associate a given brand with an origin country, this information is included in the clustering algorithm.

[28]This research is often motivated by the "variety gains" introduced in the canonical increasing returns to scale model of trade by Krugman [1979] and Krugman [1980]. Later expanded on by Melitz [2003]).

3.4 Variety Selection and Accounting for Zeros

This paper applies a two-stage grouped fixed effects estimator within a discrete choice model of demand. Given that households differ in the set of varieties they consume, this approach leads to a significant number of varieties with zero market share for any given household type, and this could be a source of bias. This section both discusses the source of this bias and accounts for this bias by implementing the method outlined in Dubé et al. [2021][29].

For any given market - that is, a (k, r, t) triplet - define the aggregate set of available varieties $\boldsymbol{J_{krt}}$ as the union of all varieties purchased by all household types. Individual household types only consider purchasing a subset of $\boldsymbol{J_{krt}}$, which I denote as the consideration set $\boldsymbol{C_{hkrt}} \subseteq \boldsymbol{J_{krt}}$. So long as entry into the consideration set is correlated with variety-specific valuation shocks, then the presence of zeros in household-type and market-specific purchasing records could lead to substantial bias. This concern is particularly relevant in the current setting, as I find that the median share of varieties with zero market share across all household types and categories is $\sim$65%[30].

I provide a detailed description of the estimator proposed by Dubé et al. [2021] in Appendix A.3. This method accounts for selection bias by estimating some propensity for varieties to enter into the consideration set. One can then identify demand parameters by estimating the model in pairwise differences and restricting the sample to variety pairs with identical consideration propensities. I refer to this estimator as the Pairwise-Differenced Weighted Ordinary Least Squares (PDWOLS) estimator. Constructing this consideration propensity while allowing for correlation between price and consideration requires a price instrument as well as a consideration instrument: $\tilde{\omega}_{jrt}$. I discuss my price instrument in Section 3.5, and the consideration instrument here.

The consideration instrument $\tilde{\omega}_{jrt}$ is relevant if it is correlated with the propensity for variety j to be included in the consideration set $\boldsymbol{C_{hkrt}}$. The exclusion restriction is that this instrument must be excluded from the choice-specific mean utility. Dubé et al. [2021] suggest using variety-specific promotions or typical "nudge" factors such as shelf placement. I augment this approach by using data on brand-level promotions. Specifically, for households in a given region r and for a variety j of brand b_j, I calculate the share of all sales of this brand in all other regions $r' \neq r$ that are flagged by NielsenIQ as being purchased with some discount or deal. This instrument is relevant in that

[29] Research estimating gravity-style models often mechanically account for zeros by implementing the Poisson Pseudo-Maximum Likelihood (PPML) estimator. This approach is only feasible when zeros are admitted as an endogenous outcome of the model, which is strictly ruled out in a discrete choice setting. Research in discrete choice settings therefore often either drop varieties with zero market share or replace the zero with a very small number. Both of these approaches, however, could lead to bias.

[30] I outline my model of selection and the subsequent origins of this bias in Appendix A.3.

it captures brand-level promotional activity which could cause consumers to consider purchasing varieties they would otherwise not have considered purchasing.

The exclusion restriction for this consideration instrument is valid so long as consumers do not derive utility explicitly from this instrument. While it may be that certain households do in fact derive utility explicitly from the act of purchasing a promoted variety, I exploit the region-level data in this paper by using the share of varieties purchased under promotion in all other regions. The key assumption is that households do not derive utility from promotions occurring in other regions, but that brand-specific promotions are often coordinated at the aggregate nation-level rather than in response to region-specific demand shocks[31].

By limiting my sample to bilateral variety pairs with identical consideration propensities, this estimator accounts for the bias associated with dropping zeros and allows for a flexible "fixed-effect" approach to estimating household-type-specific demand parameters. I now turn to addressing the endogeneity of price and within-nest market share in Equation 2.

3.5 Identification

As is well-discussed in the demand estimation literature, including price in Equation 2 almost certainly raises concerns of endogeneity. I implement an instrument variables approach to dealing with this endogeneity. Specifically, I exploit the cross-section of markets across US regions and implement an instrument similar to the "Hausman-Nevo" instrument (Hausman et al. [1994]; Nevo [2001]): I use the prices of comparable products in all regions $r' \neq r$ as an instrument for prices in each individual region r. Ideally, this approach would compare prices of the same barcode across regions, but a significant number of varieties are region-specific. Instead, I instrument p_{jrt} using the average price of all other barcodes within the same cluster $\gamma(j)$ but sold in all other regions.

The identifying assumption of this instrument is that demand shocks for variety j are uncorrelated across regions. In terms of relevance, there are two reasons as to why this instrument provides useful information about p_{jrt}. First, the nests in this paper explicitly group varieties based on their observable similarities such as price, weight, and text on the packaging. The assumption is that such characteristics are useful in determining strong and weak substitutability, thereby providing a theory-driven justification for the use of all other prices in the nest $\gamma(j)$: a price-setting model of oligopolistic competition suggests that firms charge the same price for all competing goods (close

[31] I find that the first-stage consideration estimates relating my consideration instrument to d_j are all positive and significant at a level of 1%. These estimates have a median value of 0.84. More details of this estimation procedure and results can be found in Appendix A.3.

substitutes) in equilibrium.

In addition, the text analysis discussed in Section 3.3 may provide insights into commonality across varieties in terms of their inputs, which in turn might suggest common cost shocks within nests[32]. To the extent that such differences in packaging capture true differences in production technology and inputs, rather than pure branding, it is reasonable to believe that within-cluster cost shocks may be correlated.

Lastly, the within-nest market share suffers from endogeneity as well. I instrument for the within-nest market share by assuming that retailers make all stocking decisions before each period and therefore do so with no knowledge of the region- and period-specific demand shocks associated with each variety. I then calculate, for each variety, the aggregate number of retail chains at which that variety was purchased in the data, and take the ratio of this value to the aggregate within-nest sum of all variety-retailer pairs. The relevance of this instrument derives from the fact that the more retailer-variety pairs there are within a given nest, the greater the competition facing any individual variety. The exclusion restriction follows directly from the previous assumption: if retailers make national-level stocking decisions before the idiosyncratic valuation shocks are revealed, then this instrument is uncorrelated with the error term in Equation 2.

The final estimator used in this paper, then, is an augmented GFE Two-Stage Least Squares estimator (AGFE2SLS).

4 Estimation Results

4.1 Results

This section provides descriptive features of the 150 separate demand systems estimated in this paper. For the sake of exposition, this section omits detailed results for all regressions, however I provide the distribution of key parameter estimates in Figure A.3 in Appendix A.1. The first-stage F-statistics of the main AGFE2SLS estimator have a minimum value of 32. The median value is 560 and the mean is 720. The bias associated with price endogeneity is generally towards more inelastic estimates of α, which suggests that retailers adjust prices to positively correlate with demand shocks. The bias associated with the within-nest market share is also generally positive, which is expected given that the within-nest market share contains information directly correlated

[32] As an example, specific flavours or scents are common to the clusters created by the PAM algorithm, which suggests some correlation of inputs across varieties within that cluster.

with the dependent variable.

The price disutility estimates ($\hat{\alpha}$) are for the large part positive and significantly different from zero at a level of 1%[33]. The median value of $\hat{\alpha}$ is 0.50, with a standard deviation of around 0.25. All estimates of σ lie within the required interval of $[0, 1]$, with an average estimate of 0.47 and a standard deviation of 0.13. I find that the bulk of import-specific fixed effect estimates are negative implying that, all else equal, consumers have negative utility associated with purchasing an imported variety compared to a domestic variety. This result is in line with the "home-market" preference bias discussed in Coşar et al. [2018]. Across all categories, I find that households exhibit a domestic variety premium equivalent to approximately 19%. That is, households are indifferent between purchasing a foreign variety for \$1.00 and an otherwise identical domestic variety for \$1.19[34].

I find that high income households have systematically lower disutility of price[35]. Interestingly, I also find that high-income households have larger estimates of σ, implying that wealthier households have stronger preferences for substitution *within* nests when compared to lower-income households. Similar patterns hold for population density, suggesting that households in more urban areas are less price sensitive and that urban households have stronger preferences for substitution *within* nests.

Both α and σ combine to determine the own-price elasticity for any given variety[36]. The distribution of $\hat{\alpha}$ implies that wealthy and urban households have more inelastic demand than lower income households. However the estimates of $\hat{\sigma}$ work in the opposite direction: since wealthy and urban households have stronger preference for *within*-nest substitution, each variety within a given nest has more correlated preferences and this leads to greater substitution *within nests*. The notion that these two forces might work against each other in determining the average elasticity of demand for different households is a novel finding of this paper and suggests that using a random coefficients demand system is important for capturing heterogeneity across household types since

[33] There are two notable exceptions: both "dental care" and "bathing accessories" have estimates of α that are predominantly negative across all household types. I drop these two categories moving forward as the negative estimates of α confuse the welfare analysis to follow however the aggregate results are qualitatively similar with these categories included.

[34] Coşar et al. [2018] refer to this term as a measure of home-market bias in preferences although, as I discuss in Section 4.4, this may be due to where multinationals decide to produce varieties of different quality. Without a more rigorous definition of quality - and especially a measure that is not isomorphic with home-market bias - it is difficult to discern the underlying mechanism.

[35] Figure A.4 plots the average values of $-\hat{\alpha}$ and $\hat{\sigma}$ against the two dimensions of household heterogeneity studied in this paper: income and population density.

[36] The own-price elasticity for any variety j in the nested logit demand system is given by: $\epsilon_{jj} = \frac{-\alpha}{1-\sigma}(1 - \sigma n s_j - (1-\sigma)s_j)$.

this nuance would be lost in a product-based demand system such as the CES.

Lastly, I use the nested logit own-price elasticity formula to calculate an own-price elasticity for each household-type-barcode combination: $e_{jj,h}$. Table A.4 provides the median elasticity for each category and household type. The categories studied here are generally in the more inelastic range, however this is perhaps intuitive: one might expect cosmetics and personal care products to have more inelastic demand than, for example, ready-to-eat cereal[37]. I find that the variation in $\hat{\alpha}$ explains 83% of the variation in the median own-price elasticities across household types, with variation in $\hat{\sigma}$ accounting for the remaining 17%.

4.2 Selection and Bias

This paper applies a two-stage grouped fixed effects estimator to a discrete choice demand setting while adjusting for zeros using the method in Dubé et al. [2021]. In order to study the bias associated with consideration and zeros in the data, I estimate the AGFE2SLS estimator but without imposing any consideration restriction. On average, the unrestricted elasticity estimates are 20% more inelastic than in the restricted model[38].

Interestingly, I do not find any systematic bias across household types. This is perhaps not surprising as the key mechanism behind this selection bias is simply the strength of correlation between consideration and valuation. It is not immediately obvious why this correlation should differ systematically across household types. However I do find significant heterogeneity across categories in the mean level of bias, and this heterogeneity is reassuringly driven by differences in the prevalence of zeros across product categories. In order to provide a benchmark for the relationship between zeros and the median own-price elasticity within each category, I estimate a simple linear model with the logarithm of elasticity bias as the dependent variable and the logarithm of the share of all varieties with zero market share as the independent variable. I find that a 10% increase in the share of varieties with zero market share increases the bias of elasticity estimates by 11%.

[37] To put these estimates in perspective, a median elasticity of -2.6 for ready-to-eat cereals is common in the discrete choice literature (Nevo [2001]).

[38] Dubé et al. [2021] find a larger bias associated with dropping zeros but it is of the same direction. They use store-level weekly scanner data which may explain the discrepancy: the bias associated with dropping zeros is driven by the correlation between consideration and valuation. This correlation is likely stronger in shorter time windows and within a given store. This paper uses larger time windows for each market which may weaken this correlation as households are exposed to variety sets multiple times within a "market".

4.3 Specification Test

The attribute-cluster model used in this paper marks a departure from more standard techniques often used in the trade and international macroeconomics literature. The key distinction is that traditional approaches often assume the production location of a given variety is the key attribute to understanding low and high substitutability across varieties. This section provides the results of two separate specification tests implemented to test the importance of considering the production origin of goods when estimating import substitution patterns. A full description of both tests, along with detailed results, can be found in Appendix A.4.

First I implement a Vuong test of non-nested models to test whether alternative nesting structures are able to fit the underlying consumption patterns better than the attribute-cluster model. I compare the attribute-cluster model to both a model with only two nests based on whether a variety was produced domestically or abroad as well as a model in which each production country of origin constitutes its own segment[39]. Across all product categories, I reject the null hypothesis that either of these models approximates the underlying consumption data as effectively as the attribute-cluster model at a significance level of 99%.

I then analyze whether the production origin of a good provides useful information when added to the attribute-cluster model, which constitutes the second set of specification tests. In this case, I begin with the attribute-cluster model for each product category and add an additional layer of nesting *within* each of the attribute clusters based on whether or not a variety was produced domestically or abroad. I find that this augmented model fails to meaningfully provide a better fit of the data than the attribute-cluster model alone. The opposite statement, however, does not hold: when the attribute-cluster nesting structure is used to augment a model based on the foreign/domestic dichotomy, the attribute-cluster sub-nests provide substantial improvements in the fit of the model. Again, more details are provided in Appendix A.4.

While there is no determinative approach to selecting the "correct" market segments, this paper provides a battery of evidence suggesting that models which base substitution patterns off of the production origin of goods are difficult to reconcile with consumption data covering the categories studied in this paper. This is true both of the formal specification tests provided here, as well as the novel evidence suggesting that the bulk of import expenditure accrues to varieties that are designed and branded as "American". As a final exercise, I provide an analysis of variety quality

[39]The difference between these two models being that in the first case, all imported varieties are in the same segment, whereas in the second case "Mexico" and "Canada" would constitute two separate segments.

in the following section and study how the confluence of firms and production locations shape the perceived quality of imported and domestic varieties.

4.4 Quality, Firms, and Countries

I leverage the uniquely detailed data of this study to estimate differences in the perceived quality of goods imported from different origin countries. Given that I estimate a nested logit demand system, this measure of quality is simply the unobserved valuation of each variety j[40].

As a first step, I estimate the average quality percentile associated with each country's exports to the US based on production location. These estimates therefore capture what one would find in a study using customs data. Figure A.5 plots the average quality percentile of each origin country against GDP per capita. Reassuringly, I find a strong positive relationship between average quality and GDP per capita. There are a number of countries which seem to break this trend, however. Specifically, Mexico, the Philippines, Morocco, Slovakia, Thailand and Togo all exhibit average quality rankings which far exceed what one would predict given the relationship between quality and income. These results are explained almost entirely by the presence of non-domestic firms using these countries as export platforms to the USA and I turn to a study of these effects now.

Table A.6 provides the average quality rank for the top ten production origin countries in my dataset. However in this case I decompose these quality rankings into four parts. Column (1) provides the average quality ranking of varieties produced and designed in that country. Columns (2) and (3) provide the quality rank of varieties designed in that country but produced in, respectively, other non-USA countries and the USA. Column (4) provides the average quality rank of varieties produced in that country, designed elsewhere, and exported to the USA. I find that varieties produced in the USA by American brands have an average quality rank of 52 yet varieties imported to the USA by American firms with off-shored production have a quality rank of only 42. The vast majority of these off-shored varieties are produced in Mexico, China, Philippines, and Thailand. I find that these off-shored American varieties tend to have a greater quality ranking than varieties produced in the same countries but by firms associated with those countries. As far as I am aware, this paper is the first to identify the role that multinationals may play in shaping the estimates of product quality across origin countries, although this discussion has occurred with respect to aggregate productivity (Alviarez [2019]; Alviarez et al. [2020]).

[40] Note that this definition of quality is identical to that used in Khandelwal [2010]. While similar analyses appear in Schott [2008]; Hallak and Schott [2011]; and Crozet et al. [2012], this paper provides a unique study of how brand and production location interact to determine the quality of a variety.

This relationship is most stark when studying Mexico: varieties produced in Mexico by non-Mexican firms and exported to the USA have an average quality rank of 63 and yet Mexican varieties produced by Mexican firms and exported to the USA have an average quality rank of only 35[41]. Similar results, albeit less dramatic, hold for Chinese exports. For wealthier European countries, I find that the highest quality goods sold by firms associated with these countries are in fact produced in the USA. The key takeaway is that only Columns (1) and (4) will be visible to the researcher when using customs data - indeed they will often be indistinguishable - even though for wealthy countries it is Columns (2) and (3) which provide the most important part of the quality story. Estimating quality ladders using customs data may therefore underestimate the difference in quality between wealthy and poor countries - the estimates for poor countries will be buoyed by the often overwhelming presence of multinationals and the estimates for wealthy countries may be dampened by the fact that their highest quality "exports" to the USA are in fact produced there. Given that production can relocate, it is important to disentangle "fundamental" country-specific productivity from the productivity associated with firms producing in any given location.

These results serve to reinforce one of the key conceptual contributions of this paper: imported varieties consist of a set of attributes including a brand and that brand's perceived quality. Varieties from different origin countries have attributes that reflect both their country of origin but also the origin of the firm producing them. Thus researchers should exercise caution when estimating models that assume *a priori* any nesting structure based on a dichotomy - foreign versus domestic - which may overstate the relevance of a given good's production location.

5 Distributional Costs of Tariffs

The following two sections discuss welfare and substitution outcomes associated with changes in import prices across a range of policies. In each case, a change in trade policy will simply be represented by some percentage change in the price of certain varieties[42]. In order to estimate the welfare effects of specific tariff policies, I use the standard discrete choice approach from Small and Rosen [1981]. Define the aggregate uncompensated demand for product k by households of type h

[41] It is impossible to tease out this "fundamental" quality ranking for Morocco, the Philippines, and Thailand as these countries do not have any domestic brands producing domestically and exporting to the USA. All Slovakian exports are associated with German firms.

[42] As an example, a 10% increase in the tariff rate applied to Chinese imports would simply be modeled as a 10% increase in the price of all Chinese varieties in the dataset. Note that this precludes any pass-through effects associated with wholesalers or price changes by producers.

as Q_{hk}. Aggregate utility for that product-household type pair can then be written as:

$$U_{hk} = -\frac{Q_{hk}}{\hat{\alpha}_{hk}} \left[\ln \left(\sum_{j \in J_{hk}} e^{\hat{\delta}_{hj}} \right) \right]$$

where $\hat{\delta}_{hj}$ represents the mean estimated utility associated with variety j for households of type h. In order to calculate changes in uncompensated demand, I assume that households have some Cobb-Douglas aggregate utility function across categories, which implies that expenditure within each category remains constant[43].

Distributional Costs of Tariffs: Category-specific changes in welfare associated with a 10% increase in the price of all imported varieties leads to welfare decreases ranging from 0.6% to 3.0%. These results are intuitive given that imports generally constitute around 15% of expenditure in the categories studied. When aggregating to include the entire consumption basket these welfare effects register at around 0.2% which is again expected as these categories only constitute $240 of expenditure per household-year. In general, these policies are regressive in the sense that poorer households spend a greater share of expenditure on the affected categories[44]. However, the slope of regressivity depends crucially on the specifics of the policy being implemented, and I turn to these distributional outcomes in the next section.

Table 3 provides an overview of how targeting different origin countries with a 10% tariff leads to differential outcomes across US household types. Each column represents a separate tariff target and the rows represent the two dimensions of household heterogeneity studied in this paper: income and population density of ZIP code. Each cell provides the relative welfare costs of the top three income (population density) household types compared to the three lowest income (population density) household types. The top panel presents tariffs based on production location while the bottom panel supposes that a government could feasibly tax all varieties associated with specific foreign firms, even if the varieties in question were produced domestically. All results are weighted by the relative population size of each household type.

The goal of this exercise is to study relative outcomes of different tariff policies across household

[43] If the aggregate utility function can be expressed as $\ln U_h = \sum_k \kappa_{hk} \ln U_{hk}$, then a change in U_{hk} for any k is scaled by the Cobb-Douglas exponent κ_{hk}.

[44] The categories studied in this paper often have expenditure shares for poorer households that are double those of wealthier households. This is somewhat greater than the aggregate share of expenditure on tradeables discussed in Carroll and Hur [2020], who find that the difference in aggregate tradeable expenditure shares to be 36% for poor households and 33% for wealthier households.

Table 3: Relative Welfare Costs of Origin-Specific 10% Price Increase

		(1) All Imports	(2) NAFTA	(3) Europe	(4) China	(5) Top Half	(6) Bottom Half
By Production Location:							
Income	Top 3 / Bottom 3	1.00	1.02	1.11	0.78	1.12	0.84
Pop. Density	Top 3 / Bottom 3	1.04	1.06	1.16	0.82	1.11	0.95
By Design Location:							
Income	Top 3 / Bottom 3	1.06	0.87	1.08	0.68	1.07	1.03
Pop. Density	Top 3 / Bottom 3	1.03	1.08	1.09	1.10	1.04	0.99

Table 3 provides relative welfare costs across household types associated with a 10% tariff. The first row provides the relative welfare costs for the top three income types versus the lowest three income types, and the second row provides a similar analysis but for population density. All estimates are relative to a flat 10% sales tax. Each cell should then be read as a statement of how taxing goods from specific origins lead to outcomes that are either more or less progressive and/or anti-rural than a flat 10% sales tax on all varieties. The column "Top Half" refers to a 10% increase in the price of all imports from countries in the top half of the income distribution (other than NAFTA), and the column "Bottom Half" provides the same analysis but for the lowest income half of import origins. The top panel provides estimates for tariffs by production location, and the bottom panel provides estimates for tariffs by design location.

types. I therefore create a double comparison for each counterfactual policy: I estimate the welfare cost of each specific tariff change as well as the costs associated with a flat 10% sales tax increase on all varieties. I then compare across household types the extent to which country-specific price changes affect certain households more than others, relative to a flat sales tax. Table 3 can therefore be read in the following context: if a government were required to raise a fixed revenue through either sales taxes or tariffs, how would targeting specific origin countries with tariff increases lead to differential welfare outcomes across household types when compared to the default option - a flat sales tax.

I find that a 10% tariff increase on all import origins by production location is effectively neutral in terms of the distribution of costs across household types. While such a tariff would have 4% higher relative costs for urban versus rural households, this policy is neutral across income types. Similar results hold for a tariff increase on NAFTA - such a policy would be slightly progressive and anti-urban but these relative cost differences are, respectively, 2% and 6%. Aggregating costs across the top/bottom three population density types masks some of the heterogeneity associated with these policies. The most urban household type has a relative welfare cost compared to the most rural household type of 14% for tariffs on all imports and 17% for tariffs on NAFTA. As discussed in Section 2, however, household heterogeneity in terms of tariff exposure is mainly driven by country-specific expenditure differences, rather than aggregate import expenditure.

With respect to a 10% tariff increase on all European varieties, the top three income types have an average welfare cost that is 11% greater than the poorest three income types. The three most urban household types have a welfare cost that is 16% greater than the three most rural household types. The difference between the most urban and the most rural household types is even greater, with the most urban households facing relative welfare costs that are 39% greater than for the most rural households[45].

All of these results are flipped when studying a tariff of 10% applied to Chinese imports. The three wealthiest household types face an average welfare cost that is 22% *lower* than the three poorest household types, and the most urban household types face a welfare cost that is 18% *lower* than the three most rural household types. In the extreme, I find that households of type (5) face welfare costs that are 61% greater than households of type (9). Referring to Table A.3, households of type (5) are low-income and rural, whereas households of type (9) constitute the highest-income household type in the data. Columns (5) and (6) provide aggregate analyses for a tariff change on both the upper half of origin countries by income and the lower half, excluding Canada and Mexico. Similar patterns hold in that a tariff on the wealthiest origin countries is progressive and anti-urban, whereas a tariff on the poorest half of origin countries is regressive and anti-rural.

The pattern of distributional costs associated with tariffs targeting varieties based on design origin differ slightly from tariffs based on production origin[46]. These results are shown in the bottom panel of Table 3. I find that tariffs on all varieties sold by foreign brands are slightly more progressive and anti-urban than a tariff on all imports by production origin. These results suggest that wealthy and urban households consume foreign brands to a greater extent than they consume imported varieties. Similar results hold for tariffs on Europe and all other high-income countries: I find that when these tariffs are applied by design origin the extent of progressivity is mitigated. This result suggests that European firms may be tailoring their production of varieties in Mexico and the USA in order to reach a broader consumer base, whereas varieties produced in Europe are, to a greater extent, purchased by wealthier households. Lastly, I find that a tariff on all imports belonging to domestic firms is regressive and anti-rural (0.94 and 0.97, respectively) which suggests that many American firms off-shore production in order to import lower quality varieties targeting lower income households. This story also matches the quality estimates provided in Section 4.4.

[45] It should be noted that the two household types at the extreme ends of the rural/urban divide in my data have almost identical incomes. These are household types 2 and 6 in Table A.3. These results emphasize that conditional on income, geography is a key determinant of tariff exposure.

[46] I will continue to use the word "tariff" even though many of the varieties targeted are in fact produced domestically.

As a final exercise in studying the distributional costs of different tariff policies, I map these counterfactual welfare results directly into US counties. Specifically, I aggregate across all households within a county using the NielsenIQ projection weights in order to arrive at a county-specific average welfare cost[47]. Figure A.6 to Figure A.11 in Appendix A.1 provide, respectively, maps of county-specific relative welfare costs across the contiguous USA for a 10% increase in prices of all imports, NAFTA origin imports, European varieties, Chinese varieties, varieties from origin countries in the top half of GDP per capita, and varieties from origin countries in the bottom half of GDP per capita. These results are all based on the production origin of varieties.

These maps show the strong geographic divide across households in terms of who pays for tariffs on specific origin countries. While the costs of a tariff on all imports as well as policies specifically targeting NAFTA, Europe, and wealthier countries fall predominantly on the major urban centers of the USA, a tariff on Chinese varieties is borne almost entirely by rural counties. These maps illustrate two key points. First, these maps suggest that tariffs on different origin countries have vastly different outcomes across households. As an example, consider the correlation in country-specific costs between the welfare costs of a tariff on the wealthiest half of origin countries versus a tariff on the poorest half of origin countries. This correlation is -0.50, which highlights that the counties paying the most for tariffs on wealthy countries are almost the exact opposite of the counties which pay the most for a tariff on poor countries. When comparing which counties pay for European tariffs and Chinese tariffs, the correlation in costs is -0.59. The key takeaway is that policymakers must be aware of how tariffs to any specific origin are likely to have significant distributional consequences across consumers within their country.

The second key point of interest relates to the US-China trade war of 2018. The categories studied in this paper were all included in the third round of tariffs introduced by the US government on Chinese imports. These tariffs were increased by 10ppt in September, 2018, and then increased by a further 15ppt in May of 2019. The maps just discussed show a striking relationship between counties which are most exposed to the costs of exactly this tariff policy and support for President Trump. Figure A.12 and Figure A.13 provide estimates of a simple regression at the county level. The dependent variables are the county-specific welfare costs of a 10% tariff on, respectively, Europe and China. These estimates are then regressed on deciles of county-level Trump vote share in the

[47] For counties with less than 10 households in the NielsenIQ data, I estimate the relative welfare costs associated with that county using the results from all other counties. I do this by estimating a linear model with county welfare costs as the dependent variable, and county-specific income and population density as the independent variables. I then predict the welfare costs of those counties not covered by NielsenIQ, which account for approximately 10% of all counties.

2016 US Presidential election[48]. I find that a 10% tariff on Chinese varieties has welfare costs that increase monotonically with the level of Trump support in each county. At the extremes, counties in the highest decile of Trump support have welfare costs that are over 15% greater than counties in the lowest decile of Trump support. Interestingly, had Trump decided to place a 10% tariff on European varieties instead, the effect would have been reversed: those counties with the greatest Trump support would have faced the lowest relative costs, and vice versa.

This section has highlighted the extent to which differing tariff policies have differential outcomes across US households. However households also differ in their response to these price shocks and these differences follow systematic patterns. As discussed in Section 4, wealthy and urban households generally have more inelastic demand than poorer and more rural households. Since the patterns of exposure to trade policy differ along these dimensions as well, it follows that tariff policies targeting different origin countries will also differ in their incidence on American households and, by extension, foreign producers. I discuss these patterns in the following section.

6 Welfare and Substitution

This section provides an overview as to how varying tariff policies differ along two dimensions: their ability to raise government revenue and their ability to generate substitution towards domestically produced goods.

Substitution & the Marginal Value of Public Funds: This section discusses the efficacy with which different policies raise revenue and/or generate substitution towards domestically produced goods. In order to provide a meaningful comparison across different tariff policies, I make use of two metrics. The first is the Marginal Value of Public Funds (MVPF) as outlined in Finkelstein and Hendren [2020][49]. The second metric is the elasticity of domestic substitution with respect to aggregate welfare costs - that is, the percentage change in market share of domestically produced varieties divided by the percentage change in US consumer welfare. I denote this elasticity as Ω, with larger values of Ω denoting policies which lead to greater domestic output at a lower cost to consumers. Table 4 provides both measures for a number of different policies.

As one might expect, all tariff policies have an MVPF greater than that associated with a sales

[48] These estimates control for state-level fixed effects and standard errors are clustered at the state level.

[49] This is simply a measure of the welfare costs per dollar revenue generated. The denominator is aggregate revenue generated. The numerator is the compensation required to return consumers to their pre-policy utility.

Table 4: Revenue and Substitution Effects of Tariff Policies

	(1) Sales Tax	(2) All Imports	(3) NAFTA	(4) Europe	(5) China	(6) Top Half	(7) Bottom Half
MVPF	1.21	1.24	1.22	1.25	1.59	1.23	1.36
Ω	0.00	0.61	0.65	0.44	0.51	0.49	0.54

Table 4 provides the marginal value of public funds (MVPF) and the elasticity of domestic substitution with respect to welfare costs (Ω) for different tariff policies, represented as columns. All policies represent a price increase of 10% for each respective origin. Column (6) represents a 10% tariff increase for all varieties from non-NAFTA origin countries in the top half of GDP per capita. Column (7) provides the same but for all origin countries in the bottom half of GDP per capita. The elasticity Ω is simply the percentage change in domestic market share divided by the percentage change in utility for US consumers associated with a given policy.

tax. This result arises from the fact that one must compensate consumers for substituting away from their preferred imported varieties in the case of a tariff whereas there is no substitution under a flat sales tax (hence $\Omega = 0$ for a sales tax). To put these numbers in context, Finkelstein and Hendren [2020] show that the range of MVPF estimates associated with raising revenue through an income tax is generally between $1.20 - 1.50$. In general, I find that tariffs targeting higher-income origin countries (Columns (4) and (6)) generate revenue at a lower cost when compared to tariffs targeting lower-income countries (Columns (5) and (7))[50]. At the same time, tariffs targeting wealthier countries tend to generate less domestic substitution per unit cost than tariffs targeting lower income countries. In fact Columns (4) - (7) suggest that there is an inherent trade-off between generating revenue and generating domestic substitution.

This result is intuitive when one considers the representative consumer of each origin country: high-income and urban households are less sensitive to price and more likely to purchase varieties from wealthier origin countries. This in turn suggests that demand for European varieties is more inelastic. Since wealthier American consumers are less willing to substitute away from these import price changes, these policies are effective at generating government revenue but less effective at generating substitution. A corollary of this statement is that tariffs on wealthy origin countries will carry a higher incidence on domestic consumers than a tariff levied on poorer countries. By decomposing welfare costs into a substitution effect and an income effect, I find that the share of welfare costs associated with substitution is generally 12% higher for tariffs on Chinese varieties

[50] Finkelstein and Hendren [2020] use results from Amiti et al. [2019] to estimate the MVPF associated with the 2018 tariff hike by the US on Chinese imports. They estimate a MVPF of 1.50 which is remarkably similar to the estimate found in this paper of 1.59. These results offer a useful comparison given that Amiti et al. [2019] find a price pass-through of nearly 100% associated with these policies.

versus European varieties.

Interestingly, the estimates for NAFTA suggest that this policy can both generate revenue at a welfare cost similar to a flat sales tax *and* generate greater substitution towards American varieties than any other tariff policy studied in Table 4[51]. In fact so long as the objective function of a policy-maker features some convex combination of raising revenue and generating domestic substitution, then a tariff on NAFTA imports is likely more attractive to this policy-maker than raising the sales tax. While this result is of course dependent on such policies being enacted unilaterally, the cost-effectiveness of a tariff on NAFTA imports marks an important feature of the demand system used in this paper, and warrants further discussion.

The substitutability of foreign varieties is crucial to understanding the costs to domestic consumers of changes in trade policy. Within the nested logit framework of this paper, aggregate substitution is largely determined by the dispersion of varieties across nests. So long as consumers have *within-nest* alternatives substitution is higher and costs are lower. The remainder of this section provides a simple empirical exercise to show that varieties from Mexico and Canada are more highly dispersed across market segments (nests) within the US than varieties from other origins, and that this is driven by the presence of US multinationals producing in Mexico and Canada and exporting back to the US market.

Consider the number of nests, as determined by the clustering process described in Section 3.3, for a given product category k: N_k^g. For any country c, define the set of nests which contain at least one variety from c as N_{kc}^g. Equation 3 then provides a simple model to study how standard gravity variables (distance, GDP, GDP per capita, contiguity, and common language) lead to a greater or lower dispersion of varieties from c across market segments.

$$N_{kc}^g = \alpha + f(N_{kc}) + \boldsymbol{\beta^{gra}} \boldsymbol{X_c^{gra}} + \phi_k + v_{kc} \tag{3}$$

In order to isolate dispersion, Equation 3 includes the aggregate count of varieties from country c sold in product k (N_{kc}) and I model this relationship in various ways[52]. The estimates $\hat{\boldsymbol{\beta}}^{\boldsymbol{gra}}$ therefore represent the effect of gravity variables $\boldsymbol{X_c^{gra}}$ on the dispersion of varieties across nests conditional on the aggregate number of varieties from c sold in the US, and I estimate this model

[51] As a comparison, a tariff targeting all imported varieties from non-NAFTA countries yields $MVPF = 1.29$ and $\Omega = 0.56$, both of which are less attractive, from a policy standpoint, than the associated estimates for a NAFTA tariff.

[52] These include: a simple linear relationship, a third-order polynomials, and a category-specific linear relationship. All three approaches are provided in Table 5.

Table 5: Gravity Variables and the Dispersion of Imported Varieties across Market Segments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance	−0.330***	0.060	−0.045	0.105	−0.110	0.254**	−0.115	−0.163*	−0.180
	(0.075)	(0.106)	(0.042)	(0.080)	(0.070)	(0.098)	(0.128)	(0.097)	(0.121)
GDP	0.292***	0.285***	0.143***	0.145***	0.188***	0.179***	0.251**	0.132***	0.248***
	(0.060)	(0.046)	(0.031)	(0.030)	(0.039)	(0.041)	(0.042)	(0.032)	(0.043)
GDP p.c.	−0.096	0.024	−0.023	0.010	−0.012	0.098*	0.034	−0.087*	0.052
	(0.064)	(0.058)	(0.042)	(0.045)	(0.053)	(0.053)	(0.064)	(0.046)	(0.066)
Contiguity		0.944***		0.313*		0.923***	0.034	−0.184	0.013
		(0.189)		(0.170)		(0.202)	(0.236)	(0.162)	(0.238)
Language		0.007		0.116		0.008	−0.033	0.023	−0.024
		(0.120)		(0.088)		(0.109)	(0.122)	(0.094)	(0.122)
N_{kc}	0.005***	0.005***					0.030***		
	(0.002)	(0.001)					(0.002)		
N	213	213	213	213	213	213	150	150	150
R^2	0.634	0.652	0.849	0.859	0.817	0.839	0.806	0.904	0.867
ϕ_k	X	X	X	X	X	X	X	X	X
$f()$: Linear	X	X					X		
$f()$: Polynomial			X	X				X	
$f()$: Category-specific					X	X			X
Drop US IM							X	X	X

Table 5 provides parameter estimates of the model outlined in Equation 3 and estimated using PPML. The dependent variable N^{γ}_{kc} for all models represents the number of market segments within category k that have a non-zero number of varieties from country c. The standard gravity variables are included in logarithms, whereas the variety count of barcodes from country c in category k N_{kc} is included as a count. Columns (1) - (6) are estimated using the full dataset, whereas Columns (7) - (9) drop all imported varieties belonging to American firms. Columns (1), (2), and (7) model the relationship between N_{kc} and N^{γ}_{kc} as an aggregate linear model. Columns (3), (4), and (8) model this relationship as a third-order polynomial. Columns (5), (6), and (9) model this relationship as a category-specific linear relationship. All models include category-specific fixed effects ϕ_k. Columns (2), (4), and (6) include the additional regressors of contiguitiy and common language. Standard errors are clustered at the category level and provided in parentheses. (***) p<0.01: (**) p<0.05: (*) p<0.10.

using PPML.

Table 5 provides estimates of various specifications of Equation 3. Columns (1) - (6) provide estimates for the aggregate dataset, whereas in Columns (7) - (9) I drop all imported varieties produced abroad by US multinationals. That is - to follow the terminology of this paper - all varieties designed in the US but produced abroad. The key result to take away from Table 5 is that the dummy variable for contiguity – essentially a dummy variable for Mexico and Canada – is significant and positive when the entire dataset is included, but this effect completely disappears once American multinational imports are removed from the dataset. In other words imported varieties from Mexico and Canada are dispersed across the attribute space to a greater extent than one would expect given the volume of varieties imported from Mexico and Canada, but this

dispersion is driven almost entirely by the presence of American multinationals. This dispersion in turn decreases the cost of tariffs as it implies that these imports consist of attributes which are common in the domestic market.

The notion that multinationals may allow exports from a given country to enter new market segments is a relatively understudied phenomenon. Fajgelbaum et al. [2011] and Fajgelbaum et al. [2015] show that a Linder-style model of FDI predicts increased FDI between countries of similar income levels due to demand for quality and they provide empirical evidence to support this result. While this paper also finds that the majority of production by foreign firms within the US consists of firms from similarly wealthy countries, the findings in this section suggest an alternative impetus for FDI. If large multinationals are able to overcome country-specific productivity shortcomings due to their own high productivity, then these large firms may be able to take advantage of factor price differences across countries and match market-segment-specific technology to favourable factor prices. For example, when producing vertically differentiated products a large US multinational might produce high quality varieties in the US where skilled labour is relatively abundant but low-quality varieties in Mexico where unskilled labour is relatively abundant. The fact that, in aggregate, Mexico may have lower productivity than the US can be overcome by the productivity that is inherent to the firm, while the scale of this productive firm allows it to cover the fixed costs of FDI.

Notice that when generalized to allow for market segments which differ both horizontally and vertically, this sketch of a model would suggest that countries with a large number of US multinationals producing off-shore for export back to the US would in turn increase the average substitutability of imported goods as these varieties would enter into a broader set of market segments . This paper therefore suggests an alternative mechanism driving the model discussed in Lind and Ramondo [2018]. They provide a model in which cross-country technology may be correlated for countries which are close to each other due to technology dispersion across space. This paper provides evidence that this technology dispersion may in fact represent FDI for export back to the firm's domestic country, in this case the USA, and that the subsequent increase in substitutability has consequences for the outcomes of tariff policy.

The findings discussed in this section emphasize the key conceptual framework of this paper: imported varieties consist of attributes and will carry larger gains for consumers if their attributes are scarce in the domestic economy. In practice, however, the attributes of a good are determined by both the firm which produced that good and the country in which the good was produced. To

the extent that trade is mainly driven by large firms matching blueprints to factor prices via FDI, then the substitutability of imported goods must be estimated with caution. The final exercise in this paper consists of directly comparing the welfare costs of tariffs under an Armington nesting structure versus the substitution model used in this paper. I turn to these results now.

Comparison with Armington: This paper argues that import substitution models based on the production origin of varieties may be mis-specified. Imported varieties are often designed by domestic firms and domestically produced varieties are often designed by foreign firms. The key question for this section is whether or not such model mis-specifications lead to meaningfully different economic outcomes.

Figure 2 highlights the relative welfare effects of using the Armington nesting structure versus the attribute-clustered model. Each point on the graph represents a triplet of a tariff rate, a household type, and a product category. For each household type and product category, I estimate the welfare costs of four tariff increases on all imported varieties: 10%, 25%, 50%, and 100%[53]. The y-axis provides the ratio of welfare costs under the Armington structure to welfare costs using the attribute-cluster model. The x-axis provides the median own-price elasticity across all varieties within each household-category pair. For each policy, I have included a lowess-estimated fit along with 99% confidence intervals.

There are a number of key findings to discuss from Figure 2. First, I find that as the median own-price elasticity of a category increases, the extent to which the Armington nesting structure overestimates the costs of a tariff change also increases. This result is intuitive: a larger own-price elasticity suggests that demand for this category is price-sensitive. This in turn leads to larger welfare costs when the ability of consumers to substitute away from imports is constricted, as in the Armington nesting structure. This relationship is useful in that the categories studied in this paper are generally in the more inelastic range, and while caution should of course be exercise when extrapolating, I find that for a 10% tariff on a category with a median elasticity of 2.5[54] an Armington nesting structure would lead to welfare estimates 30%-40% greater than those found using the attribute-cluster model.

In addition, I find that for a given own-price elasticity, a larger tariff increase leads to a pro-

[53] When estimating welfare costs of these policies, I keep all parameters the same between the attribute-cluster nesting structure and the Armington nesting structure. These results then serve to highlight the role of the nesting structure only. I find that estimating a full model under the Armington assumption often leads to parameter estimates incompatible with utility maximization in the nested logit model and so I opt for this comparison instead.

[54] Ready-to-eat cereal own-price elasticities are often estimated to be in the ballpark of 2.5.

Figure 2: Comparison of Welfare Costs in Armington versus Clustering

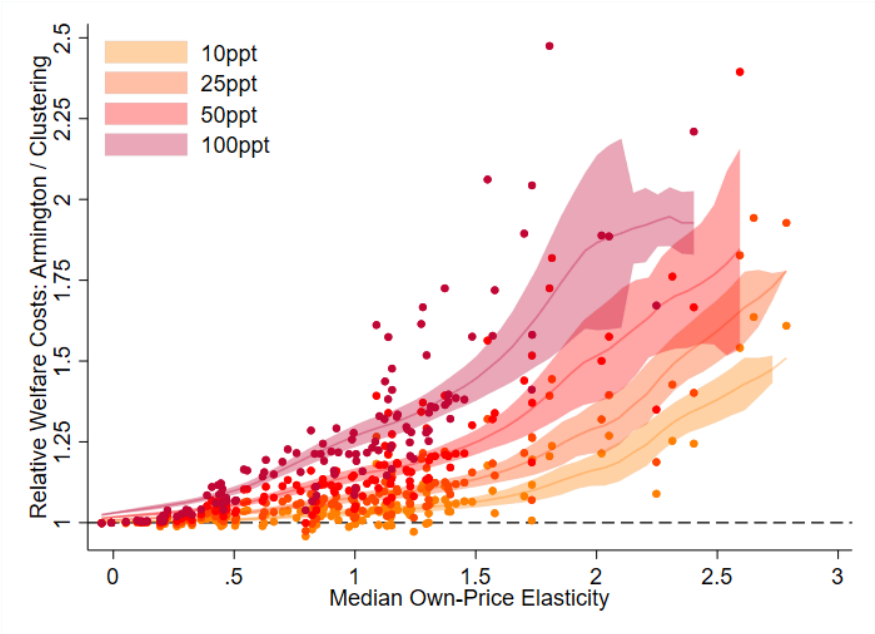

Figure 2 provides relative welfare costs of various tariff rate increases on all imported varieties as calculated using the Armington nesting structure and the clustering nests used in this paper. The y-axis provides relative welfare costs of these policies under the two configurations. Each observation is a household-type-category-tariff-rate triple ($hk\tau$). The x-axis provides the median own-price elasticity for each hk-pair. A lowess-estimated line is provided using an Epanechnikov kernel with a bandwidth of 0.25 as well as 99% confidence intervals.

portionately larger overestimate of welfare costs. As an example, for a category with a median own-price elasticity of 1.5, the overestimate of welfare costs effectively doubles for every tariff increase studied in Figure 2. These results are again intuitive: as tariffs become more binding in their stringency the desire to substitute becomes all the greater. Consumers forced to comply with an Armington-style Hessian matrix will have less recourse to do so and therefore the costs of increased tariff rates will increase. Table A.5 provides the relative elasticity of aggregate import demand at the category level for the Armington nesting structure and the attribute-clustering nests. As highlighted in Column (3), I find that the elasticities of import substitution associated with the attribute-cluster model are often double the elasticities found when using the Armington structure.

These results provide one of the first attempts to estimate the extent to which an Armington model might overstate the costs of tariffs and, by extension, the gains from trade more generally. While this paper cannot state that attribute-based nests map perfectly into the "true" underlying market segments, the main takeaway is simply that researchers must use caution when placing varieties into somewhat arbitrary *a priori* nests of strong and weak substitutes. If one takes the

attribute models of industrial organization seriously - such as the nested logit or mixed logit demand systems - then one can only justify the use of nesting varieties as "domestic" and "foreign" if one believes that this dichotomy constitutes the defining attribute of varieties in the eyes of consumers. Similarly, one must assume that an estimate of the elasticity between these two nests represents a meaningful economic parameter. While it may be that both of these assumptions are true, this section argues that these assumptions are far from innocuous and may lead researchers to substantially over-estimate the gains from trade.

7 Conclusion

A lack of detailed data has hindered efforts to establish credible estimates of the distributional costs of tariffs. This paper makes use of a novel dataset in order to make headway on the two key components of any welfare calculation in this setting: the distribution of country-specific expenditure across households and the assumptions underlying models of import substitution.

I find that tariffs in general have higher relative costs for urban centers as opposed to rural counties, however these differences are small. Tariffs on high-income countries are both progressive and anti-urban, with the opposite statement holding for tariffs on low-income countries. Tariffs on Chinese imports specifically target rural and low-income households, as these households are most likely to shop at dollar stores which enhances their exposure to such policies, conditional on a battery of household characteristics. I show that standard Armington models of substitution may be misspecified when estimating the costs of tariff policy. I find that a model of import substitution based on attribute similarity more readily fits the underlying consumption data when compared to an Armington-style model. This has implications for our understanding of the gains from trade as the attribute-based model estimates that imports are more substitutable with domestically produced alternatives than one might assume under an Armington model.

This paper provides groundwork for a plethora of future research options. The most pressing relate to the extent to which retail markets dictate the consumption baskets of households, how firms choose which varieties to produce in varying countries within narrow product categories, and the extent to which the results concerning over-estimates of the gains from trade in standard models extend to other sectors.

References

Milena Almagro and Elena Manresa. Data-driven nests in discrete choice models. *No full draft available*, Slides, 2021.

Vanessa Alviarez. Multinational production and comparative advantage. *Journal of International Economics*, 119:1–54, 2019.

Vanessa I Alviarez, Javier Cravino, and Natalia Ramondo. Firm-embedded productivity and cross-country income differences. Technical report, National Bureau of Economic Research, 2020.

Mary Amiti, Stephen J Redding, and David E Weinstein. The impact of the 2018 tariffs on prices and welfare. *Journal of Economic Perspectives*, 33(4):187–210, 2019.

Mary Amiti, Mi Dai, Robert C Feenstra, and John Romalis. How did china's wto entry affect us prices? *Journal of International Economics*, 126:103339, 2020.

Raphael Auer, Ariel Burstein, Sarah Lein, and Jonathan Vogel. Unequal expenditure switching: Evidence from switzerland. 2021.

Liang Bai and Sebastian Stumpner. Estimating us consumer gains from chinese imports. *American Economic Review: Insights*, 1(2):209–24, 2019.

Emek Basker and Pham Hoang Van. Imports "r" us: Retail chains as platforms for developing-country imports. *American Economic Review*, 100(2):414–18, 2010.

Andrew B Bernard, J Bradford Jensen, Stephen J Redding, and Peter K Schott. Wholesalers and retailers in us trade. *American Economic Review*, 100(2):408–13, 2010.

Andrew B Bernard, J Bradford Jensen, Stephen J Redding, and Peter K Schott. The empirics of firm heterogeneity and international trade. *Annu. Rev. Econ.*, 4(1):283–313, 2012.

Andrew B Bernard, J Bradford Jensen, Stephen J Redding, and Peter K Schott. Global firms. *Journal of Economic Literature*, 56(2):565–619, 2018.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

Bruce A Blonigen and Anson Soderbery. Measuring the benefits of foreign product variety with an accurate variety set. *Journal of International Economics*, 82(2):168–180, 2010.

Stéphane Bonhomme and Elena Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. Discretizing unobserved heterogeneity. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2019-16), 2017.

Kirill Borusyak and Xavier Jaravel. The distributional effects of trade: Theory and evidence from the united states. Technical report, National Bureau of Economic Research, 2021.

N Scott Cardell. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, pages 185–213, 1997.

Daniel R Carroll and Sewon Hur. On the heterogeneous welfare gains and losses from trade. *Journal of Monetary Economics*, 109:1–16, 2020.

A Kerem Coşar, Paul LE Grieco, Shengyu Li, and Felix Tintelnot. What drives home market advantage? *Journal of international economics*, 110:135–150, 2018.

Javier Cravino and Andrei A Levchenko. The distributional consequences of large devaluations. *American Economic Review*, 107(11):3477–3509, 2017.

Javier Cravino and Andrei A Levchenko. The geographic spread of a large devaluation. In *AEA Papers and Proceedings*, volume 108, pages 562–66, 2018.

Matthieu Crozet, Keith Head, and Thierry Mayer. Quality sorting and trade: Firm-level evidence for french wine. *The Review of Economic Studies*, 79(2):609–644, 2012.

Jean-Pierre Dubé, Ali Hortaçsu, and Joonhwi Joo. Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*, 2021.

Benjamin Faber and Thibault Fally. Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *Review of Economic Studies*, 2021.

Pablo Fajgelbaum, Gene M Grossman, and Elhanan Helpman. Income distribution, product quality, and international trade. *Journal of political Economy*, 119(4):721–765, 2011.

Pablo Fajgelbaum, Gene M Grossman, and Elhanan Helpman. A linder hypothesis for foreign direct investment. *The Review of Economic Studies*, 82(1):83–121, 2015.

Pablo D Fajgelbaum and Amit K Khandelwal. Measuring the unequal gains from trade. *The Quarterly Journal of Economics*, 131(3):1113–1180, 2016.

Robert C Feenstra, Philip Luck, Maurice Obstfeld, and Katheryn N Russ. In search of the armington elasticity. *Review of Economics and Statistics*, 100(1):135–150, 2018.

Amy Finkelstein and Nathaniel Hendren. Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–67, 2020.

Surabhi Ghai and Colin Hottman. Exchange rates, product variety, and substitution in us scanner data. In *2019 Meeting Papers*, number 1059. Society for Economic Dynamics, 2019.

Pinelopi Koujianou Goldberg. Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica: Journal of the Econometric Society*, pages 891–951, 1995.

Paul E Green, Jonathan Kim, and Frank J Carmone. A preliminary study of optimal variable weighting in k-means clustering. *Journal of Classification*, 7(2):271–285, 1990.

Laura Grigolon and Frank Verboven. Nested logit or random coefficients logit? a comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics*, 96(5):916–935, 2014.

Juan Carlos Hallak and Peter K Schott. Estimating cross-country differences in product quality. *The Quarterly journal of economics*, 126(1):417–474, 2011.

Jessie Handbury. Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities. *Econometrica*, 89(6):2679–2715, 2021.

Matthew Harding and Michael Lovenheim. The effect of prices on nutrition: comparing the impact of product-and nutrient-specific taxes. *Journal of Health Economics*, 53:53–71, 2017.

Jerry Hausman, Gregory Leonard, and J Douglas Zona. Competitive analysis with differenciated products. *Annales d'Economie et de Statistique*, pages 159–180, 1994.

Keith Head and Thierry Mayer. Brands in motion: How frictions shape multinational production. *American Economic Review*, 109(9):3073–3124, 2019.

Colin Hottman and Ryan Monarch. Estimating unequal gains across us consumers with supplier trade data. *Journal of International Economics*, 2021.

Dawn Iacobucci, Phipps Arabie, and Anand Bodapati. Recommendation agents on the internet. *Journal of Interactive Marketing*, 14(3):2–11, 2000.

Amit Khandelwal. The long and short (of) quality ladders. *The Review of Economic Studies*, 77 (4):1450–1476, 2010.

Paul Krugman. Scale economies, product differentiation, and the pattern of trade. *The American Economic Review*, 70(5):950–959, 1980.

Paul R Krugman. Increasing returns, monopolistic competition, and international trade. *Journal of international Economics*, 9(4):469–479, 1979.

Lerong Li. Trade policy shocks and consumer prices. *Available at SSRN 3467808*, 2019.

Nelson Lind and Natalia Ramondo. Trade with correlation. Technical report, National Bureau of Economic Research, 2018.

Daniel McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4): 303–328, 1974.

Marc J Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.

Aviv Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2): 307–342, 2001.

Girish Punj and David W Stewart. Cluster analysis in marketing research: Review and suggestions for application. *Journal of marketing research*, 20(2):134–148, 1983.

Natalia Ramondo. A quantitative approach to multinational production. *Journal of International Economics*, 93(1):108–122, 2014.

Peter K Schott. The relative sophistication of chinese exports. *Economic policy*, 23(53):6–49, 2008.

Kenneth A Small and Harvey S Rosen. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, pages 105–130, 1981.

Felix Tintelnot. Global production with export platforms. *The Quarterly Journal of Economics*, 132(1):157–209, 2017.

Quang H Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

A Appendix

This section provides four appendices to the previous chapter. Appendix A.1 provides additional tables and figures in the order in which they were referenced in the main text. Appendix A.2 provides more detail pertaining to the text similarity and clustering algorithms introduced in Section 4. Appendix A.3 provies an overview of the procedure used to model consideration at the variety level (from Section 3.4) and Appendix A.4 provides further details and results related to the specification tests outlined in Section 4.3.

A.1 Additional Tables and Figures

Table A.1: Descriptive Statistics by Product Category

Category	Expenditure *$M*	Quantity *units*	Brands #	Varieties #	Import Share %
Hair Care	3.75	577,628	209	4,492	7.0
Body Soap	2.70	461,470	195	2,718	8.5
Skin Care	2.27	256,086	252	2,820	22.3
Deodorant	2.25	403,441	207	1,702	20.1
Toothpaste	2.03	410,837	33	589	19.7
Dental Care	1.76	262,278	53	602	13.1
Facial Care	1.40	128,627	101	1,417	19.2
Hair Styling	1.20	194,534	110	1,296	15.1
Hand Soap	0.76	191,531	135	1,298	7.9
Face Cosmetics	0.66	71,896	54	2,110	16.6
Eye Cosmetics	0.57	75,452	34	1,172	15.7
Shaving Care	0.42	105,243	39	354	13.5
Lip Cosmetics	0.39	94,346	93	1,064	14.0
Nail Cosmetics	0.26	48,580	90	1,639	16.2
Bath Accessories	0.14	24,725	84	533	17.7

Table A.1 provides summary statistics for the 15 product categories studied in this paper, including aggregate expenditure by all households over three years, aggregate units purchased, the number of unique brands and varieties, and finally the expenditure import share. Back to Section 2.

Table A.2: Descriptive Statistics by Origin Country

	Expenditure %	Imports %
USA	85.55	-
Canada	5.30	36.7
Mexico	4.84	33.5
China	0.80	5.5
Ireland	0.66	4.6
Germany	0.46	3.2
France	0.45	3.1
Australia	0.24	1.7
Israel	0.22	1.5
UK	0.19	1.3
Italy	0.19	1.3
Morocco	0.16	1.1
Spain	0.16	1.1
Thailand	0.14	1.0
Taiwan	0.13	0.9
Philippines	0.08	0.6

Table A.2 provides the aggregate expenditure shares by origin country. Not shown (in descending order of %): India, Belgium, Turkey, Brazil, Switzerland, Luxembourg, Poland, Japan, Finland, Greece, Czech Republic, Netherlands, Slovakia, Sweden, Togo, Jamaica, Hungary, Chile, New Zealand, Guatemala, Ghana, Bulgaria, Malaysia, Tunisia, Peru, Dominican Republic, Sri Lanka, Colombia. Back to Section 2.

Table A.3: Household Type Descriptive Statistics

h	(1) Count	(2) **Exp. per Person**	(3) Income Decile	(4) **Pop. Density**	(5) Density Decile	(6) %White	(7) %Black	(8) %College
	#	***Mean***	*Mean*	***Mean***	*Mean*	*Share*	*Share*	*Share*
1	4,139	**5,243**	3.89	**279**	4.72	0.81	0.10	0.40
2	2,450	**8,661**	5.87	**7,681**	10.00	0.51	0.24	0.39
3	4,536	**12,342**	8.10	**141**	3.57	0.87	0.05	0.47
4	5,657	**13,561**	8.44	**1,540**	8.47	0.74	0.11	0.50
5	3,360	**4,817**	3.46	**79**	2.72	0.86	0.08	0.35
6	3,344	**8,499**	5.85	**12**	1.00	0.91	0.04	0.32
7	5,337	**5,817**	4.44	**772**	6.76	0.74	0.14	0.44
8	4,719	**4,526**	3.20	**1,814**	8.88	0.67	0.16	0.37
9	4,367	**14,359**	8.69	**476**	5.71	0.82	0.08	0.51
10	4,160	**11,513**	7.55	**40**	1.86	0.89	0.04	0.39
Aggregate	42,069	9,110	6.04	1,059	5.50	0.79	0.10	0.42

Table A.3 provides descriptive statistics for the household clusters h used in this paper. Column (1) provides the number of households in each cluster, and Columns (2) and (4) provide the mean expenditure per household member and mean ZIP code population density of households in each cluster. Columns (2) and (4) are in bold as these are the only two dimensions used to construct the types h. These types are constructed using the k-medians clustering algorithm applied to expenditure per household member and ZIP code population density. Columns (3) and (5) provide the same variables as (2) and (4) but in terms of deciles. Columns (6) and (7) provide, respectively, the share of households in each cluster that are white and black. Column (8) provides the share of households in each cluster with at least one household head possessing a college degree. Back to Section 3.2.

Table A.4: Median Elasticity Estimates

Income	3.89	5.87	8.10	8.44	3.46	5.85	4.44	3.20	8.69	7.55
Pop. Density	4.72	10.00	3.57	8.47	2.72	1.00	6.76	8.88	5.71	1.86
h	1	2	3	4	5	6	7	8	9	10
	Median Own-Price Elasticity: ϵ_{jj}									
Category										
Hair Care	-1.37	-1.14	-1.81	-1.82	-1.27	-1.70	-1.55	-1.28	-1.58	-1.73
Body Soap	-1.30	-1.12	-0.72	-0.81	-1.15	-0.99	-1.15	-1.49	-0.63	-0.83
Skin Care	-0.45	-0.39	-0.45	-0.43	-0.44	-0.45	-0.55	-0.44	-0.41	-0.62
Deodorant	-1.05	-0.78	-1.31	-1.21	-1.45	-1.37	-1.40	-1.15	-1.22	-1.42
Toothpaste	-1.10	-0.93	-0.98	-1.10	-1.30	-1.01	-1.39	-1.31	-1.12	-1.18
Facial Care	-0.70	-0.54	-0.76	-1.09	-0.86	-1.00	-0.87	-0.62	-0.92	-0.66
Hair Styling	-1.23	-0.87	-1.39	-1.31	-1.31	-1.15	-1.15	-1.07	-1.24	-1.33
Hand Soap	-1.23	-1.08	-1.17	-1.14	-1.23	-0.47	-1.73	-1.57	-0.90	-1.01
Face Cosmetics	-0.36	-0.10	-0.45	-0.42	-0.30	-0.46	-0.51	-0.30	-0.31	-0.47
Eye Cosmetics	-0.13	0.05	-0.14	0.01	-0.21	-0.11	-0.12	-0.20	0.01	-0.12
Shaving Care	-0.82	-0.80	-1.13	-0.91	-1.00	-1.00	-0.92	-0.84	-1.03	-0.84
Lip Cosmetics	-2.79	-1.74	-2.65	-2.02	-2.05	-2.25	-2.60	-2.41	-2.31	-2.70
Nail Cosmetics	-0.21	-0.19	-0.22	-0.21	-0.50	-0.33	-0.28	-0.26	-0.05	-0.46

Table A.4 provides the median elasticity for each household type and product category. The category of "Bathing Accessories" is omitted due to a weak first-stage. Household characteristics are given in the top two rows. Back to Section 4.

Table A.5: Attribute Clustering Versus Armington Import Elasticity of Demand

Category	(1) $e^{cluster}$	(2) $e^{armington}$	(3) $e^{cluster}/e^{armington}$
Hair Care	-1.30	-0.56	2.31
Body Soap	-0.73	-0.38	1.92
Skin Care	-0.33	-0.15	2.12
Deodorant	-0.80	-0.48	1.68
Toothpaste	-0.74	-0.37	2.01
Facial Care	-0.56	-0.24	2.31
Hair Styling	-0.90	-0.61	1.48
Hand Soap	-0.75	-0.41	1.82
Face Cosmetics	-0.25	-0.18	1.35
Eye Cosmetics	-0.05	-0.05	1.06
Shaving Care	-0.67	-0.51	1.32
Lip Cosmetics	-1.28	-0.74	1.72
Nail Cosmetics	-0.13	-0.09	1.45

Table A.5 provides estimates of the Armington elasticity for different nesting models. These elasticites are the change in market share of all imports with respect to a change in the relative price of foreign varieties. Column (1) provides these estimates for the attribute clustering approach used in this paper. Column (2) provides the same estimates for the "Armington" assumption of nesting foreign and domestic. Column (3) provides the ratio of clustering to Armington. Back to Section 6.

Table A.6: Quality Decomposition by Production and Design Location

Design Production	Domestic (1)	Domestic non-USA (2)	USA (3)	Foreign Domestic (4)
USA	52.0	42.4	-	55.3
	(0.73)	(0.10)	-	(0.17)
Canada	33.7	29.3	33.5	49.9
	(0.02)	(0.00)	(0.02)	(0.95)
Mexico	34.6	11.6	23.4	62.6
	(0.01)	(0.00)	(0.01)	(0.99)
China	27.2	-	26.9	35.5
	(0.01)	-	(0.00)	(0.98)
France	39.4	55.2	62.3	42.2
	(0.03)	(0.18)	(0.72)	(0.07)
UK	41.0	38.2	46.2	37.6
	(0.11)	(0.23)	(0.63)	(0.02)
Switzerland	37.0	48.0	61.6	47.4
	(0.01)	(0.03)	(0.94)	(0.02)
Germany	55.1	56.1	47.8	43.0
	(0.07)	(0.63)	(0.19)	(0.10)
Australia	46.9	67.5	66.4	47.8
	(0.08)	(0.37)	(0.54)	(0.01)
Italy	39.3	39.9	41.5	45.8
	(0.01)	(0.16)	(0.37)	(0.47)

Table A.6 provides estimates of average quality for different countries based on production and design location. Column (1) provides the average quality of varieties design and produced in that country and exported to the USA. Columns (2) and (3) provide average quality for varieties designed in that country but produced in, respectively, any other non-USA country and the USA. Column (4) provides the average quality of foreign designed varieties produced in that country and exported to the USA. Expenditure shares for each country are provided in brackets. Columns (1) and (4) combine to provide what one would typically find in a quality estimate based on customs data. Back to Section 4.4.

Figure A.1: Retail Format Trip Propensity by Population Density

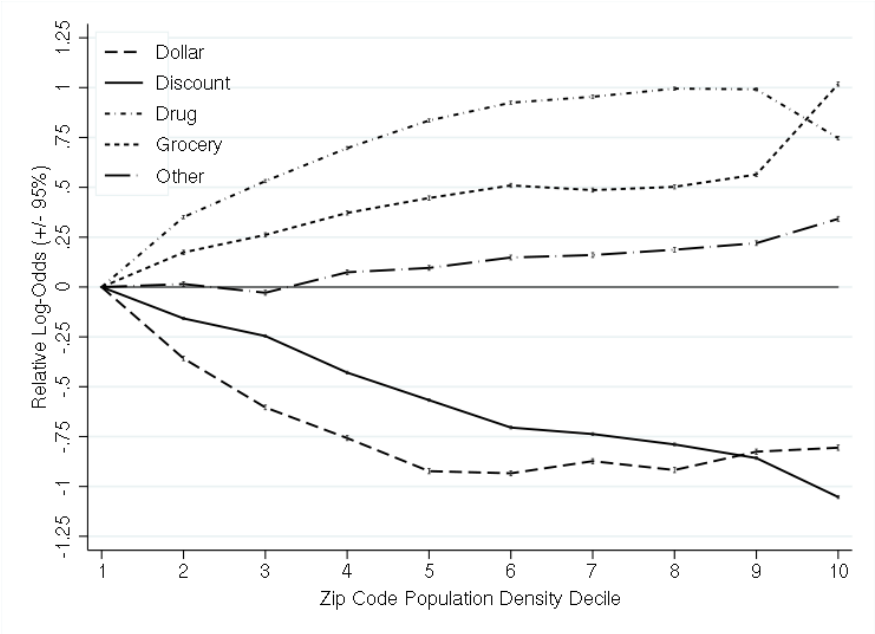

Figure A.1 provides estimates of a logit regression for each retail format. 95% confidence intervals are provided. Each points is an estimate associated with the propensity of a given population density decile (x-axis) to shop at a given retail format. All regressions include controls for household income decile, education, race, age, presence of children, married household heads, region, and ZIP code latitude and longitude. Back to Section 2.

Figure A.2: Origin Purchase Porpensity by Population Density

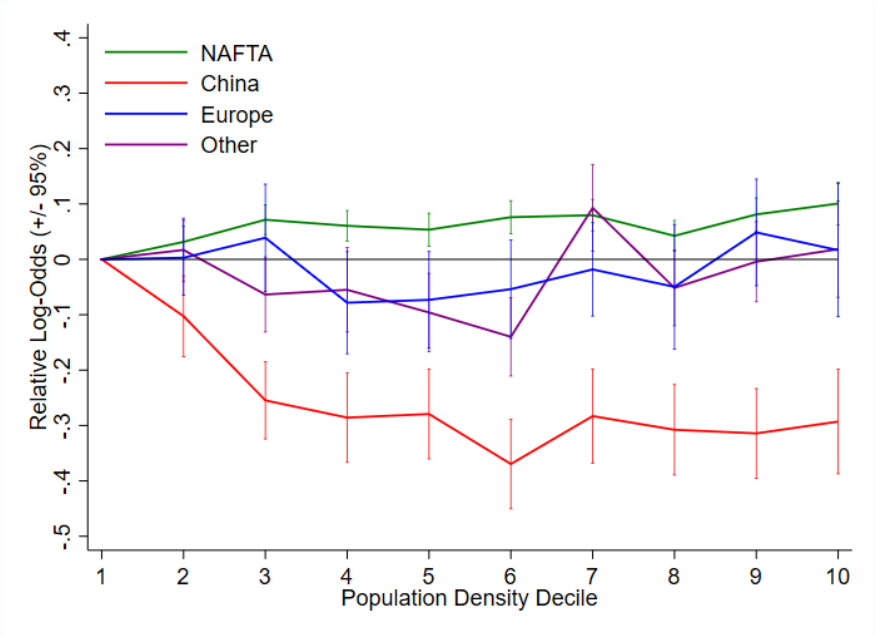

Figure A.2 provides estimates of country-specific purchase propensities across population density deciles. 95% confidence intervals are provided, and standard errors are clustered at the product-region-period level (that is, a (k, r, t) triplet). Each point is an estimate associated with the propensity of a given population density decile (x-axis) to purchase varieties from an origin. Each estimates represents log-odds of a purchase relative to the lowest-income decile. All regressions include controls for income decile, education, race, age, presence of children, married household heads, and ZIP code latitude and longitude. I also include region, half-year, and product category fixed effects. Back to Section 2.

Figure A.3: Distribution of estimates for $\hat{\alpha}$ (top-left), $\hat{\sigma}$ (top-right), $\hat{\phi}_m$ (bottom-left), and median elasticity (bottom-right)

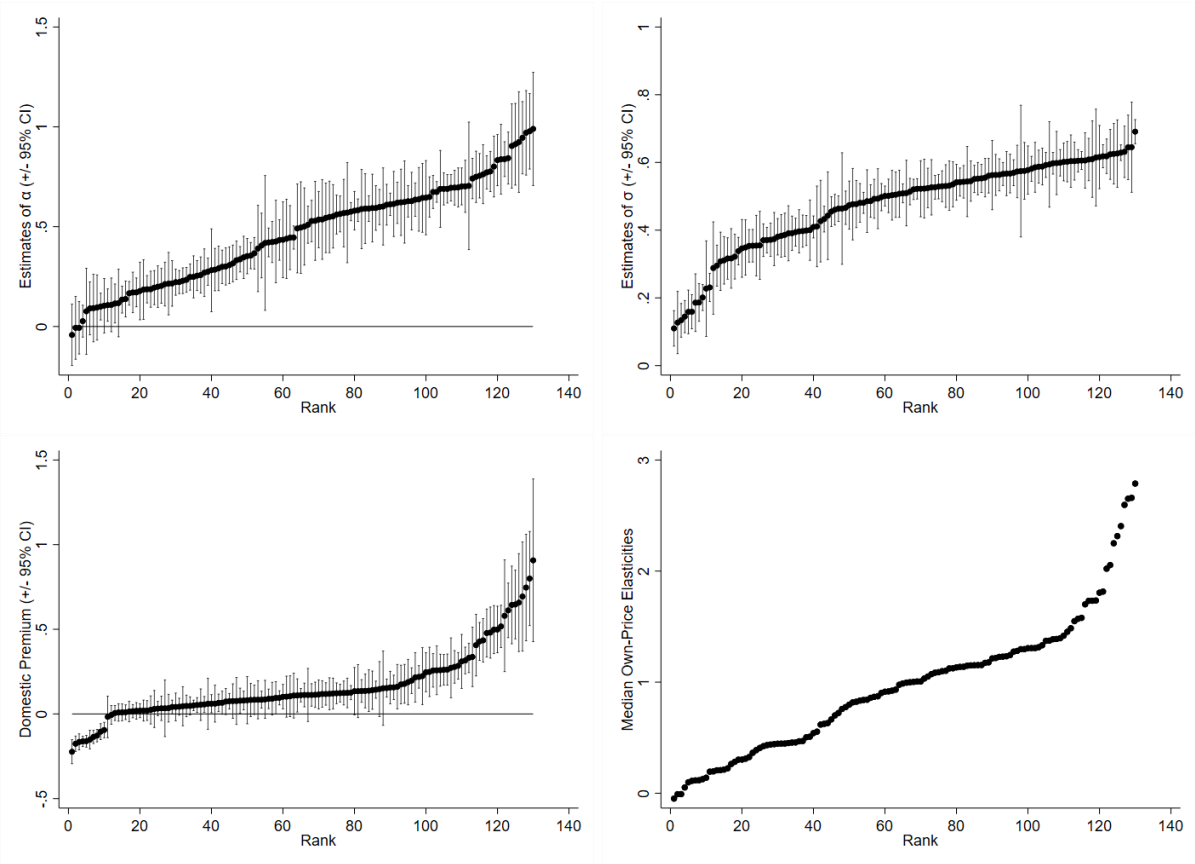

Figure A.3 provides distributions of the key parameters estimated within the demand system. Each figure contains data for 130 estimates (10 household types times 13 product categories). From top-left clockwise, these distributions are for $\hat{\alpha}$, $\hat{\sigma}$, the median elasticity across all barcodes, and $\hat{\phi}_m$. Back to Section 4.

Figure A.4: Estimates of $\hat{\alpha}$ (left) and $\hat{\sigma}$ (right) versus household income decile (top) and household population density decile (bottom)

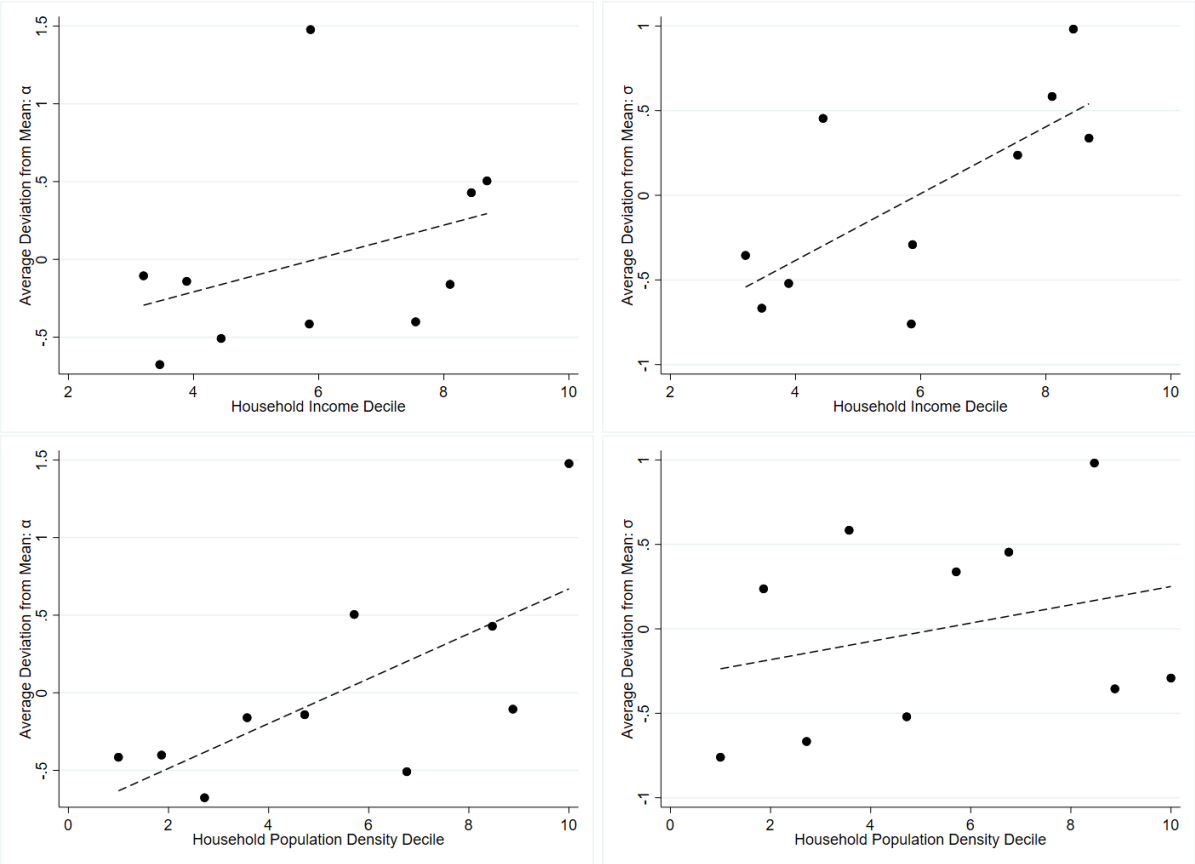

Figure A.4 provides estimates of α (left) and σ (right) for each household type. The x-axis is the average income decile of each type for the top panel and the average population density decile of each type for the bottom panel. All estimates are standardized *within* each category, and so these estimates represent the average standard deviation from category mean for each household type. Back to Section 4.

Figure A.5: Median Quality versus Origin GDP per Capita

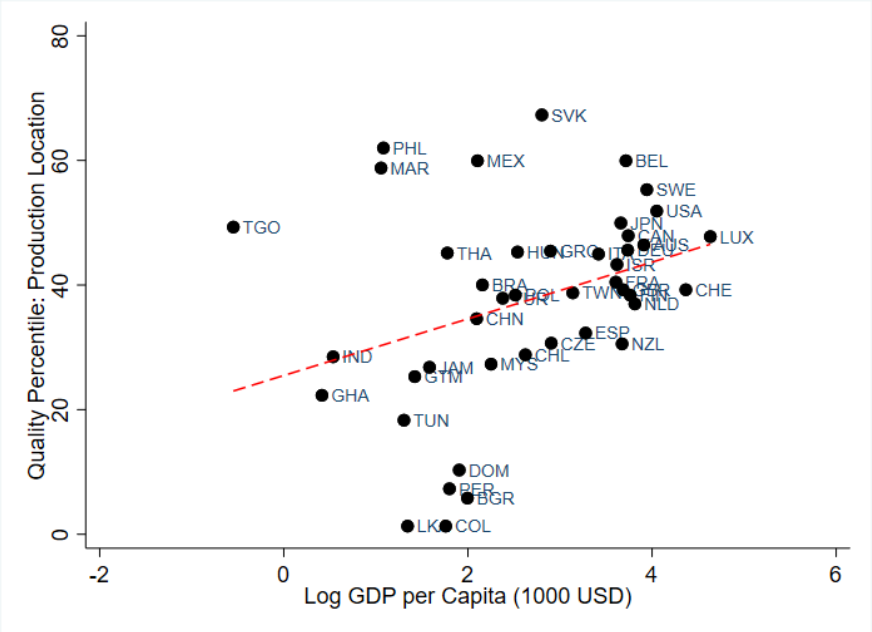

Figure A.5 provides estimates of the median quality percentile of each production origin country plotted against the GDP per capita of that country. The dotted line presents a linear fit of the data. Back to Section 4.4.

Figure A.6: County-Level Welfare Costs of 10% Tariff on all Imports

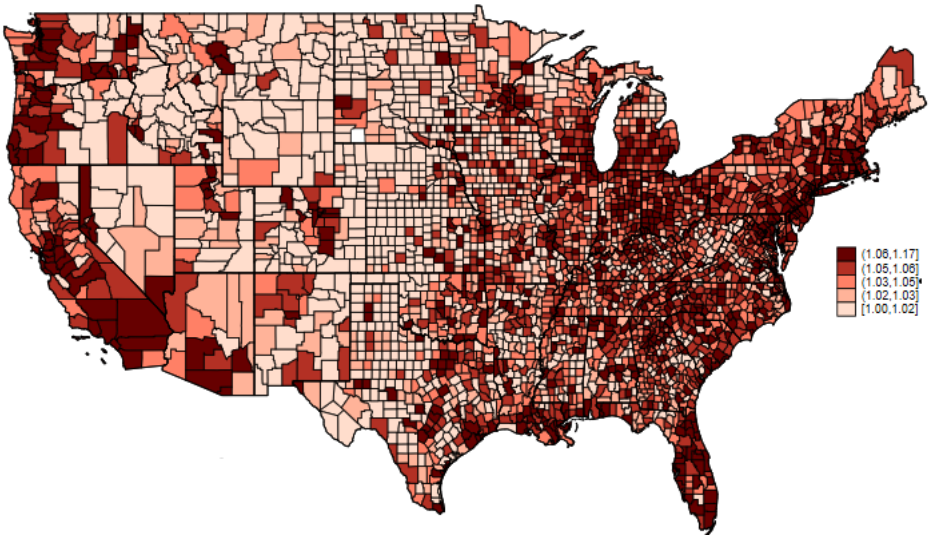

Figure A.6 provides estimates of the county-level relative welfare costs of a 10% tariff increase for all imported varieties. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.7: County-Level Welfare Costs of 10% Tariff on NAFTA

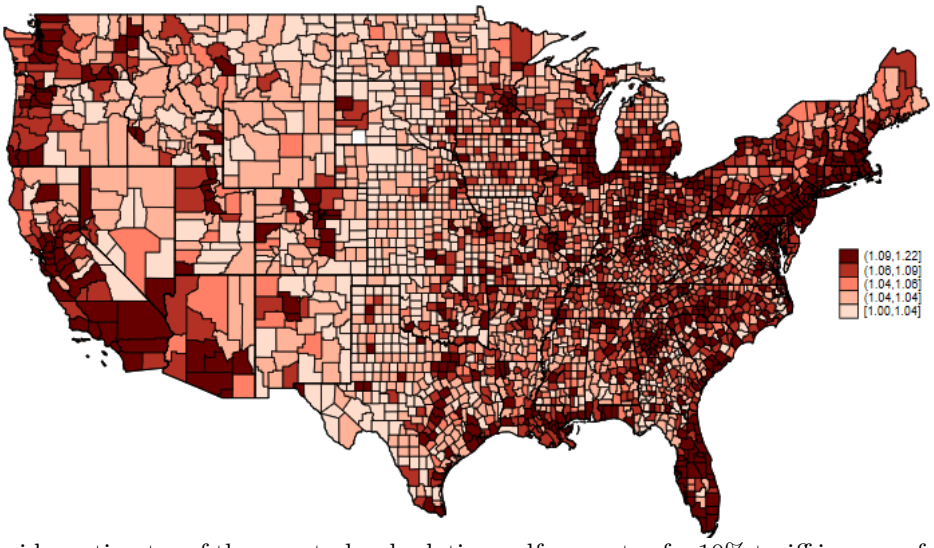

Figure A.7 provides estimates of the county-level relative welfare costs of a 10% tariff increase for all NAFTA varieties, excluding the USA. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.8: County-Level Welfare Costs of 10% Tariff on European Varieties

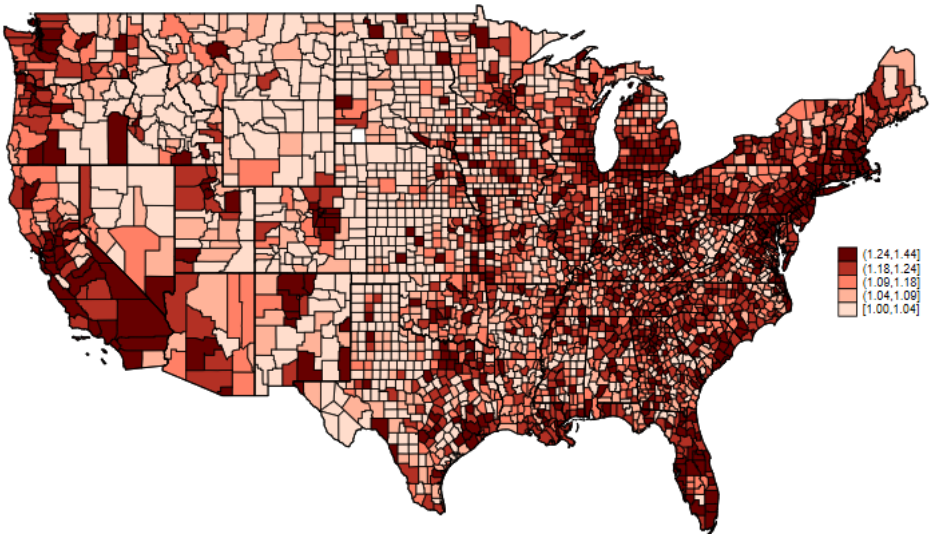

Figure A.8 provides estimates of the county-level relative welfare costs of a 10% tariff increase on all European varieties. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.9: County-Level Welfare Costs of 10% Tariff on Chinese Varieties

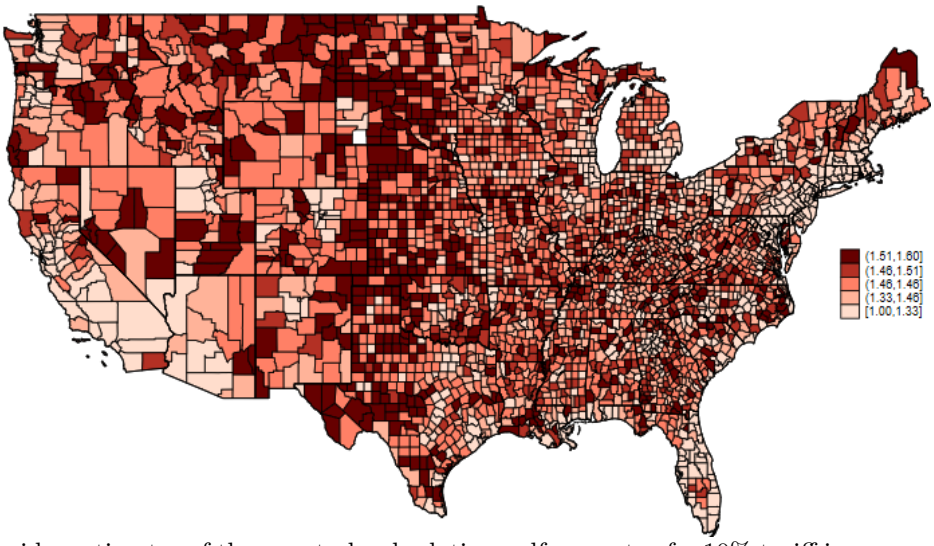

Figure A.9 provides estimates of the county-level relative welfare costs of a 10% tariff increase on all Chinese varieties. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.10: County-Level Welfare Costs of 10% Tariff on Top Half of Origins by Income

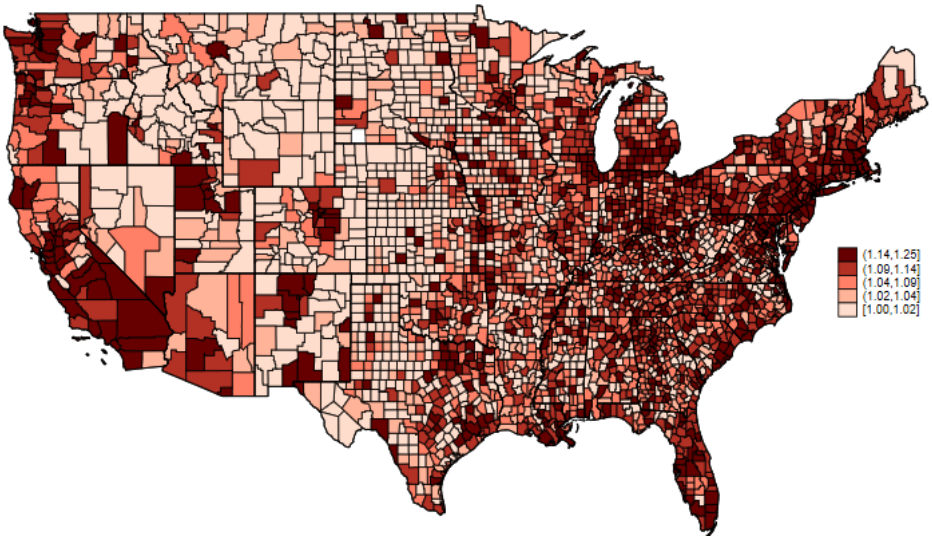

Figure A.10 provides estimates of the county-level relative welfare costs of a 10% tariff increase for all imported varieties originating in countries from the top half of the income distribution, excluding the USA and NAFTA. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.11: County-Level Welfare Costs of 10% Tariff on Bottom Half of Origins by Income

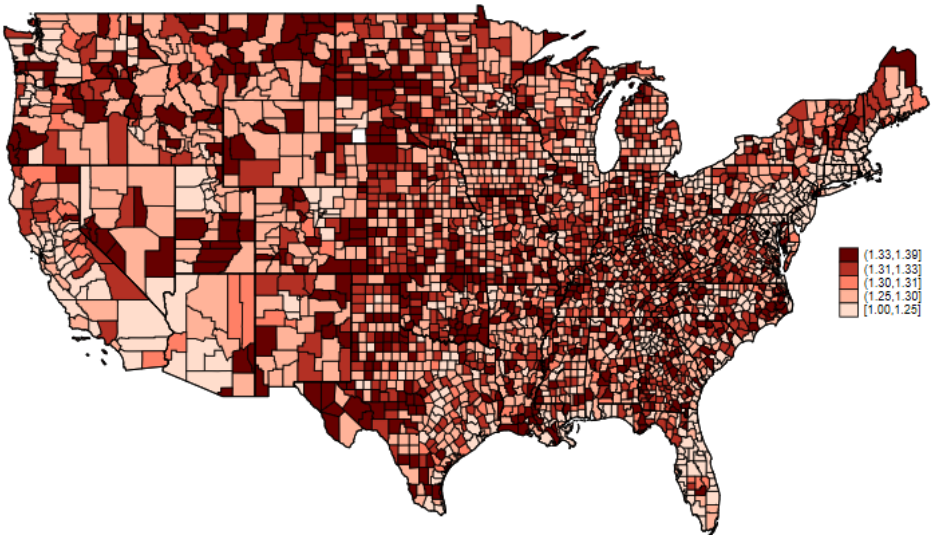

Figure A.11 provides estimates of the county-level relative welfare costs of a 10% tariff increase for all imported varieties originating in countries from the bottom half of the income distribution, excluding the USA and NAFTA. Counties are categorized into quintiles of increasing relative costs. Back to Section 5.

Figure A.12: Welfare Costs of 10% Tariff on European Varieties by Decile of Trump Vote Share (2016)

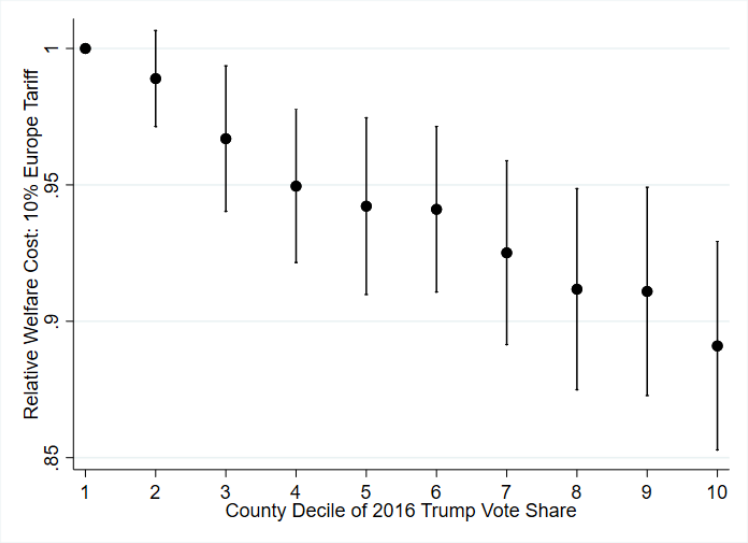

Figure A.12 provides estimates of the county-level relative welfare costs of a 10% tariff increase on European varieties. Each point represents an estimate from a regression of county-specific welfare costs on the decile of county-level Trump vote share in 2016. State-level fixed effects are included in the regression, with state-clustered 95% confidence intervals provided. Back to Section 5.

Figure A.13: Welfare Costs of 10% Tariff on Chinese Varieties by Decile of Trump Vote Share (2016)

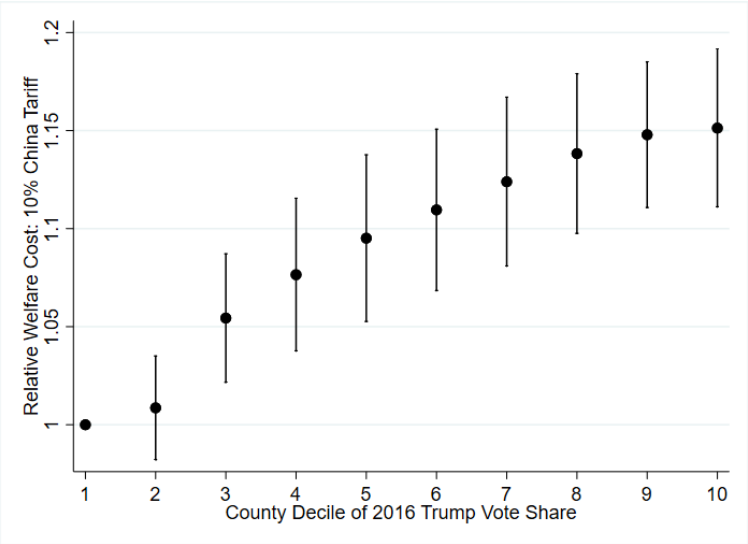

Figure A.13 provides estimates of the county-level relative welfare costs of a 10% tariff increase on Chinese varieties. Each point represents an estimate from a regression of county-specific welfare costs on the decile of county-level Trump vote share in 2016. State-level fixed effects are included in the regression, with state-clustered 95% confidence intervals provided. Back to Section 5.

A.2 Text Analysis and Clustering Algorithm Details

This Appendix provides a description of the method used to turn label packaging into a continuous distance measure across all bilateral variety pairs within the same category. Label Insight utilizes an AI to read off all text from the packaging of the consumer packaged goods in their database. These text data then constitute strings of various length that provide the brand of each barcode as well as a description/branding. In order to clean the data, I remove all references to the brand of each variety that appear in the label packaging, before then adding the brand name back to the beginning of each description. The goal is to have the brand of each variety be considered in a similar way and only appear once in the text description.

I then apply the Optimal String Alignment (OSA) method of turning two strings into a continuous measure of (dis)similarity. The OSA algorithm calculates the minimum number of insertions, deletions, symbol substitutions, and transpositions (swapping two adjacent symbols) in order to change one string into another. For the application in this paper, this measure then calculates the minimum number of actions one would have to perform on a given variety description in order to arrive at the text description of a different barcode.

Consider the following three examples from the toothpaste category:

- **A**: "COLGATE ANTICAVITY ANTIGINGIVITIS TOOTHPASTE"
- **B**: "COLGATE TOTAL PROTECTION TOOTHPASTE: SPEARMINT"
- **C**: "SENSODYNE TOOTHPASTE FOR SENSITIVE TEETH: COOL MINT"

The distance measures for these three descriptions are the following: $\Sigma^t_{AB} = 32$, $\Sigma^t_{AC} = 39$, and $\Sigma^t_{BC} = 35$. As one might expect, **A** and **B** are deemed the most similar due to being of the same brand. The "mint" connection between **B** and **C**, however, makes them more similar than **A** and **C**, which share very little in common other than the word toothpaste. In order to understand how brand labels play a role, consider that the distance between **B** and **C** decreases to 27 when the brand of **C** is changed from "Sensodyne" to "Colgate". Similarly the distance between **A** and **B** decreases from 32 to 26 when the word "MINT" is added to the end of **A**'s description.

I know provide a description of the Partitioning Around Medoids (PAM) clustering algorithm as well as the Silhouette Method of selecting the "optimal" number of clusters to implement. The PAM algorithm takes as input a dataset of N observations with a distance measure between each

observation i and j of Σ_{ij}, and a pre-determined number of clusters K. The PAM algorithm is similar to the K-Means/K-Medians algorithm in that the PAM algorithm seeks to define a centroid for each cluster and then minimize the distance between each observation and the centroid of that observation's assigned cluster. The PAM algorithm differs from K-Means in that the centroids in the PAM algorithm consist of data points - which are called "medoids" - whereas in the K-Means approach the centroid of each cluster is a point in the observation space, rather than an observation itself. In this case, given that I am feeding into the algorithm a dissimilarity measure rather than a set of real observations, the PAM algorithm provides a viable alternative to the K-Means algorithm, which requires real data points from which to calculate the centers of each cluster.

The algorithm begins by "greedily" selecting K of the N observations to act as medoids, and assigns all other observations to be in a cluster with their closest medoid. The algorithm then iteratively considers swapping a non-medoid observation with a medoid observation and considers the cost of doing so. In this case the cost refers to the distance between all observations and their respective medoids once the new clusters have been formed. This process continues iteratively until their are no additional swaps that would lower the overall cost of the clusters formed.

As mentioned in the text, this process requires both a pre-determined distance measure between all observations as well as pre-specified number of clusters. In order to select the number of clusters that leads to the best "fit" of the data, I employ the Silhouette Method. This method iterates over a range of cluster counts determined by the researcher and selects the number of clusters that maximize the "Silhouette Width".

Define the Silhouette Width for any given observation as $s(i)$. To calculate $s(i)$, the algorithm first calculates the average distance between an observation i and all observations belonging to the same cluster. Define this average distance as $a(i)$. The algorithm then calculates the additional average distances from observation i to the observations that form every other cluster. This leads to a separate measure of average distance for observation i to every other cluster. Define $b(i)$ as the average distance from i to the "nearest" cluster other than that which i has been assigned to[55]. The Silhouette Width $s(i)$ can then be calculated as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

As shown, $s(i)$ will be closest to 1 when the difference $b(i) - a(i)$ is large, meaning that obser-

[55] Think of this as the average distance from i to i's nearest neighboring cluster.

Table A.7: Attribute Cluster Summary for Deodorant ($k = 4$)

Cluster #	N_g #	s_g %	s_g^{im} %	$\bar{p}_g$ \$	$\bar{w}_g$ OZ	Key Text
15	128	13.3	1.9	5.51	2.70	Antiperspirant; clear
5	120	12.1	6.1	4.21	2.56	Antiperspirant/deodorant; fresh
16	55	10.9	31.6	4.03	2.72	Invisible; solid; women
6	71	9.5	9.2	5.03	5.89	Body spray; dry; men
13	44	7.1	16.2	3.89	2.89	Degree; Suave; women; girl; long-lasting
12	123	7.0	3.8	5.20	2.96	Deodorant; stick; natural
9	133	6.8	63.3	6.99	3.84	Body spray; dry; men
22	52	5.1	4.8	6.22	3.62	Aluminum zirchonium trichlorohydrex; antiperspirant
4	64	4.4	11.3	6.06	2.74	Old Spice; extra strong; sweat defense
11	59	3.8	1.8	7.72	6.39	Advanced care; extra dry; antiperspirant

Table A.7 provides summary statistics for the clusters created using the PAM algorithm and Silhouette method outlined in this section. These clusters are for the category deodorant. Each row represents a cluster, with the columns providing, respectively: the number of barcodes within that cluster, the aggregate market share of cluster, the import market share of cluster, the average price and weight of that cluster, and the key defining words from the packaging which unite all barcodes within that cluster.

vation i is closer to the observations within it's cluster compared to observations in the "nearest neighbouring" cluster. The Silhouette Method then averages across all $s(i)$ for all observations $i \in N$ and calculates a total score that is bounded by $[0, 1]$, with a value of 1 signifying clusters that are perfectly distinct from one another. By iterating over possible cluster counts K, I use the Silhouette Method to then select the number of clusters that returns the largest Silhouette Score. Given that the algorithm is computationally demanding, I only search over cluster counts that are multiples of five. I also begin at $K = 20$ and increase the cluster number by five until $K = 60$, the Silhouette Method then selects which cluster count provides the closest fit to the data.

As an example from deodorant category, Table A.7 provides a basic description of the top ten clusters by aggregate cluster market share. From left to right, Table A.7 provides the number of barcodes in each cluster, the aggregate cluster market share, the within-cluster import expenditure share, cluster average price and weight, and key words found in the text of most barcodes placed in that cluster. This table outlines the importance of clustering based on text attributes, price, and weight. Notice, for example, that the first two clusters seem similar from a text perspective, but occupy vastly different price segments. Similarly, clusters 6 and 11 are of similar size, but cluster 6 constitutes spray-on deodorants whereas cluster 11 constitutes roll-on deodorants for extra dry skin. A random coefficients model with the size of each barcode entering as a random coefficient might estimate high substitutability between these nests, when in fact they represent remarkably

differentiated nests of barcodes. Text analysis allows therefore for remarkably subtlety in production differentiation while incorporating the intuition that varieties closer in price and weight are likely closer substitutes.

A.3 Identification of Selection Propensity

Consider the nested logit demand system for variety j:

$$\ln\left(\frac{s_j}{s_0}\right) = -\alpha \ln p_j + \boldsymbol{\beta X_j} + \sigma \ln(ns_j) + \zeta_j$$

Define the propensity for variety j to be selected into the consideration set $\boldsymbol{C}$ as: $\chi_j = \boldsymbol{E}[d_j]$, where d_j represents an indicator function for whether or $j \in C$ or otherwise. Assume that there exists some continuous and strictly *increasing* function $\iota()$ such that we can rewrite the unobserved utility of variety j as: $\zeta_j = \iota(\chi_j) + \lambda_j$. Where λ_j is a random shock with distribution $G_\eta()$. We can then identify the parameters of this demand system by estimating the difference between variety pair j and j'.

$$\ln\left(\frac{s_j}{s_{j'}}\right) = \alpha \ln\left(\frac{p_j}{p_{j'}}\right) + \boldsymbol{\beta} \ln\left(\frac{\boldsymbol{X_j}}{\boldsymbol{X_{j'}}}\right) + \sigma \ln\left(\frac{ns_j}{ns_{j'}}\right) + \iota(\chi_j) - \iota(\chi_{j'}) + \lambda_j - \lambda_{j'}$$

Notice that for any two varieties with $\iota(\chi_j) = \iota(\chi_{j'})$, or, equivalently, $\chi_j = \chi_{j'}$, the differenced demand system is identified if $\lambda_j - \lambda j' = 0$. This identifying restriction can be written explicitly for any two varieties with observed non-zero market shares (i.e. $d_j = d_{j'} = 1$):

$$\boldsymbol{E}[\lambda_j - \lambda_{j'} \mid \tilde{p}_j, \tilde{p}_{j'}, \chi_j = \chi_{j'}, d_j = d_{j'} = 1] = 0$$

This restriction is at the core of the estimation procedure in this paper. Next, we turn to identifying $\chi_j \ \forall j \in \boldsymbol{J}$. First, we can define χ_j formally as the expectation over the consideration set indicator d_j, conditional on prices, observable variety characteristics, a price instrument, and the observed consideration instrument $\tilde{\omega}_j$: $\chi_j = \boldsymbol{E}[d_j \mid p_j, x_j, \tilde{p}_j, \tilde{\omega}_j]$. In order to identify this expression, we will first recover the residuals from the first-stage price instrument regression. That is, the residuals recovered from a regression of a variety's price on the instrumented price and observable characteristics of that variety. Define these estimated residuals as $\hat{\nu}_j$.

In order to estimate χ_j using the selection instrument and the recovered first-stage price resid-

uals, Dubé et al. [2021] assume the following distribution restriction on the distribution of η_j:

$$Pr(\eta_j < \eta \mid p_j, \boldsymbol{X_j}, \tilde{p}_j, \tilde{\omega}_j) = Pr(\eta_j < \eta \mid \nu_j, \tilde{\omega}_j) = G_\eta(-\phi(\omega_j))$$

The exclusion restriction assumption in this case is that any dependence of η_j on prices, observed variety characteristics, and the price instrument, is characterized by the residual ν_j. Dubé et al. [2021] then implement a multivariate kernel regression to non-parametrically estimate the function $\phi()$. I find that in practice, simply estimating this function as a linear probability model provides almost identical results at much lower computational cost. Therefore I estimate the following linear probability model using OLS:

$$d_j = \beta_1 \tilde{\omega}_j + \beta_2 \hat{\nu}_j + \eta_j$$

With the estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ in hand, I can then calculate the consideration propensity $\hat{\chi}_j = \hat{\beta}_1 \tilde{\omega}_j + \hat{\beta}_2 \hat{\nu}_j$. I find that all estimates of $\hat{\beta}_1$ are positive and significant at a level of 1%. This is reassuring, as one would expect promotional activity to increase the propensity for consideration. Between the two possible explanatory variables used for estimating consideration, I find that the consideration instrument, rather than the price instrument, provides the bulk of the explanatory power. The price instruments are important, however, as they illustrate how different household groups in their consideration response to price shocks.

I find that poorer household groups and households in more rural areas are more sensitive to price shocks and respond more strongly than wealthy households in the face of price shocks. That is, a price shock to any variety has a greater probability of causing that variety to drop out of the consideration set of a poor household when compared to the effect of this price shock on the consideration set of a wealthy household. Indeed, the consideration sets of wealthy household types seem almost invariant to price shocks, with $\hat{\beta}_2$ often estimated as statistically indistinguishable from zero. These results are reassuring as they suggest that modelling consideration across different household types is important and that household differences in consideration propensities follow intuitive patterns. Figure A.14 provides scatter-plots of the estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$ against household income and household type population density. Note that the mean estimate of $\hat{\beta}_2$ is negative, implying that more positive estimates are closer to zero and represent a weaker relationship between price and propensity, hence the positive relationship between $\hat{\beta}_2$ and income.

With this estimate of $\hat{\chi}_j$, it is possible to then estimate the differenced demand system by comparing variety pairs satisfying $\hat{\chi}_j = \hat{\chi}_{j'}$. The last adjustment that must be made is that, by

Table A.14: $\hat{\beta}_1$ (top) and $\hat{\beta}_2$ (bottom) versus Income (left) and Population Density (right)

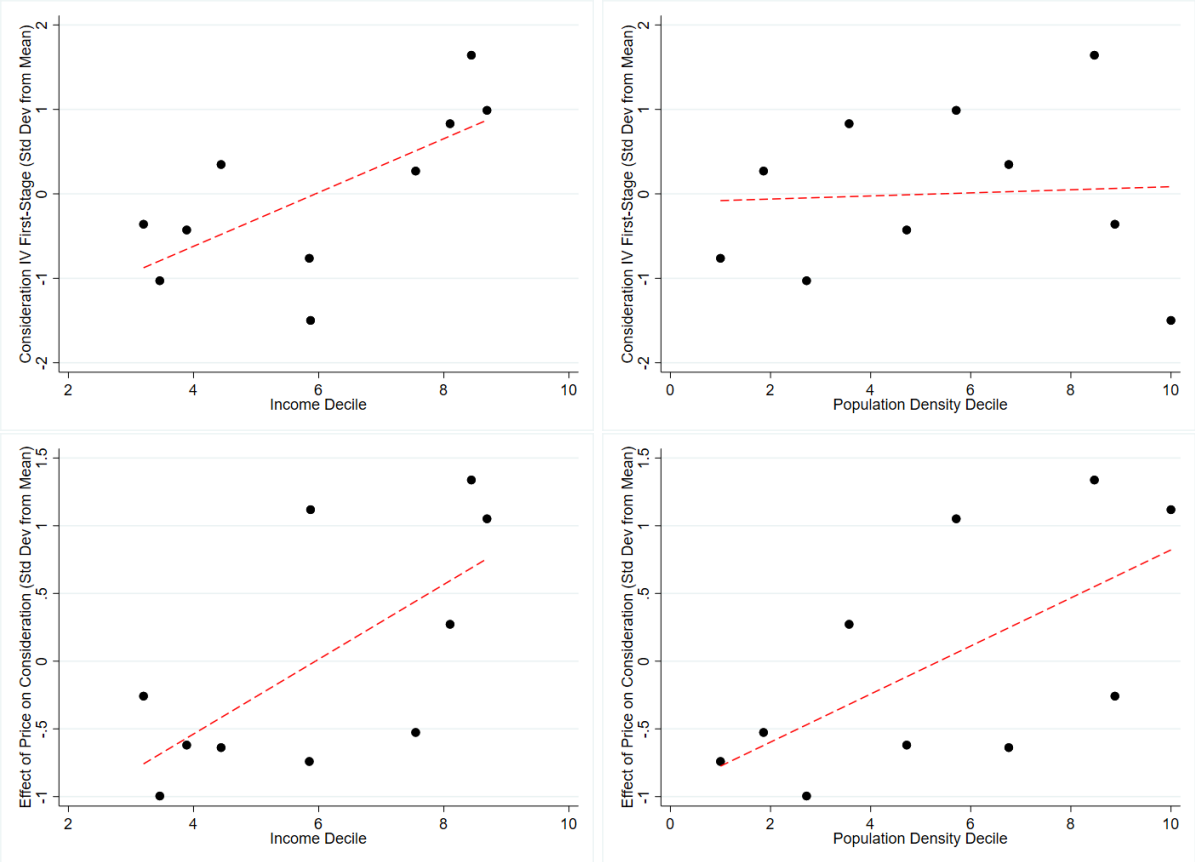

Figure A.14 provides average estimates $\hat{\beta}_1$ (top) and $\hat{\beta}_2$ (bottom) by household type plotted against income (left) and population density deciles (right). All estimates have been standardized at the product category level, and therefore represent the average number of standard deviations from the mean of each product category by household type.

construction, this measure of $\hat{\chi}_j$ is continuous and therefore the probability that $\hat{\chi}_j = \hat{\chi}_{j'}$ for any two varieties is zero. I therefore estimate the differenced estimation equation for all bilateral pairs of varieties with non-zero observed market shares and weight each observation by how similar varieties are in their propensity for non-selection. To do so, I calculate the weight of each observation (an observation being a first-differenced bilateral variety pair) as the probability density of a normal distribution. This ensures that if $\hat{\chi}_j = \hat{\chi}_{j'}$, the value of $\hat{\pi}_{j,j'}$ is maximized. For each product-household pairing, I first standardize all pair-wise differences of the consideration propensity before fitting a normal distribution to the data.

A.4 Specification Test Results and Additional Details

This section outlines the specification tests used in this paper. I provide two different sets of tests in order to evaluate the attribute-cluster model used in this paper when compared to models which segment varieties based on their production origin, as is typical in the trade literature. I first provide a likelihood ratio test of non-nested models as outlined in Vuong [1989]. I show that the attribute-cluster model provides a closer fit of the underlying data than a model based on the foreign/domestic dichotomy, as well as a model in which varieties are segmented by their origin country (so Mexico and Germany would be two separate segments). I then study whether the foreign/domestic dichotomy can provide a closer fit of the data when added as a second layer of nesting within the attribute-cluster model of this paper. Again, I find that the foreign/domestic dichotomy struggles to provide a meaningfully better fit.

Vuong Test of Non-Nested Models: In order to test the attribute-cluster model compared to alternatives which segment varieties based on their production origin, I perform a Vuong test of non-nested models for each product category in my data. I first pool all household types in order to alleviate the concern of zeros and selection discussed earlier. I then estimate three alternative models which differ in their market segmentation structure: the attribute-cluster model, a model with two segments based on foreign/domestic production origins, and a model in which each segment is a specific production origin country. I perform two tests: the first tests whether the attribute-cluster model is arbitrarily close to the foreign/domestic model in terms of their ability to fit the underlying data, and the second tests the attribute-cluster model against the country-specific nesting structure.

For all three models, I drop the weight and imported variety characteristics and instead include variety-specific fixed effects. I do this in order to isolate the role of the nesting structure of each model, rather than interactions between the nesting structure and other variety attributes. Table A.8 provides the Vuong test statistic for these two tests. In each case the hypothesis being tested is whether the attribute-cluster model is closer to the true fit when compared to the comparison model. A negative test statistic therefore states that we can reject the hypothesis that this alternative model is arbitrarily as close to the true data as the attribute-cluster model. As shown in Table A.8, across all categories the alternative model is rejected in favour of the attribute-cluster model at significance levels greater than 99%. It is important to note that these results are also

Table A.8: Vuong Test of Non-Nested Models

	(1)	(2)	(3)	(4)
	Model: Foreign/Domestic		Model: Country of Origin	
	Test Statistic	p-val	Test Statistic	p-val
Hair Care	-47.5	0.00	-58.6	0.00
Body Soap	-57.1	0.00	-44.3	0.00
Skin Care	-41.8	0.00	-59.8	0.00
Deodorant	-32.4	0.00	-42.9	0.00
Toothpaste	-44.0	0.00	-51.2	0.00
Facial Care	-31.1	0.00	-68.5	0.00
Hair Styling	-79.7	0.00	-111.4	0.00
Hand Soap	-45.1	0.00	-43.5	0.00
Face Cosmetics	-29.6	0.00	-58.2	0.00
Eye Cosmetics	-13.2	0.00	-45.1	0.00
Shaving Care	-46.3	0.00	-64.8	0.00
Lip Cosmetics	-22.2	0.00	-25.9	0.00
Nail Cosmetics	-50.9	0.00	-78.6	0.00

Table A.8 provides the test statistics of two Vuong tests of non-nested models. Column (1) provides the statistic for a test comparing the attribute-cluster model with a model based on the foreign/domestic segmentation. Column (3) provides the same statistic but for a test of the attribute-cluster model compared with a model in which segmentation is based on the production country of origin of each good. A negative test statistic states that one can reject the hypothesis that the alternative model fits the data arbitrarily as well as the attribute-cluster model. In all cases, one can reject this hypothesis at significance levels greater than 99%. Back to Section 4.3.

robust to including variety-region-specific fixed effects.

Augmented Attribute-Cluster Model: This test is implemented in order to study whether the Armington model is able to provide a useful improvement in the fit of the nested logit demand system when compared to the attribute-clustering model used in this paper. I perform this test in two ways. In each case, the test performed will be a simple likelihood ratio test as to whether or not an additional layer of nesting adds a significant improvement in the fit of the model. The test in question will then align with the following model:

$$\ln s_{jrt} = -\alpha \ln p_{jrt} + \beta^w w_j + \sigma^0 \ln ns^0_{jrt} + \sigma^1 \ln ns^1_{jrt} + \zeta_{jrt}$$

In this case σ^0 represents the nested logit parameter for the upper-tier layer of nesting, and σ^1 represents the nested logit parameter for the lower-tier layer of nesting. In each case I will test

whether $\sigma^1 = 0$. The key to this test is that I perform it twice for each product category: I first test whether the Armington model can add a significant improvement in fit as a sub-tier nesting structure added to the attribute-cluster model. I then provide the same test but with the Armington model as my upper-tier nesting structure and the sub-tier nests according to the attribute-cluster model.

In performing these tests, I first aggregate data across all household types for each category. This in effect does away with the issue of zeros and selection, and so I run these tests by estimating the model under OLS. I also do away with the instruments for price and the within-nest market shares ns^0_{jrt} and ns^1_{jrt}. I do this for a number of reasons. First, these would require an additional instrument to be constructed for the within-nest market share of the lower-tier nests and would ultimately require estimating a model with three instruments and three endogenous variables. Given that we are concerned with improvements in the fit of these models, I find that using instruments leads to often meaningless estimates of the key parameters. This is partially due to the fact that instrumenting for the within-nest market share of the Armington model often leads to weak instruments. I therefore perform the tests outlined above twice for each product category: once with the upper-tier nesting structure defined using the clusters in this paper and once using the Armington structure as the upper tier. These results are provided in Table A.9. In general, I find that one can reject the hypothesis $\sigma^1 = 0$ for both the Armington and attribute-cluster sub-nests. The test statistics, however, are often ten times greater when the sub-nest is attribute-cluster as opposed to Armington. The superiority of the attribute-cluster model can further be seen by studying the estimated values for σ^0 and σ^1 under the two different tiering structures: the estimates of σ^0 under the Armington results are negative for two product categories and generally are quite close to zero.

Table A.9: Specification Test of Attribute-Clustering versus Armington

	$\hat{\sigma}^0$: Attribute-Cluster $\hat{\sigma}^1$: Armington					$\hat{\sigma}^0$: Armington $\hat{\sigma}^1$: Attribute-Cluster				
Category	(1) $\hat{\sigma}^0$	(2) $\hat{\sigma}^1$	(3) R^2	(4) LR	(5) p	(6) $\hat{\sigma}^0$	(7) $\hat{\sigma}^1$	(8) R^2	(9) LR	(10) p
Hair Care	0.77		0.646	147	0.00	0.11		0.042	5,440	0.00
	(0.00)					(0.01)				
	0.77	0.05	0.647			0.15	0.43	0.126		
	(0.00)	(0.00)				(0.00)	(0.01)			
Body Soap	0.64		0.485	138	0.00	0.20		0.048	4,227	0.00
	(0.00)					(0.01)				
	0.63	0.08	0.487			0.26	0.40	0.160		
	(0.00)	(0.02)				(0.01)	(0.03)			
Skin Care	0.80		0.738	126	0.00	-0.12		0.035	1,196	0.00
	(0.00)					(0.01)				
	0.80	-0.06	0.739			-0.11	0.30	0.069		
	(0.00)	(0.00)				(0.01)	(0.01)			
Deodorant	0.64		0.431	458	0.00	0.04		0.021	5,367	0.00
	(0.00)					(0.01)				
	0.64	0.19	0.441			0.05	0.46	0.208		
	(0.00)	(0.01)				(0.01)	(0.01)			
Toothpaste	0.82		0.681	74	0.00	-0.33		0.031	1,182	0.00
	(0.01)					(0.03)				
	0.82	-0.13	0.683			-0.14	0.48	0.125		
	(0.01)	(0.02)				(0.03)	(0.01)			
Facial Care	0.67		0.596	132	0.00	0.19		0.056	778	0.00
	(0.00)					(0.01)				
	0.66	0.06	0.599			0.22	0.19	0.089		
	(0.01)	(0.01)				(0.01)	(0.01)			
Hair Styling	0.59		0.374	36	0.00	0.03		0.043	4,751	0.00
	(0.01)					(0.01)				
	0.58	0.06	0.375			0.10	0.50	0.265		
	(0.01)	(0.01)				(0.01)	(0.01)			
Hand Soap	0.55		0.329	30	0.00	0.26		0.064	3,658	0.00
	(0.01)					(0.01)				
	0.54	0.07	0.331			0.34	0.51	0.293		
	(0.01)	(0.01)				(0.01)	(0.01)			

Table A.9 provides results for the second specification test with multiple tiers of nesting. Likelihood ratio test statistics follow a Chi-2 distribution with $df = 1$. Likelihood ratio test statistic of the unrestricted model versus the restricted model is given in Columns (4) and (9). Associated p-values are given in Columns (5) and (10). Back to Section 4.3.