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

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

September 3, 2024

Abstract

Do immigrants make goods from their origin country more accessible to their non-immigrant neighbors? We augment U.S. grocery scanner data to include the origin country of both households and products, thereby enabling the first direct estimate of how local immigrant presence affects import penetration. Using a quantitative model of trade, we show that immigrants increase the grocery import expenditure share by 8%. Three quarters of this effect is attributable to immigrants' own disproportionate preferences for imported goods. Immigrants therefore raise import expenditures primarily through their own consumption, with muted benefits for their non-immigrant neighbors. The benefits that do accrue to natives are concentrated within high-income and urban households.

JEL Categories: F22, J31, J61, R11.

*We are grateful to Samuel Bowles, Daniel Chiquiar, Wookun Kim, Munseob Lee, Peter Morrow, and Jonathan Vogel for helpful comments, as well as participants at various conferences and seminars. We thank Ilaria Malisan and Andrea Montalbano for research assistance. The analysis in this paper is our own, calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are our own and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All errors are our own.

[†]Collegio Carlo Alberto, email: brett.mccully@carloalberto.org

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

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

1 Introduction

Immigration policy sits at the center of contentious debates in receiving countries. While the content of these debates tends to focus on how immigrants affect the nominal wages of non-immigrant households, quantifying the welfare effects of immigrants on natives also requires an understanding of how immigrants affect local prices and product availability.[1] This paper uses novel data to study whether immigrants in the U.S. promote international trade by increasing the availability of imported consumption goods, and the extent to which this benefits their non-immigrant neighbors.

Immigration may raise the local expenditure share on imported products in two ways. First, immigrants themselves may exhibit stronger preferences for imports, thereby increasing local import expenditure with no effect on native households.[2] Second, the presence of immigrants may increase native import expenditure via spillover effects such as lowering the cost of importing foreign goods, increasing market size, or shifting native preferences. Hence, quantifying the consumption welfare implications of immigration on natives due to increased imports requires data which can separately measure local import expenditure by household nativity, yet such data are exceedingly rare.

We overcome this challenge by augmenting U.S. grocery scanner data to include the origin country of both households and products. This dataset allows for an empirical and quantitative study of the effects of immigrants on both county-level import volumes and native household consumption welfare. We introduce immigrants and the comprehensive set of mechanisms described above—changing local preference composition, reducing trade costs, and increasing market size—to a standard Melitz (2003)-Chaney (2008) quantitative model of trade and quantify the contribution of each mechanism. Removing the specific trade-creating effects of immigrants increases the U.S. grocery import expenditure share by 8%, which is roughly equivalent to the effect of reducing prevailing tariffs applied to grocery goods by half.

[1] For a recent review of the academic literature on immigrants' effect on wages, see Dustmann et al. (2016). We use the terms "natives" and "non-immigrants" interchangeably.

[2] That immigrants may promote trade via their preferences is discussed by Felbermayr et al. (2015) and is also suggested in the seminal works by Gould (1994) and Head and Ries (1998). To our knowledge, there are no direct estimates of the relative import demand of immigrants compared to natives.

Three quarters of the immigrant-import elasticity is driven by immigrant preferences. Immigrants exhibit both a home-bias in preferences for goods imported from their origin country and stronger preferences than natives for all imported goods, regardless of origin. We are the first to directly observe and quantify these two preference channels. Due to this observed preference heterogeneity, a naive county-level application of the seminal welfare formula derived by Arkolakis et al. (2012) over-estimates the welfare benefits to native households of immigrant-induced imports by a factor of four. The relative strength of immigrant import demand also serves to increase the exposure of immigrant households to all trade shocks, immigrant-induced or otherwise. A counterfactual increase in variable trade costs applied to all imported grocery goods decreases the welfare of immigrants in the U.S. by 29% more than that of natives, with college-educated immigrants facing welfare costs that are over 50% greater than native households without a college degree.[3]

In an additional counterfactual exercise, we remove both the specific trade-creating effects of immigrants and their associated expenditure, thus highlighting the market size benefits of immigrants. Aggregate import volumes decrease by 26% and native welfare decreases by 1%. These welfare gains for natives are highly concentrated among high-income and urban households, which is equally attributable both to positive location sorting between high-income urban natives and immigrants, and to a positive elasticity of import demand with respect to income.[4]

The linchpin for our analysis is a novel dataset of household grocery purchases, in which we observe the country of origin of both households and the products they purchase. The data has three key 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 both product and household origin countries within household-level scanner data, which we describe in detail in Section 2.

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

[4]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.

We estimate a general 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 as well as the effects of specific household characteristics, such as immigrant status, on import demand. Our estimating equation nests a wide range of standard micro-foundations in the trade literature (Head and Mayer 2014) and we make use of the instrumental variables from Burchardi et al. (2019) to generate exogenous variation in origin-specific immigrant population shares across U.S. counties.[5] Immigrants spend 26% more on imports from all origins than their within-county, non-immigrant neighbours, and 132% more on imports specifically from their origin country. Spillovers are also significant, as a 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 1.15%.

Since immigrants may alter the preferences of natives—what we term *preference diffusion*—it is not clear whether a positive spillover effect necessarily implies a welfare gain for native households. We therefore develop a model of trade in Section 4 which allows us to separately identify the various channels by which immigrants may increase import expenditures in their county of residence. In particular, we extend the heterogeneous-firms model of Melitz (2003) and Chaney (2008) to allow for immigrant effects via (i) a shift in the composition of local preferences, (ii) a shift in the import preferences of native households, (iii) an increase in market size due to stronger preferences for foreign goods, and (iv) a reduction in variable and/or fixed costs of trade.

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. That is, we estimate the elasticity of barcode-specific prices to immigrant population shares, the barcode count elasticity of import expenditure to immigrants, and the aggregate elasticity of import expenditure to immigrants. Collectively, these estimates pin down the strength of the variable cost reduction channel, fixed cost reduction channel, and preference diffusion channel.

Our estimates highlight the gap between the trade-creating effects of immigrants and

[5]Our identifying variation leverages decades of immigration data and cross-sectional variation, rather than high-frequency short-run shocks, such as exchange rates. We therefore interpret our results as pertaining to the long-run effects of immigrants.

welfare changes associated with this trade: channels that are welfare-neutral from the perspective of native households constitute almost three quarters of the aggregate immigrant-import elasticity. Specifically, 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, the market size channel accounts for 11% of the aggregate immigrant-import elasticity.[7] We discuss model estimation results in Section 5.

The aggregate counterfactual results described earlier mask substantial variation across origin countries, geographies and income groups. Unsurprisingly, Mexican grocery imports to the U.S. are the most affected by our primary counterfactual exercise, exhibiting a 13% decrease in response to removing immigrant effects on trade. The benefit of all immigrant effects, including the increase in market size due to immigrants' expenditure, is more than five times higher in high-immigration counties, such as Queens, NY, than in the average U.S. county. Across the income distribution of native households, top earners obtain a 60% higher welfare gain than households at, or below, the median income. Low-income, less-educated, and native-born U.S. households face the lowest consumer costs associated with policies which increase barriers to either immigration or imports. We discuss all counterfactual results in Section 6.

The findings in this paper provide the first empirical validation of the concern voiced in Felbermayr et al. (2015): caution is needed when interpreting immigrant-induced changes in import penetration as akin to changes in welfare for native households.

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

[6]This finding is consistent with the assumption made by Peri and Requena-Silvente (2010), but we are the first to provide a direct empirical test of this assumption.

[7]We calibrate the Pareto shape parameter from which firms draw productivity and the demand elasticity of substitution to values in the literature, rather than estimating these parameters directly.

variety.[8] 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 uniquely detailed household-level data on consumption expenditures by product origin to study how immigrants affect local import penetration, allowing us to quantify the contribution of a comprehensive set of mechanisms and resulting welfare impacts across heterogeneous 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; Steingress 2018; Burchardi et al. 2019) and, more recently, firm-level data (Ottaviano et al. 2018; Cardoso and Ramanarayanan 2022; Ariu 2022).

The closest paper to ours is Bonadio (forthcoming), who allows for a broad set of mechanisms by which immigrants may affect trade, including home-biased preferences, reduced trade costs, and increased market size. Our study advances beyond Bonadio (forthcoming) in four key ways: (i) we observe and identify home-biased immigrant preferences directly, rather than recovering these preferences indirectly via aggregate data and strong functional-form assumptions; (ii) we are the first to introduce an immigrant preference for imports from all origin countries; (iii) we explore the distributional effects of immigration on consumption across heterogeneous native households; and (iv) we separately identify immigrants' effect on variable trade costs, fixed trade costs, and native preferences.

Lastly, our paper contributes to the literature on spatial variation in the local cost of living (Diamond 2016; Handbury and Weinstein 2015), local product variety (Couture 2016; Hottman 2021), and variation in the cost of living between skill groups within cities (Davis et al. 2019; Diamond and Moretti 2021; Handbury 2021; Su 2022).[9] By estimating consumer heterogeneity in exposure to trade shocks—immigrant-induced or otherwise—we contribute

[8] Two prior papers have explored this margin—Mazzolari and Neumark (2012) and Chen and Jacks (2012)—but lack the data and exogenous variation to causally identify potential mechanisms; moreover, they do not quantify the effect on native welfare. Iranzo and Peri (2009), Di Giovanni et al. (2015), and Aubry et al. (2016) study the aggregate variety effects of immigration but with a focus on migrants expanding production in high-productivity locations.

[9] Existing work by Lach (2007), Cortes (2008), and Zachariadis (2012) finds immigrant effects on region-level price indices but cannot quantify (i) the mechanisms that drive these estimated effects, and (ii) heterogeneity in these effects both across and within immigrants and native households.

to a growing literature studying the heterogeneous consumer outcomes associated with trade shocks.[10] This paper is the first to document the extent to which import expenditure is particularly concentrated in immigrant households, thus increasing the exposure of these households to trade shocks.

2 Data and Stylized Facts

2.1 Expenditure on tradable nondurable products

We use two datasets to link household characteristics—including country of birth—to grocery import expenditures: the NielsenIQ household panel scanner dataset and barcode country-of-origin data from Label Insight Inc.

NielsenIQ Household Panel Scanner Data: These data consist of a panel covering approximately 90,000 U.S. households and all grocery purchases at the barcode level. Detailed household demographic information and 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 to a single cross-section at the household level.

For a subset of NielsenIQ households, we observe their country of birth. In 2008, NielsenIQ distributed the "Tell Me More About You" survey, which included questions about respondents' birth place, and 19,700 (40%) of these households remain in the 2014-2016 sample used here.[11] Households may have mixed nativity, and we use the following allocation rules when assigning households to an origin country. When only one member of the household was born abroad and all others were born in the U.S., we assign the household to the country of the immigrant member. When a household has more than one foreign-born member, we assign the household to the larger country of origin as measured by total population.

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

[10] See 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).

[11] See Bronnenberg et al. (2012) for more details regarding this survey. Immigrants have a marginally lower rate of survival to our final dataset at 39%, although this difference is not significant at the 90% level. For households with purchase records in 2014-2016, 23% can be linked to the 2008 nativity survey. This rate increases marginally to 24% for the top half of households by income.

and organizing information found on the labels of consumer packaged goods.[12] Label Insight uses a computer vision algorithm to extract text from the packaging for thousands of barcodes sold across major retail chains in the U.S. Since imported goods in the U.S. are required to contain some statement equivalent to "Made in ..." on their labels, the algorithm incidentally recovers the origin country for each collected barcode.[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 therefore make use of data on the origin country for over 600,000 barcodes in these grocery product categories. Given the universality of barcodes, these data can be directly merged with the household-level purchase records from NielsenIQ. Figure C.1 documents the distribution of production origin countries in the merged scanner data with barcode origins. As expected, Mexico and Canada constitute just over half of all import expenditure, with Thailand, China, and Italy rounding out the top five product origins. There are 73 other origins with positive imports, which constitute about a third of import expenditure. The average import expenditure share is approximately 8%.[14]

Household-Level Coverage of Import Expenditure: Our final merged dataset covers \$764 billion USD of expenditure and is at the household-import origin level of aggregation. When compared to estimates from the BEA Consumer Expenditure Survey (CEX), the grocery categories studied in this paper account for approximately a third of all expenditure on tradeables, with this share increasing to almost half if one excludes passenger vehicles and energy products. Within groceries, the merged household-level expenditure data used here exhibit an average expenditure per household-year of \$2,200 USD, which is around 60% of the predicted expenditure on groceries in the CEX.

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

[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".

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

2.2 Immigration Data

We use the decadal Censuses from 1880-1930 and 1970-2000, as well as the pooled 2006–2010 sample of the American Community Survey (ACS) to obtain population counts of immigrants by origin.[15] We compute immigrant inflow measures for each available decade between 1880 and 2000. These inflows are used in the first stage of our instrumental variables strategy outlined in Section 3.4. Our main explanatory variable is the share of the local population born in country o. We provide additional details on data construction in Appendix A.1.

2.3 Stylized Facts

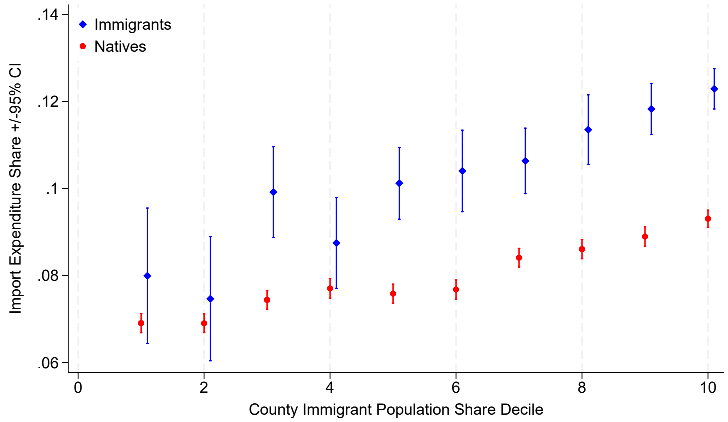
The combined NielsenIQ datasets described above constitute the first direct measurement of import expenditure by country of birth. We leverage this novel feature to demonstrate three stylized facts which characterize import consumption heterogeneity by household origin.

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

Fact 2: Aggregate import expenditure shares are 38% greater for immigrant households compared to non-immigrant households. In addition to the positive correlation between import expenditure and immigrant population shares, Figure 1 provides evidence of a stark contrast in import expenditure between immigrants and native households. Within each county decile, immigrant households exhibit significantly higher import expenditure than native households. The estimates provided in Figure 1 suggest that immigrants exhibit stronger import demand than native households even within the same county.

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

Figure 1: Immigrants, Natives, and Import Expenditure



Notes: The figure plots estimates from a linear regression at the household level in which household import expenditure shares are regressed on fixed effects at the immigrant-by-county-decile level. Counties are placed into deciles based on the immigrant population share of that county, and households are grouped into two categories: immigrants (blue) or native (red). 95% confidence intervals are provided, and all observations are weighted by the NielsenIQ projection factors.

We quantify this difference in mean import expenditure by regressing the household-level import expenditure on a dummy for whether a household is an immigrant household. Table C.1 provides the estimates from this exercise, and we find an unconditional mean difference in import expenditure between immigrants and natives of +3.1 percentage points. When compared to the average import expenditure share of non-immigrant households, this estimate represents a 38% differential.[16] Columns 3 to 6 of Table C.1 display results with additional controls in order to mitigate the potential bias associated with immigrants sorting into high-import counties or differing in other observable characteristics from natives. Even when county-level fixed effects and a suite of socio-economic household characteristics are included, the estimated differential between immigrants and natives in their average import expenditure remains highly significant and constitutes a gap of +2.8 percentage points.[17]

Fact 3: Immigrants spend over twice as much as natives on goods from their origin country. A key advantage of the data used here is that we are the first to simultaneously observe the specific origin country of households and products. This allows us to test directly whether the differential import expenditure associated with immigrants is driven by expenditure on all import origins, or goods specifically from that household's origin country.

[16]Figure C.2 provides a raw histogram of import expenditure shares for native and immigrant households.
[17]We add controls for income bins, household size, marital status, and head of household age and gender.

We turn to this analysis as our final stylized fact.

For each origin country o, we calculate the share of expenditures on goods from o by both households from o and natives. Figure 2 depicts this relationship by comparing 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. The 45-degree line in red plots where natives and immigrants from o would exhibit identical expenditure on imports from o.

We find that almost all origins lie above the 45-degree line, suggesting immigrants do in fact exhibit disproportionately stronger demand for imports from their specific country of origin. 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 median 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.[18] To our knowledge, this paper is the first to provide direct evidence that the preference persistence documented in Logan and Rhode (2010) and Atkin (2016) exists for U.S. immigrants with respect to demand for goods imported from their origin country.

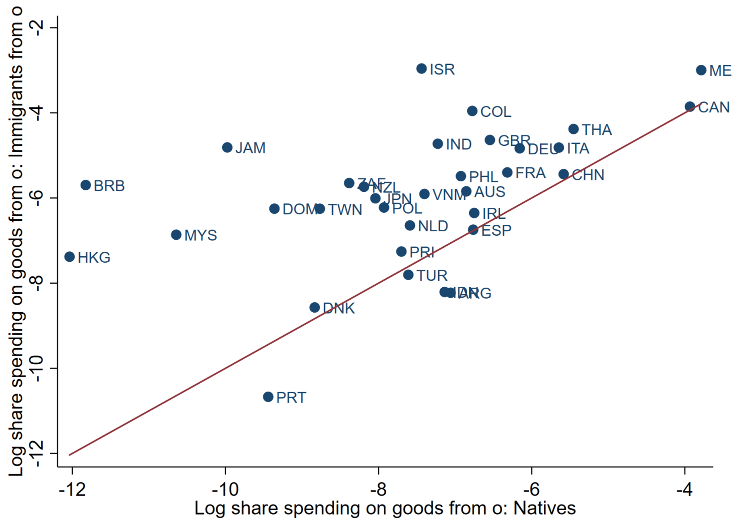
The preceding three stylized facts suggest that immigrants and natives have different demand for tradables, i.e., that there exists what we call a preference composition effect. In our subsequent empirical and theoretical exercises we will quantify the importance of this channel as well as any spillovers from immigrants onto native consumption.

3 Immigrants and Imports: General Gravity Model

In this section, we estimate a general gravity model allowing immigrants to reduce trade barriers and to affect local preferences. We show how, using household-level data, we can separately identify the direct effect of immigrants' home-biased preferences—what we term the preference composition effect—from the spillover effect of immigrants on native consumption within a general framework which nests a broad class of trade models.

[18] 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 2: 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 NielsenIQ Household Panel 2014-2016. NielsenIQ projection factor weights used to construct expenditure shares.

3.1 Immigrants and Gravity

We begin by considering a general gravity model, as defined by Head and Mayer (2014). In this model, import expenditure in county c on goods from origin country o is

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

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.[19] S_c is a measure of real demand in county c, given by $S_c = X_c/\Phi_c$, where X_c is aggregate grocery 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 and preference similarity. It is important to note that the structure outlined above nests the standard quantitative trade models, including those of Eaton and Kortum (2002), Krugman (1980), and Melitz (2003)-Chaney (2008).

[19]In practice, the country-specific term α_o also captures any features specific to the bilateral relationship between the U.S. and country o, such as shared language or culture, as well as bilateral trade policy.

The conventional interpretation of ϕ_{oc} is that it captures bilateral trade costs.[20] We generalize the standard gravity model by allowing for a bilateral affinity term, whereby consumers in c may exhibit preferences for the goods from specific origin countries. Formally, we decompose the bilateral term ϕ_{oc} into two 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.[21] We then re-write our general 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 defined relative to the analogous term for U.S. 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/or increase bilateral affinity. Instead, we make use of household-level import expenditure data, allowing us to separately identify the effects of immigrants on trade costs and preferences.

3.2 Preference Heterogeneity and Household-Level Gravity

Each household h living in c faces the same bilateral trade costs ϕ^b_{oc}, but households differ in their total 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 micro-foundation regarding the household 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. Household h's real expenditures are thus X_h/Φ_h, where Φ_h denotes h's price

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

index. This yields the household-level gravity equation as

$$X_{oh} = \alpha_o \frac{X_h}{\Phi_h} \phi^b_{oc} z_{oh}. \tag{2}$$

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

$$X_{oc} = \sum_{h \in \Lambda_c} X_{oh} = \alpha_o \phi^b_{oc} \sum_{h \in \Lambda_c} \frac{X_h}{\Phi_h} z_{oh} = \alpha_o S_c \phi^b_{oc} \underbrace{\sum_{h \in \Lambda_c} \kappa_h z_{oh}}_{\phi^z_{oc}},$$

where Λ_c is the set of households living in c, $S_c = \sum_{h \in \Lambda_c} X_h / \Phi_h$ is real aggregate expenditure, and κ_h household-specific real expenditure weights.[22] The bilateral affinity term ϕ^z_{oc} is therefore the expenditure-weighted average of bilateral preferences among households in c.

3.3 Estimating Spillover and Preference Composition Effects

To render equation (2) tractable for estimation, we normalize all expenditure volumes X_{oh} by expenditure on U.S. 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 U.S. counties.[23]

We define any variable $\tilde{a}_{oh}$ as the value of a for origin o divided by the equivalent value for U.S. 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 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

[22] Formally, $\kappa_h \equiv (X_h/\Phi_h)/S_c$. We note that this definition of S_c is consistent with the county-level gravity model, since the county-level price index Φ_c is equal to nominal over real expenditure, X_c/S_c.

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

that were born in country o:[24]

$$\phi_{oc}^{b} = \exp\left(\rho d_{oc} + \beta^{b} I_{oc} + \eta_{oc}^{b}\right). \tag{4}$$

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_{oc}^{b} captures the unobserved component of origin-county-specific import accessibility.

Lastly, we provide a functional form for the preference vector $\boldsymbol{z_h}$. We consider two components of preferences: first, immigrants may affect the preferences of nearby households, and second, a component that relates observed socioeconomic household characteristics to import demand. For a given household and origin country, we therefore assume the following functional form for z_{oh}:

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

J_h represents a vector of observed household characteristics such as income, education, and race. Motivated by our stylized facts, ζ_1 captures the strength of immigrants' taste for goods from all foreign countries, and ζ_2 captures the strength of immigrants' home-biased preferences à la Atkin (2016) and Logan and Rhode (2010).

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 preference composition effect of changes in I_{oc}. β^z, on the other hand, captures 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.[25]

Natives may share ancestry from the same origin country as the immigrants, and furthermore ancestors and immigrants from the same origin may geographically sort to live close to

[24] d_{oc} includes the log distance between o and c and the latitude difference between o and c. Measuring the immigrant variable a share of the population conforms to the functional form choice of Ottaviano et al. (2018).

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

one another, which would lead us to overestimate the effect of immigrants on native preferences. We address this concern in two ways. First, adjusting for race and ethnicity captures some of the likely demand heterogeneity for imports among native ancestors. Second, in the next section we introduce an instrumental variable strategy in which we leverage variation which is plausibly exogenous to native location sorting.

Plugging equations (4) and (5) into equation (3), we derive our estimating equation:

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

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 onto import expenditure for all households, but cannot distinguish between the supply and demand effects of this spillover.

3.4 Identification and Instrumental 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, low bilateral trade costs between New York and Italy may independently expand the set of types of pasta available locally, which thereby draws in Italian immigrants who tend to have a strong taste for pasta. To deal with such origin-by-county specific confounders, we adopt the instrumental variable approach of Burchardi et al. (2019).[26]

The approach works as follows. To predict the origin-by-county immigrant population in 2010, we generate a vector of exogenous immigration flows from the origin and into the county using 130 years of historic data.The immigration flow instrument is based on the intuition that historic immigration flows from an origin country to a U.S. county are more likely to occur when the origin is sending many immigrants at the same time the destination county is attracting immigrants from all origins.

[26]We provide only a brief description of the instrumental variable strategy here, as our approach follows closely that of Burchardi et al. (2019). We refer the interested reader to Appendix A.2 for more details.

The instrument interacts the arrival into the U.S. of immigrants from origin country o (the push) with the attractiveness of destination d to all immigrants (the pull) during a given historical decade D. To deal with potential spatial correlation in confounders, we leave out both the continent of origin country o when computing the pull component and leave out the Census region of county c when constructing the push component. Formally, the instrument is defined as

$$IV_{o,c}^{D} = \underbrace{L_{o,-r(c)}^{D}}_{\text{Push}} \times \underbrace{\frac{L_{-\mathcal{C}(o),c}^{D}}{L_{-\mathcal{C}(o)}^{D}}}_{\text{Pull}}, \tag{7}$$

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

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

We use equation (7) to predict immigrant inflows into the U.S. for decades from 1880 to 2000, and document the first-stage estimates in Appendix Table A.1. The push-pull instrument strongly predicts the contemporary bilateral immigrant population share.

Given the prevalence of zeros in household consumption expenditure shares $\tilde{X}_{oh}$, we use pseudo-Poisson maximum likelihood (PPML) to estimate equation (6) (Silva and Tenreyro 2006). To implement the instrumental variables strategy, we account for the non-linearity of PPML by using a control function approach to generating exogenous variation in the immigrant population (Petrin and Train 2010; Morten and Oliveira 2024). In particular, we add the residuals from the first-stage instrumental variable regressions as controls for our main specifications.[27]

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

Table 1: Household Gravity Estimates

	Dependent variable: Exp. share on goods from o relative to US	
	(1)	(2)
Immigrants/Pop. 2010	1.29***	1.15***
	(0.22)	(0.24)
First-stage residuals		0.18
		(0.31)
=1 if immigrant from anywhere	0.23***	0.23***
	(0.030)	(0.030)
=1 if immigrant from origin o	0.60***	0.61***
	(0.069)	(0.071)
N	1,461,130	1,461,130
Country FE	✓	✓
Household controls	✓	✓
Distance & latitude difference	✓	✓
1st-stage F-statistic		19.5

Notes: The table presents estimation results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share in column 2. Observations are weighted using NielsenIQ 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 General Gravity Results

We show the results of estimating equation (6) using PPML in Table 1. The first column shows estimates without first-stage residuals, and indicates that a higher immigrant population share corresponds to higher spending on goods from the immigrants' origin country. In column 2, we add the first-stage residuals and 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 1.15 percent (SE=0.24).

Comparing the immigrant population share coefficients between columns 1 and 2, we find that the estimate falls by about 11% 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. In terms of the preference parameter coefficients, we find that immigrants spend 26% more on imports from any origin than natives do, and

132% more on imports specifically from the immigrant's origin country.[28]

The results summarized in Table 1 provide two key takeaways. First, immigrants' preferences—the preference 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 coefficient in the first row—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 return to this distinction in Section 3.7 within the context of the welfare formula derived in Arkolakis et al. (2012). First, however, we discuss the robustness of our main results.[29]

3.6 Robustness and Heterogeneity

We conduct a variety of robustness checks and heterogeneity analyses in Appendix A.3. In Section A.3.1, we re-weight households to match the distribution of immigrant country origins observed in U.S. Census data. In Section A.3.2, we show that our main results are virtually unchanged when allowing immigrants to exhibit specific preferences for countries geographically or culturally similar to their origin country. In Section A.3.3, we assess the relative importance of the extensive versus intensive margins. In Section A.3.4, we find that our baseline results are robust to controlling for the instrumental variable mean as recommended by Borusyak and Hull (2023) for instruments which combine different sources of variation according to a known formula.

[28] $\exp\hat{\zeta}_1 - 1 = \exp 0.23 - 1 = 0.26$ and $\exp(\hat{\zeta}_1 + \hat{\zeta}_2) - 1 = \exp(0.23 + 0.61) - 1 = 1.32$.

[29] Burchardi et al. (2019) find no effect of immigrants on trade using state-level trade data and state-specific fixed effects. When aggregated to the state level and estimated alongside state-specific fixed effects our headline estimates remain positive and significant. We provide further details regarding these differing results in Appendix A.3.5.

3.7 Immigrants, Imports and Welfare

Arkolakis et al. (2012) (henceforth ACR) develop the following simple welfare formula for a large class of trade models:

$$d \ln W_c = (d \ln X_{us,c})^{1/\theta},$$

where W_c is real expenditures or, equivalently, welfare in county c; and θ is the trade elasticity. In this subsection we show that applying the ACR formula to recover the aggregate welfare effects of an immigration shock can significantly over-estimate the benefits to native households when immigrant preferences differ from natives. We then derive a multiplicative adjustment factor which allows one to recover welfare effects on natives due to an immigration trade shock using both the aggregate $d \ln X_{us,c}$ and our estimated parameters.

An immigration shock which affects trade costs deviates from ACR for two key reasons. First, if native households exhibit weaker preferences for imported goods than immigrants (ζ_1,$\zeta_2 > 0$), then natives are less sensitive to trade shocks than immigrant households. Second, immigrant-induced changes in county-level import expenditure will only translate into welfare gains for native households if associated with changes in trade barriers (i.e., ϕ^b_{oc}) rather than changes in local preferences (ϕ^z_{oc}).

With a few simplifying assumptions we derive an explicit adjustment to the standard ACR formula which allows us to express the gap between implied welfare gains from an aggregate change in import expenditure to the change in import expenditure associated with native households. For the purposes of this exercise, we assume that there exist only two regions: the United States (us) and the rest of the world. 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 $d \ln X_{us,c}$.[30] Given the general 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

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

expenditure of native households using the following transformation:[31]

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

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. With equation (8) in hand, one can then follow ACR and compute native welfare as $d \ln W_n = (d \ln X_{us,n})^{1/\theta}$.

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 that is welfare-relevant to native households: $d \ln \phi^b_{oc} / d \ln \phi_{oc}$. If at least one of β^z or ζ is positive, and both non-negative, this second term is less than one. Therefore, changes in native household welfare should be discounted when compared to the implied aggregate welfare effects. If $\beta^b = 0$, immigrant-induced changes in domestic expenditure may be large in the aggregate, but will have zero effect on the welfare of native households, and $d \ln X_{us,n} = 0$.[32]

Our estimates from Table 1 imply that immigrants exhibit stronger preferences for imports from their origin country than natives, 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 the spillover effect of immigrants is due to supply factors (lowering trade costs) or demand factors (influencing the preference of neighbors). 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.

[31] The details of the derivation can be found in Appendix B.1.

[32] We also note that assuming no preference heterogeneity ($\zeta = 0$) and no effect of a shock on preferences ($\beta = \beta^b$) eliminates the multiplicative terms in equation (8).

4 Microfounding a Model of Immigration and Imports

This section uses the Melitz (2003)-Chaney (2008) micro-foundation to expand upon the general 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}.

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 affect the supply of varieties locally (Iranzo and Peri 2009; Di Giovanni et al. 2015; Aubry et al. 2016). Second, the structure of the Melitz-Chaney heterogeneous firms model allows us to fully leverage our data 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.

4.1 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 a continuum of differentiated varieties $\Omega_{o,c(h)}$ associated with each origin country $o \in \mathcal{O}$. As in the previous section, we allow for household heterogeneity in income Y_h and origin-specific preferences denoted by $z_{oh} \in \boldsymbol{z_h}$. Preferences for the differentiated sector are represented by the following CES utility function:

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

with $\sigma > 1$ denoting the elasticity of substitution among differentiated varieties and μ_0 capturing the expenditure share on the homogeneous good, which 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 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-by-origin specific iceberg trade costs τ_{oc} and fixed costs of exporting f_{oc}. Each firm draws some productivity φ from a Pareto distribution with shape parameter $\theta > \sigma - 1$. The set of potential entrant firms in each origin is proportional to the size of that origin Y_o.[33] The cost of providing q units to destination county c by a firm in origin o with productivity φ is therefore:

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

We assume that all entry and pricing decisions are made 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 Appendix B.2, 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 general gravity model.[34] The average county-level preferences $z_{o,c(h)}$ are also the same as our definition for the bilateral affinity term introduced in Section 3.2 (ϕ^z_{oc}) and are an expenditure-weighted average of the preference shifter z_{oh} across all households in c. Household-level

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

[34] 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.

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:

$$X_{oc} = \lambda_4 Y_o S_c (w_o \tau_{oc})^{-\theta} \Big(\frac{f_{oc}}{S_c z_{oc}} \Big)^{-(\frac{\theta}{\sigma-1}-1)} z_{oc} \equiv \alpha_o S_c^{\frac{\theta}{\sigma-1}} \phi_{oc}^b \phi_{oc}^z. \tag{13}$$

Notice that we now have a micro-foundation for each term used in the model of Section 3. ϕ_{oc}^z remains unchanged, whereas real expenditures S_c is now raised 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 formally defined as $\alpha_o = Y_o w_o^{-\theta}$.

The preference shifter z_{oc} represents the novel extension beyond the standard Chaney (2008) framework. Preferences also contribute to a market size effect associated with county-level average preferences. As preferences shift towards goods from origin o, more firms from o are able to cover the fixed costs of supplying county c, which further enhances the market penetration of imports from o to county c.

As in Section 3.3, 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 its U.S. equivalent, we can express the normalized household-by-origin level expenditure equation as

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

When estimating the various channels of import demand, it will 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 $\bar{z}_{oh}$.

4.2 Immigrant Channels and Decomposition

We complete the model introduced in Section 4.1 by introducing functional-form assumptions for variable and fixed trade costs. As in Section 3.3, 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.

Our functional-form assumptions regarding the variable and fixed components of ϕ^b_{oc} closely follow the assumptions made in the general 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:[35]

$$\tilde{\tau}_{oc} = \exp[-\frac{1}{\theta}(\rho^{\tau} d_{oc} + \beta^{\tau} I_{oc} + \eta^{\tau}_{oc})], \tag{15}$$

$$\tilde{f}_{oc} = \exp[-(\frac{\sigma - 1}{1 + \theta - \sigma})(\rho^{f} d_{oc} + \beta^{f} I_{oc} + \eta^{f}_{oc})], \tag{16}$$

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. β^{τ} captures the strength of the the variable cost reduction channel of immigrants and β^{f} 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 yields the following decomposition of the county-level partial elasticity of import expenditure with respect to the immigrant population share:

$$\begin{aligned}\frac{\partial \ln \tilde{X}_{oc}}{\partial I_{oc}} &= \frac{\partial \ln \phi^b_{oc}}{\partial I_{oc}} + \frac{\partial \ln \phi^z_{oc}}{\partial I_{oc}} \\ &= \underbrace{\left[\beta^{\tau} + \beta^{f}\right]}_{\substack{\text{Trade cost} \\ \text{channel}}} + \underbrace{\left[\frac{\theta}{\sigma - 1} - 1\right]\left(\beta^{z} + \frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}\right)}_{\text{Market size channel}} + \underbrace{\beta^{z}}_{\substack{\text{Preference} \\ \text{diffusion} \\ \text{channel}}} + \underbrace{\frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}}_{\substack{\text{Composition} \\ \text{channel}}} .\end{aligned} \tag{17}$$

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

Expression (17) illustrates the channels through which immigrants affect county-level import expenditure from a given origin. The first two channels—trade costs and market size—represent changes in the supply-side effects of immigrants, or ϕ_{oc}^{b}. These include the variable cost reduction effect, the fixed cost reduction effect, and the preference-driven market size effect, which is associated with changes in 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. The strength of this effect is governed 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 preference composition channel 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 previous discussion regarding the general gravity model. The only welfare relevant channels of immigrant-induced import penetration are those associated with ϕ_{oc}^{b}: the trade cost and market size channels.

An important feature of this microfoundation 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 return to the household-level gravity model discussed previously but accommodating the microfoundation described here:[36]

$$\begin{aligned}\ln \tilde{X}_{oh} = \alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)} + \ln \bar{z}_{o,c(h)}^{\frac{\theta}{\sigma-1}-1} \\ + \delta J_h + \zeta_1 \mathbf{1}\left[o(h) \neq US\right] + \zeta_2 \mathbf{1}\left[o(h) = o\right] + \eta_{o,c(h)} + \eta_{oh}^{z}\end{aligned} \tag{18}$$

with the following definitions:

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

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

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

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

The specification here reveals three identification concerns. First, as discussed in Section 3.4, 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 instrumental variables strategy. Second, and perhaps more concerning, is that the preference diffusion effect of immigrants and the preference composition effect of immigrants are not separately identified: county-level preferences $\bar{z}_{oc}$ were not implied by the general gravity model 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}.

4.3 Identifying the Channels of Immigrant-Induced Imports

In this section we provide a three-step identification strategy which allows us to separately identify each of the various channels by which immigrants affect imports.

Identification of Exogenous Preferences: We start by identifying how household's socioeconomic characteristics—such as race, education, household size, and income—affect the demand for imports. To do so, we control for origin-by-county factors which affect trade costs and product availability, and then project those socioeconomic characteristics onto the residual variation.

We collect all terms affected by 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, which is identical to equation (18) save for the collection of terms into ψ_{oc}:

$$\ln \tilde{X}_{oh} = \psi_{o,c(h)} + \delta J_h + \zeta_1 \mathbf{1}[o(h) \neq US] + \zeta_2 \mathbf{1}[o(h) = o] + \eta^z_{oh}. \tag{19}$$

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 and firm selection 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 exogenous 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 Section 5.2 to follow.

Estimating $\boldsymbol{\beta}$: With the exogenous preference estimates in hand, we can difference out both $\bar{z}_{oh}$ 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 to obtain the following equation:

$$\ln \frac{\tilde{X}_{oh}}{\mathcal{Z}_{oh}} = \alpha_o + \rho d_{o,c(h)} + \beta I_{o,c(h)} + \eta_{o,c(h)} + \eta^z_{oh}, \tag{20}$$

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

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

We therefore arrive at an estimating equation that is reminiscent of the general gravity model estimated earlier (see equation 6), and we make use of the same instrumental variables strategy and again implement PPML with a control function approach.

Estimating $\boldsymbol{\beta^\tau}$: Even after adjusting for market size effects, we still cannot disentangle the components of β: β^τ, β^f, and β^z. We leverage model restrictions and data characteristics which allow us to identify this decomposition.

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:

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

By aggregating our data to the barcode-by-county level we can estimate the price equation directly, having incorporated the functional form assumption of $\tilde{\tau}_{oc}$ from equation (15):

$$\ln p_{\omega(o),c} = \psi_c + \psi_\omega - \frac{\beta^\tau}{\theta} I_{oc} - \frac{\rho^\tau}{\theta} d_{oc} - \frac{1}{\theta}\eta^\tau_{oc}, \tag{21}$$

where ψ_c and ψ_ω represent county and barcode-level fixed effects.[37] Since our dataset is at the barcode level, equation (21) allows us to identify the effect of immigrants on the price of imported varieties rather than the effect of immigrants on the average price of imported goods. This is an important benefit of our dataset, as our model predicts that an average price measure would conflate the effect of immigrants on within-variety prices and the effect of immigrants on entry and market composition. As with our baseline specification, we instrument for the immigrant population share to account for $cov[I_{oc}, \eta^\tau_{oc}] \neq 0$.

Estimating $\boldsymbol{\beta^f}$ and $\boldsymbol{\beta^z}$: If $\beta^\tau \approx 0$, we can isolate the effect of fixed costs from the effect of preference diffusion on import expenditure by comparing the intensive margin (expenditure shares, as in our baseline) with the extensive margin (counts of varieties). We show in Section 5.3 that immigrants have no discernible effect on variable trade costs.

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 have 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 (20) 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 Appendix B.3, 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:

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

[37] 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.

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

The intuition is as follows. On the extensive margin, firms enter a new market if and only if they can cover their fixed costs. Immigrants can therefore facilitate firm entry by either (i) reducing fixed costs, or (ii) increasing sales-per-variety in their market, either by bringing their own preferences for imported goods or influencing their neighbors' preferences.

On the intensive margin, expenditure can only increase in a given export market if preferences become more favourable to the exports in question. Reducing fixed costs has no impact on how much a firm sells to a given market, conditional on already selling to that market. This intuition is reflected in the system of equations provided above: as the aggregate trade elasticity and the variety trade elasticity converge, it must be the case that $\beta^z \rightarrow 0$, as only $\beta^f = \beta$ could satisfy the extensive margin playing such a large role.

5 Melitz-Chaney Model Estimation Results

5.1 Estimating Preference Terms

We construct estimates for household preferences $\bar{z}_{oh}$ by estimating equation (19) using our NielsenIQ 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.2.

We find that import expenditure is generally increasing in income, albeit noisily, with a similar pattern of import expenditure increasing in household education. 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 closely 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.

Appendix Figure C.3a shows the distribution of $\ln \hat{\bar{z}}_{oh}$, which we recenter around zero. Household preferences for imports appear quite symmetrically and closely distributed around the mean, with only a handful of household-origin pairs exhibiting very strong or very weak preferences for imported goods.

To impute county-level preferences, $\hat{\hat{z}}_{oc}$, one must observe the characteristics of all households in a given county. Given the small sample size of the NielsenHQ data, we turn to the 2013-2017 ACS sample to compute $\hat{\hat{z}}_{oc}$. We do so in two steps. First, we take the preference parameters estimated using the NielsenHQ data—$\hat{\delta}$, $\hat{\zeta}_1$, and $\hat{\zeta}_2$—and predict each ACS household's preference term $\hat{\hat{z}}_{oh}$. Then we sum across households to construct county-level preferences: $\hat{\hat{z}}_{oc} = \sum_{h' \in \Lambda_c} \hat{\hat{z}}_{oh'} \kappa_{h'}$. We compute the household weights $\kappa_{h'}$ as being the share of overall income in county c that is earned by household h'.

Appendix Figure C.3b 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 Estimates of the Immigrant Spillover Effects

Next we estimate the total effect of immigrants on imports using equation (20), in which expenditure is deflated by household and county-level preferences. Recall that in 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.[38]

Columns 1 and 2 of Table 2 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 $\hat{\beta}$ =1.36.

The modest effect of correcting for county-level preferences likely reflects two off-setting

[38]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 2: Estimates of Household Gravity Equation

	$\tilde{X}_{oh}/\mathcal{Z}_{oh}$		$\tilde{N}_{oh}/\mathcal{Z}_{oh}$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.50***	1.36***	1.19***	1.14***
	(0.22)	(0.29)	(0.11)	(0.15)
First-stage residuals		0.18		0.069
		(0.38)		(0.21)
N	1,461,130	1,461,130	1,461,130	1,461,130
Country FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		20.2		20.2

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

sources of bias. Given that our initial general gravity model omitted the spillover effect of immigrants due to the market size channel ($\bar{z}_{oc}^{\frac{\theta}{\sigma-1}-1}$), we would expect the unadjusted estimates of Section 3.5 to be biased upwards. However the same logic applies to all other origin countries too: we find evidence that immigrants increase demand for all import origins which, via the market size effects adjusted for in this specification, also serve to increase imports from all origins. That is, identifying β from unadjusted cross-origin variation is potentially mis-specified when immigrants from origin o also affect native household expenditure from all other origins o'. Thus, there is a possibility of both negative and positive bias in the adjusted estimates of β.

5.3 Decomposing Spillovers into Trade Costs and Preferences

We start by leveraging the price information that we observe in the NielsenIQ Homescanner data in order to estimate equation (21). We show our estimates in Table 3, which shows 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. To address concerns about products sold in only a handful of countries driving our results, we also restrict the sample to barcodes which we

Table 3: Estimates of Variable Cost Parameter using Variation in Prices

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

Notes: The table presents two-stage least square regression results at the barcode-county level. The instrumental variables strategy is described in Section 3.4. Standard errors are clustered at the barcode and country level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

observe in at least 100 counties in the NielsenIQ data in columns 3 and 4. In columns 2 and 4 we instrument for the bilateral immigrant-population share using the leave-out push-pull instrumental variables defined in equation (7).

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 point increase in the share of the local population which is born in country o raises prices by 0.01 percent, and suggests that $\hat{\beta}^{\tau} = -0.085$. We therefore conclude that $\hat{\beta}^{\tau} \approx 0$.

We then estimate equation (20) 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. Columns 3 and 4 of Table 2 provides estimates of the extensive margin effect of immigrants on import expenditure.

Solving for β^f and β^z using the elasticity estimates from columns 2 and 4 of Table 2, we recover β^f =1.09 and β^z =0.22. Since our estimate of β from Table 2 is 1.36, 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

We conduct three primary counterfactuals. First, we remove the channels through which immigrants affect the import expenditure of households in the U.S. That is, we set $\zeta_1 = \zeta_2 = \beta_f = \beta_z = 0$ and recalculate heterogeneous preference-induced market size term $\bar{z}_{oc}$ accordingly.[39] We remove each individual channel separately in order to quantify the contribution of each. In a second counterfactual, we additionally remove the grocery expenditure associated with immigrant households, which corresponds to removing immigrants from the U.S. population altogether, thus capturing the market size benefits of immigrants. Finally, we simulate a 10% increase in the variable cost associated with all imported goods and quantify the relative cost of this policy on all households—immigrant and native.[40]

We show how we compute counterfactual trade flows and utility in Appendix B.4.

6.1 Aggregate Effect of Immigrants on Imports & Native Welfare

In our first two counterfactuals, we quantify the impact of immigrants on import volumes and on native welfare via consumption. We remove all the effects distinctive to immigrants in our first counterfactual, whereas we simulate removing all immigrants from the U.S. entirely in the second counterfactual.

To generate values which are representative of the United States as a whole, as well as meaningful counterfactual values for each county, we leverage the pooled 2013-2017 ACS sample. In particular, we use the results from estimating equation (19) with the NielsenIQ 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 NielsenIQ, which matches estimates from the Consumer Expenditure Survey (CEX).[41] Finally, we use the crosswalks provided by Burchardi et al. (2019) to generate county-specific immigrant population shares based on the PUMA of residence.

[39] We do not change β_z when computing the welfare change in the counterfactual scenario as a change in preferences would make welfare comparisons between the scenarios impossible.

[40] Note that our counterfactuals exclusively allow for partial-equilibrium adjustment in the consumption space. The production-side effects of immigrants are outside the scope of our framework.

[41] When we discuss the distributional consequences of immigrants, this assumption of equal grocery expenditures across the income distribution ensures that those results are not driven by assumptions we make about differential grocery spending.

Table 4: Counterfactual Results Summary

Counterfactual exercise: Removing ...	(1) Change (%) import expenditure	(2) Change (%) welfare natives	(3) Change ($) welfare per native HH
... all immigrant channels	-7.7	-0.033	-2.4
... trade cost channel	-1.6	-0.028	-2.1
... market size channel	-0.3	-0.005	-0.3
... composition channel	-5.7	–	–
... homophily channel	-1.4	–	–
... preference diffusion channel	-0.4	–	–
... all immig. channels + expenditure	-26	-0.925	-69

Notes: This table shows the change in outcomes under various counterfactual scenarios. The baseline scenario in the first row removes all channels through which immigrants affect import expenditures—i.e. we set $\zeta_1 = \zeta_2 = \beta_f = \beta_z = 0$ and recalculate $\bar{z}_{oc}$—but keeps total household expenditure constant. In the next rows, we keep all parameters unchanged except the following: β_f (fixed trade costs); $\bar{z}_{oc}$ (market size); ζ_1 and ζ_2 (composition); ζ_2 (homophily component of composition channel); β_z (preference diffusion). In the last row, we remove all immigrant channels and the expenditures made by immigrants, equivalent to removing immigrants from the U.S. population. Note that we always keep β_z at its estimated value for the calculation of the changes in welfare (columns 2 and 3) so that preferences remain fixed.

In our first counterfactual, we remove all immigrant effects (fixed trade cost reduction, market size, preference composition and preference diffusion) altogether and subsequently shut down each effect one-by-one. We summarize the results in Table 4. Note that in order to be able to meaningfully interpret welfare impacts, we do not remove the preference diffusion effect when computing the utility changes for native households shown in columns 2 and 3.

The results from our baseline counterfactual scenario in which all immigrant effects are removed appear in the first row. Averaging across households, we find that aggregate U.S. expenditures on imports of grocery and personal care items fall by 7.7%. We further find that removing all immigrant spillover effects yields an average welfare loss of 0.033%, amounting to a welfare-equivalent fall of $2.4 per household.

We now turn to quantifying the effect of each channel in turn. Removing the effect of immigrants on fixed trade costs reduces import expenditure by 1.6%, implying that this channel contributes around 20% to the total effect. While removing the preference-driven market size channel, i.e. the change in $\bar{z}_{oc}$, causes a negligible decline in import expenditures, removing the composition channel, which is associated with immigrants' preference parameters ζ_1 and ζ_2, causes a decline of 5.7%. Thus, this channel contributes three quarters of the total effect of all immigrant channels. The fifth row shows the impact of only removing immigrants' preferences for goods from their own origin, i.e. setting $\zeta_2 = 0$. The resulting fall in import expenditure is one quarter of that resulting from removing both ζ_1 and ζ_2, suggesting that immigrants' preferences for imports from any origin drive the bulk of the composition channel effect. In terms of welfare, we find that the total effect is almost entirely driven by the trade cost channel.[42]

The last row of Table 4 shows the second counterfactual, in which we additionally remove all grocery expenditures associated with immigrants. In terms of the model, this corresponds to a reduction in $S_{c(h)}$, the real market size of county c, as opposed to only changing the preference-driven market size $\bar{z}_{oc}$ in the previous counterfactual. Accounting for this additional market size effect leads to an overall decline in import expenditure by 26%. The average loss in grocery consumption welfare for natives in this scenario is 0.93% or a welfare-equivalent fall of \$69 per household. Thus, removing immigrants' expenditures in our counterfactual scenario leads to a non-negligible welfare effect on native households.[43]

A key result highlighted in Table 4 is that the change in import expenditure associated with removing immigrant channels is generally larger than the associated change in *welfare*. Applying the ACR formula to the 7.7% change in import expenditure, as well as our calibration of $\theta = 5$, suggests a naive welfare cost of removing immigrant effects of approximately -0.15%.[44] Yet the welfare cost associated with this decrease in import expenditure derived in Table 4 is only 0.03%: just over a four-fold difference in magnitude.

While these estimates may seem small in absolute magnitude, they are naturally bounded below by the cost of autarky, which in this case is a 1.90% decrease in welfare. The naive

[42]Note that the composition/homophily channels are not directly welfare-relevant as any change in ζ_1 or ζ_2 affects native's utility only indirectly via the market size channel.

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

[44]Given an initial import expenditure share of 9%, and a decrease in this share by 7.7%, we derive $-0.0015 = 1 - (0.910/0.917)^{1/\theta}$. The calculation for autarky is identical save for $0.917 = 1.00$.

ACR application therefore yields a welfare cost estimate of removing immigrant-induced trade that is 8% of autarky costs, versus a mere 2% of these costs associated with our model.

Still, the average estimates discussed here mask considerable heterogeneity across space and across the income distribution. We explore such heterogeneity next.

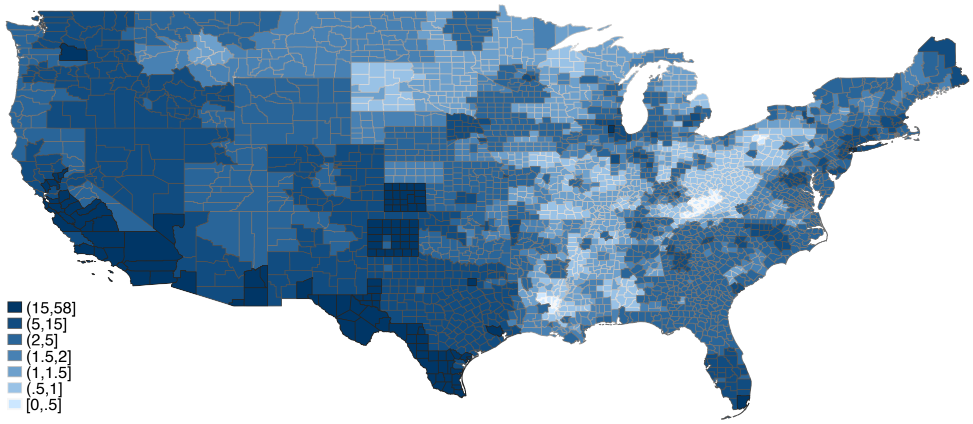
Immigrant effects by import origin country. We illustrate the unequal change in import volumes across origin countries associated with our first counterfactual exercise in Appendix Figure C.4. The expenditure share on Mexican imports falls the most, as expected, by 13.11%. Mexico is followed by China, India, the Philippines and Vietnam with expenditure share decreases between 5.98% and 4.95%.

In general, we find that the trade-creating effects of immigrants are pronounced for countries which are proximate to the U.S. and large, in terms of population. This is unsurprising, as these countries are well-represented in the immigrant populations of the U.S., but we can use our counterfactual exercises to quantify the extent to which origin country characteristics are correlated with immigrant-induced trade volumes. For every 10% reduction in distance between the U.S. and a given origin o, the counterfactual trade-creating effect of immigrants increases in magnitude by 1.6%. For every 10% increase in population of an origin country, this same effect increases in magnitude by 0.6%. While these estimates are correlations, they are both significant at levels of 99% within a sample of only 72 countries.

Immigrant effects across U.S. counties. We graphically depict the substantial variation in the fall of import volumes associated with removing immigrant effects across U.S. counties. Figure 3 maps the import expenditure changes associated with our first counterfactual at the county level. Appendix Figure C.5 maps the average dollar-equivalent change in utility associated with our second counterfactual: removing immigrants entirely. In both cases, the impact of immigrants on imports and welfare is remarkably concentrated in the Southwest, West Coast, and East Coast of the U.S., as well as almost all major cities.

The counties experiencing the largest drop in import expenditures under our first counterfactual are El Paso, TX (-44%); Los Angeles, CA (-27%); Kern, CA (-25%); Riverside, CA (-23%); and Fresno, CA (-23%). Assuming an initial MFN tariff rate of 2.5% applied to the

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



Notes: This chart plots the percent decrease in the value of grocery and personal care imports when the trade-creating effects of immigrants are removed following the procedure outlined in Appendix Section B.4.

grocery goods studied here, as well as our calibration of $\theta = 5$, these estimates are equivalent to Los Angeles facing a three-fold increase in tariff rates. For our second counterfactual, welfare effects are large and heterogeneous across space. In terms of annual dollar-equivalent welfare effects for large counties, the most affected are Queens, NY ($386); Dade, FL ($356); Hudson, NJ ($309); Santa Clara, CA ($292); and Los Angeles, CA ($277).

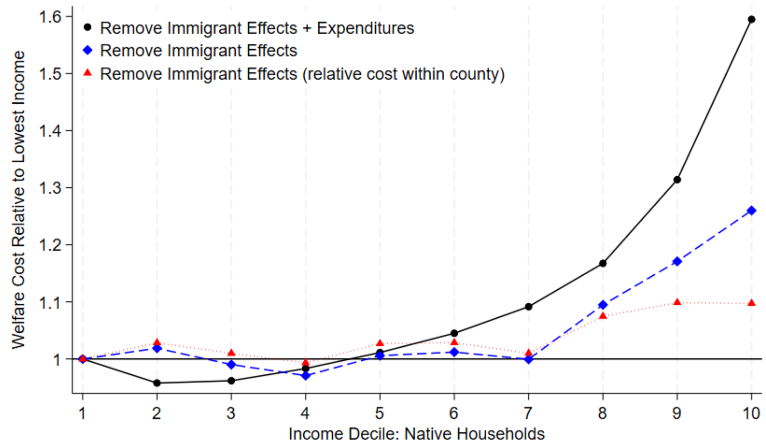
Immigrant effects by native household income. While prior literature has emphasized the distributional consequences of immigrants in the labor market (e.g., Dustmann et al. 2013), we are the first to do so looking at the consumption side, enabled by our highly detailed household-level data and approach.

Figure 4 depicts the welfare losses across the income distribution in different counterfactual scenarios, each computed relative to the lowest income decile. In all cases, there is very little difference in welfare effects associated with the first six income deciles. By contrast, the welfare gains among the top four deciles are monotonically increasing.

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

To understand the sources of the unequal gains from immigrants in our first counterfac-

Figure 4: Percent Change in Grocery Welfare by Income Decile



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

tual, we conduct a second exercise. The red dotted line depicts the relative welfare effects of our first counterfactual but calculated within counties, thus isolating the role of native preferences in shaping the cost of this counterfactual across deciles (rather than geographic sorting of natives and immigrants). Just under half of this differential is driven by greater preferences for imported goods exhibited by the highest income households, with geographic sorting of immigrants with high-income households explaining the remaining half.

The black solid line depicts our second counterfactual in which we additionally remove immigrant expenditure. This reduction in market size has far larger distributional consequences, with households in the highest-income decile facing costs that are 60% greater than households at or below the median income level.

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, counties, and income groups within the U.S. Of particular note is the striking pattern of high-income native households benefiting substantially more from immigrants—even within the same county—relative to lower-income natives.

6.2 Differential Impact of Trade Shocks by Nativity, Education

In our final counterfactual, we examine the unequal effects of a trade shock across households. Specifically, we simulate a 10% increase in variable trade costs and plot the welfare costs of this shock stratified by education and nativity in Figure 5. Three key findings emerge.

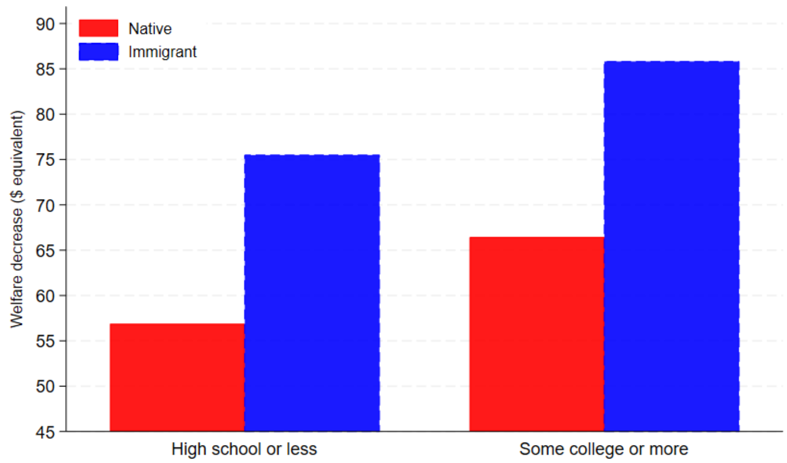
First, immigrants suffer much greater welfare losses than natives, regardless of how much education they have. On average, immigrants would lose the welfare-equivalent of about $78 compared to about $58 for natives. This result is driven by immigrants' greater preference for imports in their consumption baskets, as documented in Section 2 and via the estimates of ζ_1 and ζ_2 discussed in Section 3.5.

Second, more highly educated households lose out more from a trade shock. In particular, high-educated immigrant households are clearly the most negatively affected demographic, with welfare costs that are about 29% greater than those faced by similarly educated native households, and over 50% greater than less-educated native households.

These first two results demonstrate that a seemingly nativity-neutral trade shock results in highly disparate impacts depending on education and nativity. Low-income, less-educated, and native-born U.S. households face the lowest consumer costs associated with an increase in barriers to either immigration or imports. This paper therefore suggests a novel factor which may contribute to the well-documented lack of support for increased immigration among this demographic (Card et al. 2012; Alesina and Tabellini 2024). To back up this claim, we show in Appendix A.4 that a county's vote share for Trump in recent presidential elections correlates negatively with the local consumption welfare benefits of immigrants.

Our third finding from the trade shock counterfactual is that immigrants make natives more vulnerable to trade shocks due to preference diffusion. On average, this effect is quite small: the average native pays a 0.2% higher utility penalty with immigrants' preference diffusion relative to a scenario in which natives' import expenditures are reduced due to setting $\beta_z = 0$. The unequal spatial distribution of immigrants, however, means that 11% of natives experience a utility penalty 1% or greater with preference diffusion relative to without preference diffusion. For natives in Los Angeles, this utility penalty rises to 1.6%.

Figure 5: Effect of Variable Trade Cost Shock by Demographic Group



Notes: The chart depicts the counterfactual effect of a 10 percent increase in variable trade costs 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.

7 Conclusion

This paper has four key findings. First, maintaining aggregate expenditure but removing immigrant effects on local preferences and trade costs reduces the grocery import expenditure share by 8%, with surprisingly little consequence for native welfare.[45] Second, the same mechanism which yields muted welfare effects for natives—the strength of immigrant import preferences—increases the extent to which immigrants are more vulnerable to all trade shocks than their native neighbours. Third, the positive co-location of immigrants and higher-income natives concentrates the welfare benefits of immigrants within high-income native households. And fourth, removing immigrants' expenditure—and therefore the bulk of their impact on local market size—generates significant welfare costs for native households.

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.

[45]While this import-immigrant elasticity is a counterfactual result, and therefore not directly comparable to the existing empirical literature, it is worth noting that our elasticity lies in between the range of estimates surveyed by Felbermayr et al. (2015) (0.12–0.15) and the null result reported in Burchardi et al. (2019).

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.

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

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

Auer, Raphael, Ariel Burstein, Sarah Lein, and Jonathan Vogel, "Unequal expenditure switching: Evidence from Switzerland," *Review of Economic Studies*, 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 Peter Hull, "Nonrandom exposure to exogenous shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

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

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

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

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

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

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

_ , 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*, 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.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales, "How segregated is urban consumption?," *Journal of Political Economy*, 2019, *127* (4), 1684–1738.

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," *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," *Quarterly Journal of Economics*, 2016, *131* (3), 1113–1180.

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

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," 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," *Review of Economic Studies*, 2018, *85* (3), 1852–1896.

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

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

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," *American Economic Journal: Applied Economics*, 2024, *16* (2), 389–421.

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*, 2010, *43* (4), 1433–1459.

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

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

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.

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

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

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.

Online Appendix

Contents

A Empirical Appendix

A.1 Data Construction

We aggregate decennial census waves across individuals aged 16 and above to the county-by-origin level, applying the Census' individual sample weights. Immigrants are defined as those born outside the U.S. and not citizens by birth. To compute decadal migrant inflows from origin o into destination county c between two census years $t-10$ and t, denoted L_{oc}^{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 L_{oc}^{1880} includes all those that are either first-generation immigrants from o or second-generation immigrants whose parents were born in o.

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

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

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

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

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

$$IV_{o,c}^{D} = L_{o}^{D} \times \frac{L_{c}^{D}}{L^{D}},$$

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

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 $L^D_{o,c}$ flows when constructing the instrument for the county-country pair oc.

However, there might also be spatial correlation in confounding variables. For example, both Belgian and French immigrants and goods may go to Chicago for the same reason: many flight connections out of Paris, which is very accessible to potential Belgian migrants by train. Leaving out Belgium-to-Chicago immigration flows when computing the instrument predicting these same immigration flows is therefore not sufficient, because now the French immigration flows to Chicago (used to predict Belgium-to-Chicago flows) are also contaminated with the confounding flight connections. To avoid such endogeneity, we again follow Burchardi et al. (2019) and leave out both the set of countries which share a continent with origin country o, $\mathcal{C}(o)$, and the Census region of county c, $r(c)$, to construct the instrumental variable that we defined in equation (7).

A violation of the identification assumption may occur if, say, immigrants skilled at importing goods from France 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 the French are a large fraction of immigrants settling in Chicago, and if South Korean immigrants are a large fraction of the immigrants settling in Miami.

We use equation (7) 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 to a battery of tests. Following Burchardi et al. (2019), we include five principal component terms which capture the variation of interactions of the instruments within county-country pairs and across decades.[46]

While the push-pull instrument may bear a passing resemblance to a standard shift-share instrument, we note two key differences. First, shift-share instruments are typically summed

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

over a dimension (e.g., across origins), whereas the push-pull is not summed and thus retains two dimensions of variation. Second, the 'share' component of the push-pull is not lagged, unlike in the canonical shift-share style instrument, such as the ethnic enclave instrument proposed by Card (2001).

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

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

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

Table A.1: First stage regression

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

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

A.3 Robustness of Gravity Results

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

A.3.1 Sample weights

As discussed by Feenstra et al. (2023), the Nielsen household sample may not be perfectly representative of the U.S. population in terms of income and price sensitivity. We also find that the Nielsen data is not representative of the distribution of immigrant origin countries. Mexican-born immigrants, for example, make up 15% of Nielsen households but 30% of households in the ACS. This may be driven by a combination of two factors. First, cross-sectionally Nielsen HomeScanner may miss some households if, for example, the survey module is not available in the language an immigrant household speaks. Secondly, there may be differential attrition across immigrant origins between 2008, when the "Tell Me More About You" was distributed, and the sample period of 2014-16.[47]

To gauge the importance of Nielsen's lack of representativeness in driving our results, we adjust the survey weights so that the weighted aggregate population shares of natives and immigrants of each origin reflect those measured in the pooled 2013-2017 ACS sample. We show the results in Table A.2. Similar to our baseline results, immigrants have a positive and statistically significant effect on consumption of others on goods from their origin. Moreover, the magnitude of the estimated coefficient increases by almost 36% to 1.59. We therefore conclude that our baseline estimates are not driven by Nielsen's lack of representativeness.

A.3.2 Alternative measures of origin country connectedness

In our baseline specification, equation 6, we allow households to have specific preferences for (i) all imports, and (ii) imports specifically from the household's origin country. Households may additionally exhibit specific preferences towards goods from countries close—geographically or culturally—to their origin country.

We test the importance of such specific household preferences in Table A.3. We start by exploring whether immigrant households exhibit a specific preference for goods from

[47]The Nielsen survey is voluntary so households may drop out at any time.

Table A.2: 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 NielsenIQ 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.

their continent of origin (in addition to their country of origin) in column 1. We find that immigrant households spend 12% more on goods imported from their origin continent. The coefficient on the country-of-origin dummy falls slightly to 0.52 from 0.61.

We also consider whether non-immigrants with ancestry from a given origin region exhibit a specific preference for goods from that region. While we cannot observe ancestry for every household in the Nielsen HomeScanner data, we do observe Hispanic ethnicity. We leverage this variable to assess whether Hispanic households exhibit greater demand for foods imported from Latin America in column 2. We find a positive but statistically insignificant relationship between Hispanic background and demand for imports from Latin America.

Immigrant households may prefer goods from similar cultures, or from countries with similar colonial backgrounds, not merely countries geographically proximate to their origin. We test whether such cultural or colonial ties affect immigrant product demand in column 3. We proxy for cultural similarity between an immigrant's origin country and an import origin country with a dummy for the two foreign countries sharing the same language as measured by Conte et al. (2022). We measure current or past colonial relationships also

Table A.3: Gravity regressions with additional preference terms

	Dependent variable: Exp. share on goods from o relative to US			
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	1.19***	1.14***	1.19***	1.17***
	(0.24)	(0.24)	(0.24)	(0.24)
First-stage residuals	0.17	0.17	0.17	0.15
	(0.31)	(0.31)	(0.31)	(0.31)
=1 if immigrant from anywhere	0.20***	0.23***	0.22***	0.20***
	(0.035)	(0.030)	(0.033)	(0.037)
=1 if immigrant from origin o	0.52***	0.59***	0.62***	0.54***
	(0.085)	(0.069)	(0.072)	(0.085)
=1 if immig. from continent of o	0.12*			0.092
	(0.066)			(0.063)
=1 if hispanic and o in Latin America		0.073		0.075
		(0.059)		(0.057)
=1 if common official or primary language			0.026	0.0051
			(0.057)	(0.058)
=1 if ever in colonial or dependency relationship			0.11	0.14
			(0.10)	(0.10)
=1 if currently in colonial or dependency relationship			0.35	0.37
			(0.84)	(0.83)
N	1,461,130	1,461,130	1,460,552	1,460,552
Country FE	✓	✓	✓	✓
Household controls	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic	19.5	19.5	19.5	19.4

Notes: The table presents regression results at the household-country level. Observations weighted using NielsenIQ household weights. The dummies indicating common language or colonial relationship are taken from the CEPII Gravity Database (Conte et al. 2022) with country pairs being based on household origin and import expenditure origin country. Standard errors clustered two-ways at the household and county-country levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

using the database of Conte et al. (2022). We find no statistically significant effect of a shared language or colonial relationships on product demand.

Finally, we include all aforementioned variables in a single specification in column 4 and find no statistically significant effect of any variable on imported product demand.

A.3.3 Extensive and intensive margin

Do immigrants induce natives to purchase more intensively from immigrants' origin countries or to drive diversification of product origins for native households? We answer this question

Table A.4: Extensive and Intensive Margin of Household Import Expenditures

	=1 if $X_{oh} > 0$		$\ln(\tilde{X}_{oh})$	
	(1)	(2)	(3)	(4)
Immigrants/Pop. 2010	0.14*	-0.42**	1.87***	1.63***
	(0.081)	(0.21)	(0.24)	(0.41)
=1 if immigrant from anywhere	0.021***	0.022***	0.28***	0.28***
	(0.0037)	(0.0037)	(0.026)	(0.026)
=1 if immigrant from origin o	0.19***	0.20***	0.58***	0.59***
	(0.021)	(0.022)	(0.066)	(0.068)
N	1,461,130	1,461,130	270,869	270,869
Country FE	✓	✓	✓	✓
Household controls	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		19.5		11.4

Notes: The table presents regression results at the household-country level. The dependent variable in columns 1 and 2 is a dummy for whether the household spends a positive amount on goods imported from o. The dependent variable in columns 3 and 4 is the log of the relative expenditure share $\tilde{X}_{oh}$, dropping all 0s. Columns 1 and 3 display OLS results, while columns 2 and 4 display two-stage least squares results. Observations weighted using NielsenIQ 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.

in Table A.4. We first estimate equation (6), replacing the dependent variable with a dummy for whether household h makes any purchases of imports from o in columns 1 and 2. We separately estimate equation (6) while dropping all observations with household-by-origin expenditures equal to 0 in columns 3 and 4.

We obtain two results. First, we estimate that the spillover effect of immigrants onto natives is driven primarily by natives more intensively purchasing goods from local immigrants' origin countries, as indicated by the first row. In column 2, we surprisingly find that a 1 percentage point higher share of immigrants from a given origin living locally actually reduces the likelihood of households purchasing goods from that origin by 0.42 percentage points. In contrast, an identical increase in the immigrant population share from a given origin country increases the expenditure share on goods from that origin by over 16 percent, conditional on exhibiting positive expenditures otherwise (column 4).

Our second main finding from Table A.4 is that immigrants' preferences for imports (rows 2 and 3) affect both the likelihood and intensity of purchasing imported products.

A.3.4 Controlling for instrumental variable mean

In a recent paper, Borusyak and Hull (2023) argue that in a formula instrument combining variation from different sources, some units may be more exposed to exogenous variation than other units due to, for example, fixed geographic features. In our context, combining the push and pull shocks to construct our push-pull instrumental variable, one might worry that some origins typically send large numbers of immigrants to the U.S. and some counties typically receive large numbers of immigrants, both due to factors such as geography. To address this issue, we follow the advice of Borusyak and Hull (2023) and compute the average instrument value across each permutation of push and pull shocks within country-county pair.[48]

When controlling for the instrument mean, the coefficient on the immigrant population share falls from 1.15 to 0.56 while remaining statistically significantly different from 0. We note, however, that while the mean instrument term may capture potentially endogenous geographic factors as outlined above, it may also absorb some desirable, exogenous variation in immigration across county-country pairs. For this reason, we do not control for the mean instrument in our baseline estimation. In particular, immigrants may be historically and consistently drawn to particular metro areas due to factors unrelated to grocery imports.

A.3.5 Comparison with the results of Burchardi et al. (2019)

Burchardi et al. (2019) estimated a null effect of immigrants on trade, in contrast to our own results.[49] In this appendix section, we consider three possible explanations for our diverging results. First, the primary explanatory variable of Burchardi et al. (2019) is the log of the number of individuals with a given ancestry measured in thousands and plus one instead of the foreign-born population share that we use. Second, Burchardi et al. (2019) use state-level data whereas we leverage household-level data. Third, Burchardi et al. (2019) use a two-step Heckman estimation strategy to account for selection into bilateral trading, while we apply the pseudo-Poisson maximum likelihood (PPML) estimation strategy. We find that the choice of estimation strategy explains the difference between our results and those

[48] In practice, we reshuffle push and pull shocks across years according to a uniform distribution. Given the 10 decades for which we can construct our instrument, we therefore generate 100 potential instruments (10 of them factual and 90 counterfactual) for each county-country pair and take the arithmetic mean.

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

Table A.5: Household Gravity Estimates with Ancestry

	Dependent variable: Exp. share on goods from o relative to US			
	(1)	(2)	(3)	(4)
Ancestry/Pop. 2010	0.70***	0.64***		
	(0.10)	(0.20)		
Log Ancestry 2010			0.037***	0.033***
			(0.0038)	(0.0038)
First-stage residuals		0.073		0.0061*
		(0.22)		(0.0033)
=1 if immigrant from anywhere	0.24***	0.24***	0.21***	0.22***
	(0.030)	(0.030)	(0.030)	(0.030)
=1 if immigrant from origin o	0.60***	0.61***	0.63***	0.64***
	(0.069)	(0.071)	(0.066)	(0.070)
N	1,421,640	1,421,640	1,422,000	1,422,000
Country FE	✓	✓	✓	✓
Household controls	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		25.0		16.5

Notes: The table presents estimation results at the household-country level. We estimate each specification using pseudo-Poisson maximum likelihood estimation. The first-stage residual term is taken from a first-stage regression of all the instruments on the immigrant-population share in column 2. Observations are weighted using NielsenIQ 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.

of Burchardi et al. (2019) and explain why we prefer PPML over Heckman selection.

We start by testing whether the choice of explanatory variable (log ancestry in thousands plus one vs. immigrant population share) can explain our results. To do so, we first replace our previous explanatory variable, *Immigrants/Pop. 2010*, with *Ancestry/Pop. 2010*. Next, we take the functional form used in Burchardi et al. (2019), *Log Ancestry 2010*. Table A.5 shows that we still obtain positive and significant coefficients with these alternative measures using our household-level data and estimation strategy.

Next, we test whether the level of data aggregation or the estimation can resolve our diverging results. In Table A.6 we mimic the specification in Burchardi et al. (2019) more closely by aggregating our data to the state level. We run regressions first using our PPML approach (columns 1 and 2). As explanatory variables we employ both *Log Ancestry 2010* and our preferred measure *Immigrants/Pop. 2010*. As in our baseline household-level results, we continue to find a significantly positive effect of immigrants and ancestors on import

Table A.6: State-level Gravity estimates

	Dependent variable: Exp. share on goods from o relative to US			
	PPML + control fct		Heckman correction	
	(1)	(2)	(3)	(4)
Log Ancestry 2010	0.055***		-0.086	
	(0.0097)		(0.057)	
Immigrants/Pop. 2010		2.46***		-4.31
		(0.44)		(3.27)
N	3,626	3,626	2,922	2,922
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.

volumes. Turning to columns 3 and 4, we apply the Heckman correction strategy of Burchardi et al. (2019). Here we obtain negative and insignificant coefficients. As a result, we conclude that the choice of estimation approach is important for the contrasting results between our study and Burchardi et al. (2019).

We argue that Poisson pseudo-maximum likelihood (PPML) estimation is more appropriate in our setting. In a widely cited article, Silva and Tenreyro (2006) demonstrate that PPML performs quite well across a variety of settings, accommodating heteroskedasticity and measurement error; Santos Silva and Tenreyro (2011) provide further simulation results in support of PPML even when the proportion of zeros is very high, as in our data. Santos Silva and Winkelmann (2024) shows that PPML performs well even when the conditional expectation function is misspecified. Fally (2015) shows that PPML is the only estimation strategy which satisfies the adding-up constraints of structural gravity.

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

Table A.7: Immigrant-induced welfare increase and voting outcomes in 2016 and 2020

	Dependent variable: Republican votes (%)			
	(1)	(2)	(3)	(4)
Welfare change ($)	-0.16***	-0.17***	-0.10***	-0.069***
	(0.0085)	(0.0091)	(0.018)	(0.018)
Log population			-4.33***	-4.96***
			(0.18)	(0.18)
Native unemployment rate			-101.5***	-80.5***
			(12.0)	(12.0)
Immigrant share			-8.48	-14.7
			(9.07)	(9.03)
N	3,038	3,038	3,038	3,038
Election year	2016	2016	2016	2020
State FE		✓	✓	✓

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

A.4 Consumption Welfare Benefits and Voting

In this section, we show that higher consumption welfare benefits of immigrants correspond to less support for Donald Trump in recent presidential elections. To do so, we regress the county-level share of votes that the Republican party received with Donald Trump as candidate in the presidential elections of 2016 and 2020 on the immigrant-induced welfare increase based on our counterfactual in the last row of Table 4. The results in Table A.7 show that for every dollar increase in immigrant-induced welfare, the unconditional vote share received by the Republican party in 2016 decreased by 0.16 percentage points. Adding further controls (log county population, native unemployment rate and immigrant population share) decreases the coefficient to 0.10, while it remains highly significant.[50] The corresponding coefficient for the 2020 election is around 0.07 and thus somewhat lower, possibly because the immigration debate was less salient during the 2020 election campaigns.[51] These results suggest that the consumption benefits of immigrants may serve as an additional factor driving voting decisions, though further research is needed to confirm this hypothesis.

[50]The immigration population share control addresses concerns about a mechanical correlation whereby immigrants become naturalized citizens and disproportionately vote for the Democratic candidate.

[51]`https://www.vox.com/2020/10/8/21507407/trump-immigration-debate-2020-biden-pence-harris`

B Theory Appendix

B.1 Deriving Adjusted ACR Welfare Formula

In this Appendix section, we derive equation (8) from Section 3.7 which adjusts the standard ACR welfare formula to account for heterogeneous preferences.

We start with the identity that $X_c = X_{us,c} + X_{m,c}$, where X_c is total expenditure in county c, $X_{us,c}$ is total county expenditure on domestic goods, and $X_{m,c}$ is total county expenditure on imports. Given a shock to trade costs in which X_c remains constant, we have:

$$d\ln X_c = d\ln(X_{us,c} + X_{m,c}) = 0$$

To simplify this expression, we use the fact that for any variable x, $d\ln x = dx\frac{1}{x}$. This yields:

$$\begin{aligned}
d\ln X_c &= \frac{dX_{us,c}}{X_c} + \frac{dX_{m,c}}{X_c} \\
&= \frac{X_{us,c}d\ln X_{us,c}}{X_c} + \frac{X_{m,c}d\ln X_{m,c}}{X_c} \\
&= s_{us,c}d\ln X_{us,c} + s_{m,c}d\ln X_{m,c} = 0 \\
&\implies d\ln X_{us,c} = -\frac{1-s_{us,c}}{s_{us,c}}d\ln X_{m,c}
\end{aligned}$$

The last step uses the identity $s_{us,c} + s_{m,c} = 1$, since $s_{us,c}$ and $s_{m,c}$ are simply the expenditure shares of domestic and foreign goods. The same derivation can be used to generate a similar expression for $d\ln X_{us,n}$, the change in domestic expenditure of only native households.

To proceed we note that only part of the expression for X_{oc}, equation (1), is relevant for welfare analysis. In particular, when immigrants affect the bilateral affinity term ϕ^z_{oc}, and therefore local preferences, the welfare criterion becomes endogenous, making conventional welfare analysis infeasible. We therefore fix ϕ^z_{oc} and focus on shifts in the supply-side accessibility of imports, $d\ln\phi^b_{m,c}$.

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

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

Multiplying the numerator and denominator of the right-hand side of this expression by $d\ln X_{us,c}$, we obtain the following relationship between changes in native household domestic expenditure and county-level changes in domestic expenditure:

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

The first two terms in parentheses capture differences in baseline domestic expenditure between natives and the county average, and the second two terms in brackets capture the share of immigrant-induced trade which is welfare-relevant to native households.

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

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

where immigrant households are denoted by f and I_c is the share of the population who are immigrants. We also 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 $s_{us,f} = 1 - s_{m,f} = 1 - e^{\zeta} s_{m,n}$ and $s_{us,n} = 1 - s_{m,n}$. Combining the above 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 rearranging the previous equation:

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

We therefore obtain:

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

Deriving the Welfare-Relevant Component of Trade Shocks. Assume that all households are identical except for their immigrant-derived preferences governed by ζ, therefore $\kappa_h=\kappa\equiv\frac{1}{|\Lambda_c|}$ and $J_h=J$. We also assume that with enough households, the average idiosyncratic component of preferences $\eta^z_{m,h}$ is equal to zero.

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

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

since $\zeta=0$ for any household that is not an immigrant. As a final step, we derive the partial elasticity of this term with respect to I_c in order to show that:

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

As long as $\zeta>0$, this expression is positive and the composition effect of immigrants has a positive effect on county-level import expenditure.

Notice that the final two terms of equation (B.1) reduce to the ratio $\frac{d\ln\phi^b_{m,c}}{d\ln X_{m,c}}$. We have an explicit expression for this ratio from the general gravity model from Section 3:

$$\frac{d\ln\phi^b_{m,c}}{d\ln X_{m,c}}=\frac{d\ln\phi^b_{m,c}/dI_c}{d\ln X_{m,c}/dI_c}=\frac{\beta^b}{\beta^b+\beta^z+\frac{e^{\zeta}-1}{I_c(e^{\zeta}-1)+1}}=\frac{\beta^b}{\beta+\frac{e^{\zeta}-1}{I_c(e^{\zeta}-1)+1}} \tag{B.3}$$

Plugging equations (B.2) and (B.3) into equation (B.1), we obtain equation (8).

B.2 Deriving Heterogeneous Firms Model Equations

Deriving equations (11) and (12). Taking the ratio of the household's first-order condition for two varieties ω_1 from country o and ω_2 from country o', we obtain

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

Define

$$P_h\equiv\left(\sum_{o\in\mathcal{O}}z_{oh}\int_{\omega\in\Omega_{o,c(h)}}p_{o,c(h)}(\omega)^{1-\sigma}d\omega\right)^{\frac{1}{1-\sigma}}\tag{B.4}$$

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

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

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

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

$$x_{oh}(\varphi)=(1-\mu_0)X_hz_{oh}(p_{o,c(h)}(\varphi)/P_h)^{1-\sigma}\tag{B.7}$$

$$\tag{B.8}$$

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

$$p_{o,c(h)}(\varphi)=\frac{\sigma}{\sigma-1}\frac{w_o}{\varphi}\tau_{oc(h)}\tag{B.9}$$

Substituting equation (B.9) into equation (B.8), summing across all households in $c(h)$, and defining $\lambda_1\equiv(1-\mu_0)\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}$, we obtain the following expression for county expenditure on imports from firm with productivity φ in o:

$$x_{oc}(\varphi)=\lambda_1(w_o\tau_{oc})^{1-\sigma}\varphi^{\sigma-1}\Big(\sum_{h'\in\Lambda_c}z_{oh'}X_{h'}P_{h'}^{\sigma-1}\Big)\tag{B.10}$$

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

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

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

$$\pi_{oc}(\varphi) \geq f_{oc} \tag{B.11}$$

For a firm at the cutoff productivity, (B.11) holds with equality, resulting in the following equation for φ^*_{oc}, where $\lambda_2 \equiv \frac{\sigma}{\sigma-1}\left(\frac{\sigma}{1-\mu_0}\right)^{\frac{1}{\sigma-1}}$:

$$\varphi^*_{oc} = \lambda_2 w_o \tau_{oc} \left(\frac{f_{oc}}{\sum_{h' \in \Lambda_c} z_{oh'} X_{h'} P_{h'}^{\sigma-1}}\right)^{\frac{1}{\sigma-1}} \tag{B.12}$$

Returning to equation (B.4) and replacing varieties ω with productivity φ (since firms with identical productivity charge identical prices), we get:

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

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

Plugging in the equilibrium price, equation (B.9), we have

$$P_h = \frac{\sigma}{\sigma-1}\left(\sum_{o \in \mathcal{O}} z_{oh} \left(w_o \tau_{o,c(h)}\right)^{1-\sigma} M_{o,c(h)} \int_0^{+\infty} \varphi^{\sigma-1} g_{o,c(h)}(\varphi) d\varphi\right)^{\frac{1}{1-\sigma}} \tag{B.13}$$

Turning to the gravity equation, we integrate over equation (B.8) to obtain

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

Given the equilibrium price (B.9), we can substitute the last term with

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

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

$$X_{oh} = \lambda_1 z_{oh} X_h P_h^{\sigma-1} (w_o \tau_{o,c(h)})^{1-\sigma} M_o \frac{\theta}{\theta+1-\sigma} (\varphi^*_{o,c})^{\sigma-\theta-1} \tag{B.14}$$

To obtain equation (11) from (B.13) and equation (12) from (B.14), we then:

- substitute (B.12) for $\varphi^*_{o,c}$
- assume $M_o = \gamma Y_o$, where Y_o is the value of production in country o as in Section 3.1.
- define $S_c = \sum_{h' \in \Lambda_c} X_{h'} P_{h'}^{\sigma-1}$ and $z_{oc} = \sum_{h' \in \Lambda_c} z_{oh'} \frac{X_{h'} P_{h'}^{\sigma-1}}{S_c}$ as in Section 3.2.
- 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 Identification of Fixed Cost and Preference Diffusion Channels

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.[52,53]

[52]We assume throughout that $\beta^\tau = 0$, which implies that immigrants do not affect variables trade costs. This assumption derives from the results discussed in Table 3.

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

We begin by fully differentiating $\tilde{X}_{oh}$ from Equation (14) into terms associated with fixed costs f_{oc}, county-level preferences z_{oc}, and household-level preferences z_{oh}:

$$\begin{aligned} d\tilde{X}_{oh} = & \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial \tilde{f}_{oc}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial {\varphi^*}_{oc}}{\partial \tilde{f}_{oc}} \right] d\tilde{f}_{oc} \\ & + \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oh}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial \varphi^*_{oc}}{\partial z_{oh}} \right] dz_{oh} \\ & + \left[\int_{\varphi^*_{oc}}^{+\infty} \frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oc}} dG(\varphi) - \tilde{x}_{oh}(\varphi^*_{oc}) G'(\varphi^*_{oc}) \frac{\partial \varphi^*_{oc}}{\partial z_{oc}} \right] dz_{oc} \end{aligned} \tag{B.15}$$

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

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

$$\tilde{x}_{oh}(\varphi) = (\tilde{w}_o \tau_{oc})^{1-\sigma} z_{oh} \varphi^{\sigma-1} \left(\int_{\varphi^*_{us,c}}^{+\infty} (\varphi^{'})^{\sigma-1} dG(\varphi^{'}) \right)^{-1} \tag{B.16}$$

whereas the productivity cut-off is equation (B.12).

It is clear from inspecting Equation B.16 and equation B.12 that f_{oc} and z_{oc} only affect φ^*_{oc}, and therefore each variety's extensive margin, whereas household-level preferences z_{oh} only affect the household-specific intensive margin of demand via $\tilde{x}_{oh}$. We can therefore apply the following restrictions: $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial f_{oc}} = 0$; $\frac{\partial \tilde{x}_{oh}(\varphi)}{\partial z_{oc}} = 0$; and $\frac{\partial \varphi^*_{oc}}{\partial z_{oh}} = 0$.

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:

$$\frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} = \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oh}} \frac{\partial \ln z_{oh}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} \tag{B.17}$$

$$\frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}} \frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}} \frac{\partial \ln z_{oc}}{\partial I_{oc}} \tag{B.18}$$

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

$$\begin{aligned}\beta &= \frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial I_{oc}} \\ &= \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}}\frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oh}}\frac{\partial \ln z_{oh}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}}\frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}}\frac{\partial \ln z_{oh}}{\partial I_{oc}} \\ &\quad - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}}\frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}\end{aligned} \tag{B.19}$$

$$\begin{aligned}\beta^N &= \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial I_{oc}} \\ &= \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln f_{oc}}\frac{\partial \ln f_{oc}}{\partial I_{oc}} + \frac{\partial \ln \tilde{X}_{oh}}{\partial \ln z_{oc}}\frac{\partial \ln z_{oc}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oh}}\frac{\partial \ln z_{oh}}{\partial I_{oc}} - \frac{\partial \ln \mathcal{Z}_{oh}}{\partial \ln \bar{z}_{oc}}\frac{\partial \ln \bar{z}_{oc}}{\partial I_{oc}}\end{aligned} \tag{B.20}$$

We evaluate the expressions (B.19) and (B.20) using the definition of $\tilde{X}_{oh}$ provided in equation (14) and the definition of $\mathcal{Z}_{oh}$ provided in Section 4.3. Specifically, we reduce expressions (B.19) and (B.20) in three steps:

1. Fixed costs and the extensive margin:

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

2. County-level preferences and the extensive margin:

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

3. Household-level preferences and the intensive margin:

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

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

$$\beta = \frac{\partial \ln \tilde{X}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1}\right)\beta^z \tag{B.21}$$

$$\beta^N = \frac{\partial \ln \tilde{N}_{oh}}{\partial I_{oc}} = \beta^f + \left(\frac{\theta}{\sigma - 1} - 1\right)\beta^z \tag{B.22}$$

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.

Immigration shocks. We start with the counterfactuals from Section 6.1 relating to removing immigrants' effects and removing immigrants expenditures. From equation (12), we obtain the proportional change in household-origin import expenditures:

$$\hat{X}_{oh} = \hat{P}_h^{\sigma-1} \hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} \left(\hat{z}_{o,c(h)} \hat{S}_{c(h)}\right)^{\frac{\theta}{\sigma-1}-1} \hat{z}_{oh} \tag{B.23}$$

where changes in household imports by origin depend on the change in the household's price level $\hat{P}_h$, changes in fixed costs with the origin $\hat{f}_{o,c(h)}$, changes in average household-level preferences for the origin's products $\hat{z}_{o,c(h)}$, changes in total local expenditures $\hat{S}_{c(h)}$, and changes in the household's preferences for the origin's products $\hat{z}_{oh}$. When o is the United States, equation (B.23) reduces to

$$\hat{X}_{us,h} = \hat{P}_h^{\sigma-1} \hat{S}_{c(h)}^{\frac{\theta}{\sigma-1}-1} \tag{B.24}$$

Hence we use equations (B.23) and (B.24) as well as $\hat{f}_{o,c(h)}^{-\left(\frac{\theta}{\sigma-1}-1\right)} = e^{-\hat{\beta}^f I_{o,c(h)}}$ and $\hat{z}_{oh} = e^{-\hat{\beta}^z I_{o,c(h)}}$ to obtain our counterfactual ratio as a function of observable or calibrated values:

$$\frac{X'_{oh}}{X'_{us,h}} = \frac{X_{oh}}{X_{us,h}} \left(e^{-I_{o,c(h)}(\hat{\beta}^f + \hat{\beta}^z)}\right) z_{o,c(h)}^{\left(\frac{\theta}{\sigma-1}-1\right)} \tag{B.25}$$

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

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

$$\hat{U}_h = \hat{P}_h^{\mu_0 - 1} \tag{B.26}$$

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

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

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

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

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

Trade cost shock. Now we turn to the counterfactuals from Section 6.2 involving an increase in variable trade costs. We proceed similarly as with the immigration shock counterfactual derivations outlined above.

The proportional change in utility, with preferences constant at their observed level, is

$$\hat{U}_h = \left(\frac{X_{us,h}}{X_h} + \sum_{o\neq us}\frac{X_{oh}}{X_h}\hat{\tau}_{o,c(h)}^{-\theta}\right)^{\frac{1-\mu_0}{\sigma-1}}$$

Fixing preferences at their counterfactual level given no preference diffusion yields

$$\hat{U}_h^{\beta^z=0} = \left(\frac{\frac{X_{us,h}}{X_h} + \sum_{o\neq us}\frac{X_{oh}}{X_h}\hat{z}_{o,c(h)}^{\left(\frac{\theta}{\sigma-1}-1\right)}\exp(-\hat{\beta}^z I_{o,c(h)})}{\frac{X_{us,h}}{X_h} + \sum_{o\neq us}\frac{X_{oh}}{X_h}\hat{z}_{o,c(h)}^{\left(\frac{\theta}{\sigma-1}-1\right)}\exp(-\hat{\beta}^z I_{o,c(h)})\hat{\tau}_{o,c(h)}^{-\theta}}\right)^{\frac{\mu_0-1}{\sigma-1}}$$

Online Appendix References

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

Borusyak, Kirill and Peter Hull, "Nonrandom exposure to exogenous shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

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

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

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

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

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

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

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

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

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

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

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

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

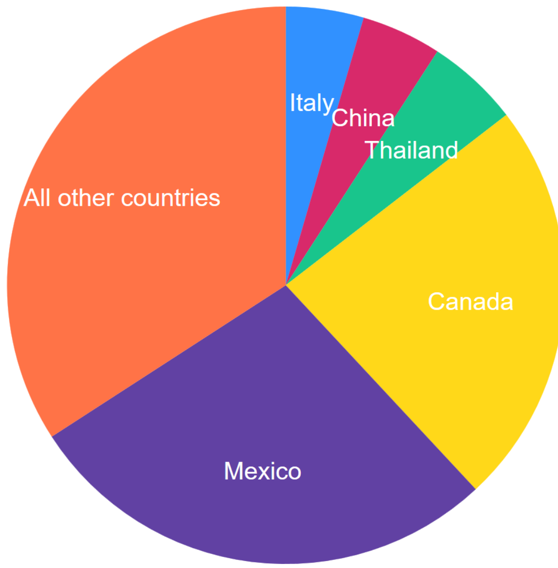
_ **and Silvana Tenreyro**, "The log of gravity," *Review of Economics and statistics*, 2006, pp. 641–658.

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

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

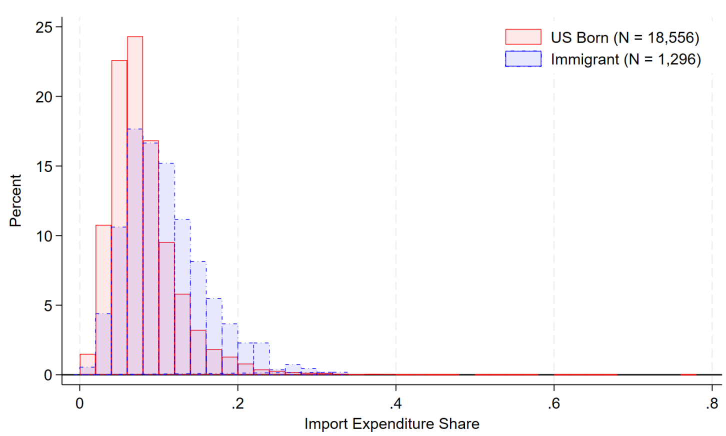
C Additional Tables and Charts

Figure C.1: Spending on Imports by Origin Country



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

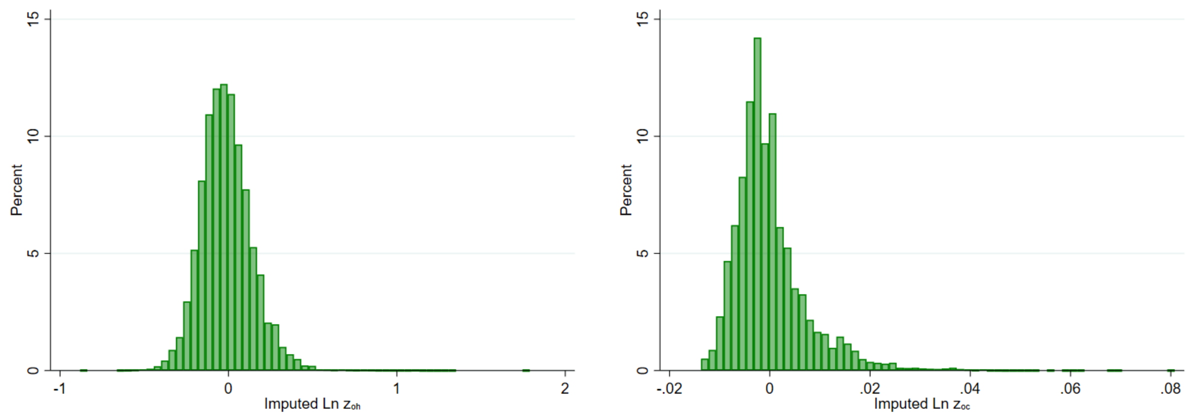
Figure C.2: Distribution of Household-level Import Expenditure Share by Nativity



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

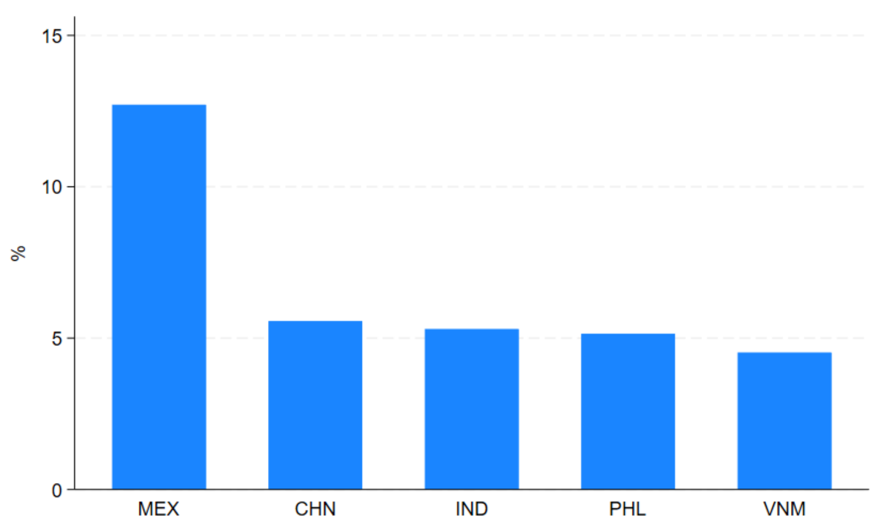
Figure C.3: Distribution of Imputed Preference Terms

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



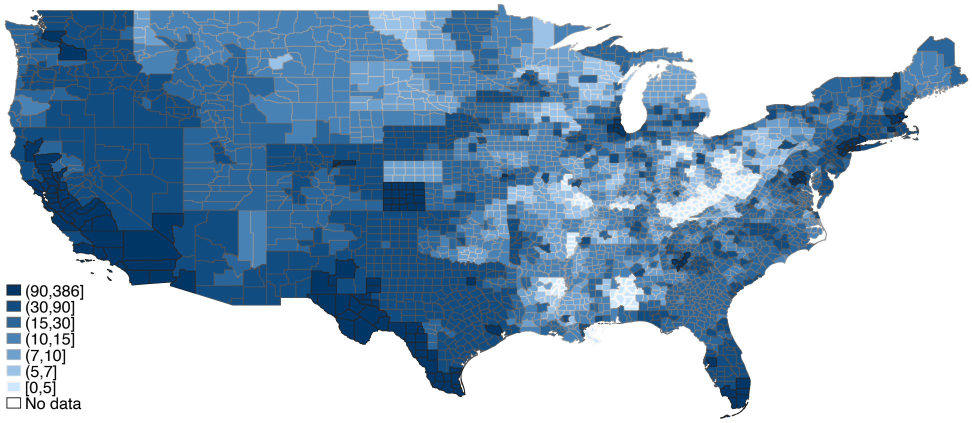
Notes: Figure (a) plots the distribution across NielsenIQ household-origin pairs of the log of $\hat{\hat{z}}_{oh} = \exp(\hat{\delta} J_h + \hat{\zeta}_1 \mathbf{1}[o(h) \neq US] + \hat{\zeta}_2 \mathbf{1}[o(h) = o])$, where the terms $\hat{\delta}$, $\hat{\zeta}_1$, and $\hat{\zeta}_2$ are estimated from equation (19). Figure (b) plots the distribution across county-origin pairs of the log of $\hat{\hat{z}}_{oc} = \sum_{h' \in \Lambda_c} \hat{\hat{z}}_{oh'} \kappa_{h'}$, computed using data from the 2012-2017 American Community Survey.

Figure C.4: 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.

Figure C.5: Spatial Distribution of Fall in Welfare due to Removing Immigrants



Notes: This chart plots the dollar decrease in the dollar-equivalent grocery welfare the trade-creating effect of immigrants and immigrant expenditure are removed following the procedure outlined in Appendix Section B.4.

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

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

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

Table C.2: 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 NielsenIQ household weights. Standard errors clustered two-ways at the household and origin-by-destination levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.