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

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

June 20, 2024

PRELIMINARY AND INCOMPLETE: DO NOT CITE OR CIRCULATE

Abstract

Do immigrants make imported goods from their origin country more accessible to their non-immigrant neighbours? Using US scanner data for grocery goods, we show that immigrants increase the aggregate import expenditure share by 8%. We observe both household and product-specific origin countries, allowing us to provide the first direct evidence that three quarters of this effect is driven by immigrant preferences for goods from their origin country. A naive application of standard welfare formulas substantially over-estimates the gains to native households of immigrant-induced imports. The benefits that do accrue to natives are disproportionately concentrated among high-income and urban households.

JEL Categories: F22, J31, J61, R11.

Keywords: Price index, product variety, distributional effects

*We are grateful to Munseob Lee for helpful discussions. We thank Ilaria Malisan for research assistance and audience members at various seminars and conferences. 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 Nielsen. Nielsen 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, email: brett.mccully@carloalberto.org

‡Vancouver School of Economics, email: tjaccard@mail.ubc.ca

§Collegio Carlo Alberto, email: christoph.albert@carloalberto.org

1 Introduction

More people than ever before live in a different country than the one they were born (United Nations 2024), putting immigration at the center of contentious political debates in receiving countries. The content of these debates tends to center on how immigrants affect the nominal wages of non-immigrant households.[1] Quantifying the welfare effects of immigrants on natives, however, also requires an understanding of how immigrants affect the consumption choices and prices faced by natives.

Immigrants may reduce the cost of accessing goods from the immigrants' origin country by, for example, reducing information frictions between importers in the U.S. and exporters in the origin country. Yet merely looking at aggregate trade flows, as virtually all past studies have done,[2] is insufficient to evaluate aggregate welfare impacts or the distributional consequences of immigrants on consumption for two reasons. First, on the assessment of aggregate welfare impacts, aggregated data make it challenging to discern whether any immigrant-induced import expansion is driven by a *spillover effect* on native households or a *composition effect* associated with the specific preferences of immigrants for imported goods. Second, on the distributional consequences, past studies have lacked detailed data on household characteristics, including income, linked with import expenditures. We overcome these two key problems by linking novel data on household import expenditures with household country of birth.

This paper empirically and quantitatively studies the impact of immigrants on import volumes and on native welfare. We provide evidence that the presence of immigrants in the US increases household's grocery import expenditure share by 8%, which is roughly equivalent to reducing the prevailing tariffs applied to these goods by half. By merging households and goods to an origin country, we provide the first direct empirical evidence for a substantial *composition effect*: almost three quarters of our estimated import-immigrant elasticity is driven by immigrant preferences for both goods from their origin country and

[1]For a recent review of the academic literature in this vein, see Dustmann et al. (2016). Following the literature, we use the terms "natives" and "non-immigrants" interchangeably.

[2]A number of studies document how immigrants affect imports using aggregate trade data (Gould 1994; Head and Ries 1998; Combes et al. 2005; Peri and Requena-Silvente 2010; Burchardi et al. 2019). We discuss them more in the 'related literature' section below.

imports in general.

We document modest welfare benefits for native households associated with immigrant-induced import expenditure. A naive county-level application of the welfare formula derived in Arkolakis et al. (2012) over-estimates the welfare benefits to native households by a factor of three. The welfare gains that do accrue to natives are highly concentrated in urban, wealthy, and educated households. Approximately 55% of this pro-wealthy and pro-urban bias is driven by positive location sorting between high-income natives and immigrants, whereas 45% is associated with import preferences increasing in income.[3]

We highlight that the estimated strength of immigrant preferences for imported goods suggests immigrants are more exposed than native households to all trade shocks, immigrant-induced or otherwise.[4] A counterfactual increase in variable trade costs on all imported grocery goods by 10% decreases welfare of immigrant households in the US by 25% more than natives, with college-educated immigrants facing welfare costs that are nearly 33% greater than native households without a college degree.[5] We are the first to document that immigrants exhibit stronger demand than natives for both their origin country and all other import origins, thus enhancing the trade-creating effect of immigrants beyond their specific origin country and increasing the exposure of immigrant households to trade shocks.

This paper therefore highlights the role of heterogeneous preferences in shaping both the substantial trade effects of immigrants and the modest welfare effects of immigrant-induced trade for native households. We find that from the perspective of a quantitative trade model, immigrants should be considered primarily as a local preference shock rather than a standard trade cost shock. In addition, we are the first to characterize heterogeneity across observable native household characteristics in the gains from immigrant-induced trade and we document a strong positive bias in these gains towards high-income and urban households.

The linchpin for our analysis is a novel dataset of household consumption in which we observe the country of origin of both households and the products purchased by households.

[3] 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.

[4] This result echoes the derivation in Borusyak and Jaravel (2021) that, to a first-order, the distribution of consumer welfare effects associated with a trade shock is approximated by the distribution of import expenditure shares across consumers.

[5] Increasing all import tariffs to 10% is in fact a policy Donald Trump is campaigning on in 2024.

This dataset comprises three distinct components: (i) household-level scanner data for nearly 20,000 U.S. households, (ii) detailed country-of-origin data for over half a million grocery barcodes, and (iii) survey responses eliciting the country of birth of each household. We are the first to link product and household origin countries within a household-level expenditure dataset, and describe the data in Section 2.

We begin by estimating a structural gravity model at the household-origin level in Section 3, and in doing so separately quantify the effect of immigrants on import accessibility for all households and the specific household characteristics, such as immigrant status, which are correlated with import demand. Our estimating equation nests a wide range of standard microfoundations in the trade literature (Head and Mayer 2014) and we make use of the instruments from Burchardi et al. (2019) to generate exogenous variation in origin-specific immigrant population shares across US counties. We estimate a significant *composition effect*: immigrants spend 28% more on imports from all origins than their within-county non-immigrant neighbours, and 134% more on imports specifically from their origin country. Spillovers are also significant, as a 1 percentage point increase in the share of immigrants from a given origin increases the expenditure share of all households on goods from that origin by 1.17 percent.

Because any native increase in import consumption due to immigrants may be driven by immigrants changing native preferences—*preference diffusion*—or immigrants reducing trade costs with their origin, we develop a model of trade which allows us to cleanly separate out the various channels by which immigrants may increase imports. In particular, we extend the heterogeneous-firms model of trade developed in Melitz (2003) and Chaney (2008) to allow for four possible channels through which immigrants may induce import expenditure spillovers above and beyond the *composition effect* of immigrant preferences discussed above. First, we allow for the aforementioned *preference diffusion channel*, in which immigrants affect import preferences of their non-immigrant neighbors. Second, immigrants' preferences and the resulting higher import demand allows more firms from their origin to cover the fixed cost of exporting, thereby increasing the number of imported varieties available to non-immigrants via a *market size channel*. Finally, we allow immigrants to affect both variable and fixed trade costs specific to their origin country which we term, respectively, the *variable* *cost reduction channel* and the *fixed cost reduction channel*.

The structure of the heterogeneous firms model employed in this paper allows us to fully leverage the available data and separately identify each channel using observable moments in the data. Specifically, the elasticity of barcode-specific prices to immigrant population shares, the extensive margin elasticity of import expenditure to immigrants, and the aggregate elasticity of import expenditure to immigrants collectively pin down the *variable cost reduction channel*, *fixed cost reduction channel*, and *preference diffusion channel*.

We find no evidence that immigrants reduce variable trade costs.[6] The *fixed cost reduction channel* accounts for 80% of the aggregate *spillover effect* with the *preference diffusion channel* accounting for the remainder. Finally, we illustrate that the *market size channel* depends on the magnitude of two parameters: the ratio of the Pareto shape parameter from which firms draw productivity, θ, and the CES demand elasticity, $\sigma - 1$. Given our calibration of this ratio to values provided in the literature, we find that the market size channel accounts for 12% of the aggregate immigrant-import elasticity.[7] When combined with the estimated *composition effect*, these estimates highlight the gap between the trade-creating effects of immigrants and the welfare-inducing effect of this trade, since the *composition effect* and *preference diffusion* channel – both of which are welfare-neutral for natives – comprise 53% of the aggregate immigrant-import elasticity.

Our counterfactual immigrant-import elasticity of 0.07 falls in between the range of estimates surveyed by Felbermayr et al. (2015) of 0.12–0.15 and the null result report in Burchardi et al. (2019).

While our primary counterfactual exercise only removes the preference and trade cost effects of immigrants, we also highlight the aggregate market size benefits to natives of immigrant presence by removing both immigrant effects and expenditure. Removing immigrants reduces grocery imports by 26%, and the associated aggregate welfare effect is -0.92%.[8] The annual grocery dollar-equivalent benefit of all immigrant effects, including

[6] This is the assumption made by Peri and Requena-Silvente (2010), but we are the first to empirically test such an assumption.

[7] We calibrate $\theta/(\sigma - 1) = 1.25$, which is consistent with $\{\theta, \sigma\} = \{5, 5\}$ or $\{\theta, \sigma\} = \{10, 9\}$. Given that we do not estimate this ratio directly, we provide a sensitivity analysis for all counterfactual exercises in which we vary this ratio between 1.05 and 2.00. We discuss our range of findings below.

[8] Piyapromdee (2021) estimates that a counterfactual 25% increase in the immigrant stock would increase native welfare by on average 1.3% when considering labor and housing market effects. Albert and Monras

expenditure, is as high as \$385 in Queens, NY and \$275 in Los Angeles, CA. However for much of the Appalachian region and the Midwest, the consumption benefits of immigrant presence are practically zero.

The findings in this paper emphasize that caution is needed when interpreting immigrant-induced changes in import penetration as akin to changes in welfare for native households. Low-income, less-educated, and native-born US households face the lowest consumer costs associated with policies which increase barriers to immigration and trade. This paper therefore suggests a novel factor which may contribute to the well-documented lack of support for increased immigration among this demographic.[9]

Related literature. This paper 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: by increasing local product variety.[10] 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.

Our study is the first to leverage household-level data on import expenditures, allowing us to both quantify the contribution of a more complete set of mechanisms and to explore how the impact of immigrants varies across different households. By contrast, a vast literature on the immigration-trade nexus uses data on region-to-region trade flows (Gould 1994; Head and Ries 1998; Combes et al. 2005; Peri and Requena-Silvente 2010; Parsons and Vézina 2018; Burchardi et al. 2019) and, more recently, firm-level data (Ottaviano et al. 2018; Cardoso and Ramanarayanan 2022; Ariu 2022). A small set of papers seeks to quantify the aggregate effects of immigration (Iranzo and Peri 2009; Di Giovanni et al. 2015; Aubry et al. 2016),

(2022) compute a 1.6% average welfare increase for natives resulting from immigrants' different consumption patterns.

[9] See, for example, Card et al. (2012). For a recent review of the literature on the determinants of voter preferences on immigration policy, see Alesina and Tabellini (2024).

[10] Two prior papers have explored this margin–Mazzolari and Neumark (2012) and Chen and Jacks (2012)–but lack detailed data and exogenous variation in immigrant populations, and thus cannot separate out the various potential mechanisms nor say anything about native welfare.

but only allowing for a circumscribed set of mechanisms: productivity differs across space so immigrants moving to high productivity locations expand the set of varieties produced globally, and immigrants local demand contributes to a market size effect.[11]

The closest paper to the present one is Bonadio (forthcoming), who allows for a broad set of mechanisms by which immigrants may affect trade, including home-biased preferences, reducing trade costs, and increasing market size. Our study advances beyond Bonadio (forthcoming) in three key ways: (i) we explore distributional effects of immigration on consumption; (ii) we can relax the functional form assumption imposed by Bonadio (forthcoming) to identify home-biased preferences; (iii) we identify an additional, economically significant preference parameter, which is the demand immigrants have for imports from all origins; and (iv) we can separately identify immigrants' effect on variable versus fixed trade costs.

Our paper also contributes to the literature on spatial variation in the local cost of living (Diamond 2016; Handbury and Weinstein 2015) and local product variety (Couture 2016; Hottman 2021) by suggesting a novel force that reduces local cost of living via greater goods variety. By exploring the unequal effects of immigrants across the income distribution, we also contribute to the understanding for why there is such substantial variation in the cost of living between skill groups in U.S. cities (Su 2022; Diamond and Moretti 2021; Handbury 2021). Existing work by Lach (2007), Cortes (2008), and Zachariadis (2012) shows immigration affects aggregate local price indices constructed by statistical agencies, but the level of aggregation (at the region-level) precludes understanding (i) the mechanisms that drive the estimated effects; (ii) how much natives gain from lower prices relative to natives; (iii) which households gain.

Finally, by estimating consumer heterogeneity in exposure to trade shocks–immigrant-induced or otherwise–this paper contributes to a growing literature studying the heterogeneous consumer outcomes associated with trade shocks (Fajgelbaum and Khandelwal 2016; Bai and Stumpner 2019; Amiti et al. 2020; Hottman and Monarch 2021; Borusyak and Jaravel 2021; Faber and Fally 2022; Auer et al. 2023; Jaccard 2023). This paper is the first to document the extent to which import expenditure is particularly concentrated in immigrant

[11] Moreover, these studies also rely on aggregated region-to-region data and take immigrant stocks as exogenous.

households, thus increasing the exposure of these households to trade shocks.

2 Data and Stylized Facts

2.1 Expenditure on tradable nondurable products

To obtain Information on local expenditure on imported nondurable products by origin, we use three datasets: the NielsenIQ retail and household scanner datasets, and barcode country-of-origin data from Label Insight Inc., which we describe in detail below.

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. Detailed household demographic information – including county of residence – are included along with barcode-level expenditure, price, date, and store for each purchase. We restrict our analysis to the years 2014-2016 and aggregate these data to a single cross-section at the household level.

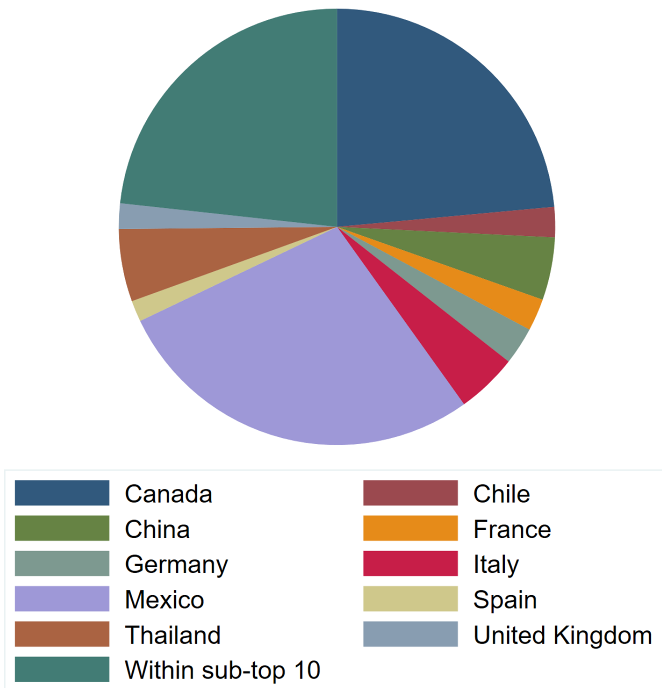
We also observe the country of birth for a subset of Nielsen households. In 2008, Nielsen distributed the "Tell Me More About You" Survey, which included questions about respondents' birth place. 80,077 individuals in 48,951 households responded to the survey for a response rate of 65 percent; more details on the survey can be found in Bronnenberg et al. (2012). Among the households surveyed in 2008, 19,700 remain in the 2014-2016 sample.

Households may have mixed nativity status. When only one member of the household was born abroad and all others were born in the US, we consider the household to be born in the country of the immigrant member. When a household has more than one foreign-born member, we assign the household to the larger country of origin as measured by the number of respondents in the survey.

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.[12] Label Insight uses an computer vision algorithm to extract the ingredients, branding, and any other text

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

Figure 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 Nielsen Household Panel 2014-2016.

information from the packaging for thousands of barcodes sold across major retail chains in the US. Since imported goods in the U.S. are required to contain some statement equivalent to "Made in ...", the Label Insight algorithm incidentally recovers a country of origin for each barcode they collect.[13] 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 use data on the origin country for over 600,000 barcodes. Given the universality of barcodes, these data can be directly merged with both the household and store-level NielsenIQ datasets.

We show the distribution of product origin countries in Figure 1. We find that the US' two neighbors and NAFTA partners, Mexico and Canada, make up over half of expenditures on imported goods. Nevertheless, substantial variation remains. China, Germany, Chile,

[13]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".

France, Italy, and Spain fill out the top 9 product origins among imported nondurables. Products from other origins make up about a quarter of expenditures on imported goods in our data.

Household Data Coverage: We merge the raw household-level purchase records with the Label Insight data and aggregate expenditure across households to the origin country-by-household level.

To get a sense for how well the Nielsen data covers U.S. consumer expenditures, we use the Nielsen-provided projection factors, which are a measure of representativeness of each household assigned by NielsenIQ.[14] Our final merged dataset covers \$764 billion USD of expenditure spanning ∼600,000 unique barcodes. There are 78 origin countries represented in the final dataset and 8.06% of all expenditure is on imported goods (\$62 billion USD in total).

We make use of the BEA Consumer Expenditure Survey (CEX) to compare the product categories covered in this paper with aggregate expenditure on tradeable sectors. The categories covered by Label Insight 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, two sectors which are not covered by Nielsen. The merged household-level dataset in this paper therefore amounts to an average expenditure per household-year of \$2,200 USD, which is just around 60% of the predicted aggregate expenditure on the categories of food and beverage, alcohol, personal care products, and cosmetics.[15]

2.2 Data of migrants and ancestry

We use the decadal Censuses from 1880–1930 and 1970–2000, as well as the pooled 2006–2010 sample of the ACS to obtain population counts of immigrants by origin. We then aggregate across individuals aged 16 and above to the county-by-country level, applying the Census'

[14]Note that these weights are not shares, but rather a population projection of the representativeness of each household. The weights sum to 120 million households, which generally matches the aggregate total for the US.

[15]If one assumes an average income of \$35,000 USD at the household level, as well as estimates from the trade literature that the tradeable sector amounts to approximately 35% of all expenditure, then the predicted aggregate tradeable expenditure is \$12,250 USD. Our categories cover approximately a third of this predicted tradeable expenditure, amounting to \$4,085 USD.

individual sample weights.[16] Immigrants are defined as those born outside the U.S. and not citizens by birth. To compute decadal migrant inflows from origin o into destination county d between two census years $t-10$ and t, denoted I_{od}^{t}, we count only those respondents who migrated to the U.S. between $t-10$ and t. Following Burchardi et al. (2019), in the first sample year the measure I_{od}^{1880} includes all those that are either first-generation immigrants from o or second-generation immigrants whose parents were born in o. The inflow measures are used in the first stage of our instrumental variables strategy outlined in Section 3.4.

Our main explanatory variable is the share of the local population who was born in country o. Destination regions d 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 (historic counties until 1940, county groups in 1970/1980, and public-use micro areas, or PUMAs, subsequently).

After merging the resulting population aggregates to the Nielsen data, we obtain a dataset at the origin country-destination county level covering 78 origin countries and 2769 destination counties.

2.3 Stylized Facts

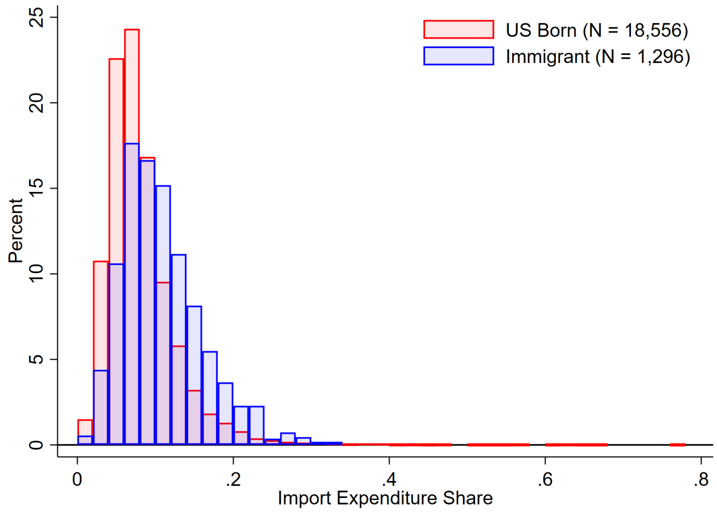
With our data, we are able to calculate expenditure by country of origin for both products and households. We leverage these novel features of our data to demonstrate three stylized facts which characterize import consumption by household origin.

Fact 1: Aggregate import expenditure shares are 38% greater for immigrant households compared to non-immigrant households. Figure 2 characterizes the distribution of raw household-specific import expenditure shares across all households in our data for which we can define a household origin country. The distribution of import expenditure shares for native-born households is given in red and foreign-born households in blue.

As is clear in Figure 2, the distribution of import expenditure shares for immigrant households appears right-shifted relative to the native-born distribution, with an average import expenditure share that is almost three percentage points greater than the average

[16] The 1940, 1950 and 1960 samples cannot be used due to missing information of the year of immigration.

Figure 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 Nielsen Household Panel 2014-2016. We exclude households who spent less than $1,000 over the 3 year sample period.

import share of non-immigrant households.

We demonstrate this difference in means by placing the household-level raw import expenditure shares in a regression framework and regressing import expenditure on a dummy for whether a household is an immigrant household. We show our results in columns 1 and 2 of Table 1, where column 1 reports the unweighted estimate and column 2 the weighted estimate. All sampling weights are provided by NielsenIQ with the goal of making the sample more representative, and are therefore our preferred estimates in this section. As shown in column 2, immigrant households exhibit an import expenditure share that is, on average, 3.1 percentage points greater than non-immigrant households. When compared to the weighted average import expenditure share of non-immigrant households, this estimate represents a 38% increase.

Of course, immigrants and imports may be more attracted to places which are better connected to international origins. To address this concern, we re-estimate this simple model but with additional county fixed effects as controls (columns 3 and 4). While these fixed effects control for geographic sorting of immigrants and non-immigrants, it comes with the draw-

Table 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. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

back of limited within-county variation: with just 19,700 households in our sample spread across the United States, some 593 households are alone in their county and hence dropped when we add in county fixed effects. Moreover, many counties have very few households, with the median county having just 3 households. Hence, we proceed cautiously with specifications which include county fixed effects, both here and when we conduct our baseline estimation in Section 5.

We show our estimates including county fixed effects in columns 3 and 4, which are the unweighted and weighted estimates, respectively. Consistent with the conditional means recovered in columns 1 and 2, we find that immigrants consume disproportionately more imports relative to natives. The magnitude falls modestly relative to the uncontrolled coefficient, to about 2.7 percentage points.

Finally, immigrant and native households even within the same county may differ in terms of income, household size, and other socioeconomic characteristics. We therefore add controls for income bins, household size, marital status, and head age and gender in columns 5 and 6. This has a negligible effect on the quantitative results, with the average immigrant household still exhibiting a conditional mean expenditure share on imported goods that is 2.8 percentage points higher than non-immigrant households.

But what characteristic of immigrant households drives this difference in import expenditure shares? One possibility is that immigrants do not favour all imported goods more than non-immigrant households, but simply those goods from their specific origin country.

The detailed nature of our data allow us to test this possibility directly, and we turn to this analysis as our second stylized fact.

Fact 2: Immigrants spend over twice as much on goods from their origin country as the non-immigrant population. For each origin country o, we calculate the share of expenditures on goods from o by both households from o and natives. We depict the resulting relationship in Figure 3.

We compare the share of expenditures on goods from o by natives on the x-axis to that of immigrant households from o on the y-axis. In both cases this aggregate share is calculated using the NielsenIQ sampling weights. The red line is the 45-degree line and indicates the points for which natives and immigrants from o allocate the same proportion of spending on goods from o. Hence, points above this line suggest that immigrants disproportionately consume goods from their origin relative to natives.

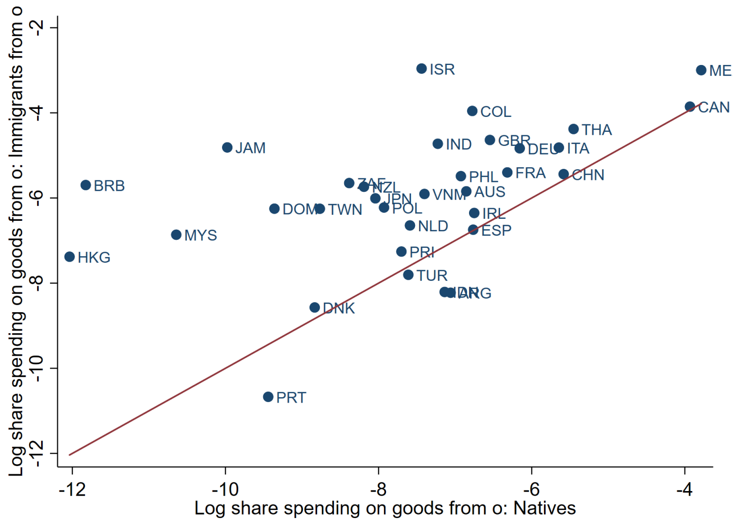
We find that most origins lie above the 45-degree line, and indeed the relationship is quite strong. For the 33 countries in our sample with non-zero expenditure by both immigrant households from that origin country and the native-born population, the average relative expenditure share on goods from origin o by immigrants from o is 2.2 times greater than the expenditure on goods from o by non-immigrant households.[17]

The relative difference in expenditure between households from o and the non-immigrant population is decreasing in the non-immigrant expenditure share associated with each origin country. A simple linear regression across the observations plotted in Figure 3 suggests that for every 10% increase in a country's expenditure share among the non-immigrant population, the relative expenditure of household's from that origin to non-immigrant households decreases by 5.81%.

This figure suggests that immigrants do indeed bring their preference for home-country goods with them to their new destination. Hence, naively estimating the relationship between immigrants and imports using aggregate data would conflate a change in the composition of demand with any reduction in trade costs, making conventional welfare analysis impossible (Felbermayr et al. 2015).

[17]Note that this estimate represents the weighted median relative expenditure across origins. The mean estimate is 30.9, but this is driven by outliers. When weighted by origin-specific aggregate expenditure shares, the mean difference is 3.4. Thus the median estimate of 2.2 represents a conservative figure.

Figure 3: Immigrants Tend to Spend more on Goods from their Origin



Notes: The figure shows the relationship between spending on goods imported from one's 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 assigned as discussed in Section 2.1. Data come from the Nielsen Household Panel 2014-2016.

As a final motivating exercise, we aggregate our data to the county-level and study the correlation between county-level immigrant population shares and the county-level import expenditure shares calculated using scanner data.

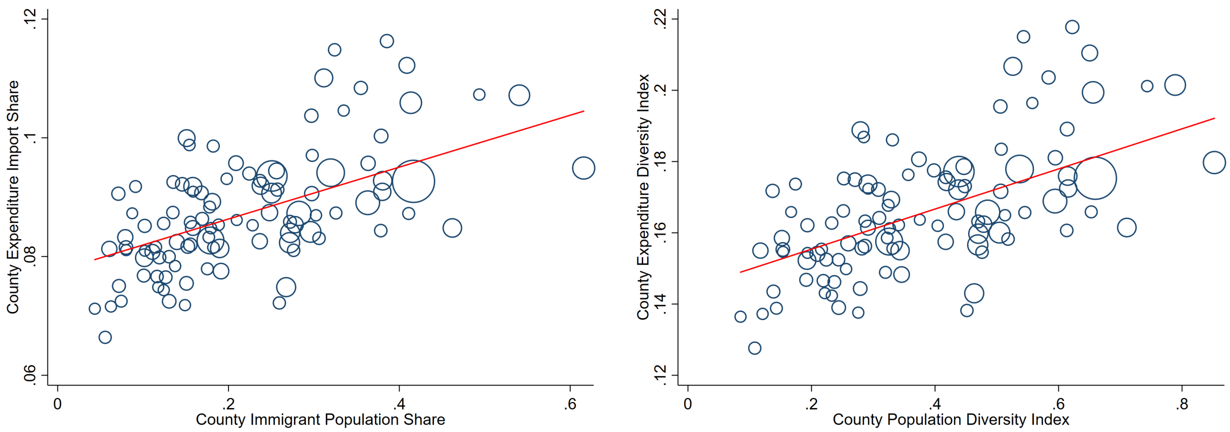
Fact 3: Both the magnitude and diversity of local immigrant population shares are correlated with the magnitude and diversity of local import expenditure shares. This fact represents the county-aggregated version of the primary relationship of interest in this paper. We show this fact in two ways in Figure 4. In both exercises we restrict our sample to counties with a population of greater than 500,000 and in which we observe more than 100 households in the Nielsen data.

First, we relate the county-level immigrant population share to the county-level import expenditure share in Figure 4a. We find that a higher share of immigrants in the population corresponds to a higher share of imports in household expenditures.

Second, we calculate measures of diversity in birth and product origins for counties with at least 100 households in our dataset. We find a strongly positive relationship between

Figure 4: Relationship between Immigrant and Product Origins

(a) Population and Expenditure Shares (b) Birthplace and Product Origin Diversity



Notes: Figure (a) Figure plots the aggregate import expenditure share at the county level against the county immigrant population share. Figure (b) plots the relationship between the diversity of product origins (y-axis) and the diversity of population birthplace origins (x-axis). We compute diversity according to Simpson's diversity index. In both subfigure, each marker size represents county population. Only counties with at least 500,000 people and at least 100 households in the Nielsen data are included. Regressions are weighted using county population measured in 2017.

immigrant birthplace diversity and product origin diversity, as shown in Figure 4b.[18]

When placed in a regression framework and estimated as logarithms, the datapoints plotted in Figure 4 suggest an elasticity of import expenditure shares to county-level immigrant population shares of 0.12 and an elasticity of import diversity to county population diversity of 0.13. Both estimates remain identical in magnitude and statistically distinguishable from zero at confidence levels of 99% when the logarithm of county population is included as a control.

When we calculate county-level import expenditure shares using only purchase records of non-immigrant households, the elasticity of import expenditure to immigrant population shares decreases to 0.1, suggesting again that composition may play a role in shaping the correlations described in Figure 4.

[18] We calculate diversity across birth places and product origins using Simpson's diversity index. The Simpson index computes the probability that two randomly sampled immigrants or products will be from the same country: $Diversity\ Index = 1 - \sum_o n_o(n_o - 1)/(N(N-1))$, where n is the number of immigrants/products from origin country o and N is the total number of immigrants/products from all origin countries. $Diversity\ Index = 0$ implies that there is no diversity of expenditure across origin countries, while a $Diversity\ Index = 1$ indicates infinite diversity (one immigrant/product per each origin).

Reassuringly, the import expenditure elasticity of 0.12 described above is remarkably similar to the estimates found in Gould (1994), Head and Ries (1998), and Felbermayr et al. (2015), who estimate aggregate import elasticities of immigration flows to be in the range of 0.10 - 0.15.

Fact 3 provides suggestive evidence that immigrants may play a role in expanding the set of available imported varieties in a location. However, our rudimentary empirical analysis is subject to two primary critiques. First, other confounders such as the strength of the local labor market may drive both import penetration and immigration to a locality. Second, even if there are no confounders, one cannot distinguish the channels through which immigrants raise imports, and hence cannot say anything about the consumption-related welfare impacts of immigration. This is particularly important given the strength of the composition effect suggested by all three stylized facts presented here. We therefore develop a model of heterogeneous firms and consumers to allow us to fully leverage the detailed data introduced in this paper and quantify the welfare effects of immigrant-induced import expenditure.

3 Immigrants and Imports: the General Gravity Model

In this section, we estimate a structural gravity model with both immigrant effects and preference heterogeneity. We show how, using household-level data, we can separately identify the direct effect of immigrants' home-biased preferences—what we term the *composition effect*—from the *spillover effect* of immigrants onto native consumption within a framework which nests a broad class of trade models.

3.1 Immigrants and Structural Gravity

We begin by considering a structural gravity model, as described in Head and Mayer (2014), with the intent of modeling import expenditure in county c on goods from origin country o, X_{oc}. The structural gravity model associated with this flow of goods is the following:

$$X_{oc} = \alpha_o S_c \phi_{oc}$$

where α_o captures some model-adjusted size of origin o, with $\alpha_o = Y_o/\Omega_o$. Y_o measures the value of production in o and Ω_o some aggregate deflator of size in production, such as marginal cost or remoteness. S_c is a measure of real demand in county c, given by $S_c = X_c/\Phi_c$, where X_c is aggregate expenditures in c and Φ_c is some price index, which we formally define below. ϕ_{oc} captures the set of bilateral factors which affect trade, such as distance, trade policy, and preference similarity.

Since we only have data for county imports, and not exports, we will from here on take this structural gravity model as a guide, rather than a true model of bilateral trade flows. It is important to note that the structure outlined above nests the key models used in modern quantitative studies of international trade, including Eaton and Kortum (2002), Krugman (1980), and Melitz (2003)-Chaney (2008).

The conventional interpretation of ϕ_{oc} is that it captures bilateral trade costs.[19] We generalize the standard gravity model by allowing for a bilateral affinity term, whereby consumers in c may exhibit preferences for the goods of specific origin countries. Formally, we decompose the bilateral term ϕ_{oc} into two multiplicative components: a supply component ϕ^B_{oc} capturing bilateral trade barriers, and a preference component ϕ^Z_{oc} reflecting the county-specific appeal associated with goods from origin o.[20] We can then re-write our structural gravity model as:

$$X_{oc} = \alpha_o S_c \phi^B_{oc} \phi^Z_{oc} \tag{1}$$

To simplify future expressions, we assume without loss of generality that for any county c, $\phi^B_{us,c} = \phi^Z_{us,c} = 1$. That is, all bilateral terms are relative to the analogous term for US producers selling to consumers in county c.

In this paper we aim to quantify the welfare effects of immigrants on native households' consumption of tradables. Because immigrants may affect both trade barriers ϕ^B_{oc} and bilateral affinity ϕ^Z_{oc}, a gravity regression using data aggregated to the origin-by-county level will be uninformative about the degree to which immigrants separately reduce trade costs and increase bilateral affinity. Instead, we make use of household-level import expenditure data, which allows us to separately identify the effects of immigrants on trade costs and

[19] Head and Mayer (2014) call ϕ_{oc} "bilateral accessibility" while Chaney (2008) calls it "trade barriers".
[20] Introduced by Combes et al. (2005), Felbermayr et al. (2015) call this term "bilateral affinity".

preferences.

3.2 Preference Heterogeneity and Household-Level Gravity

Consider a household h living in county c. Each household living in c faces the same bilateral trade costs ϕ_{oc}^B, but households differ in their aggregate expenditure X_h and vector of preference shifters $\boldsymbol{z_h}$. Each element $z_{oh} \in \boldsymbol{z_h}$ represents a household-origin-specific preference shifter, with the only restriction that $z_{us,h} = 1$ for all households. While we provide a more concrete micro-foundation regarding the household-level price index in Section 4, for now we simply allow for the possibility that the interaction between trade costs and household preferences may generate price indices which vary at the household level. We therefore define real expenditures by household h as X_h/Φ_h, where Φ_h denotes household h's price index. This assumption generates the following household-level gravity equation:

$$X_{oh} = \alpha_o \frac{X_h}{\Phi_h} \phi_{oc}^B z_{oh} \tag{2}$$

In order to link the household and county-level models, we note that:

$$X_{oc} = \sum_{h \in \Lambda_c} X_{oh} = \alpha_o \phi_{oc}^B \sum_{h \in \Lambda_c} \frac{X_h}{\Phi_h} z_{oh} = \alpha_o S_c \phi_{oc}^B \underbrace{\sum_{h \in \Lambda_c} \kappa_h z_{oh}}_{\phi_{oc}^Z}$$

where Λ_c is the set of households living in county c, $S_c = \sum_{h \in \Lambda_c} X_h/\Phi_h$ is real aggregate expenditure, and κ_h household-specific real expenditure weights.[21] The bilateral affinity term ϕ_{oc}^Z is therefore an expenditure weighted average of bilateral preferences among households in county c. In order to make progress on estimating the spillover and composition effects of immigrants on import expenditure we must make functional form assumptions on ϕ_{oc}^B and z_{oh}, as we do in the next section.

[21]Formally: $\kappa_h = (X_h/\Phi_h)/S_c$. Notice that the definition of S_c fully resolves the household-level gravity model within our county-level model, since: $\Phi_c = X_c/S_c$.

3.3 Estimating Spillover and Composition Effects

In order to render equation (2) tractable for estimation, we make the following assumptions.

First, we normalize all expenditure volumes X_{oh} by expenditure on US goods at the household level. We do so to simplify our notation, dividing out county- and household-specific terms, and in anticipation of our sample having limited coverage in many US counties.[22]

We define any variable $\tilde{x}_{oh}$ as the value of this variable for origin o divided by the equivalent value for US goods. We can therefore write the household-level gravity expression as:

$$\tilde{X}_{oh} = \tilde{\alpha}_o \phi^B_{oc} z_{oh} \tag{3}$$

To estimate the supply-side effects of immigrants on county-level import expenditure from origin o, we make the the following functional form assumption, in which d_{oc} is a vector of measures of distance between o and c and I_{oc} the population share of residents in county c that were born in country o:[23]

$$\phi^B_{oc} = e^{\rho d_{oc} + \beta^b I_{oc} + \eta^b_{oc}} \tag{4}$$

The parameter ρ captures the effect of distance on supply-side accessibility of county c to producers in o, and β^b measures the strength of the supply-side effects of immigrants in shaping import accessibility from their origin country. η^b_{oc} captures the unobserved component of origin-county-specific import accessibility.

Lastly, we provide a functional form for the preference vectors $\boldsymbol{z_h}$. We consider two components of preferences: a component that captures endogenous spillovers associated with local immigrant population shares and a component that relates exogenous and observed household characteristics to import demand. For a given household and origin country, we therefore assume the following functional form for z_{oh}:

$$z_{oh} = e^{\beta^z I_{oc}} e^{[\delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o] + \eta^z_{oh}]} \tag{5}$$

J_h represents a vector of observed household characteristics such as income, education, race,

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

[23] d_{oc} includes the log distance between o and c and the latitude difference between o and c, as well as squared and cubed terms of that latitude difference.

etc., and δ maps these observables into import demand across all origins. ζ_1 captures the extent to which immigrant households have stronger preferences for goods from all origin countries, and ζ_2 captures the extent to which immigrants prefer goods specifically from their origin country à la Atkin (2016) and Bronnenberg et al. (2012).

Household-level characteristics will not respond to changes in immigrant presence in our counterfactuals, and hence the parameters ζ_1 and ζ_2 govern the composition effect of changes in I_{oc}. Parameter β^z, on the other hand, captures some measure of preference diffusion in which the presence of immigrants from a given origin affects the average preference for goods from that origin across all households in the same county.

Plugging these functional form assumptions into our expression for $\tilde{X}_{oh}$, we derive our estimating equation:

$$\ln \tilde{X}_{oh} = \ln \tilde{\alpha}_o + \tau d_{oc} + \beta I_{oc} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o] + \eta_{oh} \tag{6}$$

with $\eta_{oh} = \eta^b_{oc} + \eta^z_{oh}$ capturing idiosyncratic county and household-level deviations in import expenditure associated with origin o. The parameter $\beta = \beta^b + \beta^z$ captures spillover effects of immigrants into import expenditure for all households, but cannot distinguish between the supply and demand effects of this spillover.

As discussed in Burchardi et al. (2019), there is reason to believe that the unobserved component of import supply and demand shocks η_{oh} is correlated with the presence of immigrants, which would lead to biased estimates of β. We discuss our identification strategy in the next section.

3.4 Identification and Instrument Variables

In estimating equation (6), there may be confounders correlated with both the consumption share of a household from a specific origin and the presence of immigrants in the household's county of residence that are not captured by our baseline controls. For example, a location may have both high immigration and more import activity, implying a higher import share in the consumption of local households, because of an idiosyncratically high number of flight connections. Notice that this unobserved correlation would not be captured by our distance measures. To deal with such origin-county specific confounders, we adopt the instrumental variables approach taken by Burchardi et al. (2019).[24]

Recall that our goal is to instrument for the share of immigrants from o in the population of county c in 2010. The intuition of the instruments derived in Burchardi et al. (2019) is that an immigration flow from an origin to a destination is more likely to occur when the origin is sending many immigrants at the same time the destination is pulling in many immigrants. For example, suppose we want to predict the number of Italians settling in Chicago in a given historical decade. To do so, we calculate the number of Italians immigrating into the United States and the number of immigrants from all origin countries settling in Chicago for the same decade. The constructed instrument will predict a large number of Italians settling in Chicago if large numbers of immigrants from other countries are also settling there.

Concretely, the immigration leave-out push-pull instrument interacts the arrival into 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. We follow Burchardi et al. (2019) and enhance this specification by leaving out both the continent of origin country o and the Census region of county c when constructing the instrumental variable that we use in our baseline estimation for origin o and county c. For decade D, we can therefore write our instrumental variable as:

$$IV_{o,c}^{D} = I_{o,-r(c)}^{D} \times \frac{I_{-\mathcal{C}(o),c}^{D}}{I_{-\mathcal{C}(o)}^{D}} \tag{7}$$

where $r(c)$ is the Census region of county c, and $\mathcal{C}(o)$ the set of countries on o's continent. Therefore, $I_{o,-r(c)}^{D}$ is the number of immigrants from o settling in the U.S. outside the Census region of county c in decade D, and $I_{-\mathcal{C}(o),c}^{D}/I_{-\mathcal{C}(o)}^{D}$ is the fraction of immigrants to the U.S. from outside the continent of o who choose to settle in county c. In our running Italy-Chicago example, the instrument interacts the number of Italian immigrants settling outside the Midwest ($I_{o,-r(c)}^{D}$) with the fraction of non-European immigrants arriving in the U.S. who choose to settle in Chicago ($I_{-\mathcal{C}(o),c}^{D}/I_{-\mathcal{C}(o)}^{D}$).

[24] 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 section A in the Appendix for more details.

One advantage of the leave-out structure of the instrumental variables is that it neatly deals with concerns over reverse causality. For example, importing firms may send workers from an origin country to migrate to the counties into which they hope to import goods. However, these bilateral flows, as well as any historical bilateral flows, are not used for the prediction of the bilateral immigrant population.

The identification assumption is that any confounding factors that make a given county more attractive for both immigration 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. A violation may occur if, say, immigrants skilled at importing goods from Italy tend to settle in Chicago and immigrants skilled in importing goods from South Korea settle in Miami in the same decade and for the same reason: a large number of flight connections. This violation is only quantitatively meaningful if Italians 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 (A.1) to predict immigrant inflows into the U.S. for all decades spanning 1880 to 2000. 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 in the appendix. We find that the push-pull instrument strongly and positively predicts the contemporary bilateral immigrant population. Our first-stage estimates therefore match closely with those of Burchardi et al. (2019), which is reassuring given the extensive series of robustness exercises explored by Burchardi et al. (2019) to ensure the validity of this instrumental variable.

Given the prevalence of zeros in the consumption expenditure shares $\tilde{X}_{oh}$, we use pseudo-Poisson maximum likelihood (PPML) to estimate equation 6 without taking the logarithm (Silva and Tenreyro 2006). When implementing the instrument variables strategy introduced below, we account for the non-linearity of PPML by implementing a control function approach to generating exogenous variation in the immigrant population (Petrin and Train 2010; Morten and Oliveira 2023). In particular, we add the residuals from the first-stage instrumental variable regressions as controls for our main specifications.[25]

[25] Atalay et al. (2019) demonstrates that the control function approach generates consistent estimates when

Table 2: Household Gravity Estimates

	Dependent variable: Exp. share on goods from o relative to US (1)	(2)
Immigrants$_o$/Pop. 2010	1.33***	1.17***
	(0.22)	(0.24)
First-stage residuals		0.20
		(0.31)
=1 if immigrant from anywhere	0.25***	0.25***
	(0.030)	(0.030)
=1 if immigrant from origin o	0.60***	0.60***
	(0.068)	(0.069)
N	1,461,130	1,461,130
Country FE	✓	✓
Household controls	✓	✓
Distance & latitude difference	✓	✓
1st-stage F-statistic		20.2

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

3.5 Results

We show the results of estimating equation (6) using PPML in Table 2. The first column shows estimates without first-stage residuals, which are added in the second column. We find in both cases that immigrants significantly increase the share of consumed goods from their origin. In our preferred specification in column 2 we find that a 1 percentage point increase in the share of immigrants from a given origin increases relative expenditures on goods from that origin by 1.17 percent (SE=0.24).

Comparing the immigrant population share coefficients between columns 1 and 2, we find that the estimate falls by about 12% when adjusting for the endogenous location choices of immigrants. This is consistent with immigrants choosing their location based on where goods from their home country are more available.

using PPML. They further show that the estimates are quite close to those produced by the related GMM estimation strategy developed by Wooldridge (1997) and Windmeijer (2000).

Turning to the nativity and homophily coefficients, we find that immigrants spend 28%[26] more on imports from any origin than natives do, and 134%[27] more on imports specifically from the immigrant's origin country.

The results summarized in Table 2 provide two key takeaways. First, immigrants' preferences—the composition effect—play a significant role in shaping import expenditures. Indeed, our estimates validate the caution expressed by Felbermayr et al. (2015) in interpreting immigrants' effect on imports using aggregated data as an effect on welfare. Second, we find that spillover effects of immigrants from a given origin to the rest of the local population—captured by the immigrant-population share coefficient—are also significant.

Even controlling for immigrant preferences, the estimated spillover effect may incorporate both immigrants' effects on trade costs and on local preferences – what we call preference diffusion. While this distinction has no bearing on the trade-creating effects of immigrants, it plays a crucial role in identifying the welfare effects of immigrant-induced trade. We discuss this distinction in the following section within the context of the welfare formula derived in Arkolakis et al. (2012) (henceforth the "ACR" formula).

3.6 Immigrants, Imports and Welfare

Given that the model described here is derived from a structural gravity model and is therefore consistent with the class of models discussed in Arkolakis et al. (2012), it is trivial to show that welfare effects of a trade shock can be represented by changes in domestic expenditure.

With heterogeneous households and a potential demand-side effect of immigrants on import expenditure, however, issues arise in naively applying the ACR welfare formula to county-level aggregate changes in domestic expenditure. This is for two reasons. First, if native households exhibit weaker preferences for imported goods than immigrants (ζ_1, $\zeta_2 > 0$), then changes in aggregate domestic expenditure will over-estimate the welfare effects of any trade shock on native households. Second, immigrant-induced changes in county-level import expenditure will only translate into welfare gains for native households

[26] Equivalently, 0.25 log points (SE=0.03).

[27] Equivalently, $0.60 + 0.25 = 0.85$ log points (SE=0.069).

if driven by changes in the supply component of the bilateral term ϕ_{oc}, rather than the demand component.

With a few simplifying assumptions, we derive an explicit adjustment to the standard ACR formula which allows us to directly express the gap between implied welfare gains from aggregate changes in expenditure to the true changes in expenditure, and therefore welfare, associated with native households. For the purposes of this exercise, we assume that there exists only two countries: the United States (us) and some foreign country, denoted by m. We denote native households with n and assume all households are identical except for their immigrant status. Lastly, we collapse ζ_1 and ζ_2 into a single parameter ζ which captures the relative import preference of immigrants versus native households.

We consider some change in the immigrant population share which causes aggregate county-level domestic expenditure to change by some exogenous $d \ln X_{us,c}$.[28] Given the structural gravity model described above and estimates of β^b, β^z, and ζ, one can transform the county-level change in domestic expenditure to the welfare-relevant change in domestic expenditure of native households using the following transformation:

$$d \ln X_{us,n} = d \ln X_{us,c} \left[\frac{1}{\frac{I_c}{s_{us,c}}(e^{\zeta} - 1) + 1} \right] \left[\frac{\beta^b}{\beta + \frac{e^{\zeta}-1}{I_c(e^{\zeta}-1)+1}} \right] \tag{8}$$

where I_c is the immigrant population share in county c and $s_{us,c}$ is the pre-shock domestic expenditure share in county c.

The first term associated with this transformation adjusts $s_{us,c}$ in order to recover the unobserved native household domestic expenditure share $s_{us,n}$. So long as $\zeta > 0$, and immigrants have stronger preferences for imports than native households, this term will be less than one and for any trade shock – immigrant-induced or otherwise – native households will exhibit smaller changes in welfare than those implied by the county-level aggregate $d \ln X_{us,c}$.

The second term captures the share of the aggregate change in domestic expenditure which is welfare-relevant to native households. That is: $d \ln \phi^B_{oc} / d \ln \phi_{oc}$. If at least one of β^z or ζ is positive, and both are non-negative, this second term is less than one. Therefore changes in native household welfare should be discounted when compared to the implied

[28] We assume that aggregate expenditure X_c remains constant.

aggregate welfare effects. If $\beta^b = 0$, then immigrant-induced changes in domestic expenditure may be large in the aggregate, but will have zero effect on the welfare associated with native households, and $d \ln X_{us,n} = 0$. It is clear from our estimates discussed previously that immigrants do exhibit stronger preferences for imports from their origin country than native households, and thus $\zeta > 0$. What we cannot disentangle from our estimates is the relative magnitude of β^b and β^z: the extent to which immigrants act as a supply or demand shock to import penetration.

We therefore take these general gravity estimates as motivation for the section to follow, in which we make use of the heterogeneous firms Melitz-Chaney variant of the structural gravity class of models to run counterfactual simulations and recover the effect of immigrants on import penetration and native household welfare. We opt for the Melitz-Chaney model for two reasons. First, the increasing returns to scale nature of this model allows for market size effects, a key channel through which immigrants might affect the supply component of accessibility above and beyond trade costs (Iranzo and Peri 2009; Di Giovanni et al. 2015; Aubry et al. 2016). Second, the structure of the Melitz (2003)-Chaney (2008) heterogeneous firms model allows us to fully leverage the data we possess and separately quantify the marginal cost, fixed cost, and preference spillover effects of immigrants on native households, thus identifying the supply and demand effect of immigrants on import penetration. We turn to describing this model now, as well as our estimation/calibration of the model and subsequent counterfactual exercises.

4 A Model of Immigration and Import Expenditure

This section uses the Chaney (2008) micro-foundation to expand upon the structural gravity model of immigrant-induced trade in the previous section. We then leverage the equilibrium moments of this model and the detailed data available to separately identify the effect of immigrants on marginal costs, fixed costs of exporting, and household preferences, thus disentangling ϕ^B_{oc} and ϕ^Z_{oc}.

4.1 Micro-founding Structural Gravity: Heterogeneous Households and Firms

Households: Each household h lives in county $c(h)$ and exhibits Cobb-Douglas preferences over a homogeneous tradable good, q_0, and a differentiated good consisting of an endogenous continuum of differentiated varieties $\Omega_{o,c(h)}$ associated with each origin country $o \in \mathcal{O}$. As in the previous section, we model household heterogeneity in income Y_h and the vector of origin-specific preferences $z_{oh} \in \boldsymbol{z_h}$. Preferences for the differentiated sector are represented by the following CES utility function:

$$U_h = {q_0}^{\mu_0} \left[\sum_{o \in \mathcal{O}} {z_{oh}}^{\frac{1}{\sigma}} \int\limits_{\omega \in \Omega_{o,c(h)}} q_{oh}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}(1-\mu_0)} \tag{9}$$

with $\sigma > 1$ denoting the elasticity of substitution. The exponent μ_0 captures the expenditure share on the homogeneous good, which we assume is constant across households and therefore pins down expenditure on the differentiated sector as $X_h = (1-\mu_0)Y_h$.

We leave the functional form of z_{oh} unchanged from the previous section (see equation (5)). That is, β^z governs the endogenous spillover effect of immigrants on native household preferences for imports, δ maps exogenous household characteristics into import demand, ζ_1 governs immigrant preferences for imported goods, and ζ_2 governs immigrant preferences for goods specifically from their origin country.

Firms: Each country $o \in \mathcal{O}$ has some exogenous size Y_o and marginal cost of production w_o. Trade is characterized by county-origin-specific iceberg trade costs and fixed costs given by, respectively, τ_{oc} and f_{oc}. Each firm draws some productivity φ from a Pareto distribution with shape parameter $\theta > \sigma - 1$ and the set of potential entrant firms in each origin is proportional to the size of that origin Y_o.[29] The cost of providing q units to destination county c by a firm in origin o with productivity φ is therefore:

$$C_{oc}(q) = \frac{w_o \tau_{oc}}{\varphi} q + f_{oc} \tag{10}$$

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

and we assume that all entry and pricing decisions are made by all firms at the county level such that each county is an independent market.

Given the extent to which this model builds upon the structure introduced by Chaney (2008), we relegate the full derivation of the model to the appendix, including all definitions of constants denoted by λ.

Equilibrium: In equilibrium, the household-specific price index is given by:

$$\Phi_h = P_h^{1-\sigma} = \lambda_3 \sum_{o \in \mathcal{O}} Y_o z_{oh} (w_o \tau_{o,c(h)})^{-\theta} \Big(\frac{f_{o,c(h)}}{S_{c(h)} z_{o,c(h)}} \Big)^{-(\frac{\theta}{\sigma-1}-1)} \tag{11}$$

in which $S_{c(h)}$ is again real aggregate expenditure in county c, as defined previously in the structural gravity model.[30] The average county-level preferences z_{oc} are also the same as our definition for the bilateral affinity term introduced earlier (ϕ_{oc}^Z) and are an expenditure weighted average of the preference shifter z_{oh} across all households in Λ_c. Household-level expenditure on goods from origin o can then be expressed as:

$$X_{oh} = \lambda_4 Y_o X_h P_h^{\sigma-1} (w_o \tau_{o,c(h)})^{-\theta} \Big(\frac{f_{o,c(h)}}{S_{c(h)} z_{o,c(h)}} \Big)^{-(\frac{\theta}{\sigma-1}-1)} z_{oh} \tag{12}$$

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

$$X_{oc} = \lambda_4 Y_o S_c (w_o \tau_{oc})^{-\theta} \Big(\frac{f_{oc}}{S_{c(h)} z_{oc}} \Big)^{-(\frac{\theta}{\sigma-1}-1)} z_{oc} \equiv \alpha_o S_c^{\frac{\theta}{\sigma-1}} \phi_{oc}^B \phi_{oc}^Z \tag{13}$$

Notice that we now have a micro-founded equivalent to all terms derived in our previous structural gravity model. ϕ_{oc}^Z remains unchanged, whereas real expenditure is now enhanced by the exponent $\theta/(\sigma-1) > 1$ due to the increasing returns to scale associated with the micro-foundation of production assumed here. The real size of origin o is now defined formally as $\alpha_o = Y_o w_o^{-\theta}$.

The supply-side component of ϕ_{oc}, however, now contains a component associated with variable trade costs (τ_{oc}), a component associated with fixed costs (f_{oc}), and a component

[30] Formally: $S_c = \sum_{h' \in \Lambda_c} X_{h'} P_{h'}^{\sigma-1}$, where Λ_c is the set of households residing in county c.

associated with average county-level preferences z_{oc}. While the trade cost components are standard in any application of the Chaney (2008) framework, the preference component is novel and reflects a market size effect associated with county-level average preferences: as preferences shift towards goods from origin o, more firms are able to cover the fixed costs of supplying county c, and this further enhances the market penetration of goods from origin o.

An implication of introducing heterogeneous households within the Chaney (2008) framework is that changes in the immigrant population share might lead to increased import expenditure by native households via the preference market size effect, a channel which was missing from our original structural gravity derivation.

As before, it will be convenient when taking our main estimating equation to the data to estimate the model relative to U.S. expenditure for a given household. Using the same definition as $\tilde{x}$ from before to denote any variable relative to the US equivalent, we can express our normalized household-origin level expenditure equation as the following:

$$\tilde{X}_{oh} = \tilde{\alpha}_o(\tilde{\tau}_{o,c(h)})^{-\theta}\Big(\frac{\tilde{f}_{o,c(h)}}{z_{o,c(h)}}\Big)^{-(\frac{\theta}{\sigma-1}-1)} z_{oh} \tag{14}$$

It will also be useful to separate household preferences into a component that is endogenous to the local immigrant population share, $e^{\beta^z I_{oc}}$, and an exogenous component $\bar{z}_{oh}$, such that $z_{oh} = e^{\beta^z I_{oc}}\bar{z}_{oh}$. Given that the endogenous component is common to all households in a given county, we can provide the same distinction at the county level $z_{oc} = e^{\beta^z I_{oc}}\bar{z}_{oc}$, where $\bar{z}_{oc}$ is simply an expenditure weighted average of $\bar{z}_{oh}$.

In the following section we complete the model described here by introducing functional form assumptions for variable and fixed trade costs. As before, we allow for the possibility that immigrants might affect these costs. We then derive our main estimating equation and highlight the various channels through which immigrants affect import penetration and welfare in this model, before discussing our strategy for separately identifying each channel.

4.2 Immigrants and Imports: Identifying the Relevant Channels

Our functional form assumptions regarding the variable and fixed components of ϕ^B_{oc} closely follow the assumptions made in the structural gravity model. That is, 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. These functional form assumptions are given by the following:[31]

$$\tilde{\tau}_{oc} = exp[-\frac{1}{\theta}(\rho^\tau d_{oc} + \beta^\tau I_{oc} + \eta^\tau_{oc})]$$

$$\tilde{f}_{oc} = \exp[-(\frac{\sigma - 1}{1 + \theta - \sigma})(\rho^f d_{oc} + \beta^f I_{oc} + \eta^f_{oc})]$$

where η^τ_{oc} and η^f_{oc} represent idiosyncratic deviations in trade costs across county-origin pairs that are assumed to be mean-zero. In this case β^τ represents the *variable cost reduction channel* of immigrants and β^f represents the *fixed cost reduction channel* of immigrants on import expenditure in county c.

We can now return to our expression for $\tilde{X}_{oc}$ and plug in our functional form assumptions for z_{oh}, $\tilde{\tau}_{oc}$, and $\tilde{f}_{oc}$. Taking the logarithm of this expression and differentiating with respect to the immigrant population share yields the following decomposition of the county-level partial elasticity of import expenditure with respect to the immigrant population share:

$$\begin{aligned}\frac{\partial \ln \tilde{X}_{oc}}{\partial I_{oc}} &= \frac{\partial \ln \phi^B_{oc}}{\partial I_{oc}} + \frac{\partial \ln \phi^Z_{oc}}{\partial I_{oc}} \\ &= \underbrace{\left[\beta^\tau + \beta^f\right]}_{\substack{\text{Trade cost} \\ \text{channel}}} + \underbrace{\left[\frac{\theta}{\sigma - 1} - 1\right]\left(\beta^z + \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}\right)}_{\text{Market size channel}} + \underbrace{\beta^z}_{\substack{\text{Cultural} \\ \text{diffusion} \\ \text{channel}}} + \underbrace{\frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}}_{\substack{\text{Composition} \\ \text{channel}}} \qquad (15)\end{aligned}$$

This expression clearly illustrates the channels through which immigrants affect county-level import expenditure from a given origin. The first two channels represent changes in the supply-side effects of immigrants, or ϕ^B_{oc}. These include the *variable cost reduction effect*, the *fixed cost reduction effect*, and the *market size effect* associated with changes in

[31] The normalization terms $\frac{1}{\theta}$ and $\frac{\sigma-1}{1+\theta-\sigma}$ are not necessary but simplify notation and interpretation of our estimates later on.

local preferences. A shift in county-level preferences for goods from origin o will lead to greater entry by firms exporting from o, and given the CES preferences assumed in this model, this increased availability will lead to non-zero expenditure on these new varieties by non-immigrant households. This effect is entirely mediated by the ratio $\theta/(\sigma-1) > 1$.

The final two terms capture the extent to which immigrants affect the bilateral affinity term ϕ_{oc}^{Z}. β^{Z} captures the effect of immigrants on preferences for goods from their origin that are common to all households in county c, whereas the *composition effect* captures the extent to which increased immigrant presence shifts the composition of households towards those with non-zero values of the parameters ζ_1 and ζ_2.

From a welfare perspective, the intuition is identical to the discussion previously regarding the structural gravity model. The only welfare relevant channels of immigrant-induced import penetration are those associated with ϕ_{oc}^{B}: the *trade cost channel* and the *market size channel*.

An important feature of the micro-foundation used here is that when combined with the detailed data available, we can separately identify all parameters necessary to quantify each channel. We will again make use of our household-level data and so we return to the household-level gravity model discussed previously but adjusted for the microfoundation described here:[32]

$$\ln \tilde{X}_{oh} = \alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)} + \ln \bar{z}_{o,c(h)}^{\frac{\theta}{\sigma-1}-1} + \delta J_h + \zeta_1 \mathbf{1}\left[o(h) \neq US\right] + \zeta_2 \mathbf{1}\left[o(h) = o\right] + \eta_{o,c(h)} + \eta_{oh}^{z} \tag{16}$$

with the following definitions:

$$\rho = \rho^{\tau} + \rho^{f}$$

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

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

The specification here reveals a number of key identification concerns. First, we encounter the

[32]With some abuse of notation, we define $\alpha_o = \ln \alpha_o$.

same identification concern as in the previous section: the unobserved component of variable costs and fixed costs η^τ and η^f are likely correlated with the immigrant population share I_{oc}, and hence we make use of the same instrument variables strategy. Second, and perhaps more concerning, is that the *preference spillover effect* of immigrants and the *composition effect* of immigrants are not separately identified: the county-level preferences $\bar{z}_{oc}$ were not included in our structural gravity regression and therefore loaded on to estimates of β. Lastly, even an unbiased estimate of β would simply yield a combination of β^τ, β^f, and β^z.

In the following section we provide a three-step identification strategy which allows us to separately identify these various channels.

4.3 Identifying the Channels of Immigrant-Induced Import Expenditure

Identification of Composition Effects: We begin by collecting all terms endogenous to the local immigrant population share into an origin-county fixed effect ψ_{oc} and make use of our households-level purchase data to estimate the exogenous component of preferences $\bar{z}_{oh}$. Specifically, we estimate the following model:

$$\ln \tilde{X}_{oh} = \psi_{o,c(h)} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o] + \eta^z_{oh} \tag{17}$$

In this case it is safe to assume that the estimates $\hat{\delta}$, $\hat{\zeta}_1$, and $\hat{\zeta}_2$ are unbiased as the only error term not captured by the fixed effects is the idiosyncratic household-origin preference shock η^z_{oh}. That is, all components of the model associated with prices are captured by the origin-county fixed effects ψ_{oc}. We estimate this specification using PPML to account for the number of zeros in $\tilde{X}_{oh}$ and recover the estimates $\hat{\delta}$, $\hat{\zeta}_1$, and $\hat{\zeta}_2$. We then construct an estimate of the household-level preference term as $\hat{\bar{z}}_{oh} = e^{(\hat{\delta} J_h + \hat{\zeta}_1 \mathbf{1}[o(h) \neq US]) + \hat{\zeta}_2 \mathbf{1}[o(h)=o])}$ and plug this estimate into the county-level average preference term $\bar{z}_{oc}$ to arrive at an estimate of $\hat{\bar{z}}_{oc} = \sum_{h' \in \Lambda_c} \hat{\bar{z}}_{oh'} \kappa_{h'}$. We make use of publicly available Census data to construct the household-level weights κ_h, and we make use of calibrated values of σ and θ taken from the literature, which we discuss in the next section.

Estimating $\boldsymbol{\beta}$: With these estimates in hand, we can difference out both $\bar{z}^\sigma_{oh}$ and $\bar{z}_{oc}^{\frac{\theta}{\sigma-1}-1}$

from our main estimating equation and isolate the effect of county-level parameters on this adjusted measure of import expenditure, as shown in the following equation:

$$\ln \frac{\tilde{X}_{oh}}{\mathcal{Z}_{oh}} = \alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)} + \eta_{o,c(h)} + \eta^z_{oh} \tag{18}$$

in which we define $\mathcal{Z}_{oh} = \hat{\hat{z}}_{oh} \hat{\hat{z}}_{o,c(h)}^{\frac{\theta}{\sigma-1}-1}$ to simplify notation.

Notice that the dependent variable represents observed household-level expenditure on imports from origin o adjusted by their expected level of expenditure, given their own observable characteristics as a household but also their expected import expenditure given the observed characteristics of all other households living in their county. Applying this deflator $\mathcal{Z}_{oh}$ to household h's expenditure on goods from origin o allows us to isolate the *spillover effect* of immigrants by controlling for the composition effect directly.

We therefore arrive at an estimating question that is reminiscent of the structural gravity model estimated earlier, and we make use of the same instrumental variables strategy and again implement PPML with a control function approach. But how to separately identify the components of β? In the following paragraphs we discuss model restrictions and data characteristics which allow us to perform this decomposition.

Estimating $\boldsymbol{\beta^\tau}$: We have assumed throughout this section that firms price according to monopolistic competition, and thus set constant mark-ups. Specifically, the optimal pricing function for any variety ω from origin o in county c is the following:

$$p_{\omega(o),c} = \frac{\sigma}{\sigma-1}\frac{w_o \tau_{oc}}{\varphi(\omega)} = \frac{\sigma}{\sigma-1}\frac{w_o}{\varphi(\omega)}\tau_{us,c}\tilde{\tau}_{oc}$$

By aggregating our data to the barcode-county level we can estimate this equation directly, after having incorporated the functional form assumption of $\tilde{\tau}_{oc}$ introduced earlier:

$$\ln p_{\omega(o),c} = \psi_c + \psi_\omega - \frac{\beta^\tau}{\theta} I_{oc} - \frac{\rho^\tau}{\theta} d_{oc} - \frac{1}{\theta}\eta^\tau_{oc} \tag{19}$$

where ψ_c and ψ_ω represent county and barcode-level fixed effects.[33] Since our dataset is at

[33]Note that each barcode ω is unique to an origin country o; hence ψ_ω also captures variation in production costs w_o across origins.

the barcode level we are able to estimate β^{τ} while controlling for the composition effects our model predicts. As with our baseline specification, we also instrument for the immigrant population share to account for the likelihood that $cov[I_{oc}, \eta^{\tau}_{oc}] \neq 0$.

Estimating $\boldsymbol{\beta^f}$ and $\boldsymbol{\beta^z}$: We will show in the next section that our estimates for β^{τ} are approximately equal to zero, implying that immigrants have no effect on variable trade costs. This fact allows us to isolate the effect of fixed costs and preferences on import expenditure by comparing the expenditure import-immigrant elasticity to the *variety* import-immigrant elasticity.

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. Since $\beta^{\tau} \approx 0$, we derive two equations with two unknowns: β^f and β^z. The scanner data used in this paper provide detailed barcode count data, and so we estimate the extensive margin effect of immigrants on trade directly by replacing $\tilde{X}_{oh}$ in (18) with $\tilde{N}_{oh}$: the count of barcodes from origin o in household h's consumption basket compared to the count of barcodes from the U.S. in household h's consumption basket.

While the full derivation is provided in the appendix, it is simple to show that our functional form assumptions for β^f and β^z yield the following two expressions regarding the import expenditure elasticity and import variety elasticity, respectively:

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

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

The intuition behind the difference between these two equations is the following. A marginal increase in the immigrant share will affect the number of entering firms and therefore available varieties through reducing fixed costs, captured by β_f, and by increasing market size, captured by $\left(\frac{\theta}{\sigma-1} - 1\right)\beta^z$. These channels affect both the number of varieties purchased by each household, the *extensive margin* captured by the barcode count $\tilde{N}_{oh}$, and the overall expenditure share on goods from o given by $\tilde{X}_{oh}$ (relative to the respective US equivalents).

In contrast, a change in preferences for goods from o via the preference diffusion channel captured by β_z (see equation (15)) additionally leads to an increase in the *intensive margin* of spending. That is, it increases households' expenditure per given variety from o and hence only shows up in the equation for $\tilde{X}_{oh}$, yielding the expression $\left(\frac{\theta}{\sigma-1}\right)\beta^z$.

Notice that as the extensive margin elasticity approaches the total expenditure elasticity, β^f approaches β and β^z approaches zero, implying that immigrants affect fixed costs of importing rather than average preferences. By estimating both elasticities, we are therefore able to recover estimates of β^f and β^z.

With these three steps we can therefore provide plausibly unbiased estimates of the effect of immigrants on import expenditure, as well as separately identify the composition and spillover effect and the three channels which constitute the spillover effect.

5 Results and Discussion

5.1 Estimating Preference Terms

We construct estimates for household preferences $\bar{z}_{oh}$ by estimating equation (17) using our Nielsen household sample. We therefore recover estimates of the parameter vector δ as well as ζ_1 and ζ_2. The detailed regression results are presented in Appendix Table C.1.

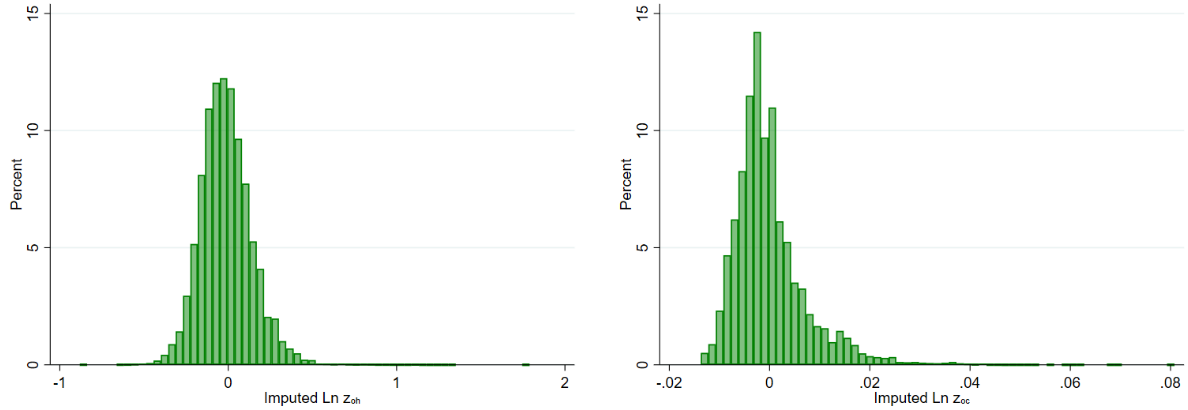
We find that import expenditure is generally increasing in income, albeit noisily, with a similar pattern of import expenditure increasing in household education.[34] We estimate that immigrant households consume more imported goods from any origin ($\hat{\zeta}_1$) with an estimated effect of 0.23 (SE=0.029), as well as more goods from their specific birth country ($\hat{\zeta}_2$), with an estimated effect of 0.64 (SE=0.069). These estimates match what we found in Section 3.5, and suggest that immigrant import expenditure is 1.26 times that of native households for all origins, and 2.39 times greater for imports from their specific origin country.

Figure 5a shows the distribution of $\ln \hat{\bar{z}}_{oh}$, which is centered around zero with a few extreme values above 1. Due to the low number of observed households in many of the

[34]Note that these estimates combine the extensive and intensive margin of import expenditure: a higher estimate associated with a given household characteristic implies that this household has greater expenditure on the same set of origins and/or consumes imports from more origins.

Figure 5: 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 Nielsen household-origin pairs of the log of $\hat{\hat{z}}_{oh} = \exp(\hat{\delta} J_h + \hat{\zeta}_1 \mathbf{1}[o(h) \neq US]) + \hat{\zeta}_2 \mathbf{1}[o(h) = o])$, where the terms $\hat{\delta}$ and $\hat{\zeta}$ are estimated from equation (17). 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.

smaller counties in the Nielsen data, we rely on the more comprehensive ACS 2012-2017 sample, which we use to construct the same set of household-level variables included in the vector J_h to compute $\hat{\hat{z}}_{oc} = \sum\limits_{h' \in \Lambda_c} \hat{\hat{z}}_{oh'} \kappa_{h'}$. In particular, we predict $\hat{\hat{z}}_{oh}$ for each household in the ACS sample and aggregate these predictions across counties and origins with the household weights $\kappa_{h'}$ being the share of overall income in county c that is earned by household h'.

Figure 5b shows the distribution of $\ln \hat{\hat{z}}_{oc}$. Due to the aggregation of household-origin expenditure shares at the county level, the distribution has a much lower range, which lies between -0.016 and 0.077. Five out of the six largest values correspond to the preference terms for products from Mexico in counties in California and Texas. Other county-origin pairs in the top 10 include preferences for Cuban products in Miami-Dade county, preferences for Chinese goods in the San Francisco and Santa Clara counties, and preferences for Indian goods in Middlesex county (NJ).

5.2 Structural Gravity Estimates of Immigrant Spillover Effects

We start by estimating the total effect of immigrants on imports using equation (18), in which expenditure is deflated by household and county-level preferences. Recall that in

Table 3: Estimates of Household Gravity Equation

Dependent variable: Relative expenditure share on goods from o	$\tilde{X}_{oh}/\mathcal{Z}_{oh}$		$\tilde{X}_{oh}$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.50***	1.34***	1.31***	1.16***
	(0.22)	(0.30)	(0.22)	(0.24)
First-stage residuals		0.21		0.20
		(0.39)		(0.31)
=1 if immigrant from anywhere			0.25***	0.25***
			(0.030)	(0.030)
=1 if immigrant from origin o			0.60***	0.60***
			(0.068)	(0.070)
N	1,461,130	1,461,130	1,461,130	1,461,130
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		20.2		19.5

Notes: The table presents regression results at the household-country level. Observations weighted using Nielsen 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.

order to deflate by the appropriate market size effect, we require parameter values for σ and θ. We assume a value for the CES elasticity parameter of σ =5. In the heterogeneous firms model used here, θ is simply the elasticity of trade with respect to variable costs, and we therefore follow Costinot and Rodríguez-Clare (2014) and calibrate θ =5.[35]

Columns 1 and 2 of Table 3 provide estimates of β with and without the use of the instrument from Burchardi et al. (2019). Our estimates broadly match, and are statistically indistinguishable from, those estimated using the general gravity model in Section 3.5. In our preferred IV specification, we estimate $\hat{\beta}$ =1.34.

The modest effect of correcting for county-level preferences likely reflects two off-setting sources of bias. First, we would expect the unadjusted estimates to be biased upwards, given that our initial structural gravity model omitted the spillover effect of immigrants via the

[35]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 θ =5 from Costinot and Rodríguez-Clare (2014) in order to match our larger value of σ =5.

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

	Dependent variable: Log Average UPC Price			
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	-0.034**	0.011	-0.057***	-0.024
	(0.014)	(0.032)	(0.016)	(0.047)
N	2,261,777	2,261,777	1,601,674	1,601,674
Barcode FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		24.3		23.7
Sample	All	All	>100 Counties	>100 Counties

Notes: The table presents regression results at the barcode-county level. Standard errors clustered at the barcode and country level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

market size channel. Second, we find evidence that immigrants have stronger preferences for all origin countries via ζ_1, suggesting a downward bias in estimates which do not control for the effect, for example, of Mexican immigrants increasing the market size in county c for imports from India. That is, identifying β from unadjusted cross-origin variation is potentially mis-specified when immigrants from origin o affect native household expenditure from all other origins o'. With our preferred estimate of $\hat{\beta}$ =1.34, we now discuss estimation results for the components of β.

5.3 Decomposing Spillovers into Trade Costs and Preferences

We start by leveraging the price information that we observe in the Nielsen Homescanner data in order to estimate equation (19). We show our estimates in Table 4, which represent estimates of $-\frac{\beta^{\tau}}{\theta}$. In columns 1 and 2, we use variation across all barcodes regardless of how regularly we observe them across counties. In column 3 and 4, we restrict the sample of barcodes to those which we observe in at least 100 counties in the Nielsen data. In columns 2 and 4 we instrument for the bilateral immigrant-population share using the leave-out push-pull instrumental variables defined in equation (A.1).

We find that the IV estimate using either sample is statistically indistinguishable from zero and very small in magnitude. The coefficient in column 2 implies that a 1 percentage

Table 5: Estimates of Household Gravity Equation using Number of Varieties

Dependent variable: Relative number of varieties from o				
	$\tilde{N}_{oh}/\mathcal{Z}_{oh}$		$\tilde{N}_{oh}$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.19***	1.13***	1.13***	1.21***
	(0.11)	(0.14)	(0.12)	(0.18)
First-stage residuals		0.087		-0.11
		(0.20)		(0.23)
=1 if immigrant from anywhere			0.17***	0.17***
			(0.017)	(0.017)
=1 if immigrant from origin o			0.56***	0.55***
			(0.046)	(0.045)
N	1,461,130	1,461,130	1,461,130	1,461,130
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		20.2		19.5

Notes: The table presents regression results at the household-country level. Observations weighted using Nielsen 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.

point increase in the share of the local population which is born in country o raises prices by 0.01 percent, and suggests that $\hat{\beta}^{\tau} = -0.06$. We therefore conclude that $\hat{\beta}^{\tau} \approx 0$.

We then estimate equation (18) but with the relative expenditure term $\tilde{X}_{oh}$ replaced with the relative variety count share $\tilde{N}_{oh}$ in order to recover the extensive margin elasticity of immigrants on import expenditure. Table 5 provides estimates of the extensive margin effect of immigrants on import expenditure under the same specifications as Table 3.

Solving for β^f and β^z using the elasticity estimates from column 2 of both Table 5 and Table 3, we recover β^f =1.07 and β^z =0.21. Since our estimate of β from Table 3 is 1.34, we therefore conclude that the primary *spillover* channel through which immigrants affect non-immigrant households is the fixed-cost channel, which accounts for approximately 80% of the overall spillover effect implied by β.

6 Counterfactual analysis

To quantify the contribution of immigrants to trade flows and native welfare, we finally use our estimated model to conduct a simple counterfactual exercise in which we remove all immigrants from the United States. We successively shut down individual channels to quantify the contribution of each in turn. Our counterfactual exclusively allows for partial-equilibrium adjustment to consumption choices, as the labor market effects of immigrants are outside the scope of our framework.

All derivations used in this section to calculate counterfactual changes in expenditure and welfare can be found in Appendix Section B.4.

6.1 Aggregate Effect of Immigrants on Imports & Native Welfare

To generate values which are representative of the United States as a whole, as well as meaningful counterfactual values for various U.S. cities, we leverage the American Community Survey (ACS). In particular, we use the results from estimating equation (17) on the Nielsen data to predict household-origin-specific expenditures for each ACS household. We further assume that each household spends $7,500 on grocery and personal care products covered by Nielsen. Finally, we use the crosswalks provided by Burchardi et al. (2019) to generate county-specific values based on the PUMAs in which households are located.

We show a summary of our results across different counterfactual scenarios in Table 6, with our baseline counterfactual scenario results appearing in the first row. Summing across households, we find that aggregate U.S. expenditures on imports of grocery and personal care items decreases by 26%, amounting to a fall of about $20 billion as shown in column 2 of Table 6. In our baseline counterfactual, we further find that removing all immigrants yields an average welfare loss from grocery and personal care consumption of 0.92%. This amounts to a welfare-equivalent fall of $69 per household.

To better understand which mechanisms drive our counterfactual results, we turn off each mechanism one at a time as shown in the remaining rows of Table 6. We start in row 2 by maintaining the same local population and income, that is, $S'_c = S_c$. The fall in aggregate import expenditures is a quarter of that in our baseline, falling by about $5 billion. The

Table 6: Counterfactual Results Summary

Counterfactual:	(1) Change (B$) aggregate import expenditures	(2) Change (%) import expenditure share	(3) Change (%) grocery welfare natives	(4) Change ($) welfare per HH
Baseline	-20.0	-6.81	0.920	69
$S_c' = S_c$	-5.48	-7.28	0.027	2
$S_c' = S_c$ and $z_{oh}' = z_{oh}$	-1.22	-1.62	0.027	2
$S_c' = S_c$ and $z_{oc}' = z_{oc}$	-4.98	-6.60	0.023	2
$S_c' = S_c$ and $f_{oc}' = f_{oc}$	-4.27	-5.65	0.004	0

Notes: This table shows the change in outcomes under various counterfactual scenario. The baseline scenario removes all immigrants from the United States. The other rows refer to counterfactuals in which each variable changes except for those listed in the left-hand column.

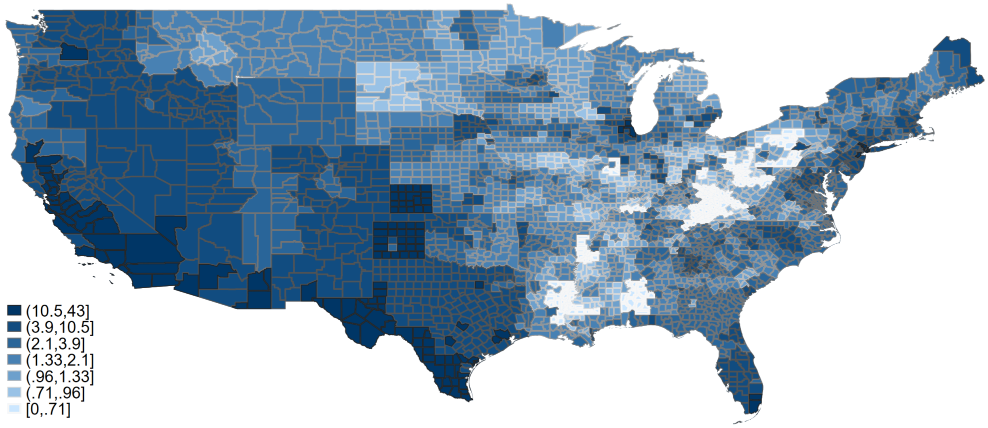
change in average import expenditure share is similar to that in our baseline. By contrast, the change in welfare is quite minimal relative to the baseline. This pattern continues for each of our other mechanisms, demonstrating that the market size – in terms of income – is the key channel through which immigrants affect native welfare. Still, the other channels, particularly that of fixed costs and preferences have a non-trivial impact on trade volumes even if their effect on welfare is small.

Our baseline estimates mask considerable heterogeneity both across origin countries and across geographies within the U.S. We first graphically depict variation across the U.S. in import volumes in Figure 6a and in dollar-equivalent utility in Figure 6b. In both maps, we see substantial spatial variation in the impact of immigrants on imports and welfare. However, the impact tends to concentrate in the Southwest, West Coast, and East Coast of the United States, as well as in big cities.

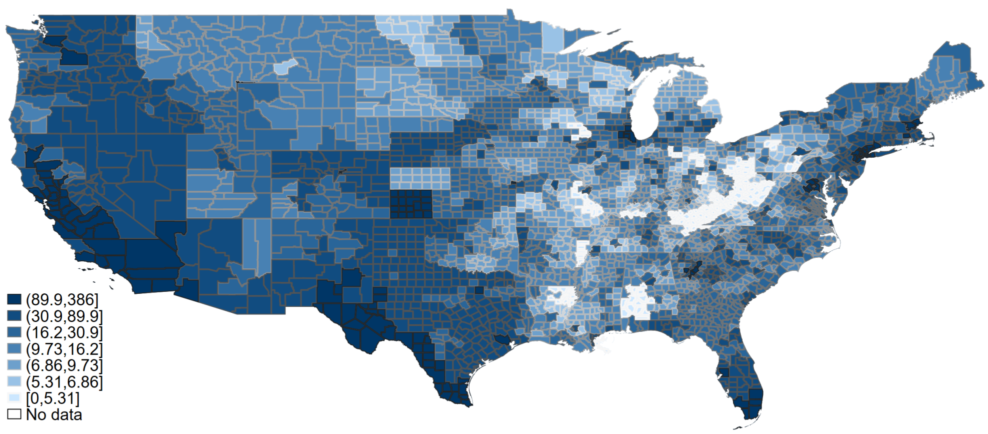
Among large counties, the most affected with respect to imports are El Paso, TX (34%); Los Angeles, CA (21%); Santa Clara, CA (21%); Kern, CA (18%); and Riverside, CA (18%). In terms of annual dollar-equivalent welfare effects for large counties, the most affected are Queens, NY ($385); Dade, FL ($356); Hudson, NJ ($308); Santa Clara, CA ($291); and Los Angeles, CA ($275). We show the counterfactual change in import volumes across origin countries in Appendix Figure C.1. We find that the expenditure share on Mexican imports

Figure 6: Spatial Distribution of the Effect of Removing Immigrants on Imports and Native Welfare

(a) Percent Change in Imports



(b) Dollar-Equivalent Change in Welfare

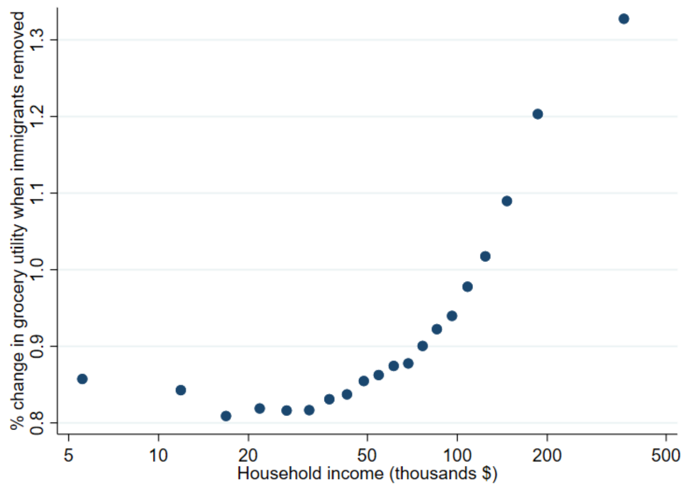


Notes: Figure (a) plots the percent decrease in the value of grocery and personal care imports when immigrants are removed following the procedure outlined in Appendix Section B.4. Figure (b) plots the dollar-equivalent welfare change for each U.S. county.

falls the most, by 12.02%. Mexico is followed by China, India, the Philippines and Germany with expenditure share decreases between 5.81% and 4.84%.

We use these counterfactual exercises to emphasize two key findings of this paper. First, while the welfare estimates presented here are surely a lower-bound in that they are only relevant for grocery products, they do shed light on the remarkable variation in the consumption gains from immigrant populations across cities and geographies within the U.S. These disparities may shed light on the strong polarization across U.S. geographies in atti-

Figure 7: Percent Change in Grocery Welfare from Removing all Immigrants Across the Income Distribution



Notes: The chart depicts a bin-scatter plot of household income (in log scale on the x-axis) relative to the percent grocery/personal care welfare gains from immigrants on the y-axis.

tudes towards immigrants and immigration policy, and represent the first estimates to date calculating this variation across U.S. geographies.

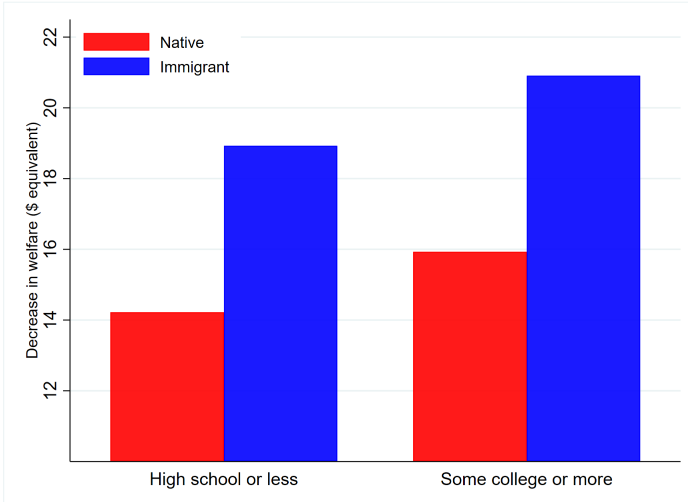
Second, we note that the change in import expenditure associated with removing immigrant effects is generally larger than the effect on *welfare*. This is due to two forces. First, we assume a Pareto distribution of firm productivities, as is standard in Melitz-Chaney models to obtain analytical tractability.

6.2 Welfare Impact Across the Income Distribution

We finally investigate the welfare consequences of immigrants across the income distribution. While the prior literature has emphasized the distributional consequences of immigrants in the labor market (e.g., Dustmann et al. 2013), we are the first to do so looking at the consumption side, enabled by our highly detailed household-level data and approach.

We depict welfare gains by income group graphically using a bin-scatter plot, shown in Figure 7. We find that up to about the median income of $57,000, the welfare gains from immigrants are flat, between 0.8% and 0.9% of grocery utility. The gains begin to rise quickly with income, with the highest income bin (those earning around $400,000) obtain

Figure 8: Effect of Variable Trade Cost Shock by Demographic Group



Notes: The chart depicts the counterfactual effect of a 10% increase in τ_{od} for all o, d pairs on natives (in red) and immigrants (in blue). Effects are grouped by education, with the effect on those with high school or less on the left and those with some college or more on the right.

a nearly 60% higher welfare gain from immigrants. These results suggest that the affluent may benefit disproportionately from immigrants through their consumption.

6.3 Differential Impact of Trade Shocks by Nativity, Education

In our second counterfactual, we consider the unequal effects of a trade shock. Inspired by the 2024 Trump Campaign's proposal to raise tariffs on all imported goods to 10%, we simulate the effects on households of a 10% rise in variable trade costs.

We show the results of such a trade shock stratified by education and nativity in Figure 8. Two key findings emerge. First, immigrants suffer much greater welfare losses than natives, regardless of how much education they have. On average, immigrants would lose the welfare-equivalent of about \$20 compared to about \$15 for natives. This result is driven by immigrants greater exposure to international imports in their consumption baskets, as documents in Section 2.

Second, more highly educated households lose out more from a trade shock. Households with some college or more experience a welfare-equivalent fall of about \$2. Our results demonstrate that a seemingly nativity-neutral policy, such as a universal tariff increase, will result to highly disparate impacts which depend on one's education and nativity.

7 Conclusion

This paper provides the first detailed decomposition of immigrant-induced import expenditure into a welfare-enhancing component and a *composition effect*. We find that the *composition effect* dominates and that within the welfare-enhancing component of immigrant-induced trade, wealthy households in large urban areas accrue the vast majority of the benefits.

A core contribution of this paper lies in separately identifying the various channels and mechanisms through which immigrants affect import expenditure. In estimating the spillover effects of immigrants on import expenditure of non-immigrant households, we make use of the leave-out push-pull instrumental variable introduced by Burchardi et al. (2019) to generate exogenous variation in origin-specific immigrant population shares across U.S. counties. By leveraging the structure inherent in the Chaney (2008) framework alongside detailed price data at the barcode level and a robust identification strategy, we are able to separately identify the effect of immigrants on price (and therefore variable trade costs), variety availability (fixed costs), and the intensive margin of demand (preferences). We are the first to provide direct evidence that among these spillover effects, the fixed cost reductions associated with immigrants are the dominant source of immigrant-induced trade.

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.

Alesina, Alberto and Marco Tabellini, "The political effects of immigration: Culture or economics?," *Journal of Economic Literature*, 2024, *62* (1), 5–46.

Amiti, Mary, Mi Dai, Robert C Feenstra, and John Romalis, "How did China's WTO entry affect US prices?," *Journal of International Economics*, 2020, *126*, 103339.

Ariu, Andrea, "Foreign workers, product quality, and trade: Evidence from a natural experiment," *Journal of International Economics*, 2022, *139*, 103686.

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.

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*, 2023, p. rdad098.

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.

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.

_ , Christian Dustmann, and Ian Preston, "Immigration, wages, and compositional amenities," *Journal of the European Economic Association*, 2012, *10* (1), 78–119.

Cardoso, Miguel and Ananth Ramanarayanan, "Immigrants and exports: Firm-level evidence from Canada," *Canadian Journal of Economics/Revue canadienne d'économique*, 2022, *55* (3), 1250–1293.

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.

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.

Cortes, Patricia, "The effect of low-skilled immigration on US prices: evidence from CPI data," *Journal of political Economy*, 2008, *116* (3), 381–422.

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.

Couture, Victor, "Valuing the consumption benefits of urban density," *University of California, Berkeley, Working Paper*, 2016.

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

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American Economic Review*, 2016, *106* (3), 479–524.

_ and Enrico Moretti, "Where is standard of living the highest? Local prices and the geography of consumption," Technical Report, National Bureau of Economic Research 2021.

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, "The impact of immigration: Why do studies reach such different results?," *Journal of Economic Perspectives*, 2016, *30* (4), 31–56.

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

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

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," *The 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," Vol. 1, Elsevier, 2015, pp. 913–1025.

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

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.

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

_ **and David E Weinstein**, "Goods prices and availability in cities," *Review of Economic Studies*, 2015, *82* (1), 258–296.

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

Hottman, Colin J, "Retail markups, misallocation, and store variety across US cities," 2021.

_ **, 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," *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4619541*, 2023.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–59.

Lach, Saul, "Immigration and prices," *Journal of political Economy*, 2007, *115* (4), 548–587.

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

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

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.

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," Unpublished manuscript 2023.

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

Ottaviano, Gianmarco I.P., Giovanni Peri, and Greg C. Wright, "Immigration, trade and productivity in services: Evidence from U.K. firms," *Journal of International Economics*, 2018, *112*, 88–108.

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/Revue canadienne d'économique*, 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," *The Review of Economic Studies*, 2021, *88* (1), 406–453.

Silva, JMC Santos 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.

Su, Yichen, "Measuring the value of urban consumption amenities: A time-use approach," *Journal of Urban Economics*, 2022, *132*, 103495.

United Nations, "Global Issues: International Migration," 2024.

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.

Zachariadis, Marios, "Immigration and international prices," *Journal of International Economics*, 2012, *87* (2), 298–311.

A 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 derived in Burchardi et al. (2019).

The immigration leave-out push-pull instrument interacts the arrival into 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

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

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

There may still exist threats to the exogeneity of the instrument as defined thus far. These threats include 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 $I^D_{o,c}$ flows when constructing the instrument for the county-country pair oc.

However, there might also be spatial correlation in confounding variables. For example, both Italian and French immigrants and goods may go to Chicago for the same reason: many flight connections. Leaving out Italy-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 Italy-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 use in our baseline estimation:

$$IV^D_{o,c} = I^D_{o,-r(c)} \times \frac{I^D_{-\mathcal{C}(o),c}}{I^D_{-\mathcal{C}(o)}} \tag{A.1}$$

Therefore, $I^D_{o,-r(c)}$ is the number of immigrants from o settling in the U.S. outside the Census region of county c in decade D, and $I^D_{-\mathcal{C}(o),c}/I^D_{-\mathcal{C}(o)}$ is the fraction of immigrants to the U.S. from outside of the continent of o who choose to settle in county c in decade D.

The identification assumption is that any confounding factors that make a given county more attractive for both immigration 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. A violation may occur if, say, immigrants skilled at importing goods from Italy tend to settle in Chicago and immigrants skilled in importing goods from South Korea settle in Miami in the same decade and for the same reason: a large number of flight connections. This violation is only quantitatively meaningful if Italians 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 (A.1) 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 extensive robustness checks for the instrument in the same setting and find that it holds up well.

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, as well as the population of county residents with ancestry from the origin country as of the 2010 Census.

Table A.1: First stage regression

	Dependent variable: Immigrants/Pop. 2010			
	(1)	(2)	(3)	(4)
$I^{1880}_{o,-r(d)} \times \frac{I^{1880}_{-c(o),d}}{I^{1880}_{-c(o)}}$	0.000063***	0.000057***	-0.00015	-0.00015
	(0.000021)	(0.000020)	(0.00015)	(0.00016)
$I^{1900}_{o,-r(d)} \times \frac{I^{1900}_{-c(o),d}}{I^{1900}_{-c(o)}}$	0.000033	0.000017	-0.00058	-0.00072
	(0.00013)	(0.00013)	(0.00072)	(0.00087)
$I^{1910}_{o,-r(d)} \times \frac{I^{1910}_{-c(o),d}}{I^{1910}_{-c(o)}}$	0.00026	0.00024	-0.00046	-0.00078
	(0.00020)	(0.00020)	(0.00048)	(0.00063)
$I^{1920}_{o,-r(d)} \times \frac{I^{1920}_{-c(o),d}}{I^{1920}_{-c(o)}}$	0.0018***	0.0018***	0.00056	0.00036
	(0.00025)	(0.00025)	(0.00070)	(0.00088)
$I^{1930}_{o,-r(d)} \times \frac{I^{1930}_{-c(o),d}}{I^{1930}_{-c(o)}}$	0.0016***	0.0016***	0.0029***	0.0031***
	(0.00017)	(0.00017)	(0.00058)	(0.00069)
$I^{1970}_{o,-r(d)} \times \frac{I^{1970}_{-c(o),d}}{I^{1970}_{-c(o)}}$	0.00086***	0.00084***	0.00084***	0.00092***
	(0.000081)	(0.000080)	(0.00023)	(0.00030)
$I^{1980}_{o,-r(d)} \times \frac{I^{1980}_{-c(o),d}}{I^{1980}_{-c(o)}}$	0.0032***	0.0032***	0.0042***	0.0047***
	(0.00028)	(0.00028)	(0.00058)	(0.00071)
$I^{1990}_{o,-r(d)} \times \frac{I^{1990}_{-c(o),d}}{I^{1990}_{-c(o)}}$	0.0023***	0.0022***	0.00093	0.0012
	(0.00025)	(0.00025)	(0.00075)	(0.00090)
$I^{2000}_{o,-r(d)} \times \frac{I^{2000}_{-c(o),d}}{I^{2000}_{-c(o)}}$	0.0015***	0.0015***	0.0015***	0.0016***
	(0.00019)	(0.00019)	(0.00029)	(0.00034)
Scores for component 1	-0.000015***	-0.000015***	-0.000015***	-0.000016***
	(0.0000026)	(0.0000026)	(0.0000033)	(0.0000039)
Scores for component 2	0.0000029*	0.0000029*	0.0000028	0.0000022
	(0.0000015)	(0.0000015)	(0.0000055)	(0.0000063)
Scores for component 3	0.000013**	0.000013**	0.000017*	0.000018
	(0.0000067)	(0.0000067)	(0.0000099)	(0.000011)
Scores for component 4	0.0000040	0.0000042	0.000026	0.000031
	(0.0000054)	(0.0000054)	(0.000020)	(0.000024)
Scores for component 5	-0.000084***	-0.000083***	-0.000095***	-0.00010***
	(0.000012)	(0.000012)	(0.000015)	(0.000017)
=1 if immigrant from anywhere		0.000051		
		(0.000075)		
=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	✓	✓	✓	✓
F-statistic	20.2	19.5	17.3	17.5

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

B Derivation Details

B.1 Deriving Adjusted ACR Welfare Formula

Begin by noting that given our assumption of constant expenditure, we can characterize the relationship between changes in domestic and import expenditure at the county level.

$$d\ln X_c = 0 \quad \implies \quad s_{us,c} d\ln X_{us,c} = -s_{m,c} d\ln X_{m,c} = -(1 - s_{us,c}) d\ln X_{m,c}$$

The derivation of this result begins with the following observation:

$$X_c = X_{us,c} + X_{m,c} \implies d\ln X_c = d\ln(X_{us,c} + X_{m,c})$$

Notice that for any variable x, $d\ln x = dx\frac{1}{x}$. We can therefore re-write the final expression above as:

$$d\ln X_c = \frac{d(X_{us,c} + X_{m,c})}{X_{us,c} + X_{m,c}} = \frac{d(X_{us,c} + X_{m,c})}{X_c} = \frac{dX_{us,c}}{X_c} + \frac{dX_{m,c}}{X_c}$$

Invoking again the transformation between $d\ln x$ and dx, we can write this expression as:

$$d\ln X_c = \frac{X_{us,c} d\ln X_{us,c}}{X_c} + \frac{X_{m,c} d\ln X_{m,c}}{X_c} = s_{us,c} d\ln X_{us,c} + s_{m,c} d\ln X_{m,c}$$

The last step is to assume that $d\ln X_c = 0$ and $s_{us,c} + s_{m,c} = 1$. Notice that the exact same steps can be used to derive the same expression for native households, although in this case we want to focus on welfare-relevant changes in domestic expenditure. Formally, this would be the component of changes in domestic expenditure which are due to changes in supply-side accessibility of imports associated with immigrants: $d\ln \phi^d_{mc}$.

The welfare relevant change in domestic expenditure for native households is therefore characterized by the following:

$$s_{us,n} d\ln X_{us,n} = -s_{m,n} d\ln X_{m,n} \frac{d\ln \phi^d_{mc}}{d\ln X_{m,n}} = -(1 - s_{us,n}) d\ln X_{m,n} \frac{d\ln \phi^d_{mc}}{d\ln X_{m,n}}$$

We can use the county-level aggregate expression to relate welfare-relevant changes in native household domestic expenditure to the county-level aggregate change in the following way:

$$d\ln X_{us,n} = d\ln X_{us,c}\left(\frac{1-s_{us,n}}{s_{us,n}}\right)\left(\frac{s_{us,c}}{1-s_{us,c}}\right)\left(\frac{d\ln X_{m,n}}{d\ln X_{m,c}}\right)\left(\frac{d\ln \phi^d_{mc}}{d\ln X_{m,n}}\right) \tag{B.1}$$

Deriving the Expenditure Share Adjustment Term. We begin by noting the following identities for county-level expenditure shares:

$$s_{us,c} = (1-I_c)s_{us,n} + I_c s_{us,f} \qquad s_{m,c} = (1-I_c)s_{m,n} + I_c s_{m,f}$$

We also make note that due to our assumption regarding the structure of preferences: $s_{m,f}/s_{m,n} = e^{\zeta}$. Given that expenditure shares must sum to one, it is trivially true that for immigrants $s_{us,f} = 1 - s_{m,f} = 1 - e^{\zeta}s_{m,n}$. For native households: $s_{us,n} = 1 - s_{m,n}$. Combining these two expressions, we derive:

$$e^{\zeta}(1-s_{us,n}) = 1 - s_{us,f} = 1 - \left[\frac{s_{us,c} - s_{us,n}(1-I_c)}{I_c}\right]$$

We can therefore solve for $s_{us,n}$ – the native domestic expenditure share – as a function of $s_{us,c}$, I_c, and ζ by re-arranging the previous expression:

$$s_{us,n} = \frac{s_{us,c} + I_c(e^{\zeta}-1)}{I_c(e^{\zeta}-1)+1}$$

Plugging this definition into the term of interest in our main equation of interest ($d\ln X_{us,n}$), we can derive our final expression:

$$\left(\frac{1-s_{us,n}}{s_{us,n}}\right)\left(\frac{s_{us,c}}{1-s_{us,c}}\right) = \frac{1}{\frac{I_c}{s_{us,c}}(e^{\zeta}-1)+1} \tag{B.2}$$

Deriving the Welfare-Relevant Component of Trade Shocks. Notice that the final two terms of the equation derived above reduce to the ratio: $\frac{d\ln\phi^d_{mc}}{d\ln X_{m,c}}$. We have an explicit expression for this ratio from the model:

$$\frac{d\ln \phi^d_{mc}}{d\ln X_{m,c}} = \frac{\beta^d}{\beta^d + \beta^z + \frac{e^{\zeta}-1}{I_c(e^{\zeta}-1)+1}} = \frac{\beta^d}{\beta + \frac{e^{\zeta}-1}{I_c(e^{\zeta}-1)+1}}$$

where the fraction in the denominator derives from evaluating:

$$\frac{d\ln[\sum \kappa_h z_{oh}]}{dI_c} = \frac{e^{\zeta} - 1}{I_c(e^{\zeta} - 1) + 1}$$

This derivation relies on the assumption that all households are identical except for their immigrant-derived preferences governed by ζ, so therefore $\kappa_h = \kappa$, and $J_h = J$ for all households. We also assume that with enough households, the average idiosyncratic component of preferences η^z_{oh} is equal to zero in expectation.

We can therefore express the average preference term $\sum \kappa_h z_{oh}$ as:

$$\sum_{h\in\Lambda_c} \kappa_h z_{oh} = \kappa e^{\delta J} \sum_{h\in\Lambda_c} e^{\zeta}$$

Since $\zeta = 0$ for any household that is not an immigrant, this expression further reduces to:

$$\sum_{h\in\Lambda_c} \kappa_h z_{oh} = \kappa e^{\delta J}[(1 - I_c) + I_c e^{\zeta}] = \kappa e^{\delta J}[I_c(e^{\zeta} - 1) + 1]$$

As a final step, we can derive the partial elasticity of this term with respect to I_c in order to show that:

$$\frac{d\ln[\sum \kappa_h z_{oh}]}{dI_c} = \frac{d[\sum \kappa_h z_{oh}]}{dI_c}\frac{1}{\sum \kappa_h z_{oh}} = \frac{\kappa e^{J}(e^{\zeta} - 1)}{\kappa e^{\delta J}[I_c(e^{\zeta} - 1) + 1]} = \frac{e^{\zeta} - 1}{I_c(e^{\zeta} - 1) + 1}$$

Notice that so long as $\zeta > 0$, this expression is positive and the composition effect of immigrants has a positive effect on county-level import expenditure. As $\zeta \uparrow$, this effect intensifies, as the derivative of the composition effect with respect to ζ is simply e^{ζ}.

B.2 Deriving Heterogeneous Firms Model Equations

Deriving equations (11) and (12). We start by deriving county-level expenditures on a variety supplied by a firm with productivity φ and imported from country o, $x_{oc}(\varphi)$.

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

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

Define

$$P_h \equiv \left(\sum_{o\in\mathcal{O}} (z_{oh})^\sigma \int_{\omega\in\Omega_{o,c(h)}} p_{o,c(h)}(\omega)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}} \tag{B.3}$$

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

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

Solving for q_{oh}, we get quantity and expenditure for a variety associated with productivity φ as

$$q_{oh}(\varphi) = (1-\mu_0)X_h z_{oh}^\sigma p_{o,c(h)}(\varphi)^{-\sigma} P_h^{\sigma-1} \tag{B.5}$$

$$x_{oh}(\varphi) = (1-\mu_0)X_h z_{oh}^\sigma (p_{o,c(h)}(\varphi)/P_h)^{1-\sigma} \tag{B.6}$$

From the firm's profit maximization problem, we obtain the price equation

$$p_{o,c(h)}(\varphi) = \frac{\sigma}{\sigma-1}\frac{w_0}{\varphi}\tau_{oc(h)} \tag{B.7}$$

Substituting this expression in the equation for $x_{oh}(\varphi)$, 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 expression for expenditure $x_{oc}(\varphi)$ given by (B.8):

$$x_{oc}(\varphi) = \lambda_1 (w_o\tau_{oc})^{1-\sigma}\varphi^{\sigma-1}\Big(\sum_{h'\in\Lambda_c} z_{oh'}^\sigma X_{h'} P_{h'}^{\sigma-1}\Big) \tag{B.8}$$

To derive the productivity cutoff term φ_{oc}^*, we start by deriving variable profits earned by a firm with productivity φ selling to market c from origin o:

$$\begin{aligned}\pi_{o,c}(\varphi) &\equiv \left(p_{o,c}(\varphi) - \frac{w_o}{\varphi}\tau_{o,c}\right)\sum_{h'\in c} q_{oh}(\omega(\varphi)) \\ &= (1-\mu_0)\left(\frac{w_o}{\varphi}\tau_{o,c}\right)^{1-\sigma}\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\sum_{h'\in c}\left(z_{oh'}\right)^{\sigma} W_{h'}\left(P_{h'}\right)^{\sigma-1} \\ &= \frac{1}{\sigma}x_{oc}(\varphi)\end{aligned}$$

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:

$$\pi_{oc}(\varphi) \geq f_{oc}$$

At the cutoff productivity, this holds with inequality, resulting in equation (B.9) for φ^*_{oc}, where $\lambda_2 = \frac{\sigma}{\sigma-1}\left(\frac{\sigma}{1-\mu_0}\right)^{\frac{1}{\sigma-1}}$:

$$\varphi^*_{oc} = \lambda_2 w_o \tau_{oc}\left(\frac{f_{oc}}{\sum\limits_{h'\in\Lambda_c} z^{\sigma}_{oh'}X_{h'}P^{\sigma-1}_{h'}}\right)^{\frac{1}{\sigma-1}} \tag{B.9}$$

Returning to equation (B.3) and replacing varieties ω with productivity φ (since firms with identical productivity charge identical prices), we get:

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

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

Plugging in our equilibrium price function $p_{o,c(h)}(\omega)$, we have

$$P_h = \frac{\sigma}{\sigma-1}\left(\sum_{o\in\mathcal{O}}\left(z_{oh}\right)^{\sigma}\left(w_o\tau_{o,c(h)}\right)^{1-\sigma} M_{o,c(h)}\int_0^{+\infty}\varphi^{\sigma-1}g_{o,c(h)}(\varphi)d\varphi\right)^{\frac{1}{1-\sigma}} \tag{B.10}$$

We derive the gravity equation using the expression for $x_{oh}(\varphi)$ as

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

Given the equilibrium price (B.7), 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

$$X_{oh} = \lambda_1 z_{oh}^{\sigma} X_h P_h^{\sigma-1} (w_o \tau_{o,c(h)})^{1-\sigma} M_o \frac{\theta}{\theta+1-\sigma} (\varphi_{o,c}^*)^{\sigma-\theta-1} \tag{B.11}$$

To obtain equation (11) from (B.10) and equation (12) from (B.11), we perform the following operations:

- substitute (B.9) for $\varphi_{o,c}^*$
- assume $M_o = \gamma Y_o$
- define $S_c = \sum_{h' \in \Lambda_c} X_{h'} P_{h'}^{\sigma-1}$ and $z_{oc} = \sum_{h' \in \Lambda_c} z_{oh'}^{\sigma} \frac{X_{h'} P_{h'}^{\sigma-1}}{S_c}$
- define $\lambda_3 \equiv \gamma \left(\frac{\sigma}{1-\mu_0}\right)^{\frac{\sigma-\theta-1}{\sigma-1}} \left(\frac{\sigma}{\sigma-1}\right)^{\sigma-\theta-1} \frac{\theta}{\theta+1-\sigma}$
- define $\lambda_4 \equiv \gamma (1-\mu_0)^{\frac{\theta}{\sigma-1}} \sigma^{\frac{\sigma-\theta-1}{\sigma-1}} \left(\frac{\sigma}{\sigma-1}\right)^{-\theta} \frac{\theta}{\theta+1-\sigma}$

B.3 Deriving Expenditure and Extensive Margin Immigrant Elasticity

In this section we fully differentiate Equation 14 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.

We begin by assuming that $\beta^\tau \approx 0$, which implies that immigrants do not affect variables trade costs. This assumption derives from the results discussed in Table 4.

We follow Chaney (2008) and fully differentiate $\tilde{X}_{oh}$ from Equation 14 into terms associated with fixed costs f_{oc}, county-level preferences z_{oc}, and household-level preferences z_{oh}. By applying Leibnitz Rule, we can separate each term into both an intensive margin and extensive margin, and the full differentiation is given by the following:

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

In all cases, 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 aggregate expenditure on U.S. goods by h - is given by the following:

$$\tilde{x}_{oh}(\varphi) = (\tilde{\omega}_o \tau_{oc})^{1-\sigma} z_{oh} \varphi^{\sigma-1} \Big(\int_{\bar{\varphi}_{us,c}}^{+\infty} \varphi^{\sigma-1} dG(\varphi) \Big)^{-1} \tag{B.13}$$

whereas the productivity cut-off expression is the following:

$$\bar{\varphi}_{oc} = \lambda_2 \omega_o \tau_{oc} \left(\frac{f_{oc}}{S_c z_{oc}} \right)^{\frac{1}{\sigma-1}} \tag{B.14}$$

It is clear from inspecting Equation B.13 and Equation B.14 that f_{oc} and z_{oc} only affect $\bar{\phi}_{oc}$, and therefore the 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: $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial \tilde{f}_{oc}} = 0$; $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oc}} = 0$; and $\frac{\partial \bar{\phi}_{oc}}{\partial z_{oh}} = 0$.

We therefore have an expression for the aggregate semi-elasticity of import expenditure with respect to immigrants and an expression for the extensive margin semi-elasticity of import expenditure with respect to immigrants. These expressions are given by, respectively:

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

$$\frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} \tag{B.16}$$

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:

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

$$\begin{aligned} \beta^N &= \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial I_{oc}} \\ &= \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}} \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}} \end{aligned} \tag{B.18}$$

While these expressions seem daunting, they are trivial to evaluate given the definition of $\tilde{X}_{oh}$ provided in Equation 14 and the definition of $\mathcal{Z}_{oh}$ provided earlier.

Specifically, we can solve these three expression in three parts:

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

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

$$\beta^N = \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1} - 1\right)\beta^z \tag{B.20}$$

B.4 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.

To obtain an expression for the change in household-origin import expenditures, we start with equation (12), and express the ratio of counterfactual to observed household-level imports from o as

$$\hat{X}_{oh} = \hat{P}_h^{\sigma-1} \hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} \left(\hat{z}_{o,c(h)} \hat{S}_{c(h)}\right)^{\frac{\theta}{\sigma-1}-1} \hat{z}_{oh} \tag{B.21}$$

where changes in household imports by origin depend on the change in overall price level, changes in fixed costs with the origin, changes in local market demand for the origin's products, and changes in average household-level preferences. When o is the United States, equation (B.21) reduces to

$$\hat{X}_{us,h} = \hat{P}_h^{\sigma-1} \hat{S}_{c(h)}^{\frac{\theta}{\sigma-1}-1} \tag{B.22}$$

Hence we use equations (B.21) and (B.22) as well as the fact that $\hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} = e^{-\hat{\beta}^f I_{o,c(h)}}$ 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:

$$\frac{X'_{oh}}{X'_{us,h}} = \frac{X_{oh}}{X_{us,h}} \left(e^{-I_{o,c(h)}(\hat{\beta}^f + \hat{\beta}^z)}\right) z_{o,c(h)}^{\left(\frac{\theta}{\sigma-1}-1\right)} \tag{B.23}$$

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

Lastly, while it is simple to show that under CES preferences, the change in welfare is given by the change in the price index, we must be careful to distinguish between welfare-relevant components of the price index and components of the price index which are not welfare relevant. To begin with, changes in utility are generally given by:

$$\hat{U}_h = \hat{P}_h^{\mu_0 - 1} \tag{B.24}$$

Notice, however, that P_h includes changes in preferences associated with β^z, which are not welfare-relevant. We therefore compute the change in the welfare-relevant price index as follows:

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

in which we purge the price index relevant for firms, and therefore containing β^z, in order to isolate the welfare-relevant component of changes in the price index.

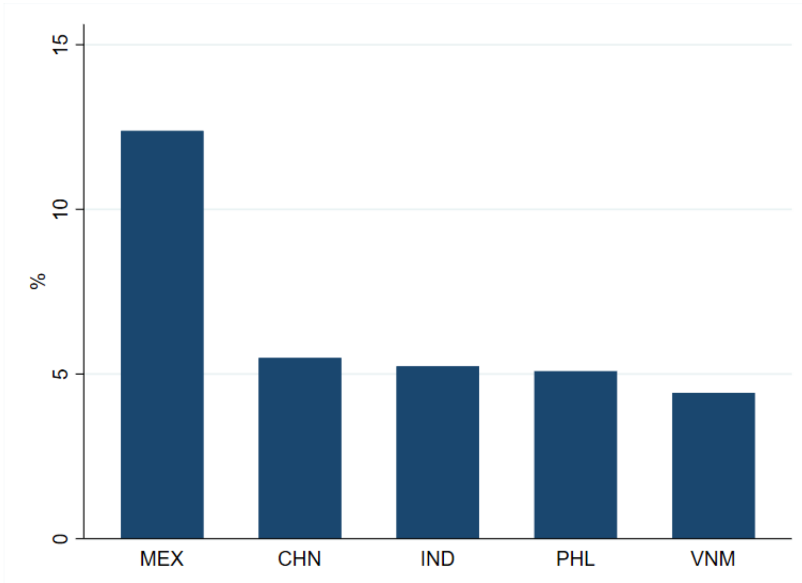
We further assume that immigrant and native households spend the same amount on grocery and personal care produces, which implies that

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

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

C Additional Tables and Charts

Figure C.1: Most Impacted Origins under Baseline Counterfactual



Notes: This chart shows the percent increase in imports by origin attributable to the presence of immigrants. We compute imports under our counterfactual scenario as discussed in Appendix Section B.4.

Table C.1: Effect of Household Characteristics on Import expenditure

	Dep. var.: Rel. expenditure share on goods from o (1)	
Immigrant from o	0.64***	(0.069)
Immigrant from anywhere	0.23***	(0.029)
Income: 10k-30k	0.031	(0.042)
Income: 30k-50k	0.011	(0.040)
Income: 50k-70k	0.074*	(0.042)
Income: 70k-100k	0.063	(0.042)
Income: >100k	0.18***	(0.043)
HH size: 2	-0.073**	(0.029)
HH size: 3	-0.10***	(0.033)
HH size: 4	-0.19***	(0.041)
HH size: >4	-0.19**	(0.085)
Children: 6-12 y.o.	-0.087	(0.088)
Children: 13-17 y.o.	-0.10	(0.092)
Children: <6 + 6-12	-0.11	(0.10)
Children: <6 + 13-17	-0.051	(0.16)
Children: 6-12 + 13-17	-0.056	(0.096)
Children: All Age Groups	-0.26**	(0.12)
No Children	-0.070	(0.084)
Some College	0.064***	(0.023)
College Degree	0.097***	(0.024)
Postgraduate Degree	0.18***	(0.027)
Widowed	0.0043	(0.036)
Divorced/Separated	-0.0026	(0.034)
Single	-0.021	(0.034)
Black	0.058**	(0.024)
Asian	0.075**	(0.035)
Other	0.097**	(0.040)
Hispanic	-0.036	(0.034)
Age	-0.018	(0.032)
Age^2	0.00025	(0.00054)
Age^3	-0.00000087	(0.0000029)
N	868,261	
County-origin FE	✓	

Notes: The table presents regression results at the household-country level. Observations weighted using Nielsen 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.2: Replication of BCH results

	Dependent variable: Log imports			
	Heckmann correction		PPML + control fct	
	(1)	(2)	(3)	(4)
log_anc	-0.033		0.53***	
	(0.058)		(0.054)	
stock_pop2010		-2.67		15.2***
		(2.81)		(1.96)
N	2,922	2,922	3,626	3,626
State FE	✓	✓	✓	✓

Notes: The table presents regression results at the state-origin level. Standard errors clustered at the state level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.3: Replication of BCH results, relative expenditure share

	Dependent variable: Relative expenditure share on imports			
	Heckmann correction		PPML + control fct	
	(1)	(2)	(3)	(4)
log_anc	-0.086		0.055***	
	(0.057)		(0.0097)	
stock_pop2010		-4.31		2.46***
		(3.27)		(0.44)
N	2,922	2,922	3,626	3,626
State FE	✓	✓	✓	✓

Notes: The table presents regression results at the state-origin level. Standard errors clustered at the state level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.4: Replication of BCH results

	Dependent variable: Relative expenditure share on imports					
	Dep. var. at state level		Dep. var. at county level		Dep. var. at HH level	
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrants/Pop. 2010	2.53***	1.46**	2.74***	4.09**	1.50***	1.34***
	(0.28)	(0.67)	(0.37)	(1.91)	(0.22)	(0.30)
First-stage residuals		1.37*		-1.74		0.21
		(0.71)		(1.97)		(0.39)
N	1,461,130	1,461,130	1,461,130	1,461,130	1,461,130	1,461,130

Notes: The table presents regression results at the state-origin level. Standard errors clustered at the state level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.5: Replication of BCH results, unadjusted dep var

	Dependent variable: Relative expenditure share on imports					
	Dep. var. at state level		Dep. var. at county level		Dep. var. at HH level	
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrants/Pop. 2010	2.86***	2.20***	2.69***	4.10**	1.31***	1.16***
	(0.26)	(0.71)	(0.34)	(1.78)	(0.22)	(0.24)
First-stage residuals		0.85		-1.81		0.20
		(0.74)		(1.84)		(0.31)
N	1,461,130	1,461,130	1,461,130	1,461,130	1,461,130	1,461,130

Notes: The table presents regression results at the state-origin level. Standard errors clustered at the state level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.6: Gravity regressions with adjusted weights

	Dependent variable: Exp. share on goods from o relative to US (1)	(2)
Immigrants/Pop. 2010	1.40***	1.59***
	(0.25)	(0.28)
First-stage residuals		-0.26
		(0.29)
=1 if immigrant from anywhere	0.26***	0.26***
	(0.037)	(0.037)
=1 if immigrant from origin o	0.60***	0.60***
	(0.086)	(0.087)
N	1,461,130	1,461,130
Country FE	✓	✓
Household controls	✓	✓
Distance & latitude difference	✓	✓
1st-stage F-statistic		20.2

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

Table C.7: Gravity regressions with additional homophily terms

	Dependent variable: Exp. share on goods from o relative to US (1)	(2)
Immigrants/Pop. 2010	1.30***	1.16***
	(0.21)	(0.24)
First-stage residuals		0.19
		(0.31)
=1 if immigrant from anywhere	0.23***	0.23***
	(0.033)	(0.033)
=1 if immigrant from origin o	0.53***	0.54***
	(0.079)	(0.080)
=1 if immig. from continent of o	0.073	0.073
	(0.061)	(0.061)
=1 if hispanic and o in Latin America	0.060	0.060
	(0.059)	(0.059)
N	1,461,130	1,461,130
Country FE	✓	✓
Household controls	✓	✓
Distance & latitude difference	✓	✓
1st-stage F-statistic		20.2

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