Immigrants, Imports, and Welfare: Evidence from Household Purchase Data*

Brett A. McCully[†] ⓡ Torsten Jaccard[‡] ⓡ Christoph Albert[§]

February 19, 2025

Abstract

Who buys imports? This question is fundamental to understanding which households are most exposed to trade shocks. By augmenting U.S. grocery purchase data to include the origin countries of both products and households, we provide the first direct evidence that immigrants exhibit substantially stronger preferences for imports than natives. Immigrants therefore disproportionately gain from trade, and higher import prices reduce consumer welfare by 42% more for immigrants than for natives. We estimate a quantitative model of trade to show that immigrants also reduce trade costs, thereby generating gains from trade for natives due to increased import supply. These gains accrue primarily to urban and high-income households. In the aggregate, however, immigrants generate nearly three times as much import expenditure via their preferences than by reducing trade costs.

Keywords: Gains from trade, heterogeneous preferences, spillover effects.

JEL Categories: F22, J31, J61, R11.

*We are grateful to Samuel Bowles, Daniel Chiquiar, Keith Head, Wookun Kim, Munseob Lee, Hong Ma, Peter Morrow, Riccardo Turati, and Jonathan Vogel for helpful comments, as well as participants at various conferences and seminars. We thank Ilaria Malisan and Andrea Montalbano for research assistance. The analysis in this paper is our 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 our own and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All errors are our own.

[†]Collegio Carlo Alberto; brett.mccully@carloalberto.org

[‡]Vancouver School of Economics; tjaccard@mail.ubc.ca

[§]Collegio Carlo Alberto; christoph.albert@carloalberto.org

1 Introduction

Who buys imports? This question is fundamental to understanding the distributional effects of trade shocks, such as tariff increases, which incur greater costs for households spending more on imported goods. The limited existing evidence focuses almost exclusively on differential exposure to international trade by consumer income (Fajgelbaum and Khandelwal 2016; Cravino and Levchenko 2017; Borusyak and Jaravel 2021; Galle et al. 2023; Auer et al. 2024).

This paper is the first to explore how the consumer gains from trade vary with nativity. Immigrants are particularly relevant in this context given that they may affect both local import demand and supply. On the demand side, immigrants' preferences and habits may differ from those of natives (Fernández and Fogli 2009; Atkin 2016; Miho et al. 2023). If such differential preferences affect their demand for goods from foreign countries, immigrants will be differently impacted by trade shocks, such as tariffs. On the import supply side, immigrants may decrease trade costs by reducing information frictions between local importers and exporters in their origin countries and hence also affect natives' gains from trade (Gould 1994, Head and Ries 1998, Parsons and Vézina 2018).

Quantifying the separate effects of immigrants on import demand and supply requires detailed data that simultaneously capture households' nativity, geographic location and the origin countries of their purchases. We construct, to our knowledge, the first dataset to contain all three elements by augmenting U.S. grocery scanner data at the household level to include the origin country of both households and products.[1]

We obtain three key results. First, immigrants have a stronger preference for imports than natives, leading them to accrue a disproportionate share of the gains from trade. Second, immigrants reduce the fixed costs of importation and increase market size, thereby increasing natives' gains from trade. Third, the demand-side effects of immigrants dominate the supply-side effects: immigrants generate nearly three times more import expenditure via their preferences than by reducing trade costs.

[1]We are unaware of any alternative data source that simultaneously provides information on household nativity and household import shares. For example, the Consumer Expenditure Survey, which may be combined with other datasets to measure import expenditure shares on cars as in Borusyak and Jaravel (2021), does not ask about respondent's nativity.

Regarding import demand, our data allow us to provide the first direct test of whether and how immigrants differ from natives in their import expenditure. Applying the general welfare formula of Arkolakis et al. (2012), immigrants accrue consumer gains from trade that are 42% greater than those of native households. Furthermore, we show that within-county variation across households in import expenditure explains the vast majority of this differential, highlighting the role of preferences rather than sorting of immigrants into locations with high import supply.

Nativity is substantially more important than income in explaining variation in the household consumption gains from trade. The difference in the gains from trade between a household earning over $100,000 a year and an observationally identical household earning $10,000 a year is approximately half as large as the difference between an immigrant and a native household. Our estimates suggest that, on average, immigrants spend 25% more on all grocery imports and 139% more on imports specifically from their own origin country when compared to an observationally identical native household living in the same U.S. county.

Regarding import supply, we estimate a heterogeneous firms model of international trade à la Melitz (2003) extended to allow immigrants to affect local import costs. We leverage unique features of our dataset, including barcode-level price and variety count data, in order to estimate the various channels through which immigrants affect local import supply.[2] Our estimates imply that immigrants reduce the fixed costs of trade but have a negligible effect on variable trade costs.[3]

We combine the estimated import demand and supply effects of immigrants within our full model and run counterfactual exercises to quantify the effect of immigrants on import expenditure and native household welfare. In our first counterfactual, we consider what would happen to imports and native welfare in the absence of immigrant effects. The national grocery import expenditure share would fall by almost 8%, which is roughly equivalent

[2] The identification challenge in our setting is similar to the one faced by Burchardi et al. (2019), in that we observe a single cross-section of origin-destination pairs. We make use of the instrumental variables developed by Burchardi et al. (2019) in order to generate plausibly exogenous variation in immigrant populations across counties.

[3] A large literature documents a positive relationship between immigrants and trade, with suggestive evidence that fixed cost reductions—such as information friction reductions—play a significant role in explaining this relationship (Head and Ries 1998; Rauch 2001; Combes et al. 2005; Peri and Requena-Silvente 2010; Parsons and Vézina 2018).

to the effect of doubling prevailing tariffs applied to grocery goods. Immigrants' stronger preference for imports explains nearly three times more of the immigrant-import elasticity as compared to changes in import supply common to all households. Thus, welfare gains to native households are only a quarter of the gains one might infer from trade flow data aggregated to the regional level and standard welfare formulas, such as Arkolakis et al. (2012). Immigrants increase local import expenditure predominantly via their own expenditure, with limited spillovers to native expenditure.

We show in a second counterfactual exercise that the presence of immigrants yields significant consumption welfare gains for natives via increased market size. If immigrants were to entirely disappear, aggregate import volumes would decrease by 26% and native welfare by 1%, primarily due to variety loss associated with the large drop in grocery expenditure.[4] These losses are highly concentrated among high-income and urban households, which is equally attributable to similar location sorting between high-income natives and immigrants, and to a positive elasticity of import demand with respect to income.[5]

This paper provides the first direct evidence that immigrants, as consumers, derive substantially larger gains from trade than native households in the U.S. Low-income natives in particular exhibit weak preferences for imported goods and limited consumption benefits from the immigrant population. Consequently, lower barriers to the international movement of both people and goods yield the smallest consumer gains for low-income, less-educated, native-born U.S. households.

Related literature. By estimating consumer heterogeneity in exposure to trade shocks we contribute to a growing literature studying the heterogeneous impact of trade across consumers (Fajgelbaum and Khandelwal 2016; Cravino and Levchenko 2017; Bai and Stumpner 2019; Hottman and Monarch 2020; Borusyak and Jaravel 2021; Faber and Fally 2022; Jaccard 2023; Auer et al. 2024). This paper is the first to document that immigrants are substantially more exposed to trade shocks than non-immigrants. When assessing international

[4]While this scenario is unrealistic, it provides a useful benchmark (comparable to the standard gains-from-trade exercise) by which to value immigration in terms of native expenditure.

[5]We estimate this positive income elasticity of import preference directly. "Preference" refers to household-level demand shifters for imported varieties, conditional on price, which Hottman et al. (2016) define as "appeal" when measuring firm-level market share.

trade exposure, nativity matters more than income, which has been the focus of prior work.

We also add to the literature on immigrants' trade impact by leveraging our unique data to unpack the mechanisms driving this relationship. Doing so allows us to compute the resulting welfare effects across heterogeneous households via immigrant-induced trade creation. In contrast, existing studies almost exclusively use aggregated data on region-to-region trade flows, such that the separate effects of immigrants on import demand and supply cannot be observed directly (Gould 1994; Head and Ries 1998; Combes et al. 2005; Peri and Requena-Silvente 2010; Parsons and Vézina 2018; Steingress 2018; Bonadio forthcoming).

Thus, our study contributes to the ongoing public discourse on the benefits and costs of immigration. A vast literature has focused on the way in which immigrants affect the labor market outcomes of native workers (e.g., Card 2001, Borjas 2003, Ottaviano and Peri 2012, Dustmann et al. 2017, Monras 2020, Burstein et al. 2020). We introduce and quantify a novel margin by which immigrants benefit natives: increasing local product variety.[6] Furthermore, while studies on the effects of immigration on the labor market carefully consider distributional effects (e.g., Dustmann et al. 2013 and Llull 2018), the consumption-side distributional effects have thus far been ignored.

This paper proceeds as follows. We describe our data and present stylized facts in Section 2. In Section 3, we characterize heterogeneity in preferences between immigrants and natives and therefore in the gains from trade. Section 4 microfounds and estimates how immigrants affect import supply. Section 5 describes our counterfactuals using the full quantitative model. Section 6 concludes.

2 Data and Stylized Facts

2.1 Expenditure on Consumer Packaged Goods

We use two datasets that link household characteristics—including country of birth—to grocery import expenditures: the NielsenIQ household panel scanner dataset and barcode

[6]Two prior papers have explored this margin—Mazzolari and Neumark (2012) and Chen and Jacks (2012)—but lack the data and exogenous variation to rigorously identify potential mechanisms. Iranzo and Peri (2009), Di Giovanni et al. (2015), and Aubry et al. (2016) study the aggregate variety effects of immigration but with a focus on immigrants expanding production in high-productivity locations.

country-of-origin data from Label Insight Inc.

NielsenIQ Household Panel Scanner Data: These data consist of a panel covering approximately 90,000 U.S. households and all grocery purchases at the barcode level. Variables include detailed household demographic information and county of residence as well as barcode-level expenditure, price, date, and store for each purchase.

In 2008, NielsenIQ distributed the "Tell Me More About You" survey, which included questions about respondents' birth place.[7] Thus, for the subset of households present in the data since 2008 we observe their country of origin. Household members may have mixed nativity, and we use the following allocation rules when assigning them to an origin country. When only one member of the household was born abroad and all others were born in the U.S., we assign the household to the country of the immigrant member. In the rare case that a household has multiple foreign-born members from different origins, we assign the household to the larger country of origin as measured by the total immigrant population in the U.S.

Barcode Country of Origin: We merge the NielsenIQ data with barcode-specific country-of-origin information purchased from Label Insight Inc., a firm that specializes in extracting and organizing information found on the labels of consumer packaged goods.[8] Label Insight uses a computer vision algorithm to extract text from the packaging for thousands of barcodes sold across major retail chains in the U.S. Since imported goods in the U.S. are required to contain some statement equivalent to "Made in ..." on their labels, the algorithm incidentally recovers the origin country for each collected barcode.[9] Naturally, Label Insight can only cover a segment of total consumption and their coverage is best for food and beverages, alcohol, personal care products, and cosmetics.

We therefore make use of data on the origin country for over 600,000 barcodes in these grocery product categories. Given the universality of barcodes, these data can be directly merged with the household-level purchase records from NielsenIQ. Figure C.1 documents the distribution of production origin countries in the merged scanner data. As expected, Mexico

[7]See Bronnenberg et al. (2012) for more details regarding this survey.

[8]See Jaccard (2023) for a more detailed discussion of this dataset.

[9]The U.S. Customs and Border Protection require that the country-of-origin printed on the label corresponds to the last country in which the good underwent a "substantial transformation."

and Canada constitute just over half of all import expenditure, with Thailand, China, and Italy rounding out the top five product origins. Overall, our sample contains 74 origins with positive imports. The average import expenditure share is approximately 8%.[10]

The sample of NielsenIQ households interviewed in 2008, and therefore with nativity information, attrits over time. On the other hand, the Label Insight data on products' country of origin has been improving in recent years. To maximize data quality in both datasets, we restrict our analysis to the years 2014-2016, which we pool to a single cross-section at the household level. Our final sample consists of 19,745 households, which are 40% of those interviewed in 2008.[11]

Household-Level Coverage of Import Expenditure: Our final merged dataset exhibits $764 billion of expenditure and is at the household-import origin level of aggregation. When compared to estimates from the BLS Consumer Expenditure Survey, the grocery categories studied in this paper account for approximately a third of all expenditure on tradeables, with this share increasing to almost half if one excludes passenger vehicles and energy products. Jaravel (2019) estimates a within-grocery coverage of NielsenIQ expenditure data of approximately 80%, while the Label Insight data covers approximately 60% of expenditure within the NielsenIQ sample.

2.2 Immigration Data

We use the decadal Censuses from 1880-1930 and 1970-2000, as well as the pooled 2006–2010 sample of the American Community Survey (ACS) to obtain population counts of immigrants by origin.[12] We compute immigrant inflow measures for each available decade between 1880 and 2000. These inflows are used in the first stage of our instrumental variables strategy outlined in Section 4.2.2 to predict county-level immigrant population shares by origin. We

[10]Throughout this paper we make use of the projection factor weights provided by NielsenIQ when presenting aggregated statistics. These weights are a population projection based on the representativeness of each household, and sum to 120 million households, which roughly matches the aggregate total for the U.S.

[11]Immigrants have a virtually identical rate of survival to our final dataset at 39%, and the difference is not significant at the 90% level. For households with purchase records in 2014-2016, 23% can be linked to the 2008 nativity survey. This rate is virtually unchanged (24%) for the top half of households by income. We find that sample representativeness and differential attrition does not significantly drive our baseline results, as discussed in Appendix Section A.3.2.

[12]The 1940, 1950 and 1960 censuses did not ask about the year of the respondent's immigration.

provide additional details on data construction in Appendix A.1.

2.3 Stylized Facts

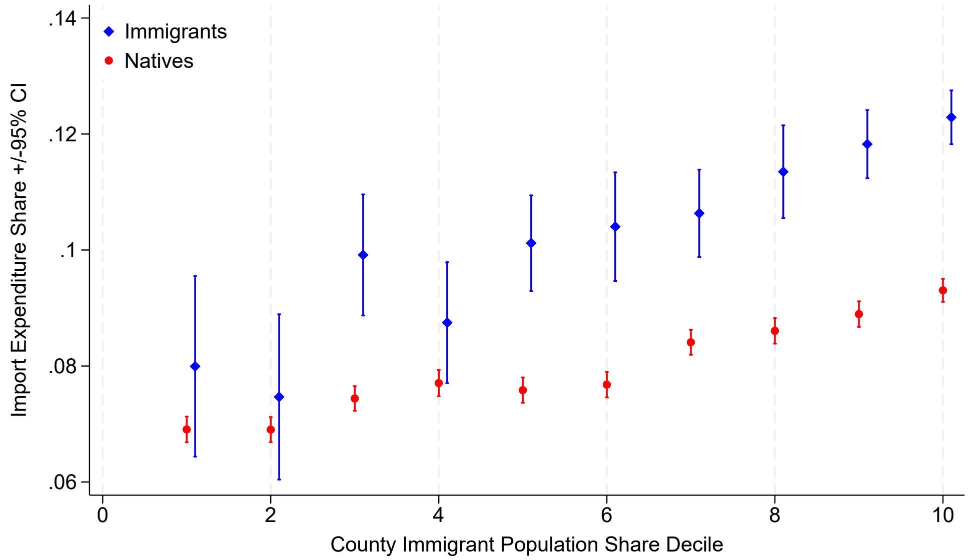
The combined dataset described above constitutes the first direct measurement of import expenditure by country of birth. We leverage this novel feature to demonstrate three stylized facts which characterize import consumption heterogeneity by nativity.

Fact 1: Household-level import expenditure share is increasing in the immigrant population share of a household's county. Figure 1 plots the average import expenditure share across county deciles based on the immigrant population share. Both native and immigrant households exhibit a strikingly positive relationship between the presence of immigrants and the propensity to purchase imported goods. Relative to the lowest decile, native households living in the most immigrant-intensive decile of counties exhibit import expenditure shares which are 35% higher on average. For immigrant households, this differential increases to +50%. The figure represents the first direct evidence of a positive correlation between household-level import expenditure and local immigrant population shares.

Fact 2: Average import expenditure shares are 38% greater for immigrant households compared to non-immigrant households. Figure 1 also shows that immigrants exhibit stronger import demand than native households even within the same county decile. We quantify this difference in mean import expenditure by regressing the household-level import expenditure on a dummy for whether a household is an immigrant household. Table C.1 provides the estimates from this exercise, and we find an unconditional mean difference in import expenditure between immigrants and natives of +3.1 percentage points. When compared to the average import expenditure share of non-immigrant households, this estimate represents a 38% differential.[13] Columns 3 to 6 of Table C.1 display results with additional controls in order to mitigate the potential bias associated with immigrants sorting into high-import counties or differing in other observable characteristics from natives. Even when county-level fixed effects and a suite of socioeconomic household characteristics are included, the estimated differential between immigrants and natives in their average import

[13]See Figure C.2 for a histogram of import expenditure shares for native and immigrant households.

Figure 1: Immigrants, Natives, and Import Expenditure



Notes: The figure plots average import expenditure shares by county immigrant population share deciles and nativity. Each point on the graph is the corresponding nativity-by-county decile fixed effect coefficient, with 95% confidence intervals plotted. These fixed effects are recovered from a linear regression at the household level in which household import expenditure shares are the dependent variable. Counties are placed into deciles based on the immigrant population share of that county, and households are grouped into two categories: immigrants (blue diamonds) or natives (red circles). 95% confidence intervals are provided, and all observations are weighted by the NielsenIQ projection factors.

expenditure remains highly significant and constitutes a gap of +2.8 percentage points.[14]

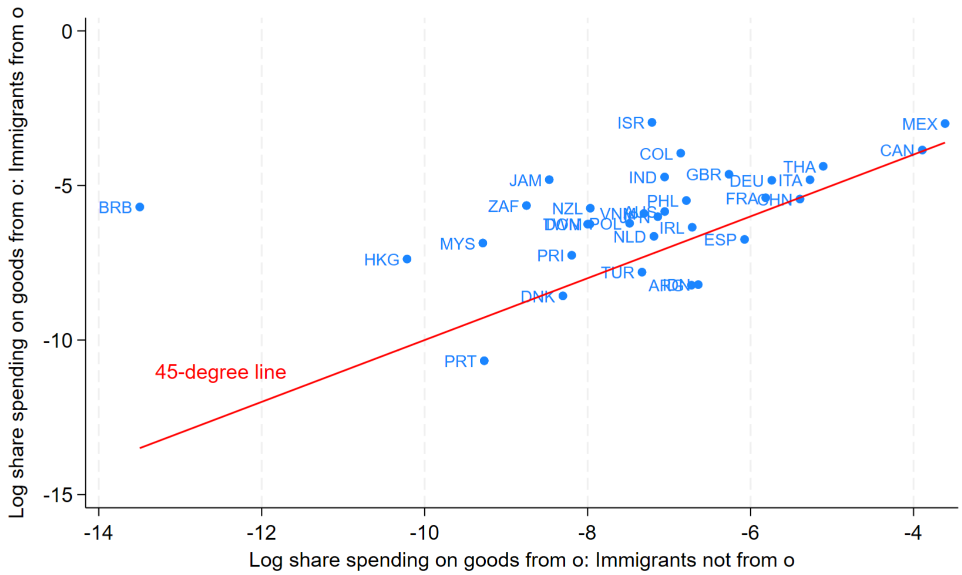
Fact 3: Immigrants spend over twice as much as on goods from their origin country as immigrants not from that origin. For each origin country o, we calculate the share of expenditures on goods from o by both immigrants from o and immigrants not from o.[15] Figure 2 depicts this relationship. The 45-degree line in red plots where immigrants not from o and immigrants from o would exhibit identical expenditure on imports from o.

We find that most origins lie above the 45-degree line, suggesting that immigrants exhibit disproportionately stronger demand for imports from their specific country of origin. For the 33 countries in our sample with nonzero expenditures by both immigrant households from that origin country and immigrants not from that origin, the median relative expenditure

[14] We add controls for income bin, household size, marital status, and household head's age and gender.

[15] We also compare immigrants from o to natives and find the same relationship, as can be seen in Appendix Figure C.3. However, because we also find that immigrants purchase more imports from all origins than natives (Stylized Fact 2 above), one worry is that the general import preference of immigrants might contaminate a comparison of immigrants' home bias relative to natives. We thank an anonymous referee for highlighting this issue.

Figure 2: Immigrants Tend to Spend more on Goods from their Origin



Notes: The figure shows the relationship between spending on goods imported from immigrants' own country (the y-axis) and spending on goods from that country by other immigrants (x-axis). The red line is the 45-degree line, which plots when there is no preference by immigrants for goods imported from their origin country relative to immigrants from other origins. Household nativity is assigned as discussed in Section 2.1. Data come from the NielsenIQ Household Panel 2014-2016 and Label Insight. NielsenIQ projection factor weights used to construct expenditure shares.

share on goods from origin o by immigrants from o is 1.9 times greater than the expenditure on goods from o by immigrant households not from o.[16] To our knowledge, this paper is the first to provide direct evidence that the preference persistence documented in Logan and Rhode (2010) and Atkin (2016) exists for U.S. immigrants with respect to goods imported from their origin country.

The preceding three stylized facts suggest that immigrants and natives have different demand for tradables. In our subsequent empirical and theoretical analysis we quantify the importance of these differential preferences for trade volumes and welfare.

[16]Note that this estimate represents the weighted median relative expenditure across origins. The mean estimate is 81.6, but this is driven by outliers. When weighted by origin-specific aggregate expenditure shares, the mean difference is 2.8. Thus the median estimate of 1.9 represents a conservative figure.

3 Immigrants and Import Demand

How are the gains from trade distributed by nativity? Little is known about the distribution of import expenditure across households, and the few studies that exist focus primarily on non-homothetic demand and therefore the extent to which import expenditure varies with income (Fajgelbaum and Khandelwal 2016; Cravino and Levchenko 2017; Borusyak and Jaravel 2021; Galle et al. 2023; Auer et al. 2024). In the exercise below, we quantify the role of nativity and country of origin in driving household import expenditure.

3.1 Import Expenditure and the Consumer Gains from Trade

We start by showing that, to a first-order approximation, immigrants are more exposed to trade shocks via their heightened expenditure on imported goods regardless of the underlying demand system.

Consider a household $h \in H$ characterized by income X_h and expenditure $X_h(\omega)$ across each variety $\omega \in \Omega$, such that $X_h = \sum_{\omega \in \Omega} X_h(\omega)$. We assume that all households face the same price for each variety $p(\omega) \in \boldsymbol{p}$. For some infinitesimally small change in prices $d \ln \boldsymbol{p}$ and total income $d \ln X_h$, the equivalent variation is defined as the change in X_h which generates the same change in indirect utility under constant prices. We refer to the equivalent variation of household h as $d \ln W_h$. As discussed in Fajgelbaum and Khandelwal (2016), this equivalent variation can be expressed as:

$$d \ln W_h = \sum_{\omega} [-d \ln p(\omega)] s_h(\omega) + d \ln X_h$$

where $s_h(\omega) = X_h(\omega)/X_h$. The first term on the right-hand side captures the expenditure channel, through which changing prices affect welfare. The second term on the right-hand side captures the income channel. We focus in this article exclusively on the expenditure channel, and therefore fix income throughout our analysis.

The key implication of this simple derivation is that households with elevated expenditure shares on imported goods are more exposed to changes in trade costs and the price of imports. Moreover, this finding holds regardless of demand system assumed. To characterize preference heterogeneity and to quantify the resulting implications on households' gains from trade, however, we must add structure to preferences, as we do in the next section.

3.2 Preferences and Import Demand

Households have Cobb-Douglas preferences over differentiated grocery goods, our focus in this paper, and a non-grocery good, q_0. Households aggregate grocery variety consumption with constant elasticity of substitution (CES) preferences. Doing so keeps our framework in line with the seminal work of Arkolakis et al. (2012), freeing us from taking a stance on microfounding the distribution of import supply across households.[17]

There is a continuum of differentiated varieties $\Omega_{o,c(h)}$ associated with each origin country $o \in \mathcal{O}$ and h's county of residence $c(h)$. We allow for household heterogeneity in income Y_h, origin-specific preferences denoted by z_{oh}, and substitution elasticity σ_h. Preferences are represented by the following utility function:

$$U_h = q_0^{\mu_h} \left[\sum_{o \in \mathcal{O}} z_{oh}^{1/\sigma_h} \int_{\omega \in \Omega_{o,c(h)}} q_{oh}(\omega)^{\frac{\sigma_h - 1}{\sigma_h}} d\omega \right]^{\frac{\sigma_h}{\sigma_h - 1}(1-\mu_h)}, \tag{1}$$

where $\sigma_h > 1$ denotes h's elasticity of substitution among grocery varieties and μ_h measures the expenditure share on the homogeneous good. Equilibrium expenditure on the differentiated (grocery) sector is therefore $X_h = (1 - \mu_h)Y_h$.

Demand at the origin-household level is then:

$$X_{oh} = z_{oh} p_{oh}^{1-\sigma_h} X_h P_h^{\sigma_h - 1}, \tag{2}$$

where p_{oh} denotes the origin-household-level and P_h the household-level price index.[18]

[17] We microfound the firm side in Section 4 to enable an exploration of how immigrants affect import supply and to facilitate running counterfactuals.

[18] We explicitly microfound price indices in Section 4, but do not need to in order to estimate the contribution of household preferences on demand as shown in the next section.

3.3 Estimating Preference Heterogeneity

To make demand defined in equation (2) estimable, including the preference term z_{oh}, we take the following five steps. First, we assume a functional form for the preference vector z_{oh} that relates observed socioeconomic household characteristics to import demand:

$$z_{oh} = \exp\left(\beta_z I_{oc} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o]\right) \times \eta^z_{oh}. \tag{3}$$

J_h represents a vector of observed household characteristics such as income, education, ethnicity, and race.[19] Motivated by our stylized facts, ζ_1 captures the strength of immigrants' taste for goods from all foreign countries, and ζ_2 captures the strength of immigrants' home-biased preferences à la Atkin (2016) and Logan and Rhode (2010).[20] I_{oc} denotes the population share in county c of immigrants from origin o, and we allow for the possibility that the local immigrant population share from o endogenously affects preferences for imports from o via the parameter β_z.[21] The goal of this section is to estimate differences in exogenous preferences between immigrants and natives, such that this endogenous preference term will be absorbed by fixed effects at the origin-county level. We return to estimate β_z in Section 4. η^z_{oh} denotes some idiosyncratic preference at the household-origin level.

Second, we assume firms decide on entry and exit as well as prices at the county level, i.e., $p_{oh} = p_{o,c(h)}$. Households within the same county therefore face the same schedule of price indices across origin countries.[22]

Third, we divide equation (2) by its domestic counterpart, i.e., $X_{US,h}$. Variables computed relative to the U.S. equivalent are denoted as $\tilde{x}_{oh} \equiv \frac{x_{oh}}{x_{US,h}}$ for any variable x. We also normalize preferences for domestic goods $z_{US,h} = 1$ for all households, such that $\tilde{z}_{oh} = z_{oh}$.[23]

Fourth, for the elasticity parameter σ, we consider two alternative possibilities. In the

[19]The full list of household characteristics represented by J_h are dummy variables indicating h's race, ethnicity, household size bins, bins for the number and age of children in the household, highest education among household members, household income bins, marital status and age.

[20]The function $o(h)$ maps each household h to its country of birth.

[21]In this way, we do not treat preferences as a primitive, but instead allow one's preferences to be at least partially determined by one's cultural and social context (Bowles 1998; Atkin et al. 2021).

[22]As implied by equation (3), all households value each variety within origin the same. But because households may value imports for each origin differently, the overall price index P_h varies across households.

[23]Head and Mayer (2014) refer to this normalization when estimating gravity models as a "ratio method."

more restricted model, the substitution elasticity is constant across households within a given county such that $\sigma_h = \sigma_{c(h)}$. In a less restricted version of this demand system, we allow the substitution elasticity to also vary by income quintile within county such that $\sigma_h = \sigma_{c(h),y(h)}$, where $y(h)$ denotes the income quintile of household h.

Fifth, we calculate μ_h according to the same income quintile bins $y(h)$ described above, so that we can denote household h's grocery expenditure share as $1-\mu_{y(h)}$. We recover these estimates using the BLS Consumer Expenditure Survey.[24]

The above steps yield the following estimable equation:

$$\tilde{X}_{oh} = \exp\left(\psi_{o,c(h)} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o]\right) \times \eta^z_{oh}, \tag{4}$$

where we have gathered all terms related to prices, entry, price sensitivity, and county-specific average preferences into the origin-county fixed effects $\psi_{o,c(h)}$. Thus, we identify preferences for imported goods by comparing import expenditure across households within a given county and origin country. In our less restrictive model, with $\sigma_h = \sigma_{c(h),y(h)}$ and demand elasticities varying at the county-income-quintile level, we make use of fixed effects at the origin-county-income quintile-level $\psi_{o,c(h),y(h)}$. In this case, preferences are estimated by comparing origin-specific import expenditure across households within the same county and income quintile.

Estimates of ζ_1 and ζ_2 therefore quantify how much immigrants differ from native households within the same county in their propensity to purchase imported goods from all origins (ζ_1) and imports specifically from their own origin country (ζ_2). Our estimation essentially projects observed import expenditure onto household characteristics, with county-level factors—such as prices, variety availability, local average preferences, etc.—controlled for via fixed effects. Our identifying assumption is that within county-origin pairs, there are no factors that simultaneously affect a household's import expenditure via the supply side and its location choice within the county.

In estimating equation (4) we make use of the pseudo-Poisson maximum likelihood

[24] In calculating these expenditure shares, we sum annual expenditure over the categories "Food at Home," "Alcohol," "Personal Care Products," and "Housekeeping Supplies," and divide by "Total Household Expenditure." We recover the following grocery expenditure shares by income bin: 0.144 (<\$10k), 0.127 (\$10k-\$30k), 0.117 (\$30k-\$50k), 0.106 (\$50k-\$70k), 0.103 (\$70k-\$100k), and 0.093 (>\$100k).

(PPML) estimator to deal with the large number of zeros observed in our outcome variable (Santos Silva and Tenreyro 2006). We provide robustness using alternative estimators (albeit ones that are not theory-consistent) in Appendix Section A.3.1.

3.4 Preference Results

Table 1 provides estimates of ζ_1 and ζ_2 for three separate specifications of equation (4).[25] Columns 1 and 2 include origin-county fixed effects without household controls J_h, columns 3 and 4 add in J_h, and columns 5 and 6 control for origin-county-income-quintile fixed effects.[26]

Across specifications, our estimates of ζ_1 and ζ_2 are remarkably stable. When estimated alongside ζ_2, ζ_1 fluctuates between 0.22 and 0.26 and is distinguishable from zero at 99% confidence levels. ζ_2 fluctuates between 0.62 and 0.66, with all estimates statistically distinguishable from zero at 99% confidence levels. To place these estimates in context, consider the estimates recovered in column 4. For two households in the same county and identical in their income, education, age, and family structure, the immigrant household spends a 25% greater share on all import origins than the native household. For imports specifically from this immigrant household's origin country, their expenditure is 139% greater.[27]

While these estimates provide a clear role for immigrant status in shaping import preferences, Table 1 does not provide the relative importance of immigrant status relative to more commonly studied dimensions of consumer heterogeneity, such as income. Table C.2 conducts just such a horse race, providing the coefficient estimates associated with income quintiles as well as those of ζ_1 and ζ_2. We additionally look at how household education affects import preferences. One of our key findings is that both income and education are less important in shaping import expenditure heterogeneity than nativity and country of birth. Concretely, a household earning more than \$100,000 a year exhibits within-grocery import expenditure shares that are 19% greater than a household in the same county earning less than \$10,000 a year. Similarly, a household with at least one post-graduate degree exhibits

[25] We show coefficients for the full vector of household characteristics with county-origin fixed effects in Table C.3.

[26] Note that when comparing households within income quintiles, we lose approximately a third of our sample size due to separation (Correia et al. 2019), hence the reduction in N in columns 5 and 6.

[27] That is, $e^{\hat{\zeta}_1} - 1 = 0.25$ and $e^{\hat{\zeta}_1 + \hat{\zeta}_2} - 1 = 1.39$.

Table 1: Estimates of Import Demand Preferences

	Dep. var.: Rel. expenditure share on goods from o					
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant from anywhere	0.31***	0.26***	0.27***	0.22***	0.29***	0.23***
	(0.028)	(0.028)	(0.029)	(0.029)	(0.031)	(0.031)
Immigrant from o		0.62***		0.65***		0.66***
		(0.069)		(0.07)		(0.081)
N	868,261	868,261	868,261	868,261	597,276	597,276
Origin-County FE	✓	✓	✓	✓		
HH Controls			✓	✓	✓	✓
Origin-County-Income FE					✓	✓
Sample Weights	✓	✓	✓	✓	✓	✓

Notes: The table presents results from estimating equation 4 using pseudo-Poisson maximum likelihood estimation at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. Household controls are dummy variables indicating: race, ethnicity, household size bins, bins for the number and age of children in the household, highest education among household members, household income bins, marital status and age groups. The estimation sample size is less than our total sample, and falls across columns, due to fixed effects causing separation in the sense of Correia et al. (2019). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

import expenditure shares that are 20% greater than a household in the same county with at most a high school education. By contrast, immigrants exhibit import expenditure shares that are at least 25% greater than those of natives. Our analysis is therefore the first to highlight a dimension of household heterogeneity that has been entirely overlooked in the import expenditure literature: household nativity.

In Appendix A.3.3, we consider a range of additional dimensions of demand heterogeneity. We first investigate whether households of Hispanic or Asian origin, the best proxies available in our data to capture second-generation immigrants, exhibit a higher demand for imports in general and specifically for those from their own continent. Households of Hispanic and Asian origin spend more on imports from Latin America and Asia, respectively, than natives (column 1 of Table A.6). Furthermore, we explore the heterogeneity of immigrant demand in terms of several alternative bilateral linkages between household and import origins, which aim to capture geographic or cultural proximity: lying on the same continent, having a common language, and having colonial history. Column 3 of Table A.6 shows that such linkages have no significant effects on import expenditure.

3.5 Import Demand and the Gains from Trade

Given that immigrants' preferences differ from those of natives, how are the gains from trade distributed between immigrant and non-immigrant households?

We make two simplifying assumptions in order to isolate variation in the gains from trade attributable to heterogeneous preferences. First, we assume that all households exhibit the same trade elasticity, θ, which we calibrate as $\theta = 5$ following Head and Mayer (2014) and Costinot and Rodríguez-Clare (2014). Second, we assume that the import supply-side generating the household-specific price index adheres to the assumptions described in Arkolakis et al. (2012).

We define the grocery consumption gains from trade for household h as the proportional change in welfare due to an external shock to foreign production and trade costs, denoted by $\hat{W}_h = W'_h/W_h$. Applying the welfare formula derived in Arkolakis et al. (2012), the household's gains from trade in groceries are a function of the observed household-level expenditure share on U.S. goods $\pi_{US,h}$:

$$\hat{U}_h = (\hat{W}_h)^{1-\mu_h} = (\pi_{US,h})^{-\frac{1}{\theta}(1-\mu_{y(h)})}$$

The values of $\mu_{y(h)}$ are calibrated using the CEX, as described in Section 3.3, with grocery expenditure shares monotonically decreasing in income. Having calibrated θ and $\mu_{y(h)}$, and directly observing $\pi_{US,h}$ in our expenditure data, we recover the cross-sectional distribution in consumer gains from grocery trade.

To understand household heterogeneity in the gains from trade, we estimate a linear regression model to recover both conditional and unconditional differences in the average gains from trade by nativity. Table 2 provides these estimates for both the within-grocery welfare gains from trade, $\hat{W}_h$ and the aggregate welfare gains from grocery trade, $\hat{U}_h$.

Columns 1 to 3 provide estimates of the gap in grocery welfare between immigrants and natives. The dependent variable is defined as $(\hat{W}_h - 1) \times 100$ and thus captures the percentage gain in welfare attributable to trade, such that each coefficient can be interpreted as a percentage point welfare gap across households. Column 1 provides the unconditional difference in the gains from trade between immigrants and natives, while column 2 provides

Table 2: Consumer Gains from Trade by Nativity

	Dep. var.: $\%\Delta W_h$			Dep. var.: $\%\Delta U_h$		
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant from anywhere	0.714***	0.635***	0.556***	0.067***	0.060***	0.057***
	(0.033)	(0.034)	(0.036)	(0.004)	(0.004)	(0.004)
Income: 10k-30k			0.037			-0.024***
			(0.047)			(0.005)
Income: 30k-50k			0.027			-0.042***
			(0.048)			(0.005)
Income: 50k-70k			0.124**			-0.050***
			(0.050)			(0.005)
Income: 70k-100k			0.109**			-0.057***
			(0.050)			(0.005)
Income: >100k			0.311***			-0.057***
			(0.050)			(0.005)
Constant	1.715***	1.728***	2.509***	0.187***	0.188***	0.372***
	(0.009)	(0.009)	(0.529)	(0.001)	(0.001)	(0.057)
N	19,750	19,158	19,158	19,750	19,158	19,158
County FE		✓	✓		✓	✓
HH Controls			✓			✓
Sample Weights	✓	✓	✓	✓	✓	✓

Notes: The table presents OLS regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

the within-county comparison, thus controlling for geographic sorting of households. Lastly, column 3 includes the entire vector of observable household controls J_h and provides the estimates associated with income bins.

Within the grocery sector, immigrants exhibit unconditional gains from trade that are, on average, 42% greater than the average non-immigrant (2.43% versus 1.72%, column 1). When controlling for county of residence, and therefore local import supply, immigrants still exhibit gains from trade that are 37% greater than their within-county non-immigrant neighbors (2.36% versus 1.73%, column 2). We interpret this within-county variation as evidence of preference heterogeneity, and these estimates highlight the significant effect such preference heterogeneity has on the variation in consumer gains from trade across households.

Column 3 highlights the extent to which immigrant effects are substantially larger than other dimensions of household heterogeneity, such as income. Households earning over $100,000 per year exhibit within-grocery gains from trade that are 0.31 percentage points greater than their within-county neighbors earning less than $10,000 a year, whereas conditional on income, immigrant households exhibit gains from trade that are 0.56 percentage points greater than an equivalent native household. The largest differential, between immigrants earning over $100,000 a year and native households earning less than $10,000 a year, is just over a percentage point, compared to an unconditional baseline of 1.72%.

Columns 4 to 6 provide the same analysis and a similar set of results, but scaled by household expenditure on groceries. Given that groceries exhibit a negative income elasticity, high-income households benefit less from the grocery-specific gains from trade than low-income households. Still, immigrant effects persist: the unconditional average gains from trade for immigrants are 36% greater than for native households.

While these estimates focus on the consumer gains from trade, the same import expenditure differential documented here also leaves immigrants more exposed to increases in trade costs, such as tariffs. As a first-order approximation, the fact that the unconditional immigrant consumer gains from trade are 42% greater than those of natives also suggests that immigrants face relative costs associated with tariff increases of the same magnitude.

This section thus provides novel insights regarding import preference heterogeneity across natives and immigrants and the heightened exposure of immigrants to trade shocks. As is common in the literature on import expenditure heterogeneity, we have assumed throughout this section that immigrants do not affect the local supply of imports. This assumption is at odds with research documenting the trade-creating effects of immigrants, which suggests that immigrants reduce trade costs with their origin country. In the following section, we develop a heterogeneous firms model of import supply, in which we allow immigrants to affect the supply of imports and therefore the distribution of the gains from trade across native households.

4 Immigrants and Import Supply

In this section we examine how immigrants affect local import supply. Specifically, we microfound the supply-side of imports in which immigrants may reduce trade costs and expand market size. Combined with the demand system introduced in Section 3, we then estimate the model and quantify the effect of immigrants through imports and local market size on trade volumes and native welfare.

4.1 A Heterogeneous Firms Model of Import Supply

In this section, we develop a heterogeneous firms model in which immigrants may affect (i) prices through variable trade costs, (ii) variety availability through fixed trade costs and market size, and (iii) native preferences for imported goods. We opt for the Melitz (2003) model to microfound the supply side for two reasons. First, the model features fixed costs and hence market size effects, a key channel through which immigrants may affect the supply of varieties locally (Iranzo and Peri 2009; Di Giovanni et al. 2015; Aubry et al. 2016). Second, the structure of the model allows us to fully leverage our data and separately quantify the variable trade cost, fixed trade cost, preference spillover, and market size effects of immigrants on native households, thus separating the supply from the demand effects of immigrants on import penetration.

Preferences: The demand-side remains a CES aggregator across grocery varieties with a country-specific preference term z_{oh} as in equation (3). Recall that exogenous import preferences are determined by the vector of observable household characteristics via the parameters δ, ζ_1, and ζ_2. In addition, household preferences may endogenously respond to the local immigrant population share from a given origin via β_z, a process we term preference diffusion.

Firms: Each country $o \in \mathcal{O}$ has some exogenous size Y_o and marginal cost of production w_o. Trade is characterized by origin-by-county iceberg trade costs τ_{oc} and fixed costs of exporting f_{oc}. Each firm draws some productivity φ from a Pareto distribution with shape parameter $\theta > \sigma - 1$.[28] The set of potential entrant firms in each origin is proportional to the size of that origin Y_o. The cost of providing q units to destination county c by a firm in origin o with productivity φ is $C_{oc}(q) = \frac{w_o \tau_{oc}}{\varphi} q + f_{oc}$. All entry and pricing decisions are made at the county level such that each county is an independent market.

We allow both types of trade costs to vary according to a vector of distance measures d_{oc}, the local immigrant population share I_{oc}, and an unobserved component:

$$\tau_{oc} = \exp[-\frac{1}{\theta}(\rho^\tau d_{oc} + \beta^\tau I_{oc})] \times \eta_{oc}^\tau, \tag{5}$$

$$f_{oc} = \exp[-(\frac{\sigma - 1}{1 + \theta - \sigma})(\rho^f d_{oc} + \beta^f I_{oc})] \times \eta_{oc}^f, \tag{6}$$

where η_{oc}^τ and η_{oc}^f represent idiosyncratic deviations in trade costs across county-origin pairs with a mean of one. β^τ captures the strength of the the variable cost channel of immigrants and β^f the fixed cost channel of immigrants.[29] We normalize domestic trade costs $\tau_{us,c}$ and $f_{us,c}$ to 1.

Firms price according to monopolistic competition and thus set constant mark-ups. The optimal pricing function for any variety ω from origin o in county c is

$$p_{\omega(o),c} = \frac{\sigma}{\sigma - 1} \frac{w_o \tau_{oc}}{\varphi(\omega)}. \tag{7}$$

The only bilateral factor affecting variety price is variable trade costs.

Gravity: Household-level expenditure on goods from origin o can be expressed as:[30]

$$X_{oh} = X_h P_h^{\sigma-1} \tau_{o,c(h)}{}^{-\theta} \Big(\frac{f_{o,c(h)}}{S_{c(h)} z_{o,c(h)}} \Big)^{-(\frac{\theta}{\sigma-1} - 1)} z_{oh} \tag{8}$$

where $S_{c(h)}$ is real grocery sector expenditure in county c;[31] average county-level preferences $z_{o,c(h)}$ are an expenditure-weighted average of the preference shifter z_{oh} across all households

[28] We assume that θ is identical across origin countries.

[29] The normalization terms $\frac{1}{\theta}$ and $\frac{\sigma-1}{1+\theta-\sigma}$ simplify notation in subsequent steps but are not necessary.

[30] We relegate the full derivation of model equations to Appendix B.1.

[31] Formally: $S_{c(h)} = \sum_{h' \in \Lambda_{c(h)}} X_{h'} P_{h'}^{\sigma-1}$, where $\Lambda_{c(h)}$ is the set of households residing in h's county of residence $c(h)$.

in $c(h)$; X_h is total spending on the grocery sector by h, and P_h is the price index faced by h. The real size of origin o relative to the U.S. is defined as $\exp(\alpha_o) \equiv \lambda Y_o w_o^{-\theta}$, where λ is a collection of parameters defined in Appendix Section B.1.

Increasing returns to scale due to fixed costs feature in equation (8) via the term raised by the exponent $\theta/(\sigma-1)-1$. Market size can be divided into two components: $S_{c(h)}$ which measures expenditure market size due to total real expenditures on goods from all origins, and country-specific preference market size due to average consumer preferences for goods in the county, $z_{o,c(h)}$.[32] The firm productivity dispersion parameter θ and consumer elasticity of substitution σ determine the strength of returns to scale in our framework.

As before, we consider a gravity equation of expenditure on imports relative to expenditure on U.S. produced goods for a given household. We express household-by-origin expenditure (relative to expenditures on domestic produced goods) as

$$\tilde{X}_{oh} = \exp(\alpha_o)(\tau_{o,c(h)})^{-\theta}\Big(\frac{f_{o,c(h)}}{z_{o,c(h)}}\Big)^{-(\frac{\theta}{\sigma-1}-1)} z_{oh}. \tag{9}$$

Plugging in our functional form assumptions, equations (3), (5), and (6), we obtain a household-level gravity equation

$$\tilde{X}_{oh} = \exp\left(\alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)} + \ln \bar{z}_{o,c(h)}^{\frac{\theta}{\sigma-1}-1} + \ln \bar{z}_{oh}\right) \times \eta_{o,c(h)} \times \eta_{oh}^z \tag{10}$$

with the following definitions:

$$\rho = \rho^\tau + \rho^f,$$

$$\beta = \beta^\tau + \beta^f + (\frac{\theta}{\sigma-1})\beta^z,$$

$$\eta_{o,c(h)} = \eta_{o,c(h)}^\tau \times \eta_{o,c(h)}^f,$$

$$z_{oc} = e^{\beta^z I_{oc}} \bar{z}_{oc} \qquad \text{and} \qquad z_{oh} = e^{\beta^z I_{oc}} \bar{z}_{oh}.$$

[32] The preference shifter $z_{o,c(h)}$ represents the heterogeneous preference extension beyond the standard Melitz (2003) framework. In particular, as preferences shift towards goods from origin o, more firms from o are able to cover the fixed costs of supplying county c, which further increases the market penetration of imports from o to county c.

4.2 Estimation

Separating the effect of immigrants on trade volumes through local preferences—captured by the terms $\bar{z}_{o,c(h)}$, $\bar{z}_{oh}$, and β^z—from their effect through trade costs—captured by β^τ and β^f—as well as market size is essential for quantifying the welfare impact of immigrants on non-immigrants. In this section we provide a three-step estimation strategy to do so. We start by estimating the total spillover effect of immigrants on import purchases. We then estimate their effect on prices. Finally, we estimate the extensive margin effect of immigrants in order to separate out their effect on fixed trade cost from preference diffusion.

4.2.1 Identifying the Channels of Immigrant-Induced Imports

Step 1: Estimating $\boldsymbol{\beta}$. The total spillover effect of immigrants on natives is captured by β in equation (10). To make this equation estimable, we must impute the unobserved county-average preference term $\bar{z}_{o,c(h)}$. We do so using the individual preferences $\bar{z}_{oh}$ estimated in Section 3.4. We denote imputed preferences terms with an apostrophe, i.e., $\bar{z}'_{oh}$. As discussed in Section 3, we calibrate $\theta = 5$. We assume a value for the CES elasticity parameter of $\sigma = 5$.[33]

With the preference terms in hand, we then difference out both $\bar{z}'_{oh}$ and $(\bar{z}'_{o,c(h)})^{\frac{\theta}{\sigma-1}-1}$ from the relative expenditure share $\tilde{X}_{oh}$ to obtain the following equation:

$$\frac{\tilde{X}_{oh}}{\mathcal{Z}_{oh}} = \exp\left(\alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)}\right) \times \eta_{oh}, \tag{11}$$

in which we define $\mathcal{Z}_{oh} = \bar{z}'_{oh}(\bar{z}'_{o,c(h)})^{\frac{\theta}{\sigma-1}-1}$ and $\eta_{oh} = \eta_{o,c(h)} \times \eta^z_{oh}$ to simplify notation.

The deflated expenditure term in equation (11), $\tilde{X}_{oh}/\mathcal{Z}_{oh}$, represents variation in import expenditure across households that is solely attributable to variation in trade costs. That is, $\tilde{X}_{oh}/\mathcal{Z}_{oh}$ represents import expenditure by household h conditional on the implied import demand associated with the observed characteristics of h and the observed characteristics of all other households within the same county. The residual variation in $\tilde{X}_{oh}$ net of $\mathcal{Z}_{oh}$

[33] Recall that $\theta > \sigma - 1$ is a restriction inherent to the model. Melitz and Redding (2015) calibrate $\theta = 4.25$ when $\sigma = 4$ and Simonovska and Waugh (2014) estimate the trade elasticity as 4.10 and 4.27, depending on specification. We opt for the relatively larger value of $\theta = 5$ from Costinot and Rodríguez-Clare (2014) in order to match our larger value of $\sigma = 5$.

therefore captures the effect of immigrants on local import expenditure that is common to all households and unexplained by observed household characteristics. Through the lens of the model described here, this residual variation is attributed to variation in local immigrant population shares via β.

We estimate equation (11) using PPML. Estimating β alone, however, does not facilitate computing the welfare effect of immigrants on natives because immigrants may change the preferences of natives. We next leverage model restrictions and data characteristics in order to separate preference diffusion β^z from the trade cost effects of immigrants.

Step 2: Estimating β^τ. Immigrants may benefit natives by reducing the price of imported goods. We test this hypothesis using the highly detailed price data available in our household purchase data.

Because the bilateral price only depends on variable trade costs net of county and variety fixed effects, we can leverage equation (7) to isolate β^τ. To do so we aggregate our data to the barcode-by-county level, incorporate the functional form assumption of τ_{oc} from equation (5), and estimate the following log-linearized price equation:

$$\ln p_{\omega(o),c} = \psi_c + \psi_{\omega(o)} - \frac{\beta^\tau}{\theta} I_{oc} - \frac{\rho^\tau}{\theta} d_{oc} - \frac{1}{\theta} \ln \eta^\tau_{oc}, \tag{12}$$

where ψ_c and $\psi_{\omega(o)}$ represent county- and barcode-level fixed effects.[34] Using data at the barcode level, equation (12) allows us to identify the effect of immigrants on the price of imported varieties rather than the effect of immigrants on the average price of imported goods. This is an important benefit of our dataset, as our model predicts that an average price measure would conflate the effect of immigrants on within-variety prices and the effect of immigrants on variety entry and exit.[35]

[34]Since barcodes are unique to origin countries, $\psi_{\omega(o)}$ absorbs variation in production costs w_o across origins.

[35]When constructing $p_{\omega(o),c}$ we simply pool all purchases of barcode ω in county c and calculate a sales-weighted average price. One concern may be that retail firms price nationally (DellaVigna and Gentzkow 2019), leaving us with little variation in price across counties. However, we are leveraging household purchase data, not store-level data. This means that the average price across retail establishments within county varies for consumers. We graphically depict the average price variation across counties relative to the average variation in immigrant population shares in Appendix Figure C.5. Consistent with the above logic, we find substantial variation in average prices (net of barcode fixed effects) across counties.

Step 3: Estimating $\boldsymbol{\beta^f}$ and $\boldsymbol{\beta^z}$. We quantitatively separate the effect of immigrants on fixed costs β^f and the effect of immigrants on local preferences β^z by comparing the total import elasticity of immigrants with the extensive margin import elasticity of immigrants.

Specifically, we follow Chaney (2008) and derive expressions for both the extensive margin elasticity of imports with respect to the immigrant population share and the total expenditure elasticity of imports with respect to the immigrant population share. When $\beta^\tau \approx 0$, which we show to be true in Section 4.2.3, this derivation yields two equations and two unknowns: β^f and β^z. The scanner data used in this paper allow us to count the number of barcodes from each origin purchased by each household. We thus estimate the extensive margin effect of immigrants on trade directly by replacing $\tilde{X}_{oh}$ in equation (11) with $\tilde{N}_{oh}$: the count of barcodes from origin o in household h's consumption basket divided by the count of barcodes from the U.S. in household h's consumption basket.[36]

While the full derivation is provided in Appendix B.2, it is straightforward to show that our functional form assumptions, equations (3) and (6), yield the following two expressions regarding the import expenditure elasticity and import variety elasticity, respectively:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1}\right)\beta^z,$$

$$\frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1} - 1\right)\beta^z.$$

The intuition follows directly from the results derived in Chaney (2008): firms enter a new market if and only if they can cover their fixed costs. Immigrants can thus facilitate firm entry by (i) reducing fixed costs, or (ii) increasing the intensive margin of sales-per-variety in their market. Yet, expenditure per barcode only increases if preferences become more favorable to the firm in question, as a reduction in fixed costs has no impact on within-firm sales, conditional on exporting. This intuition is reflected in the system of equations above: as the total trade elasticity and the variety trade elasticity converge, $\beta^z \to 0$ and $\beta^f \to \beta$.

[36] While the store-level scanner data from NielsenIQ offers a more complete count of *within-store* barcode availability, it does not fully capture the extent to which non-NielsenIQ stores are distributed across counties or the extent to which non-NielsenIQ stores differ in their stocking decisions from stores within the NielsenIQ sample. We therefore opt for the household-level data when estimating β^f in order to maintain a sample that is consistent with our estimate of β.

4.2.2 Identification and Instrumental Variables

Estimates of β, β^f, β^z, and β^τ recovered by applying OLS to, respectively, equations (11) and (12), are likely biased due to endogenous sorting of immigrants into locations with high import availability. For example, low transportation costs between New York and Italy may independently expand both the local import expenditure on Italian goods and the population of Italian immigrants. To deal with such origin-by-county specific confounders, we adopt the instrumental variable approach of Burchardi et al. (2019).[37]

The instrument interacts the arrival into the U.S. of immigrants from origin country o (the push) with the attractiveness of destination d to all immigrants (the pull) during a given historical decade D. We follow Burchardi et al. (2019) and leave out both the continent of origin country o when computing the pull component and leave out the Census region of county c when constructing the push component. Formally, the instrument is defined as

$$IV_{o,c}^{D} = \underbrace{L_{o,-r(c)}^{D}}_{\text{Push}} \times \underbrace{\frac{L_{-\mathcal{C}(o),c}^{D}}{L_{-\mathcal{C}(o)}^{D}}}_{\text{Pull}}, \tag{13}$$

where $r(c)$ is the Census region of county c, and $\mathcal{C}(o)$ the set of countries on o's continent. $L_{o,-r(c)}^{D}$ is the number of immigrants from o settling in the U.S. outside the Census region of county c in decade D and $L_{-\mathcal{C}(o),c}^{D}/L_{-\mathcal{C}(o)}^{D}$ is the fraction of immigrants arriving to the U.S. in decade D who come from outside the continent of o and choose to settle in county c.

The identification assumption is that any confounding factors that make a given county more attractive to both immigrants and importing firms from a given country do not simultaneously affect the interaction of (i) the settlement of immigrants from other continents with (ii) the total number of immigrants arriving from the same country but settling in a different Census regions.

We use equation (13) to predict immigrant inflows into the U.S. for all decades between 1880 to 2000, with Appendix Table A.1 providing the first-stage estimates. When estimating

[37] We provide only a brief description of the instrumental variable strategy here, as our approach follows closely that of Burchardi et al. (2019). We refer the interested reader to Appendix A.2 for more details, as well as to the growing literature making use of this same identification strategy: Terry et al. (2022), Choi et al. (2024), McCully (2024), and Bonadio (forthcoming).

equation (12), we use 2SLS to incorporate these instruments. When estimating equation (11), however, we encounter many zeros in household expenditure and variety shares. We again use PPML and account for the non-linearity of PPML by taking a control function approach (Petrin and Train 2010; Morten and Oliveira 2024). In particular, we add the residuals from the first-stage instrumental variable regressions as controls in our main specifications.[38]

4.2.3 Estimating Immigrant Spillover and Channel Magnitudes

We start with the estimate of the total effect of immigrants on imports using equation (11), in which expenditure is deflated by household- and county-level preferences.[39] Columns 1 and 2 of Table 3 provide estimates of β with and without the use of the instrument from Burchardi et al. (2019).[40] In our preferred control function specification, we estimate $\hat{\beta} = 1.36$. Controlling for immigrants' endogenous location choice reduces the magnitude of the immigrant spillover effect, suggesting that immigrants are in fact locating in places in which goods from their origin are more widely available.[41]

With the estimate for β in hand, we now turn to decomposing β into its constituent channels. We estimate equation (12), with the estimates appearing in Table 4. In columns 1 and 2, we use variation across all barcodes regardless of how regularly we observe them across counties. To address concerns about products sold in only a handful of counties driving our results, we restrict the sample to barcodes which we observe in at least 100 counties in the NielsenIQ data in columns 3 and 4. In columns 2 and 4 we instrument for the bilateral immigrant-population share using the leave-out push-pull instrumental variables defined in

[38] Atalay et al. (2019) demonstrate in a Monte Carlo simulation that the control function approach generates estimates when using PPML that are quite close to the true data generating process. They further show that the PPML control function estimates are very similar to those produced by the related GMM estimation strategy developed by Wooldridge (1997) and Windmeijer (2000).

[39] We plot the normalized distribution of imputed household- and county-level preferences, $\bar{z}'_{oh}$ and $\bar{z}'_{o,c(h)}$, in Appendix Figure C.4.

[40] Burchardi et al. (2019) find no effect of immigrants on trade using state-level trade data and state fixed effects. When we aggregate our household data to the state level and control for state fixed effects, our headline estimates remain positive and significant. We provide further details regarding these differing results in Appendix A.3.4.

[41] As robustness checks, we estimate equation (11) without adjusting the dependent variable for the preference term $\mathcal{Z}_{oh}$ in columns 1 and 2 of Appendix Table C.4. We also estimate equation (11) using the full sample (including those for which we do not observe nativity) in columns 3 and 4 of Table C.4. For both robustness checks, the estimated coefficient is modestly higher but still fairly close to our baseline estimate of β.

Table 3: Estimates of Household Gravity Equation

	$\tilde{X}_{oh}/\mathcal{Z}_{oh}$		$\tilde{N}_{oh}/\mathcal{Z}_{oh}$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.50***	1.36***	1.29***	1.30***
	(0.30)	(0.49)	(0.17)	(0.39)
First-stage residuals		0.18		-0.0089
		(0.49)		(0.49)
N	1,461,130	1,461,130	1,461,130	1,461,130
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		19.6		19.6

Notes: The table presents regression results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share in column 2. Observations are weighted using NielsenIQ household weights. Standard errors and first-stage F-statistics are computed over 1,000 bootstrapped samples of households, in which the z_{oh}, $z_{o,c(h)}$, first-stage, and control function are estimated on the same sample. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Estimates of Variable Cost Parameter using Variation in Prices

	Dependent variable: Log Average Barcode Price			
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	-0.041***	-0.017	-0.058***	-0.040
	(0.013)	(0.031)	(0.016)	(0.044)
N	2,261,777	2,261,777	1,601,674	1,601,674
Barcode FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		17.3		17.5
Sample	All barcodes		barcodes in >100 counties	

Notes: The table presents two-stage least square regression results at the barcode-county level. The instrumental variables strategy is described in Section 4.2.2. Standard errors are clustered at the barcode and country level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

equation (13).

The OLS estimates are statistically significantly negative, suggesting more immigrants from a given origin correspond to lower prices for goods imported from that origin. However, the estimated coefficients fall and become insignificant when adjusting for endogenous immigrant location choice in columns 2 and 4. This suggests that immigrants tend to sort into counties in which the price for goods from their origin country is lower.

The point estimate in column 2 implies that a one percentage point increase in the share of the local population from o decreases prices by 0.017 percent, implying that $\hat{\beta}^{\tau} = -0.085$.[42] Due to its small magnitude and statistical insignificance, we conclude that $\hat{\beta}^{\tau} \approx 0$. This represents the first direct test of how immigrants affect variable trade costs. The null result is consistent with the untested assumption made by Peri and Requena-Silvente (2010).

We next estimate equation (11) but replace the relative expenditure term $\tilde{X}_{oh}$ with the relative variety count share $\tilde{N}_{oh}$ in order to recover the extensive-margin import elasticity of immigrants on import expenditure. Columns 3 and 4 of Table 3 provide estimates of the extensive margin effect of immigrants on import expenditure. Our preferred coefficient estimate is 1.3, implying that a one percentage point increase in immigrants from a given origin raises the share of varieties purchased from that origin by 1.3 percent.

Solving for β^f and β^z using the elasticity estimates from columns 2 and 4 of Table 3, we recover β^f =1.28 and β^z =0.06. Since our estimate of β from Table 3 is 1.36, we conclude that the primary spillover channel through which immigrants affect non-immigrant households is the fixed cost channel, while the preference diffusion channel is quantitatively negligible. Therefore, in the subsequent counterfactual analysis, we set $\beta^z = 0$.

5 Import and Welfare Effects of Immigrants

5.1 Implementing the Counterfactuals

We leverage the model introduced in Section 4.1 and the parameters estimated in both Section 3 and Section 4.2 to run two counterfactuals:

1. **Turning immigrants into natives:** Remove the channels through which immigrants affect the import expenditure of households in the U.S. That is, set $\zeta_1 = \zeta_2 = \beta_f = 0$ and recalculate the preference-induced market size term z_{oc} accordingly.

[42] Note that the estimated immigrant population share coefficient is equal to $-\frac{\beta^{\tau}}{\theta}$.

2. **Removing all immigrants:** Set $\zeta_1 = \zeta_2 = \beta_f = 0$ and remove all grocery expenditure associated with immigrant households. This corresponds to removing immigrants from the U.S. population altogether, thus capturing the total market size benefits of immigrants.

Note that our counterfactuals only allow for partial-equilibrium adjustment in the consumption space. The local labor market effects of immigrants are outside the scope of our framework. For details on how we compute counterfactual trade flows and utility, see Appendix B.3.

To generate counterfactual results which are representative of the U.S. as a whole and meaningful for each county, we leverage the pooled 2013-2017 ACS sample. In particular, we use the results from estimating equation (4) with the NielsenIQ data to map household socioeconomic characteristics to predicted relative import expenditure shares. We use the crosswalks provided by Burchardi et al. (2019) to generate county-specific immigrant population shares based on the Public Use Microdata Area of residence. To compute dollar-equivalent utility values, we assume that each household spends \$7,500 on grocery and personal care products, which is close the value computed from the Consumer Expenditure Survey.

5.2 Aggregate Effects

We summarize the results in Table 5. The results from the first counterfactual scenario—turning immigrants into natives—appear in the first row. Averaging across households, we find that aggregate U.S. expenditures on imports of grocery and personal care items fall by 7.7%.[43] We also find that removing all immigrant spillover effects yields a negligible average welfare loss of 0.039%, amounting to a welfare-equivalent fall of \$2.9 per household-year.

To understand which mechanisms drive the counterfactual effects, we shut down each effect one at a time. The results appear in the middle section of Table 5. The key finding is that immigrants generate nearly three times more import expenditure via their preferences than by reducing trade costs.

[43] It is worth noting that our elasticity lies in between the range of estimates surveyed by Felbermayr et al. (2015), 0.12–0.15, and the null result reported in Burchardi et al. (2019).

Table 5: Counterfactual Results Summary

Counterfactual exercise:	(1) Change (%) import expenditure	(2) Change (%) welfare natives	(3) Change ($) welfare per native HH
Turning immigrants into natives	-7.7	-0.039	-2.9
Shutting down ...			
... fixed trade cost channel	-2.0	-0.035	-2.6
... market size channel	-0.3	-0.005	-0.3
... composition channel	-5.7	–	–
... homophily channel	-1.4	–	–
Removing all immigrants	-26	-0.932	-70

Notes: This table shows the change in outcomes under various counterfactual scenarios. The first counterfactual, with results shown in the first row, removes all channels through which immigrants affect grocery import expenditures—i.e. we set $\zeta_1 = \zeta_2 = \beta_f = 0$ and recalculate $\bar{z}_{oc}$—but keeps total household expenditure constant. The next four rows show the results when only the following parameters change one at a time: β_f (fixed trade cost); $\bar{z}_{oc}$ (market size); ζ_1 and ζ_2 (composition); and ζ_2 (homophily). In the last row, we remove all immigrant channels and the grocery expenditures made by immigrants (counterfactual 2), equivalent to removing immigrants from the U.S. population. For column 3, we assume that each household spends $7,500 on grocery and personal care products, which is close the value computed from the Consumer Expenditure Survey.

Removing the effect of immigrants on fixed trade costs reduces import expenditure by 2%, implying that this channel contributes nearly a quarter to the total effect. The indirect effect of immigrants' preferences through the market size term $z_{o,c(h)}$ is negligible.[44] In contrast, the preferences of immigrants have a substantial direct effect on trade volumes: removing the composition channel associated with immigrants' preference parameters ζ_1 and ζ_2 causes a decline of 5.7%, about three quarters of the total effect of all immigrant channels. The fifth row shows the impact of only removing immigrants' preferences for goods from their own origin. The significant reduction in magnitude suggests that immigrants' preference for imports from any origin (ζ_1) drive the bulk of the preference composition channel effect.

[44]The strength of the market size channel is governed by our calibration of $\theta/(\sigma-1) = 1.25$. As a sensitivity analysis, we calibrate $\theta/(\sigma-1) = 2$ and find that the welfare cost to natives of removing immigrant effects remains modest at -0.05%. In this alternative specification, the market size channel doubles in terms of its contribution to the total welfare cost.

The change in import expenditure associated with removing immigrant channels is generally larger than the associated change in *welfare*. In particular, the effect of immigrants on fixed trade costs and market size are the only welfare-relevant channels through which immigrants increase native welfare via increased import expenditure. Yet these two channels capture just over a quarter of the total effect of immigrants on local import expenditure. A naive application of the ACR welfare formula to the aggregate trade-creating effect of immigrants would therefore overestimate the welfare gains to natives by just under a factor of four.[45]

The last row of Table 5 shows the second main counterfactual, in which we remove all channels of the first counterfactual as well as immigrants' grocery expenditures. In the model, this corresponds to a reduction in $S_{c(h)}$, the real market size of county c, as opposed to only changing the preference-driven market size z_{oc} as in the previous counterfactual. Accounting for this additional expenditure market size effect leads to a decline in import expenditure by 26%. The average loss in grocery consumption welfare for natives in this scenario is 0.93% or a welfare-equivalent fall of \$70 per household-year. Thus, removing immigrants' expenditures leads to a sizeable welfare effect on natives.[46] As the scale of demand decreases, fewer varieties—both domestic and foreign—are available for native consumers.

5.3 Distributional Effects

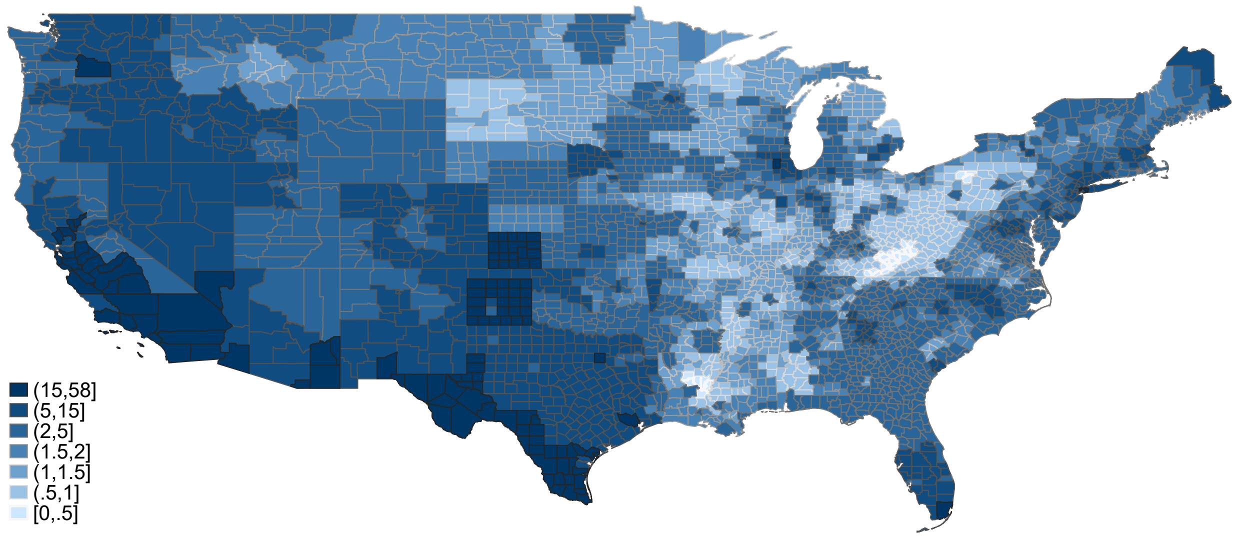
The average estimates discussed here mask considerable heterogeneity across space and across the income distribution, which we explore next.

Immigrant effects across U.S. counties. The impact of immigrants varies substantially across space, with cities much more affected than rural areas. We depict the geographic variation in import spending declines associated with Counterfactual 1 in Figure 3. Appendix Figure C.6 maps the average dollar-equivalent change in utility associated with our second counterfactual: removing immigrants entirely. In both cases, the impact of immigrants on

[45] Assuming an initial import expenditure share of 9%: $[(1 - 0.09(1 - 0.077))/(1 - 0.09)]^{(-1/\theta)} - 1 = -0.00152$. The naive welfare estimate is therefore -0.152%, as opposed to our estimated -0.039%.

[46] Piyapromdee (2021) estimates that a counterfactual 25% increase in the immigrant stock would increase native welfare by 1.3% when considering both labor and housing market effects. Albert and Monras (2022) compute a 1.6% welfare increase for natives resulting from immigrant consumption patterns.

Figure 3: Spatial Distribution of Fall in Imports due to Removing Immigrant Effects



Notes: This chart plots the percent decrease in the value of grocery and personal care imports when the trade-creating effects of immigrants are removed (counterfactual 1) following the procedure outlined in Appendix Section B.3.

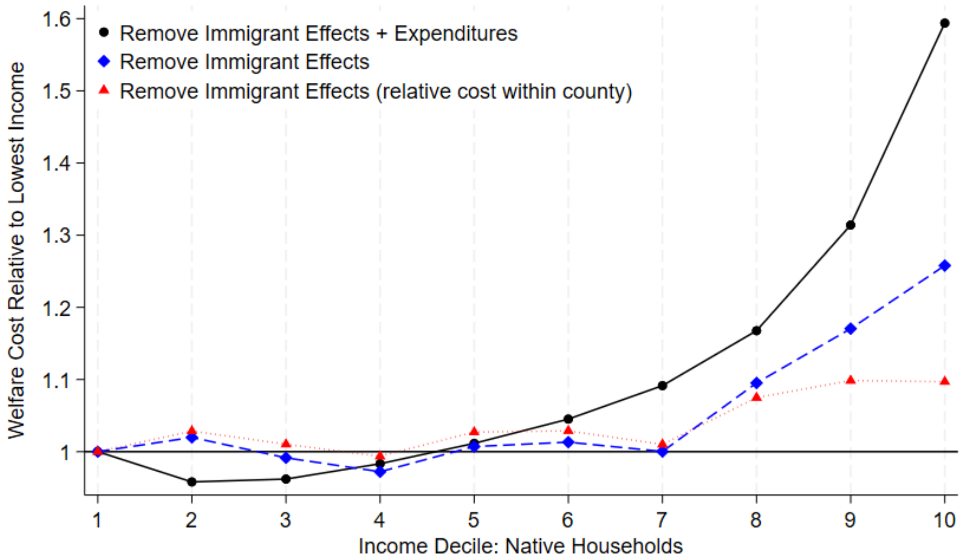
imports and welfare is remarkably concentrated in the Southwest, West Coast, and East Coast of the U.S., as well as almost all major cities.

The counties experiencing the largest drop in import expenditures under our first counterfactual are El Paso, TX (-44%); Los Angeles, CA (-27%); Kern, CA (-25%); Riverside, CA (-23%); and Fresno, CA (-23%). Assuming an initial tariff rate of 2.5% applied to the grocery goods studied here, as well as our calibration of $\theta = 5$, these estimated changes are equivalent to Los Angeles experiencing a three-fold increase in tariff rates. For our second counterfactual, welfare effects are large and heterogeneous across space. In terms of annual household dollar-equivalent welfare effects for large counties, the most affected are Queens, NY ($386); Miami-Dade, FL ($356); Hudson, NJ ($309); Santa Clara, CA ($293); and Los Angeles, CA ($280).

Immigrant effects by native household income. The effect of immigrants may vary substantially across the income distribution, as the expansive literature on labor market effects has emphasized (e.g., Dustmann et al. 2013). We are the first to look at distributional effects on the consumption side, enabled by our highly detailed household-level data.

Figure 4 depicts the welfare losses across the income distribution in different counter-

Figure 4: Percent Change in Grocery Welfare by Income Decile



Notes: The chart depicts average welfare costs at the income decile level. The solid black line depicts the welfare costs of removing immigrant spillovers and expenditure, our second counterfactual. The dashed blue line depicts the welfare costs of removing immigrant spillovers, our first counterfactual. The dotted red line calculates the average welfare differential of native households associated with our first counterfactual but within counties. All averages are then normalized relative to the lowest income decile.

factual scenarios, each computed relative to the lowest income decile.[47] In all cases, there is very little difference in welfare effects associated with the first six income deciles. By contrast, the welfare gains among the top four deciles are monotonically increasing.

The blue dashed line depicts our first counterfactual: removing immigrants' distinctive effects on import consumption but not immigrants' total expenditure. Households in the highest income decile face average costs of losing access to immigrant-induced imports that are 25% larger than households at or below the seventh income decile.

To understand the sources of the unequal gains from immigrants in our first counterfactual, we conduct a second exercise. The red dotted line depicts the welfare effects of our first counterfactual relative to county-level average effects, thus isolating the role of native preferences in shaping the cost of this counterfactual across deciles (rather than geographic sorting of natives and immigrants). Just under half of this differential is driven by stronger preferences for imports exhibited by the highest-income households, with geographic sorting

[47]Across income groups, we fix expenditure on consumer packaged goods, as implied by equation (1). Hence, variation in the welfare impact of immigrants across income groups is driven instead by the spatial sorting of immigrants and heterogeneous preferences for imported goods.

of immigrants with high-income households explaining the remaining half.

The black solid line depicts our second counterfactual, in which we additionally remove immigrant expenditure. This reduction in market size has even more pronounced distributional consequences, with households in the highest-income decile facing costs that are 60% greater than households at or below the median income level. The estimates presented here shed light on the remarkable variation in the consumer gains from immigrant populations across cities, counties, and income groups within the U.S. Of particular note is the striking pattern of high-income urban native households benefiting substantially—even within the same county—relative to lower-income natives.

6 Conclusion

This paper introduces and quantifies a novel dimension of consumer heterogeneity with respect to the gains from trade: household nativity. By linking U.S. scanner data to both barcode and household origin country data, we provide evidence that immigrants exhibit grocery consumer gains from trade that are 42% greater than those of native households. While we find evidence that immigrants also increase the local supply of imports—thereby generating consumer gains from natives—over three quarters of the aggregate effect of immigrants on import expenditure is attributable to immigrant preferences. This study therefore contributes to the literature on the costs and benefits of immigration for natives. Our estimates, paired with estimates of the labor and housing market effects of immigrants, may be useful for a general equilibrium analysis to quantify the aggregate welfare effect of immigrants.

Our results are consistent with recent political trends across socioeconomic groups in the U.S. Immigrant, high-income and highly educated households pay an outsized share of the cost of tariffs, and high-income and urban natives pay an outsized share of the cost of immigrant removals. These same households tend to oppose tariffs and anti-immigrant policies (O'Rourke and Sinnott 2006; Hanson et al. 2007; Tabory and Smeltz 2017; Stantcheva 2022). Further research is needed to understand the relationship between consumption preferences and voting behavior.

References

Albert, Christoph and Joan Monras, "Immigration and Spatial Equilibrium: The Role of Expenditures in the Country of Origin," *American Economic Review*, November 2022, *112* (11), 3763–3802.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Atalay, Enghin, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson, "How wide is the firm border?," *Quarterly Journal of Economics*, 2019, *134* (4), 1845–1882.

Atkin, David, "The caloric costs of culture: Evidence from Indian migrants," *American Economic Review*, 2016, *106* (4), 1144–1181.

_ , Eve Colson-Sihra, and Moses Shayo, "How do we choose our identity? A revealed preference approach using food consumption," *Journal of Political Economy*, 2021, *129* (4), 1193–1251.

Aubry, Amandine, Michał Burzyński, and Frédéric Docquier, "The welfare impact of global migration in OECD countries," *Journal of International Economics*, 2016, *101*, 1–21.

Auer, Raphael, Ariel Burstein, Sarah Lein, and Jonathan Vogel, "Unequal expenditure switching: Evidence from Switzerland," *Review of Economic Studies*, 2024, *91* (5), 2572–2603.

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

Bonadio, Barthélémy, "Migrants, trade and market access," *Review of Economics and Statistics*, forthcoming.

Borjas, George J, "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market," *Quarterly Journal of Economics*, 2003, *118* (4), 1335–1374.

Borusyak, Kirill and Xavier Jaravel, "The distributional effects of trade: Theory and evidence from the United States," Technical Report, National Bureau of Economic Research 2021.

Bowles, Samuel, "Endogenous preferences: The cultural consequences of markets and other economic institutions," *Journal of Economic Literature*, 1998, *36* (1), 75–111.

Bronnenberg, Bart J, Jean-Pierre H Dubé, and Matthew Gentzkow, "The evolution of brand preferences: Evidence from consumer migration," *American Economic Review*, 2012, *102* (6), 2472–2508.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan, "Migrants, ancestors, and foreign investments," *Review of Economic Studies*, 2019, *86* (4), 1448–1486.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the United States," *Econometrica*, 2020, *88* (3), 1071–1112.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Chaney, Thomas, "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–21.

Chen, Bo and David S. Jacks, "Trade, variety, and immigration," *Economics Letters*, 2012, *117* (1), 243–246.

Choi, Jaerim, Jay Hyun, and Ziho Park, "Bound by ancestors: Immigration, credit frictions, and global supply chain formation," *Journal of International Economics*, 2024, *147*, 103855.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer, "The trade-creating effects of business and social networks: Evidence from France," *Journal of International Economics*, 2005, *66* (1), 1–29.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin, "Verifying the existence of maximum likelihood estimates for generalized linear models," *arXiv preprint arXiv:1903.01633*, 2019.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade theory with numbers: Quantifying the consequences of globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

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

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in us retail chains," *Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Di Giovanni, Julian, Andrei A Levchenko, and Francesc Ortega, "A global view of cross-border migration," *Journal of the European Economic Association*, 2015, *13* (1), 168–202.

Dustmann, Christian, Tommaso Frattini, and Ian P Preston, "The effect of immigration along the distribution of wages," *Review of Economic Studies*, 2013, *80* (1), 145–173.

_ , Uta Schönberg, and Jan Stuhler, "Labor supply shocks, native wages, and the adjustment of local employment," *Quarterly Journal of Economics*, 2017, *132* (1), 435–483.

Faber, Benjamin and Thibault Fally, "Firm heterogeneity in consumption baskets: Evidence from home and store scanner data," *Review of Economic Studies*, 2022, *89* (3), 1420–1459.

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

Felbermayr, Gabriel, Volker Grossmann, and Wilhelm Kohler, "Migration, international trade, and capital formation: Cause or effect?," in "Handbook of the economics of international migration" 2015, pp. 913–1025.

Fernández, Raquel and Alessandra Fogli, "Culture: An empirical investigation of beliefs, work, and fertility," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 146–177.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi, "Slicing the pie: Quantifying the aggregate and distributional effects of trade," *The Review of Economic Studies*, 2023, *90* (1), 331–375.

Gould, David M., "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *Review of Economics and Statistics*, 1994, *76* (2), 302–316.

Hanson, Gordon H, Kenneth Scheve, and Matthew J Slaughter, "Public finance and individual preferences over globalization strategies," *Economics & Politics*, 2007, *19* (1), 1–33.

Head, Keith and John Ries, "Immigration and trade creation: econometric evidence from Canada," *Canadian Journal of Economics*, 1998, pp. 47–62.

_ **and Thierry Mayer**, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Hottman, Colin J and Ryan Monarch, "A matter of taste: Estimating import price inflation across US income groups," *Journal of International Economics*, 2020, *127*, 103382.

_ **, Stephen J Redding, and David E Weinstein**, "Quantifying the sources of firm heterogeneity," *Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Iranzo, Susana and Giovanni Peri, "Migration and trade: Theory with an application to the Eastern–Western European integration," *Journal of International Economics*, 2009, *79* (1), 1–19.

Jaccard, Torsten, "Retailers and Import Penetration: Evidence from US Dollar Stores," 2023.

Jaravel, Xavier, "The unequal gains from product innovations: Evidence from the us retail sector," *The Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

Llull, Joan, "Immigration, wages, and education: A labour market equilibrium structural model," *Review of Economic Studies*, 2018, *85* (3), 1852–1896.

Logan, Trevon D and Paul W Rhode, "Moveable feasts: A new approach to endogenizing tastes," *manuscript (The Ohio State University)*, 2010.

Mazzolari, Francesca and David Neumark, "Immigration and product diversity," *Journal of Population Economics*, 2012, *25* (3), 1107–1137.

McCully, Brett A, "Immigrants, legal status, and illegal trade," *Journal of International Economics*, 2024, *152*, 104016.

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

_ **and Stephen J Redding**, "New trade models, new welfare implications," *American Economic Review*, 2015, *105* (3), 1105–1146.

Miho, Antonela, Alexandra Jarotschkin, and Ekaterina Zhuravskaya, "Diffusion of gender norms: Evidence from Stalin's ethnic deportations," *Journal of the European Economic Association*, 2023, *22* (2), 475–527.

Monras, Joan, "Immigration and wage dynamics: Evidence from the Mexican Peso crisis," *Journal of Political Economy*, 2020, *128* (8), 3017–3089.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *American Economic Journal: Applied Economics*, 2024, *16* (2), 389–421.

O'Rourke, Kevin H and Richard Sinnott, "The determinants of individual attitudes towards immigration," *European journal of political economy*, 2006, *22* (4), 838–861.

Ottaviano, Gianmarco IP and Giovanni Peri, "Rethinking the effect of immigration on wages," *Journal of the European Economic Association*, 2012, *10* (1), 152–197.

Parsons, Christopher and Pierre-Louis Vézina, "Migrant networks and trade: The Vietnamese boat people as a natural experiment," *Economic Journal*, 2018, *128* (612), F210–F234.

Peri, Giovanni and Francisco Requena-Silvente, "The trade creation effect of immigrants: Evidence from the remarkable case of Spain," *Canadian Journal of Economics*, 2010, *43* (4), 1433–1459.

Petrin, Amil and Kenneth Train, "A control function approach to endogeneity in consumer choice models," *Journal of Marketing Research*, 2010, *47* (1), 3–13.

Piyapromdee, Suphanit, "The impact of immigration on wages, internal migration, and welfare," *Review of Economic Studies*, 2021, *88* (1), 406–453.

Rauch, James E, "Business and social networks in international trade," *Journal of economic literature*, 2001, *39* (4), 1177–1203.

Santos Silva, Joao MC and Silvana Tenreyro, "The log of gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Simonovska, Ina and Michael E Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Stantcheva, Stefanie, "Understanding of trade," Technical Report, National Bureau of Economic Research 2022.

Steingress, Walter, "The causal impact of migration on US trade: Evidence from political refugees," *Canadian Journal of Economics*, 2018, *51* (4), 1312–1338.

Tabory, Sam and Dina Smeltz, "The Urban-Suburban-Rural «Divide» in American Views on Foreign Policy," Technical Report, The Chicago Council on Global Affairs 2017.

Terry, Stephen J, Thomas Chaney, Konrad B Burchardi, Lisa Tarquinio, and Tarek A Hassan, "Immigration, Innovation, and Growth," 2022.

Windmeijer, Frank, "Moment conditions for fixed effects count data models with endogenous regressors," *Economics Letters*, 2000, *68* (1), 21–24.

Wooldridge, Jeffrey M, "Multiplicative panel data models without the strict exogeneity assumption," *Econometric Theory*, 1997, *13* (5), 667–678.

Online Appendix

Contents

A Empirical Appendix

A.1 Data Construction

We aggregate decennial census waves across individuals aged 16 and above to the county-by-origin level, applying the Census' individual sample weights. Immigrants are defined as those born outside the U.S. To compute decadal immigrant inflows from origin o into destination county c between two census years $t-10$ and t, denoted L^t_{oc}, we count only those respondents in t who immigrated to the U.S. between $t-10$ and t. Following Burchardi et al. (2019), in the first sample year the measure L^{1880}_{oc} includes all those that are either first-generation immigrants from o or second-generation immigrants whose parents were born in o.

Destination regions c are defined as 1990 counties and we use the transition matrices provided by Burchardi et al. (2019) to maintain consistent boundaries over time despite the Census providing changing geographies across waves. The U.S. geography of reference is called, "Historic counties" until 1940; then county groups in 1970/1980; and finally public-use microdata areas (PUMAs) from 1980 to the present.

The latest available transition matrix provided by Burchardi et al. (2019) is for the year 2010, in which PUMAs are based on 2000 boundaries. Thus, for the 2013-2017 ACS sample, in which PUMAs are based on 2010 boundaries, we use the crosswalk provided by the Missouri Census Data Center to transition PUMAs to 2000 boundaries before applying the corresponding transition to 1990 counties.

A.2 Instrumental Variables: Details and First-Stage Estimates

This section provides a more detailed discussion regarding our implementation of the leave-out push-pull instrumental variables introduced by Burchardi et al. (2019). The same instrument has been used by a recent crop of papers studying the effects of immigration, such as Terry et al. (2022), Bonadio (forthcoming), McCully (2024), and Choi et al. (2024).

The immigration leave-out push-pull instrument interacts the arrival to the U.S. of immigrants from origin country o (push) with the attractiveness of different destinations to immigrants (pull) measured by the fraction of all immigrants to the U.S. who choose to settle in county c. A simple version of the instrument is defined as

$$IV^D_{o,c} = L^D_o \times \frac{L^D_c}{L^D},$$

where L^D_o is the number of immigrants from origin o coming to the U.S. in decade D, and L^D_c/L^D is the fraction of immigrants to the U.S. who choose to settle in county c in that decade.

Two key threats to the exogeneity of the instrument remain: a scale component and a spatial correlation component. The scale component is the threat that a single origin o constitutes a large share of the instrument's components for a given county c. A simple solution would be to leave out the bilateral immigration $L^D_{o,c}$ flows when constructing the instrument for the country-county pair $\{o, c\}$.

However, there might also be spatial correlation in confounding variables. For example, both Belgian and French immigrants and goods may go to Chicago for the same reason: many flight connections out of Paris, which is very accessible to potential Belgian immigrants by train. Leaving out Belgium-to-Chicago immigration flows when computing the instrument predicting these same immigration flows is therefore not sufficient, because now the French immigration flows to Chicago (used to predict Belgium-to-Chicago flows) are also contaminated with the confounding flight connections. To avoid such endogeneity, we again follow Burchardi et al. (2019) and leave out both the set of countries which share a continent with origin country o, $\mathcal{C}(o)$, and the Census region of county c, $r(c)$, to construct the instrumental variable that we defined in equation (13).

A violation of the identification assumption may occur if, say, immigrants and goods from France tend to flow to Chicago and immigrants and goods from South Korea flow to Miami in the same decade and for the same reason: a large number of flight connections. This violation is only quantitatively meaningful if the French are a large fraction of immigrants settling in Chicago, and if South Korean immigrants are a large fraction of the immigrants settling in Miami.

We use equation (13) to predict immigrant inflows into the U.S. decades spanning 1880 to 2000. Burchardi et al. (2019) extensively explore the validity of this instrumental variable and conduct numerous robustness checks for the instrument in the U.S. setting and find that it holds up quite well. Following Burchardi et al. (2019), we include five principal component terms which capture the variation of interactions of the instruments within county-country pairs and across decades.[48]

While the push-pull instrument may bear a passing resemblance to a standard shift-share instrument, we note two key differences. First, shift-share instruments are typically summed over a dimension (e.g., across origins), whereas the push-pull is not summed and thus retains two dimensions of variation. Second, the 'share' component of the push-pull is not lagged, unlike in the canonical shift-share style instrument, such as the ethnic enclave instrument

[48] We compute 1,013 higher-order interaction terms, defined as $L^{D'}_{o,-r(c)} \times \cdots \times L^{D}_{o,-r(c)} L^{D}_{-\mathcal{C}(o),c} / L^{D}_{-\mathcal{C}(o)}$ for each $D' < D \leq 2000$. We then compute five principal components which capture the variation contained within those 1,013 terms.

proposed by Card (2001).

We show the first-stage results of the leave-out push-pull instruments using our Homescanner data at the household level in Table A.1. We find that the push-pull instrument strongly and positively predicts the contemporary bilateral immigrant population.

We estimate the first-stage four ways. In columns 1 and 2, the specification is at the household-by-origin level, consistent with equation (11). Here we cluster standard errors at the level of the instrumental variables—the origin-by-county level—so the estimates are equivalent to a specification at the origin-by-county level but each county weighted based on the location of NielsenIQ households within the U.S.[49] In column 1, we predict immigrant population shares without using information on household nativity. In column 2, we include household nativity variables. In both cases, the first-stage F-statistic is about 20 and surpasses conventional thresholds. Coefficients are always positive and typically statistically significant, with the exception of the early 20th century.

Columns 3 and 4 show estimates from data at the barcode-by-county level, as in equation (12). We again cluster standard errors at the origin-by-county level. The F-statistics are near 20, and most coefficients are positive and statistically significant, with the exception of the earlier decades. Note that we are predicting immigrant populations (and not ancestry populations, as in Burchardi et al. 2019), and cohorts of immigrant groups likely change their location choices over time.

[49]F-statistics differ from those appearing in the main text because we bootstrap them in the main text.

Table A.1: First stage regression

	Dependent variable: Immigrants/Pop. 2010			
	(1)	(2)	(3)	(4)
$L_{o,-r(d)}^{1880} \times \frac{L_{-c(o),d}^{1880}}{L_{-c(o)}^{1880}}$	0.000063***	0.000057***	-0.00015	-0.00015
	(0.000021)	(0.000020)	(0.00015)	(0.00016)
$L_{o,-r(d)}^{1900} \times \frac{L_{-c(o),d}^{1900}}{L_{-c(o)}^{1900}}$	0.000033	0.000017	-0.00058	-0.00072
	(0.00013)	(0.00013)	(0.00072)	(0.00087)
$L_{o,-r(d)}^{1910} \times \frac{L_{-c(o),d}^{1910}}{L_{-c(o)}^{1910}}$	0.00026	0.00024	-0.00046	-0.00078
	(0.00020)	(0.00020)	(0.00048)	(0.00063)
$L_{o,-r(d)}^{1920} \times \frac{L_{-c(o),d}^{1920}}{L_{-c(o)}^{1920}}$	0.0018***	0.0018***	0.00056	0.00036
	(0.00025)	(0.00025)	(0.00070)	(0.00088)
$L_{o,-r(d)}^{1930} \times \frac{L_{-c(o),d}^{1930}}{L_{-c(o)}^{1930}}$	0.0016***	0.0016***	0.0029***	0.0031***
	(0.00017)	(0.00017)	(0.00058)	(0.00069)
$L_{o,-r(d)}^{1970} \times \frac{L_{-c(o),d}^{1970}}{L_{-c(o)}^{1970}}$	0.00086***	0.00084***	0.00084***	0.00092***
	(0.000081)	(0.000080)	(0.00023)	(0.00030)
$L_{o,-r(d)}^{1980} \times \frac{L_{-c(o),d}^{1980}}{L_{-c(o)}^{1980}}$	0.0032***	0.0032***	0.0042***	0.0047***
	(0.00028)	(0.00028)	(0.00058)	(0.00071)
$L_{o,-r(d)}^{1990} \times \frac{L_{-c(o),d}^{1990}}{L_{-c(o)}^{1990}}$	0.0023***	0.0022***	0.00093	0.0012
	(0.00025)	(0.00025)	(0.00075)	(0.00090)
$L_{o,-r(d)}^{2000} \times \frac{L_{-c(o),d}^{2000}}{L_{-c(o)}^{2000}}$	0.0015***	0.0015***	0.0015***	0.0016***
	(0.00019)	(0.00019)	(0.00029)	(0.00034)
=1 if immigrant from anywhere		0.000022		
		(0.000072)		
=1 if immigrant from origin o		0.013***		
		(0.0032)		
N	1,461,130	1,461,130	2,261,777	1,601,674
Country FE	✓	✓		
Barcode FE			✓	✓
County FE			✓	✓
Distance & latitude difference	✓	✓	✓	✓
Household controls		✓		
Principal components	✓	✓	✓	✓
F-statistic	20.2	19.5	17.3	17.5
Sample	All counties	All counties	All counties	UPC in >100 counties

Notes: Columns 1 and 2 show regression results at the household-origin level with observations weighted using NielsenIQ household weights and standard errors clustered two-ways at the household and origin-by-county levels. Columns 3 and 4 show regression results at the barcode-county level with standard errors clustered two-ways at the barcode and origin-by-county levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

A.3 Robustness of Gravity Results

In this appendix we test the robustness of our main estimates in several ways.

A.3.1 Robustness of Preference Estimation

Even when using the pseudo-Poisson maximum likelihood (PPML) estimator in order to account for the zeros in $\tilde{X}_{oh}$ while maintaining a theory-consistent estimating equation, one may be concerned that the relative weight PPML attributes to zeros could potentially bias our estimates Breinlich et al. (2024); Tyazhelnikov and Zhou (2024). We therefore provide two robustness checks and replicate our estimates in Table 1 using OLS. Table A.2 applies OLS to estimate equation (4) with relative expenditure by h on goods from o measured in levels, and Table A.3 applies OLS to estimate equation (4) as a linear probability model, such that the dependent variable is a binary indicator for $\tilde{X}_{oh} > 0$.[50]

In both cases, the estimates of ζ_1 and ζ_2 mirror those found when estimating our empirical model with PPML. In levels, we find that immigrants exhibit import expenditure shares that are 0.03 percentage points higher than within-county natives for all origins and 0.91 percentage points higher for their specific origin country. Both estimates are significant at 99% confidence levels, and suggest an import expenditure of immigrants that is 25% greater than for native households.[51] As shown in Table A.3, the probability of purchasing any nonzero amount of groceries from a given import origin is 10% greater for immigrants than within-county natives, and almost twice as large for the immigrant's specific import origin.

A.3.2 Sample Representativeness

As discussed by Feenstra et al. (2023), the NielsenIQ household sample may not be perfectly representative of the U.S. population in terms of income and price sensitivity. We further find that the NielsenIQ data is not representative of the distribution of immigrant origin

[50] Table A.2 provides estimates of ζ_1 and ζ_2 associated with applying OLS to the following estimating equation:

$$\tilde{X}_{oh} = \Psi_{o,c(h)} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o] + \epsilon_{oh}.$$

[51] The average expenditure share across all origins is 0.12 percent, with an average probability of purchasing nonzero groceries from any origin of 0.16.

Table A.2: Estimator Robustness: ζ_1 and ζ_2 Estimated in Levels

	Dep. var.: Rel. expenditure share on goods from o					
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant from anywhere	.00045***	.00035***	.0004***	.0003***	.00043***	.00032***
	(.00005)	(.000042)	(.00005)	(.000043)	(.000056)	(.000047)
Immigrant from o		.0091***		.0091***		.0098***
		(.0021)		(.0021)		(.0025)
N	1494324	1494324	1494324	1494324	1308918	1308918
Origin-County FE	✓	✓	✓	✓		
HH Controls			✓	✓	✓	✓
Origin-County-Income FE					✓	✓
Sample Weights	✓	✓	✓	✓	✓	✓
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.3: Estimator Robustness: ζ_1 and ζ_2 as Linear Probability Estimates

	Dep. var.:$X_{oh} \neq 0$					
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant from anywhere	0.023***	0.021***	0.016***	0.014**	0.014***	0.012**
	(0.0055)	(0.0055)	(0.0058)	(0.0058)	(0.0051)	(0.0051)
Immigrant from o		0.18***		0.18***		0.18***
		(0.021)		(0.021)		(0.023)
N	1496274	1496274	1496274	1496274	1310634	1310634
Origin-County FE	✓	✓	✓	✓		
HH Controls			✓	✓	✓	✓
Origin-County-Income FE					✓	✓
Sample Weights	✓	✓	✓	✓	✓	✓
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

countries. Mexican-born immigrants, for example, make up 15% of immigrant households in NielsenIQ but 30% in the ACS.

NielsenIQ's shortcomings in representativeness across immigrant origins may be driven by a combination of two factors. First, cross-sectionally NielsenIQ HomeScanner may miss some households if, for example, the survey module is not available in the language an

Table A.4: Estimates of Import Demand Preferences with Adjusted Weights

	Dep. var.: Rel. expenditure share on goods from o					
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant from anywhere	0.32***	0.25***	0.30***	0.23***	0.30***	0.23***
	(0.032)	(0.031)	(0.035)	(0.035)	(0.037)	(0.036)
Immigrant from o		0.63***		0.67***		0.71***
		(0.088)		(0.087)		(0.100)
N	868,261	868,261	868,261	868,261	597,276	597,276
Origin-County FE	✓	✓	✓	✓		
HH Controls			✓	✓	✓	✓
Origin-County-Income FE					✓	✓
Sample Weights	✓	✓	✓	✓	✓	✓
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. Household controls are dummy variables indicating: race, ethnicity, household size bins, bins for the number and age of children in the household, highest education among household members, household income bins, marital status and age groups. The estimation sample size is less than our total sample, and falls across columns, due to fixed effects causing separation in the sense of Correia et al. (2019). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

immigrant household speaks. Secondly, there may be differential attrition across immigrant origins between 2008, when the "Tell Me More About You" was distributed, and the sample period of 2014-2016.[52]

To gauge the importance of NielsenIQ's uneven representation across immigrant origins in driving our results, we adjust the survey weights so that the weighted aggregate population shares of natives and immigrants of each origin reflect those measured in the pooled 2013-2017 ACS sample. Using these adjusted weights, we replicate our baseline estimates of import demand preferences from Table 1 in Table A.4 and our baseline household gravity equations from Table 3 in Table A.5. Both the estimates of immigrants' import demand preferences and the estimates of their spillover effects are very similar to the baseline estimates. If anything, the spillover effects are slightly larger when using the adjusted weights.

As an additional check on potentially selective attrition between 2008 and our sample period of 2014-2016, we estimate equation (11) but using the full 2014-2016 sample (which includes every household whether or not we observe its nativity) and therefore dropping the

[52]The NielsenIQ survey is voluntary so households may drop out at any time.

Table A.5: Household Gravity Equation with Adjusted Weights

	$\tilde{X}_{oh}/\mathcal{Z}_{oh}$		$\tilde{N}_{oh}/\mathcal{Z}_{oh}$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.53***	1.43***	1.32***	1.37***
	(0.24)	(0.54)	(0.20)	(0.41)
First-stage residuals		0.13		-0.056
		(0.60)		(0.45)
N	1,461,130	1,461,130	1,461,130	1,461,130
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		28.5		28.5

Notes: The table presents regression results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share in column 2. Observations are weighted using NielsenIQ household weights. Standard errors and first-stage F-statistics are computed over 1,000 bootstrapped samples of households, in which the z_{oh}, $z_{o,c(h)}$, first-stage, and control function are estimated on the same sample. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

nativity variables from the estimation. The dependent variable is thus $\tilde{X}_{oh}$ without deflating by $\mathcal{Z}_{oh}$), since we no longer observe the full vector of household characteristics. We then compare the results with the same specification using our smaller main sample in order to investigate how much attrition matters for the estimates. Table C.4 shows that the IV estimate increases modestly from 1.71 with the nativity sample (column 2) to 1.81 with the full sample (column 4). However, these two estimates are statistically indistinguishable from each other. We conclude that differential attrition is unlikely to drive our estimate of the immigrant spillover effect.

A.3.3 Ethnicity and alternative measures of origin country connectedness

In our baseline specification for estimating import demand heterogeneity, equation (4), we allow households to have specific preferences for (i) all imports and (ii) imports specifically from their origin country. Households may additionally exhibit specific preferences towards goods from countries close—geographically or culturally—to their origin country or, in the case of second-generation immigrants, the origin country of their parents.

We test the importance of such specific household preferences in Table A.6. We start

Table A.6: Estimates of Import Demand Preferences with Additional Variables

	Dep. var.: Rel. expenditure share on goods from o		
	(1)	(2)	(3)
Immigrant from anywhere	0.22***	0.21***	0.21***
	(0.029)	(0.032)	(0.035)
Immigrant from o	0.59***	0.56***	0.59***
	(0.069)	(0.084)	(0.083)
Hispanic origin	-0.043	-0.041	-0.059
	(0.055)	(0.055)	(0.052)
=1 if Hispanic and o in Latin America	0.16**	0.15**	0.17***
	(0.066)	(0.067)	(0.064)
Asian origin	0.0065	0.011	0.013
	(0.037)	(0.038)	(0.038)
=1 if Asian and o in Asia	0.38***	0.36***	0.38***
	(0.092)	(0.091)	(0.090)
=1 if immigrant from continent of o		0.038	0.0084
		(0.065)	(0.060)
=1 if common official or primary language			0.0045
			(0.056)
=1 if ever in colonial/dependency rel.			0.15
			(0.10)
=1 if currently in colonial/dependency rel.			0.34
			(0.83)
N	868,261	868,261	867,494
HH Controls	✓	✓	✓
Origin-County FE	✓	✓	✓
Sample Weights	✓	✓	✓
Estimator	PPML	PPML	PPML

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. Household controls are dummy variables indicating: race, ethnicity, household size bins, bins for the number and age of children in the household, highest education among household members, household income bins, marital status and age groups. The estimation sample size is less than our total sample, and falls across columns, due to fixed effects causing separation in the sense of Correia et al. (2019). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

by displaying the estimate of dummies indicating Hispanic or Asian origin (which are both also included in our baseline specification), our best available proxies for being a second-generation immigrant (although they might also capture some third-and-higher-generation immigrants). We further interact these variables with dummies indicating imports originating in Latin America or Asia, respectively. We find that households of Hispanic origin do not spend more than natives on imports in general, as captured by the insignificant indicator for Hispanic origin, but 16% more on imports from Latin America. Similarly, households of Asian origin spend 38% more on imports from Asian countries. In column 2, we add a dummy indicating whether immigrant households generally exhibit a specific preference for goods from their continent of origin (in addition to their country of origin), which we find to be insignificant.

In column 3, we add a set of dummy variables that capture cultural similarity between an immigrant's origin and the import origin: sharing the same language and past or current colonial relationships, all obtained from Conte et al. (2022). We find no statistically significant effect of any of these indicators on immigrants' import expenditure.

A.3.4 Comparison with the results of Burchardi et al. (2019)

Burchardi et al. (2019) estimated a null effect of immigrants on trade, in contrast to our own results.[53] In this appendix section, we consider three possible explanations for our diverging results. First, the primary explanatory variable of Burchardi et al. (2019) is the log of the number of individuals with a given ancestry measured in thousands and plus one instead of the foreign-born population share that we use. Second, Burchardi et al. (2019) use state-level data whereas we leverage household-level data. Third, Burchardi et al. (2019) use a two-step Heckman estimation strategy to account for selection into bilateral trading, while we apply the pseudo-Poisson maximum likelihood (PPML) estimation strategy. We find that the choice of estimation strategy explains the difference between our results and those of Burchardi et al. (2019) and explain why we prefer PPML over Heckman selection.

We start by testing whether the choice of explanatory variable (log ancestry in thousands plus one vs. immigrant population share) can explain our results. To do so, we first replace our previous explanatory variable, *Immigrants/Pop. 2010*, with *Ancestry/Pop. 2010*. Next, we take the functional form used in Burchardi et al. (2019), *Log Ancestry 2010*. Table A.7

[53] The focus of Burchardi et al. (2019), however, was on how immigrants shape FDI.

Table A.7: Household Gravity Estimates with Ancestry

	$\tilde{X}_{oh}/\mathcal{Z}_{oh}$			
	(1)	(2)	(3)	(4)
Ancestry/Pop. 2010	0.77***	0.65***		
	(0.10)	(0.19)		
Log Ancestry 2010			0.044***	0.039***
			(0.0039)	(0.0036)
First-stage residuals		0.14		0.0079***
		(0.22)		(0.0029)
N	1,421,640	1,421,640	1,421,640	1,421,640
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		25.5		18.7

Notes: The table presents regression results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share in column 2. Observations are weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

shows that we still obtain positive and significant coefficients with these alternative measures using our household-level data and estimation strategy.

Next, we test whether the level of data aggregation or the estimation can resolve our diverging results. In Table A.8 we mimic the specification in Burchardi et al. (2019) more closely by aggregating our data to the state-origin level. We run regressions first using our PPML approach (columns 1 and 2). As explanatory variables we employ both *Log Ancestry 2010* and our preferred measure *Immigrants/Pop. 2010*. As in our baseline household-level results, we continue to find a significantly positive effect of immigrants and ancestors on import volumes. Turning to columns 3 and 4, we apply the Heckman correction strategy of Burchardi et al. (2019). Here we obtain negative and insignificant coefficients. As a result, we conclude that the choice of estimation approach is important for the contrasting results between our study and Burchardi et al. (2019).

We argue that Poisson pseudo-maximum likelihood (PPML) estimation is more appropriate in our setting. In a widely cited article, Santos Silva and Tenreyro (2006) demonstrate that PPML performs quite well across a variety of settings, accommodating heteroskedas-

Table A.8: State-level Gravity estimates

	Dependent variable: Exp. share on goods from o relative to US			
	PPML + control fct		Heckman correction	
	(1)	(2)	(3)	(4)
Log Ancestry 2010	0.06***		-0.09	
	(0.008)		(0.057)	
Immigrants/Pop. 2010		2.46***		-4.31
		(0.377)		(3.266)
N	3,626	3,626	2,922	2,922
State FE	✓	✓	✓	✓

Notes: The table presents regression results at the state-origin level. Standard errors are computed over 1,000 bootstrapped samples of households. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

ticity and measurement error; Santos Silva and Tenreyro (2011) provide further simulation results in support of PPML even when the proportion of zeros is very high, as in our data. Santos Silva and Winkelmann (2024) shows that PPML performs well even when the conditional expectation function is misspecified. Fally (2015) shows that PPML is the only estimation strategy which satisfies the adding-up constraints of structural gravity.

Burchardi et al. (2019) follow Helpman et al. (2008) in applying the two-step Heckman estimation approach. As pointed out by Santos Silva and Tenreyro (2015), the Heckman approach makes two strong assumptions: on the distribution of the error terms and on the homoskedasticity of those errors. PPML estimation, in contrast, necessitates no assumptions about the distribution of errors and allows for heteroskedasticity. Furthermore, using the same vector of variables in both first and second stage of the Heckman approach leads to identification by functional form (Puhani 2000; Lewbel 2019).

B Theory Appendix

B.1 Deriving Heterogeneous Firms Model Equations

Deriving equation (8). Taking the ratio of the household's first-order condition for two varieties ω_1 from country o and ω_2 from country o', we obtain

$$\left(\frac{q_{o'h}(\omega_2)}{q_{oh}(\omega_1)}\right)^{-1/\sigma}\left(\frac{z_{o'h}}{z_{oh}}\right)^{1/\sigma}=\frac{p_{o',c(h)}(\omega_2)}{p_{o,c(h)}(\omega_1)}$$

Define

$$P_h \equiv \left(\sum_{o\in\mathcal{O}} z_{oh}\int_{\omega\in\Omega_{o,c(h)}} p_{o,c(h)}(\omega)^{1-\sigma}d\omega\right)^{\frac{1}{1-\sigma}} \tag{B.1}$$

as the price index faced by household h for differentiated goods. Assuming the household budget is equal to X_h, we then obtain

$$(1-\mu_0)X_h = z_{oh}^{-1}q_{oh}(\omega)p_{o,c(h)}(\omega)^{\sigma}P_h^{1-\sigma} \tag{B.2}$$

We rearrange to get quantity and expenditure for a variety associated with productivity φ as

$$q_{oh}(\varphi) = (1-\mu_0)X_h z_{oh}p_{o,c(h)}(\varphi)^{-\sigma}P_h^{\sigma-1} \tag{B.3}$$

$$x_{oh}(\varphi) = (1-\mu_0)X_h z_{oh}(p_{o,c(h)}(\varphi)/P_h)^{1-\sigma} \tag{B.4}$$

From the firm's profit maximization problem, the price equation is

$$p_{o,c(h)}(\varphi) = \frac{\sigma}{\sigma-1}\frac{w_o}{\varphi}\tau_{oc(h)} \tag{B.5}$$

Substituting equation (B.5) into equation (B.4), summing across all households in $c(h)$, and defining $\lambda_1 \equiv (1-\mu_0)\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}$, we obtain the following expression for county expenditure on imports from firm with productivity φ in o:

$$x_{oc}(\varphi) = \lambda_1(w_o\tau_{oc})^{1-\sigma}\varphi^{\sigma-1}\Big(\sum_{h'\in\Lambda_c} z_{oh'}X_{h'}P_{h'}^{\sigma-1}\Big) \tag{B.6}$$

Next, we derive variable profits earned by a firm with productivity φ selling to market c from origin o:

$$\begin{aligned}
\pi_{o,c}(\varphi) &\equiv \left(p_{o,c}(\varphi) - \frac{w_o}{\varphi}\tau_{o,c}\right)\sum_{h'\in c} q_{oh}(\varphi) \\
&= (1-\mu)\left(\frac{w_o}{\varphi}\tau_{o,c}\right)^{1-\sigma}\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\sum_{h'\in\Lambda_c} z_{oh'}X_{h'}P_{h'}^{\sigma-1} \\
&= \frac{1}{\sigma}x_{oc}(\varphi)
\end{aligned}$$

A firm with productivity φ only exports from o to c if it is profitable, i.e., if variable profits are at least as much as the fixed cost of exporting:

$$\pi_{oc}(\varphi) \geq f_{oc} \tag{B.7}$$

For a firm at the cutoff productivity, (B.7) holds with equality, resulting in the following equation for φ^*_{oc}, where $\lambda_2 \equiv \frac{\sigma}{\sigma-1}\left(\frac{\sigma}{1-\mu_0}\right)^{\frac{1}{\sigma-1}}$:

$$\varphi^*_{oc} = \lambda_2 w_o \tau_{oc}\left(\frac{f_{oc}}{\sum\limits_{h'\in\Lambda_c} z_{oh'}X_{h'}P_{h'}^{\sigma-1}}\right)^{\frac{1}{\sigma-1}} \tag{B.8}$$

Returning to equation (B.1) and replacing varieties ω with productivity φ (since firms with identical productivity charge identical prices), we get:

$$P_h = \left(\sum_{o\in\mathcal{O}} z_{oh}\int_0^{+\infty} p_{o,c(h)}(\varphi)^{1-\sigma}M_{o,c(h)}g_{o,c(h)}(\varphi)d\varphi\right)^{\frac{1}{1-\sigma}}$$

where $M_{o,c(h)}$ is the measure of firms exporting from o to $c(h)$ and $g_{o,c(h)}(\varphi)$ is the (equilibrium) density of firms from o with productivity φ that export to $c(h)$.

Plugging in the equilibrium price, equation (B.5), we have

$$P_h = \frac{\sigma}{\sigma-1}\left(\sum_{o\in\mathcal{O}} z_{oh}\left(w_o\tau_{o,c(h)}\right)^{1-\sigma}M_{o,c(h)}\int_0^{+\infty}\varphi^{\sigma-1}g_{o,c(h)}(\varphi)d\varphi\right)^{\frac{1}{1-\sigma}} \tag{B.9}$$

Turning to the gravity equation, we integrate over equation (B.4) to obtain

$$X_{oh} = \int_{\omega \in \Omega_{o,c(h)}} x_{oh}(\omega) d\omega = (1-\mu_0) z_{oh} X_h P_h^{\sigma-1} \int_{\omega \in \Omega_{o,c(h)}} p_{o,c(h)}(\omega)^{1-\sigma} d\omega$$

Given the equilibrium price expression in (B.5), we can substitute the last term with

$$\begin{aligned}\int_{\omega \in \Omega_{o,c(h)}} p_{o,c(h)}(\omega)^{1-\sigma} d\omega &= \left(\frac{\sigma}{\sigma-1} w_o \tau_{o,c(h)}\right)^{1-\sigma} M_{o,c(h)} \int_0^\infty \varphi^{\sigma-1} g_{o,c(h)}(\varphi) d\varphi \\ &= \left(\frac{\sigma}{\sigma-1} w_o \tau_{o,c(h)}\right)^{1-\sigma} M_o \int_{\varphi^*_{o,c(h)}}^\infty \varphi^{\sigma-1} g_{o,c(h)}(\varphi) d\varphi\end{aligned}$$

Finally, we use the assumption that φ is Pareto distributed with shape parameter θ so that $g_o(\varphi) = \theta/\varphi^{\theta+1}$ to obtain

$$X_{oh} = \lambda_1 z_{oh} X_h P_h^{\sigma-1} (w_o \tau_{o,c(h)})^{1-\sigma} M_o \frac{\theta}{\theta+1-\sigma} (\varphi^*_{o,c})^{\sigma-\theta-1} \tag{B.10}$$

To obtain equation (8) from (B.10), we then:

- substitute (B.8) for $\varphi^*_{o,c}$
- assume $M_o = \gamma Y_o$, where Y_o is the value of production in country o as in Section 4.1.
- define $S_c = \sum\limits_{h' \in \Lambda_c} X_{h'} P_{h'}^{\sigma-1}$ and $z_{oc} = \sum\limits_{h' \in \Lambda_c} z_{oh'} \frac{X_{h'} P_{h'}^{\sigma-1}}{S_c}$.
- define $\lambda \equiv \gamma(1-\mu_0)^{\frac{\theta}{\sigma-1}} \sigma^{\frac{\sigma-\theta-1}{\sigma-1}} \left(\frac{\sigma}{\sigma-1}\right)^{-\theta} \frac{\theta}{\theta+1-\sigma}$

Finally, county-level expenditure on goods from origin o is simply the summation over all household-level expenditure, and is given by the following:

$$X_{oc} = \lambda Y_o S_c (w_o \tau_{oc})^{-\theta} \left(\frac{f_{oc}}{S_c z_{oc}}\right)^{-(\frac{\theta}{\sigma-1}-1)} z_{oc} \equiv \alpha_o S_c^{\frac{\theta}{\sigma-1}} \phi^b_{oc} \phi^z_{oc}. \tag{B.11}$$

B.2 Identification of Fixed Cost and Preference Diffusion Channels

In this section we fully differentiate equation (9) in order to arrive at two expressions relating the total import expenditure-immigrant elasticity and the extensive margin-immigrant elasticity to two parameters: β^f and β^z.[54,55]

We begin by fully differentiating $\tilde{X}_{oh}$ from equation (9) into terms associated with fixed costs f_{oc}, county-level preferences z_{oc}, and household-level preferences z_{oh}:

$$\begin{aligned} d\tilde{X}_{oh} = & \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial f_{oc}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial {\varphi^*}_{oc}}{\partial f_{oc}} \right] df_{oc} \\ & + \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oh}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial \varphi^*_{oc}}{\partial z_{oh}} \right] dz_{oh} \\ & + \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oc}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial \varphi^*_{oc}}{\partial z_{oc}} \right] dz_{oc} \end{aligned} \tag{B.12}$$

where we applied the Leibniz Rule to separate each term into both an intensive margin and extensive margin. Within each pair of brackets, the first term captures the intensive margin effect and the second term captures the extensive margin effect.

The expression for the intensive margin—the relative expenditure by household h on a given variety from origin o relative to its total expenditure on U.S. goods—is given by:

$$\tilde{x}_{oh}(\varphi) = (\tilde{w}_o \tau_{oc})^{1-\sigma} z_{oh} \varphi^{\sigma-1} \left(\int_{\varphi^*_{us,c}}^{+\infty} (\varphi^{'})^{\sigma-1} dG(\varphi^{'}) \right)^{-1} \tag{B.13}$$

whereas the productivity cut-off is equation (B.8).

It is clear from inspecting equation (B.13) and equation (B.8) that f_{oc} and z_{oc} only affect φ^*_{oc}, and therefore each household's extensive margin, whereas household-level preferences z_{oh} only affect the household-specific intensive margin of demand via $\tilde{x}_{oh}$. We can therefore apply the following restrictions: $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial f_{oc}} = 0$; $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oc}} = 0$; and $\frac{\partial \varphi^*_{oc}}{\partial z_{oh}} = 0$.

Hence, we have an expression for the total semi-elasticity of import expenditure with respect to immigrants and an expression for the extensive margin semi-elasticity of import

[54] We assume throughout that $\beta^\tau = 0$, which implies that immigrants do not affect variables trade costs. This assumption derives from the results discussed in Table 4.

[55] We refer to the extensive margin at the level of a variety.

expenditure with respect to immigrants:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} = \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} \tag{B.14}$$

$$\frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} \tag{B.15}$$

Recall that when estimating β, we normalize $\tilde{X}_{oh}$ and $\tilde{N}_{oh}$ by $\mathcal{Z} = \bar{z}_{oh} \bar{z}_{oc}^{\frac{\theta}{\sigma-1}-1}$. That is, we normalize expenditure by the expenditure for that household which is predicted by exogenous preference terms at the household and county level. Recall further that $z_{oh} = e^{\beta^z I_{oc}} \bar{z}_{oh}$ and $z_{oc} = e^{\beta^z I_{oc}} \bar{z}_{oc}$. We can therefore explicitly derive our estimate of β and the extensive margin counterpart β^N as the following:

$$\begin{aligned} \beta &= \frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial I_{oc}} \\ &= \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} \\ &\quad - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}} \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}} \end{aligned} \tag{B.16}$$

$$\begin{aligned} \beta^N &= \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial I_{oc}} \\ &= \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}} \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}} \end{aligned} \tag{B.17}$$

We evaluate the expressions (B.16) and (B.17) using the definition of $\tilde{X}_{oh}$ provided in equation (9) and the definition of $\mathcal{Z}_{oh}$ provided in Section 4.2.1. Specifically, we reduce expressions (B.16) and (B.17) in three steps:

1. Fixed costs and the extensive margin:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} = \beta^f$$

2. County-level preferences and the extensive margin:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}} \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}} = \left(\frac{\theta - (\sigma - 1)}{\sigma - 1} \right) \beta^z$$

3. Household-level preferences and the intensive margin:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}} \frac{\partial \ln \bar{z}_{oh}}{\partial I_{oc}} = \beta^z$$

We then derive an expression for the total import expenditure semi-elasticity with respect to the immigrant population share and the extensive margin semi-elasticity of import expenditure with respect to the immigrant population share:

$$\beta = \frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1}\right)\beta^z \tag{B.18}$$

$$\beta^N = \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1} - 1\right)\beta^z \tag{B.19}$$

B.3 Deriving Counterfactual Objects

Following Dekle et al. (2007), we denote the proportional change in a variable x as $\hat{x} = x'/x$, where an apostrophe $'$ denotes the counterfactual value.

Immigration shocks. We start with the counterfactuals from Section 5 relating to removing immigrant effects and removing immigrant expenditures. From equation (8), we obtain the proportional change in household-origin import expenditures:

$$\hat{X}_{oh} = \hat{P}_h^{\sigma-1} \hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} \left(\hat{\bar{z}}_{o,c(h)} \hat{S}_{c(h)}\right)^{\frac{\theta}{\sigma-1}-1} \hat{z}_{oh} \tag{B.20}$$

where changes in household imports by origin depend on the change in the household's price level $\hat{P}_h$, changes in fixed costs with the origin $\hat{f}_{o,c(h)}$, changes in average household-level preferences for the origin's products $\hat{\bar{z}}_{o,c(h)}$, changes in total local expenditures $\hat{S}_{c(h)}$, and changes in the household's preferences for the origin's products $\hat{z}_{oh}$. When o is the United States, equation (B.20) reduces to

$$\hat{X}_{us,h} = \hat{P}_h^{\sigma-1} \hat{S}_{c(h)}^{\frac{\theta}{\sigma-1}-1} \tag{B.21}$$

Hence we use equations (B.20) and (B.21) as well as $\hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} = e^{-\hat{\beta}^f I_{o,c(h)}}$ and $\hat{z}_{oh} =$

$e^{-\hat{\beta}^z I_{o,c(h)}}$ to obtain our counterfactual ratio as a function of observable or calibrated values:

$$\frac{X'_{oh}}{X'_{us,h}} = \frac{X_{oh}}{X_{us,h}} \left(e^{-I_{o,c(h)}(\hat{\beta}^f + \hat{\beta}^z)}\right) z_{o,c(h)}^{\left(\frac{\theta}{\sigma-1}-1\right)} \tag{B.22}$$

Summing across non-U.S. origins o and holding fixed total expenditures X_h, we compute the counterfactual imports from each origin o for each household h.

Lastly, it is simple to show that under CES preferences, the change in welfare is given by the change in the price index:

$$\hat{U}_h = \hat{P}_h^{\mu_0 - 1} \tag{B.23}$$

Notice, however, that P_h includes changes in h's preferences associated with preference diffusion β^z, which significantly complicates conventional welfare analysis. In our main counterfactual we simply fix β^z and therefore z_{oh} to its observed level and do not allow it to change. The change in the welfare-relevant price index is then:

$$\hat{P}_h^{\sigma-1} = \frac{1}{\frac{X_{us,h}}{X_h} \hat{S}_{c(h)}^{\frac{\theta}{\sigma-1}-1} + \sum_{o \neq us} \frac{X_{o,h}}{X_h} \hat{f}_{oc(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} \left(\hat{z}_{oc(h)} \hat{S}_{c(h)}\right)^{\frac{\theta}{\sigma-1}-1}}$$

We further assume that immigrant and native households spend the same amount on grocery and personal care produces, which implies that

$$\hat{S}_{c(h)} = 1 - I_{c(h)}$$

where $I_{c(h)}$ is the share of the population who are immigrants in county $c(h)$.

Online Appendix References

Bonadio, Barthélémy, "Migrants, trade and market access," *Review of Economics and Statistics*, forthcoming.

Breinlich, Holger, Dennis Novy, and JMC Santos Silva, "Trade, gravity, and aggregation," *Review of Economics and Statistics*, 2024, *106* (5), 1418–1426.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan, "Migrants, ancestors, and foreign investments," *Review of Economic Studies*, 2019, *86* (4), 1448–1486.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Choi, Jaerim, Jay Hyun, and Ziho Park, "Bound by ancestors: Immigration, credit frictions, and global supply chain formation," *Journal of International Economics*, 2024, *147*, 103855.

Conte, M., P. Cotterlaz, and T. Mayer, "The CEPII Gravity database," CEPII Working Paper N°2022-05, CEPII 2022.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced trade," *American Economic Review*, 2007, *97* (2), 351–355.

Fally, Thibault, "Structural gravity and fixed effects," *Journal of International Economics*, 2015, *97* (1), 76–85.

Feenstra, Robert C, Luca Macedoni, and Mingzhi Xu, "Large firms, consumer heterogeneity and the rising share of profits," Technical Report, National Bureau of Economic Research 2023.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

Lewbel, Arthur, "The identification zoo: Meanings of identification in econometrics," *Journal of Economic Literature*, 2019, *57* (4), 835–903.

McCully, Brett, "Immigrants, legal status, and illegal trade," Technical Report 2024.

Puhani, Patrick, "The Heckman correction for sample selection and its critique," *Journal of Economic Surveys*, 2000, *14* (1), 53–68.

Santos Silva, Joao MC and Rainer Winkelmann, "Misspecified exponential regressions: Estimation, interpretation, and average marginal effects," *Review of Economics and Statistics*, 2024, pp. 1–25.

_ **and Silvana Tenreyro**, "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters*, 2011, *112* (2), 220–222.

_ **and** _ , "Trading partners and trading volumes: implementing the Helpman–Melitz–Rubinstein model empirically," *Oxford Bulletin of Economics and Statistics*, 2015, *77* (1), 93–105.

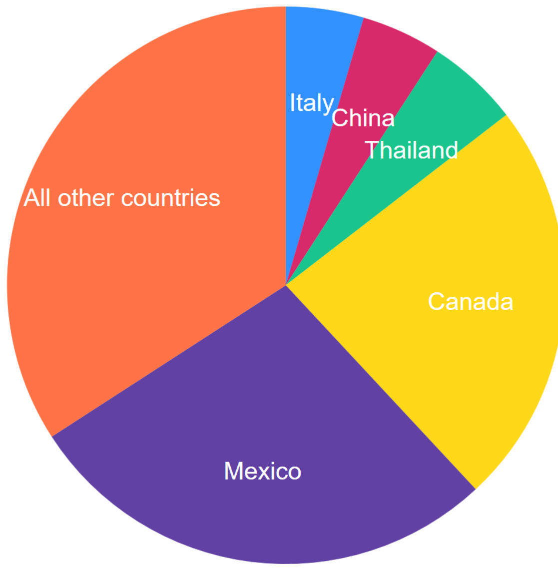
Santos Silva, Joo MC and Silvana Tenreyro, "The log of gravity," *Review of Economics and Statistics*, 2006, pp. 641–658.

Terry, Stephen J, Thomas Chaney, Konrad B Burchardi, Lisa Tarquinio, and Tarek A Hassan, "Immigration, Innovation, and Growth," 2022.

Tyazhelnikov, Vladimir and Xinbei Zhou, "PPML, Gravity, and Heterogeneous Trade Elasticities," Technical Report, Working Paper 2024.

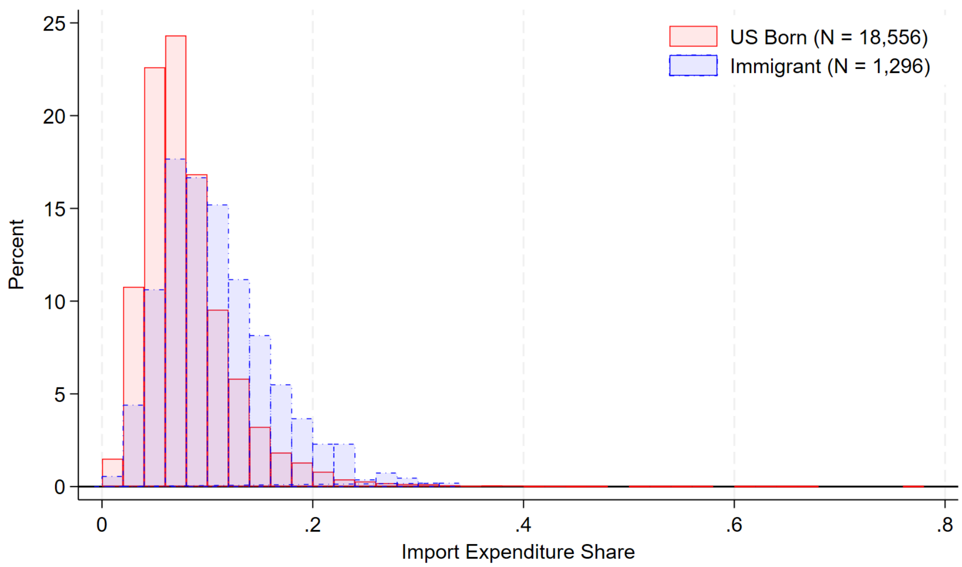
C Additional Tables and Charts

Figure C.1: Spending on Imports by Origin Country



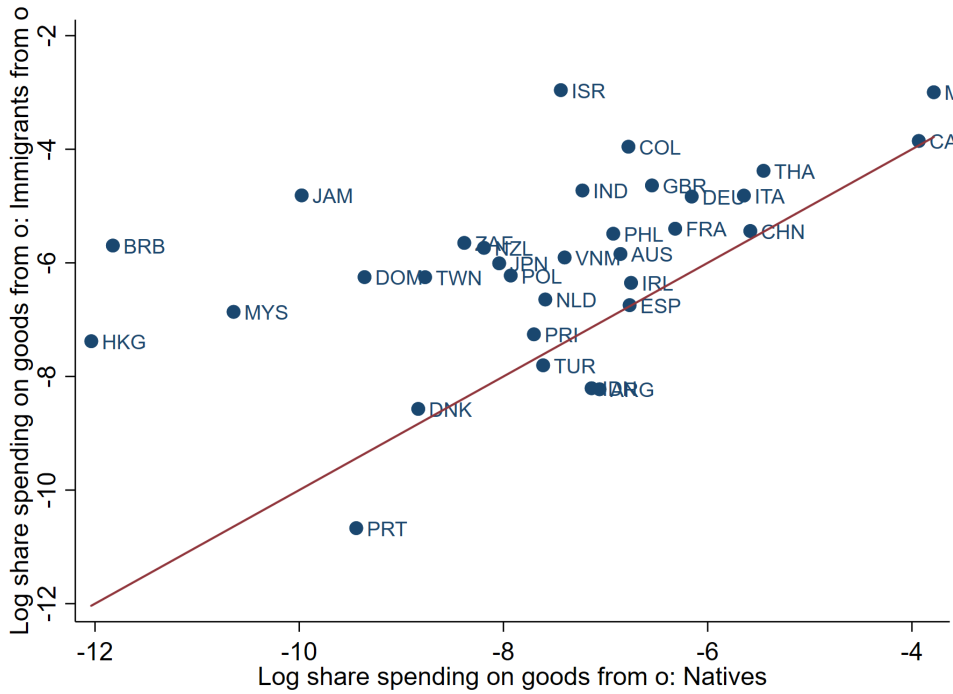
Notes: The figure shows the percent of expenditure on imports by country of origin. Data come from the NielsenIQ Household Panel 2014-2016.

Figure C.2: Distribution of Household-level Import Expenditure Share by Nativity



Notes: The figure shows the distribution of household's expenditure on imported goods, split by U.S. born (in red) and foreign-born (in blue) households. Household nativity assigned as discussed in Section 2.1. Data come from the NielsenIQ Household Panel 2014-2016. We exclude households who spent less than $1,000 over the 3 year sample period. Observations are weighted by the NielsenIQ projection factors.

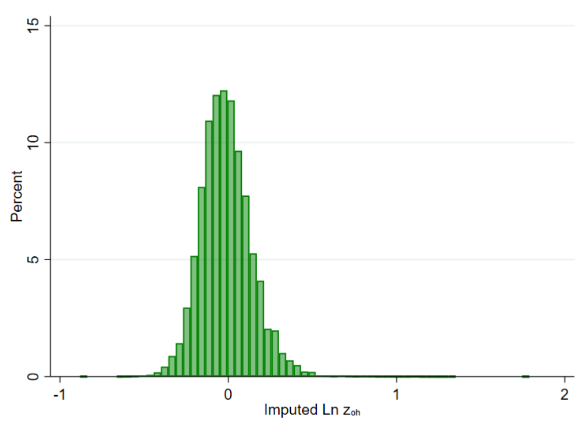
Figure C.3: Immigrants Tend to Spend more on Goods from their Origin Relative to Natives



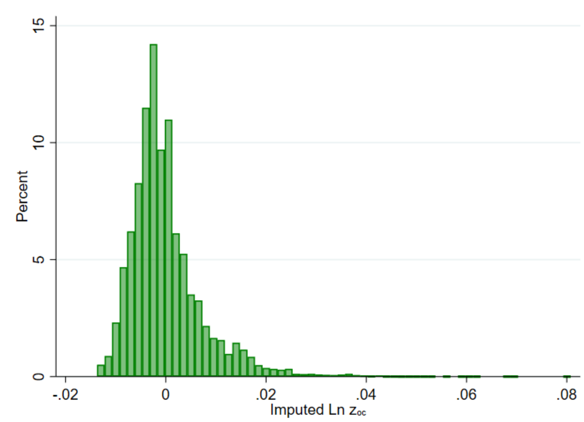
Notes: The figure shows the relationship between spending on goods imported from immigrants' own country (the y-axis) and spending by goods from that country by natives (x-axis). The red line is the 45-degree line, which plots when there is no preference by immigrants for goods imported from their origin country relative to natives. Household nativity is assigned as discussed in Section 2.1. Data come from the NielsenIQ Household Panel 2014-2016. NielsenIQ projection factor weights used to construct expenditure shares.

Figure C.4: Distribution of Imputed Preference Terms

(a) Distribution of $\ln \hat{\hat{z}}^{\sigma}_{oh}$

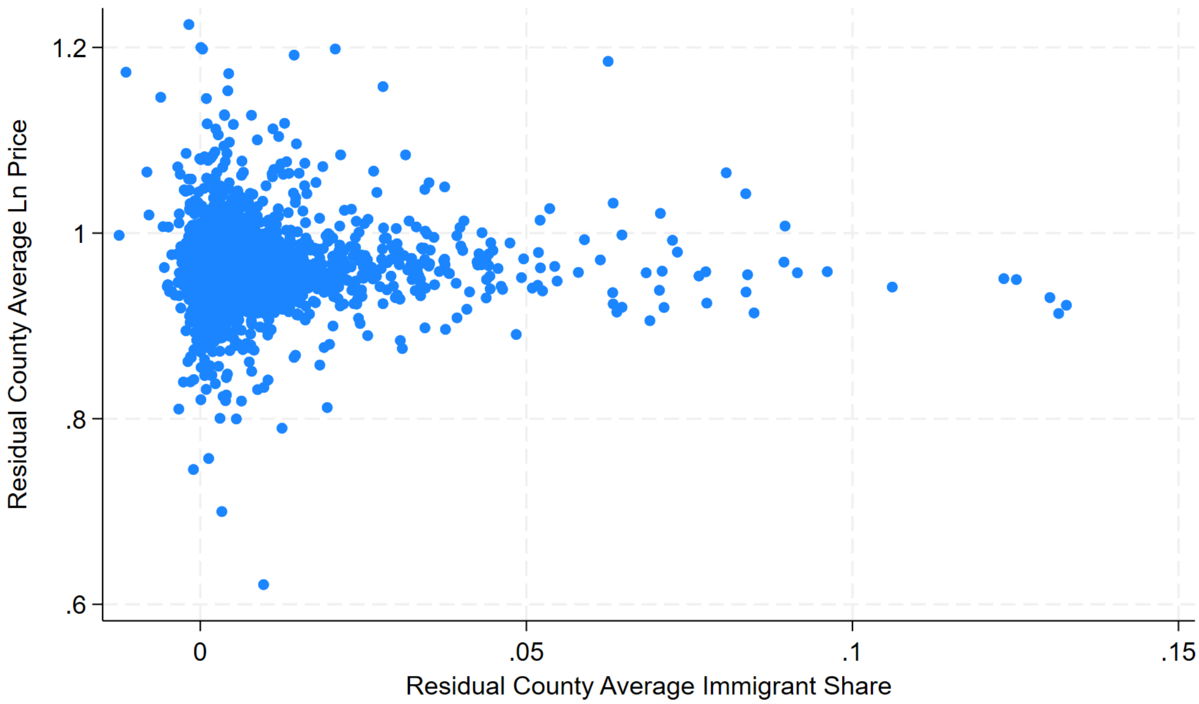


(b) Distribution of $\ln \hat{\hat{z}}_{oc}$



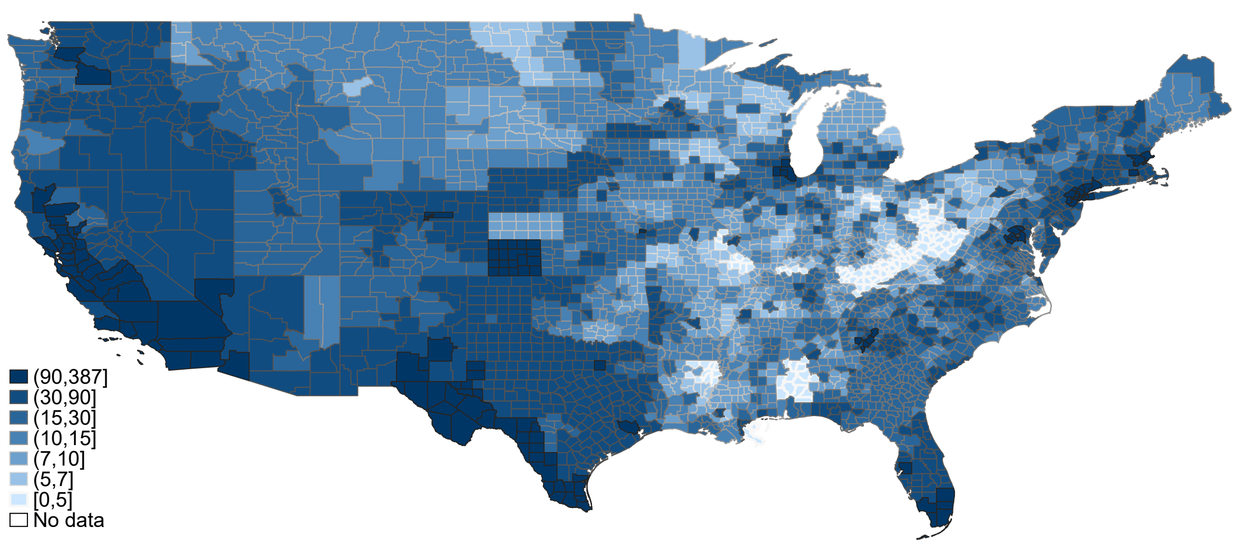
Notes: Figure (a) plots the distribution across NielsenIQ household-origin pairs of the log of $\hat{\hat{z}}_{oh} = \exp\left(\hat{\delta} J_h + \hat{\zeta}_1 \mathbf{1}[o(h) \neq US] + \hat{\zeta}_2 \mathbf{1}[o(h) = o]\right)$, where the terms $\hat{\delta}$, $\hat{\zeta}_1$, and $\hat{\zeta}_2$ are estimated from equation (4). Figure (b) plots the distribution across county-origin pairs of the log of $\hat{\hat{z}}_{oc} = \sum_{h' \in \Lambda_c} \hat{\hat{z}}_{oh'} \kappa_{h'}$, computed using data from the 2012-2017 American Community Survey.

Figure C.5: County Fixed Effects for Price and Immigrant Population Share



Notes: The figure plots average log price by county (y-axis) against average immigrant population share by county (x-axis). Each value is obtained by regressing the variable (log price or immigrant population share) on fixed effects for county and barcode, and taking the county fixed effect.

Figure C.6: Spatial Distribution of Fall in Welfare due to Removing Immigrants



Notes: This chart plots the dollar decrease in the dollar-equivalent grocery welfare the trade-creating effect of immigrants and immigrant expenditure are removed following the procedure outlined in Appendix Section B.3.

Table C.1: Relationship between Import Expenditure Shares and Immigrant Status

	Dependent variable: Import expenditure share					
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if immigrant	0.028***	0.031***	0.023***	0.027***	0.024***	0.028***
	(0.0018)	(0.0027)	(0.0017)	(0.0026)	(0.0017)	(0.0026)
N	19,700	19,700	19,107	19,107	19,107	19,107
County fixed effects			✓	✓	✓	✓
Household controls					✓	✓
Weighted		✓		✓		✓

Notes: The table presents regression results at the household level. Standard errors are clustered at the county level. Household controls are income bins, household size, marital status, and household head age and gender. Sample drops when including county fixed effects due to the 593 households living in a county with no other NielsenIQ panelists in our sample. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.2: Import Preference Heterogeneity

	Dep. var.: Rel. expenditure share on goods from o			
	(1)	(2)	(3)	(4)
Immigrant from anywhere	0.31***	0.26***	0.27***	0.22***
	(0.028)	(0.028)	(0.029)	(0.029)
Immigrant from o		0.62***		0.65***
		(0.069)		(0.070)
Income: 10k-30k			0.032	0.029
			(0.042)	(0.042)
Income: 30k-50k			0.016	0.011
			(0.041)	(0.040)
Income: 50k-70k			0.077*	0.073*
			(0.042)	(0.042)
Income: 70k-100k			0.064	0.060
			(0.041)	(0.042)
Income: >100k			0.18***	0.17***
			(0.043)	(0.043)
Some College			0.063***	0.063***
			(0.023)	(0.023)
College Degree			0.094***	0.096***
			(0.024)	(0.024)
Postgraduate Degree			0.18***	0.18***
			(0.027)	(0.027)
N	868,261	868,261	868,261	868,261
Origin-County FE	✓	✓	✓	✓
HH Controls			✓	✓
Sample Weights				
Estimator	✓	✓	✓	✓
estimator	PPML	PPML	PPML	PPML

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.3: Effect of Household Characteristics on Import expenditure

	Dep. var.: Rel. expenditure share on goods from o			
	(1)		(2)	
Immigrant from anywhere	0.22***	(0.029)	0.23***	(0.031)
Immigrant from o	0.65***	(0.070)	0.66***	(0.081)
Income: 10k-30k	0.029	(0.042)		
Income: 30k-50k	0.011	(0.040)		
Income: 50k-70k	0.073*	(0.042)		
Income: 70k-100k	0.060	(0.042)		
Income: >100k	0.17***	(0.043)		
HH size: 2	-0.073**	(0.029)	-0.078**	(0.035)
HH size: 3	-0.100***	(0.033)	-0.11***	(0.040)
HH size: 4	-0.19***	(0.041)	-0.20***	(0.049)
HH size: >4	-0.19**	(0.085)	-0.14	(0.097)
Children: 6-12 y.o.	-0.085	(0.088)	-0.098	(0.098)
Children: 13-17 y.o.	-0.098	(0.092)	-0.10	(0.10)
Children: <6 + 6-12	-0.11	(0.10)	-0.080	(0.12)
Children: <6 + 13-17	-0.054	(0.16)	-0.18	(0.19)
Children: 6-12 + 13-17	-0.054	(0.095)	-0.047	(0.11)
Children: All Age Groups	-0.26**	(0.12)	-0.26	(0.16)
No Children	-0.068	(0.084)	-0.054	(0.096)
Some College	0.063***	(0.023)	0.064**	(0.027)
College Degree	0.096***	(0.024)	0.087***	(0.028)
Postgraduate Degree	0.18***	(0.027)	0.18***	(0.030)
Widowed	0.0046	(0.036)	0.0098	(0.042)
Divorced/Separated	-0.0024	(0.034)	-0.021	(0.041)
Single	-0.022	(0.034)	-0.023	(0.040)
Black	0.057**	(0.024)	0.068**	(0.027)
Asian	0.080**	(0.036)	0.077**	(0.037)
Other	0.096**	(0.040)	0.12***	(0.045)
Hispanic	0.051	(0.035)	0.044	(0.039)
Age	-0.018	(0.032)	-0.048	(0.036)
$\text{Age}^2/100$	0.026	(0.054)	0.074	(0.060)
$\text{Age}^3/10000$	-0.0093	(0.029)	-0.035	(0.033)
N	868,261		597,276	
Origin-County FE	✓			
Origin-County-Income FE			✓	
Sample Weights	✓		✓	
Estimator	PPML		PPML	

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. The estimation sample size is less than our total sample, and falls across columns, due to fixed effects causing separation in the sense of Correia et al. (2019). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.4: Effect of Immigrants on Local Import Expenditures

	Dependent variable: Exp. share on goods from o			
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.60***	1.71***	1.47***	1.81***
	(0.22)	(0.23)	(0.16)	(0.13)
First-stage residuals		-0.14		-0.45**
		(0.34)		(0.22)
N	1,461,130	1,461,130	6,442,722	6,442,722
Country FE	✓	✓	✓	✓
Household controls	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		20.2		18.9
Sample	Nativity	Nativity	All	All

Notes: The table presents estimation results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share. Observations are weighted using NielsenIQ household weights. Standard errors clustered two-ways at the household and county-country levels. The Nativity sample refers to the set of households that we observe both in 2014-2016 and in the 2008 Tell Me More About You Survey. The 'All' sample refers to the unrestricted set of households observed in 2014-2016. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.