The Trade-Creating Effect of Immigrants: Evidence from Detailed Purchase Data*

Christoph Albert†,
Torsten Jaccard‡,
Brett A. McCully§

July 26, 2023

PRELIMINARY AND INCOMPLETE
DO NOT CITE OR CIRCULATE

Abstract

We estimate the causal effect of immigrants on imports and explore supply and demand mechanisms. We leverage extremely detailed store- and household-level data on consumption in which we observe the country of origin of purchased products. We isolate the demand channel of immigrants on imports by relying on store-level retail sales data. We do so by assuming that fixed costs of importing a good are invariant across markets within the same retail chain, whereas preferences for goods vary across markets. Finally, we examine nondurable consumption baskets for a sample of native-born households and show that they consume more imports from a given origin country when more immigrants from that country live local to the household.

JEL Categories: F22, J31, J61, R11.

Keywords: Immigration, international trade, variety effects

*We thank Ilaria Malisan for research assistance. All errors are our own.
†Collegio Carlo Alberto, email: christoph.albert@carloalberto.org
‡Department of Economics, Dartmouth College, email: torsten.s.jaccard@dartmouth.edu
§Collegio Carlo Alberto, email: brett.mccully@carloalberto.org

1 Introduction

How do immigrants affect the consumption choices of natives? If immigrants raise the number of varieties available from their origin country, natives may respond by consuming more imported varieties. Local immigration may shift the consumption baskets of natives in two ways. First, retailers may begin to stock more varieties from a foreign country when more immigrants from that country arrive locally. In this way, immigration expands the set of varieties available for purchase to natives. Second, immigrants may demonstrate how to consume novel varieties from their homeland, thus raising natives' demand for such imported varieties.

While most prior research has focused on how immigrants affect production[1], there has been scant research exploring the consumption effects of immigration. A key factor holding back progress on this question has been an absence of detailed data on the consumption of natives disaggregated into imported and domestically produced varieties. We overcome this key hurdle by drawing on detailed household- and store-level data on nondurable product purchases from Nielsen, merged with data on the country of origin for each product.

In this paper, we study the effect of immigrants on imports and the consumption choices of the native-born. To guide our empirical analysis, we develop a model of heterogeneous firms in the style of Melitz (2003) and Chaney (2008) in which immigrants affect imports. Immigrants can affect imports both by reducing information frictions and trade costs (what we term the supply-side effect of immigration) and through a preference channel (the demand-side). The model produces an estimable gravity equation of trade.

The empirical analysis proceeds in three steps. First, we estimate the standard gravity model of trade allowing immigrants to affect import volumes which is common in the literature on immigrants and trade (see, e.g., Felbermayr et al. 2015). To do so, we aggregate data on household purchases to the level of the county of residence, product country of origin, and product group. To obtain causal identification, we use the leave-out push-pull instrumental variable introduced by Burchardi et al. (2019) to generate exogenous variation in the population of immigrants from a given origin country living in a given U.S. county. We find that more local immigrants from a given origin country significantly raises the local expenditure share on imported varieties from that origin country.

This estimated effect from the gravity equation represents the total effect of immigrants on im-

[1] For example, Borjas (2003), Ottaviano and Peri (2012), Burstein et al. (2020), and Monras (2020) look at the labor market impact of immigration, while Ariu (2022) and Ottaviano et al. (2018) look at how immigrants affect exports and the imports of intermediate inputs.

ports. That is, it combines (i) the reduction of information frictions and trade costs by immigrants, (ii) the increased consumption of imported goods from a given origin by immigrants from that origin, and (iii) the increased consumption of imported goods by the native-born. Only channels (i) and (iii) correspond to effects on natives.

To isolate just the demand effect of immigrants on imports, we leverage our product-by-store-level data. We re-estimate the gravity equation at this more disaggregated level of analysis. The key assumption allowing us to isolate the contribution of immigrant-induced demand on imports is that retailers establish importing relationships at the national level, not the store level. Hence, adding retailer fixed effects to our gravity model absorbs the supply-component of the immigrant-import effect. We find that the bulk of the immigrant-import effect is driven by the demand channel.

We conclude our empirical analysis by assessing whether natives increase their consumption of imported varieties in response to an influx of immigrants. We do so by leveraging an underutilized component of the Nielsen household-level consumption surveys: the country of birth of panelists. This variable allows us to focus on native-born households. We estimate that natives significantly increase their expenditures on imported varieties from a given origin country when an exogenously higher number of immigrants from that origin country settle locally.

Literature. Taken together, our results suggest a novel mechanism by which immigrants affect natives. Prior research has tended to focus on how immigrants affect what we call the supply side, i.e., how immigrants reduce trade costs (e.g., Peri and Requena-Silvente (2010); Parsons and Vézina (2018)) by looking at how immigrants shape exports. We suggest that immigrants, solely through their demand effects, can raise affect natives by inducing a greater variety of imported products at retailers. We are able to do so because unlike all prior research, we have store- and household-level data on consumption baskets, can observe the country of origin of each product purchased, and can observe the nativity of each household. Prior studies have relied on either aggregate trade statistics or firm-level trade data.

Two existing studies have explored how immigrants affect the number of consumption varieties. Mazzolari and Neumark (2012) explore the effect of immigrants on the number of big-box stores and the variety of restaurants and Chen and Jacks (2012) estimates the effect of immigrants on the number of imported varieties. Both studies, while novel in exploring how immigrants affect product varieties, face significant shortcoming both in generating exogenous variation in immigrant populations and in measuring consumption at a sufficiently detailed level.

2 Data

2.1 Expenditure on imported food products

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

NielsenIQ Household Panel Scanner Data: These data consist of a panel covering approximately 65k US households and all grocery purchases at the barcode level. Detailed household demographic information – including county of residence – are included along with barcode-level expenditure, price, date, and store for each purchase. We restrict our analysis to the years 2014 – 2016 and aggregate these data to a single cross-section at the household level.

NielsenIQ Retail Scanner Data: These data consist of weekly barcode-level prices and units sold for over 30k stores in the US. We focus this analysis on a single year (2015) and aggregate to a store-category cross-section of data. While the household scanner data offers a wider range of products, the sheer volume of data included in the retail scanner datasets requires focusing the analysis on a narrower set of products. I therefore opted for the 70 (out of 161) product categories with the highest observed expenditure in the household-level dataset. Each store is linked to an aggregate retail chain (although the identity of that chain is unknown) and a county where that store operates.

In both the household and retail datasets, it is generally not possible to study these records on their own, as barcodes themselves do not contain information as to the product category they belong to. These datasets must therefore be supplemented by an external dataset with barcode-level information.

Barcode Country of Origin: We merge the NielsenIQ data with barcode-specific country-of-origin information purchased from Label Insight Inc. (LI), a firm that specializes in extracting and organizing information found on the labelling of consumer packaged goods. Label Insight uses an AI to extract the ingredients, branding, and any other text information from the packaging for thousands of barcodes sold across major retail chains in the US. Since imported goods in the US are required to contain some statement equivalent to "Made in ...", the Label Insight AI incidentally recovers a country of origin for each barcode they collect[2]. Naturally, Label Insight can only

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

cover a segment of total consumption and their coverage is best for food and beverages, alcohol, personal care products, and cosmetics. We purchased the origin country, ingredients list, brand, and barcode description (as read directly off package) for over 600k barcodes spanning 161 narrow product categories within these product categories. Given the universality of barcodes, these data can be directly merged with both the household and store-level NielsenIQ datasets.

Household Panel Data Summary: We merge the raw household-level purchase records with the Label Insight data and aggregate expenditure across households to the category-origin-county level (for example, expenditure on tequila produced in Mexico in Los Angeles county). When aggregating, each household is weighted by their projection factor, which is a measure of representativeness of each household assigned by NielsenIQ.[3] With the weighting structure included, this final merged dataset covers \$764 billion USD of expenditure spanning 161 narrow product categories and ∼315k unique barcodes. Note that the product categories studied here are narrow in focus, such as yogurt, beer, deodorant, or ready-to-eat cereal. There are 110 origin countries represented in the final dataset and 8.06% of all expenditure is on imported goods (\$62 billion USD in total).

We make use of the BEA Consumer Expenditure Survey (CEX) to compare the product categories covered in this paper with aggregate expenditure on tradeable sectors. The categories covered by Label Insight account for approximately a third of all expenditure on tradeable sectors, with this share increasing to almost half if one excludes passenger vehicles and energy products, for which retailer data is clearly not available. The merged household-level dataset in this paper therefore amounts to an average expenditure per household-year of \$2,200 USD, which is just around 60% of the predicted aggregate expenditure on the categories of food and beverage, alcohol, personal care products, and cosmetics.[4]

The fact that certain counties exhibit zero sales of expenditure on category-origin pairs is an important source of information for our analysis, and we therefore incorporate zeros by assigning each category a set of "feasible" origin countries N_{ko}. These countries are simply all countries exhibiting non-zero expenditure in our sample for category k. This set N_{ko} is then applied to all counties, which leads to approximately 75% of all category-county-origin observations exhibiting zero expenditure.

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

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

Store Cross-Section Data Summary: The top 70 product categories by expenditure were chosen to be studied using the store-level dataset, for reasons mentioned earlier. Approximately 80% of all recorded expenditure within these categories is captured by the Label Insight data, and I further drop all categories with only one country exhibiting non-zero expenditure and all categories with circulation at less than 5k stores in the merged dataset. These data are then aggregated across the year 2015 to a single cross-sectional dataset with observations at the store-category level.

The final store-category merged dataset consists of expenditure and variety counts for 66 narrow product categories across 30k stores amounting to \$31.6 billion USD of expenditure and $\sim$45k barcodes across 111 distinct retail chains. 82 separate origin countries are present in the data with an aggregate import expenditure share of 12.06%. We again account for zeros but this time at the store-category-origin level. That is, if a store exhibits zero sales of any varieties from product category k, we simply omit that store-category combination. This decision reflects the fact that certain stores do not sell certain product categories – such as alcohol, to give one example – and that these sourcing decisions likely do not reflect the effect of immigrants on stocking decisions by stores. However within product categories we again take the set of countries exhibiting non-zero sales of barcodes within that category across all stores as the "feasible set" of sourcing countries, which leads to again approximately 75% of all store-category-origin observations exhibiting zero expenditure.

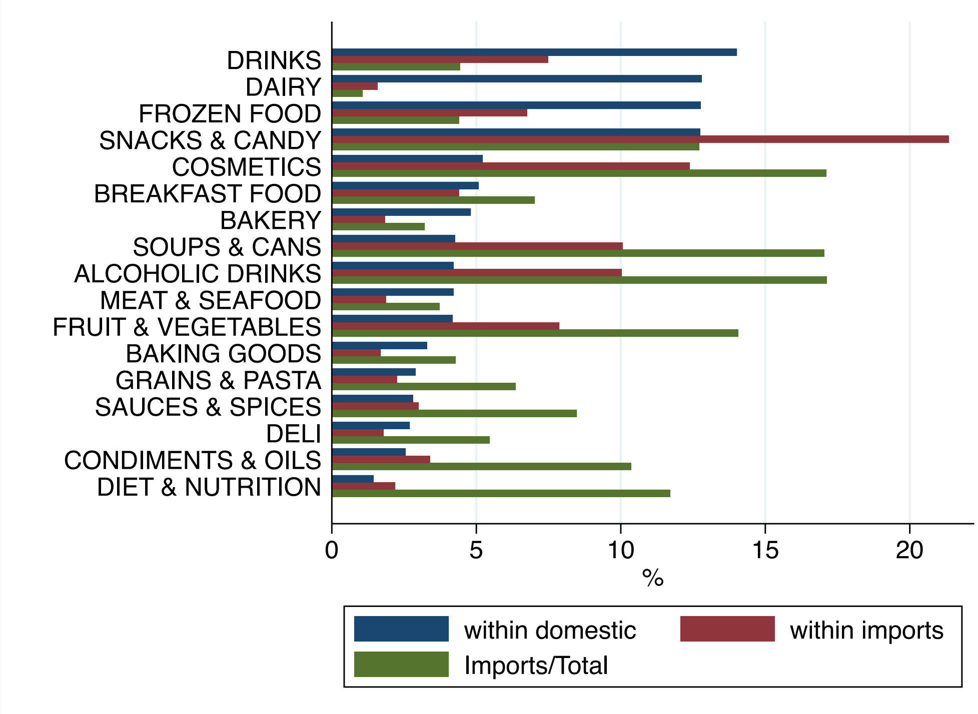
2.2 Data of migrants and ancestry

We use the Census samples of the years 1880, 1900, 1910, 1920, 1930, 1970, 1980, 1990, and 2000 waves, and the pooled 2006–2010 sample of the ACS to obtain population counts of migrants by origin by aggregating up observations of individuals aged 16 and above, applying the provided personal weights.[5] Migrants are defined as those born outside the US and not citizens by birth. To compute decadal migrant inflows from origin o into destination county d between two census years $t-1$ and t, denoted I_{od}^{t}, we county only those respondents who migrated to the US between $t-1$ and t. Following Burchardi et al. (2019), in the first sample year the measure I_{od}^{1880} includes all those that are either first-generation immigrants from o or second-generation immigrants whose parents were born in o. The inflow measures are used in the first stage of our IV strategy outlined in section 4.

We construct two alternative measures for our main explanatory variables. The first one is

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

Figure 1: Spending across product categories



Notes: The figure shows the percent of expenditure going to each product category among all domestic products (blue), among imported products (red) as well as the percent of expenditure on imports out of total expenditure within the product category. Data come from the Nielsen Household Panel 2014-2016.

the 2010 stock of all migrants from o present in county d. The second one is based on the 2010 stocks of individuals by self-reported primary ancestry o and includes both US-born individuals and immigrants, therefore being less restrictive than the measure based on a foreign country of birth.[6] Destination regions d are defined by 1990 counties and we use the transition matrices provided by Burchardi et al. (2019) to bring the aggregates at time-varying geographic definitions (historic counties until 1940, county groups in 1970/1980, and PUMAs subsequently) to the level of 1990 counties.

After merging the resulting population aggregates to the Nielsen data, we obtain a dataset at the origin-destination-product level covering 80 origin countries, 2771 destination countries and 17 product categories.

2.3 Descriptives and reduced-form evidence

Figure 1 shows the distribution of spending across the 17 broad product groups that, which we will use to define product categories throughout the remainder of the paper. Blue and red bars show the percent of spending going to each category among all domestic and imported products, respectively. Green bars show the percent of goods imported out of total expenditure on goods within the category. Categories are ordered from top to bottom by the expenditure share within domestic goods. With percentages between 12.7% and 14%, drinks, dairy, frozen food, and snacks & candy are by far the most important categories for spending on domestic goods, together making up 52% of all spending. Within imported goods, 21% of spending goes to snacks & candy, followed by cosmetics, soups & cans, and alcoholic drinks. Similarly to domestic goods, the top four categories make up just above half of all import spending. When it comes to the import share of spending, the dominating categories are those that are among the top categories among imports but have relatively little weight among domestic spending: alcohol, cosmetics and soups & cans, with over 17% each.

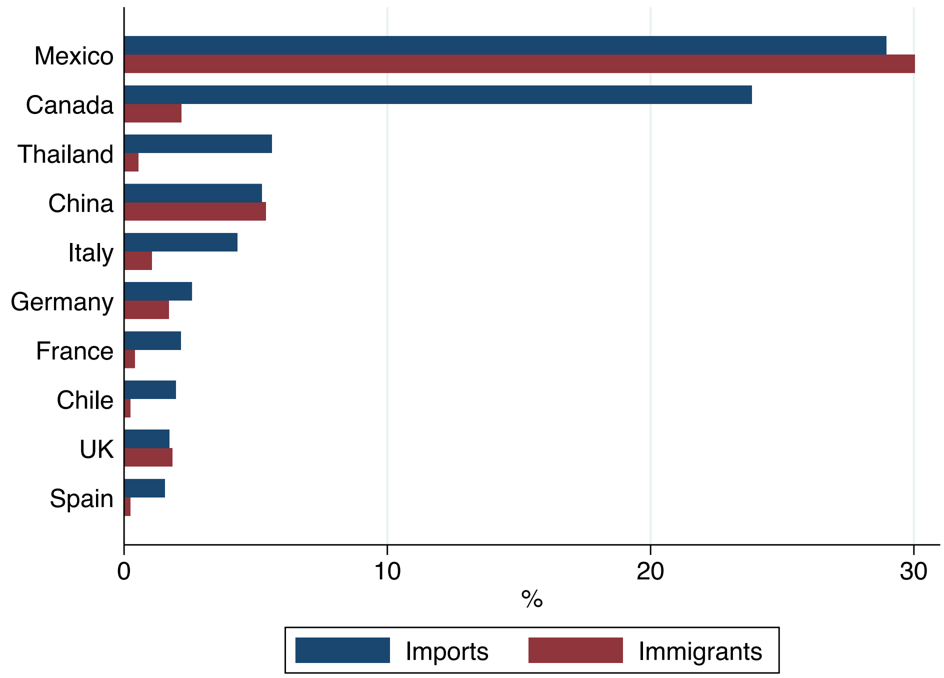
Figure 2 shows the distribution of spending on imports across the ten origins with the highest spending shares as blue bars and the share of immigrants from the respective origin residing in the US in 2010 as red bars. Not surprisingly, the by far highest shares of imports, and combined more than 50%, go to the two neighboring countries Mexico and Canada with 29% and 24%, respectively. Interestingly, the percent of immigrants born in Mexico is very similar to the import share, whereas only 2.2% of immigrants are from Canada.

In terms of import shares, the two border countries are followed by Thailand, China and Italy with around 5.6% to 4.4%. The remaining countries among the top ten have spending shares from 2.6% for German products to 1.5% for Spanish ones. Immigrant shares tend to be much smaller than import shares among these origins, with the exceptions of China and the UK, for which the two are of similar magnitude.

Figure 3 shows a first piece of reduced-form evidence for the relationship between import expenditure and the presence of immigrants at the county level. It plots the overall share of expenditure going to imports in the Household Panel Scanner data against the overall immigrant share among the population at the county level, restricted to counties with at least 500,000 inhabitants and 100 sampled households in the Nielsen data. The plot indicates a strong positive relationship between

[6] In case an immigrant reports an ancestry different to his/her place of birth, e.g., a person born in Austria reporting having German ancestry, we replace ancestry with the country of birth.

Figure 2: Import expenditure and and population share of immigrants by origin



Notes: The figure shows the percent of overall import expenditure going to imports from each origin country for the 10 origins with the highest percentages. Data come from the Nielsen Household Panel 2014-2016.

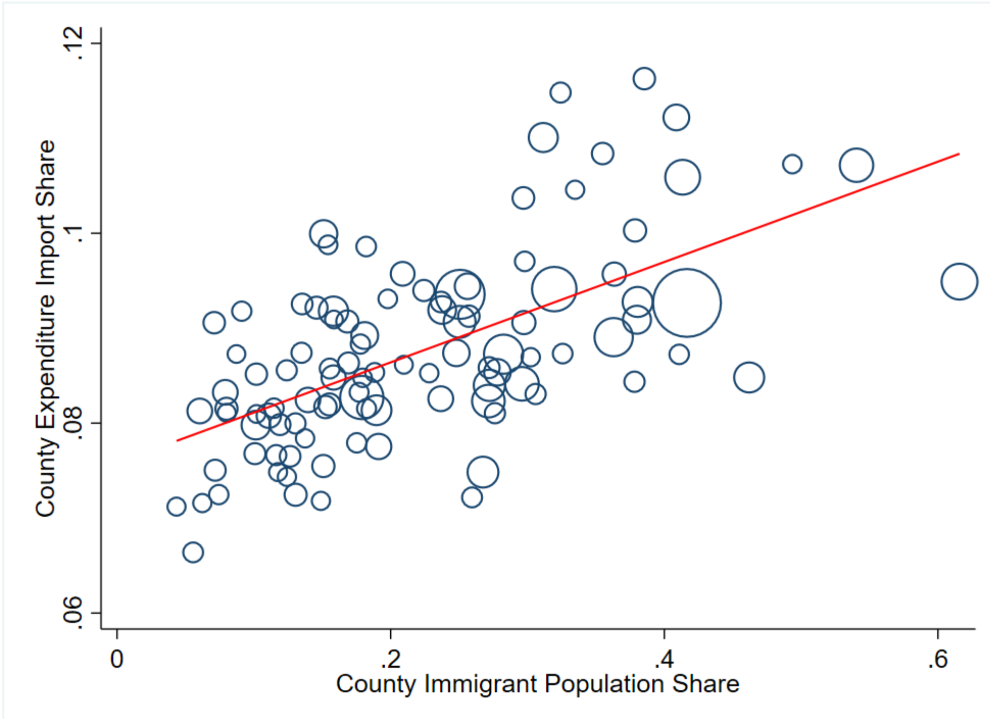
immigrant presence and expenditure on imports.

3 Theory

We build on Chaney (2008) to construct a model of demand and supply of imported varieties within the US which allows for varying immigrant stocks across US counties to play a role in shaping trade flows. The first section provides a model at the aggregated county-category-origin level whereas the second section discusses a store-level model which can be used to uncover mechanisms through which immigrants affect trade flows.

Henceforth, we denote a variety as j, an origin country as o, a destination county in the US as d, and a product category as h. When using retail data, we denote stores s as belonging to retail chain $r(s)$ and located in county $d(s)$.

Figure 3: Correlation between import expenditure and immigrant shares across counties



Notes: The figure plots the aggregate import expenditure share at the county level against the county immigrant population share. Each marker size represents county population. Only counties with at least 500,000 people and at least 100 households in the Nielsen data are included.

3.1 Aggregate Model

Preferences: Consumers living in county d exhibit quality-augmented CES preferences over all available varieties within a given product category h: Ω_h. This CES sub-utility is nested within an aggregate Cobb-Douglas preference structure over all "inside" categories, which are differentiated, and a homogenous outside good, $h = h_0$. Denote the set of all categories then as $H+1$ with the restriction $H > 1$. Assume for now that all regions exhibit non-zero expenditure on the homogenous outside good, denoted by q_0. The "quality" of a given variety $z(\cdot)$ is simply some demand-shifter associated with that variety, which we take to be exogenous.

These nested preferences can then be represented by the following utility function for consumers in destination d:

$$U_d = {q_0}^{\mu_0} \prod_{h=1}^{H} \left[\int_{\omega \in \Omega_h} [z(\omega) q^h(\omega)]^{\frac{\sigma_h - 1}{\sigma_h}} d\omega \right]^{\frac{\sigma_h}{\sigma_h - 1}\mu_h} \tag{1}$$

with the restrictions $\sigma_h > 1$ and $\mu_0 + \sum_{h=1}^{H} \mu_h = 1$. The vector $\boldsymbol{\mu}$ simply represents Cobb-Douglas exponents which are equivalent to expenditure shares in equilibrium. Denote $X_d^h = \mu_h X_d$ as the aggregate expenditure by households in d on sector h. Assume that the perceived quality of a variety is entirely dependent on where that variety was produced and the destination in which it is being consumed. Therefore $z(\omega) = z_{od}^h$.

Technology: The set of origin countries in the world is given by N with individual countries indexed by o. Each country has an exogenous size Y_o and marginal cost of production within a given category w_o^h. Trade is characterized by category-destination-specific iceberg trade costs and fixed costs given by, respectively, τ_{od}^h and f_{od}^h. Given that we focus our analysis on differences in import shares across US counties, it is useful here for notational purposes to define $\tau_{od}^h = \bar{\tau}_o^h \hat{\tau}_{od}^h$ with $\hat{\tau}_{od}^h$ representing iceberg trade costs of shipping a variety from origin o to a county within the US, conditional on that variety having reached the border. We can therefore, with some abuse of notation, subsume $\bar{\tau}_o^h$ within the marginal cost term w_o^h and assume that moving forward this term captures both marginal costs of production in o as well as iceberg costs of transporting varieties from o to the US border.

Each firm draws some productivity ϕ from a Pareto distribution with shape parameter $\gamma_h > \sigma_h + 1$, and the set of potential entrant firms in each origin is proportional to the size of that origin Y_o. The cost of providing q units to destination county d by a firm in origin o with productivity ϕ is therefore:

$$c_{od}^h(q) = \frac{w_o^h \tau_{od}^h}{\phi} q + f_{od}^h \tag{2}$$

Equilibrium: In equilibrium, we can solve for aggregate expenditure in region j originating from country i as the following, which is from Chaney (2008):

$$X_{od}^h = \mu^h \frac{Y_o Y_d}{Y} \left(\frac{w_o^h \hat{\tau}_{od}^h}{\theta_d^h} \right)^{-\gamma^h} (z_{od}^h)^{\gamma^h} (f_{od}^h)^{-[\frac{\gamma^h}{\sigma^h - 1} - 1]} \tag{3}$$

where Y_o, Y_d, and Y capture, respectively, the size of o, the size of d, and aggregate global output. We assume that county-level expenditure is equal to income in equilibrium, such that $X_d = Y_d$ for all counties j[7]. The term θ_d captures the "remoteness" of market j, and is formally given by:

$$\theta_d^{h^{-\gamma_h}} = \sum_k^N \frac{Y_k}{Y} (w_k^h \hat{\tau}_{od}^h)^{-\gamma_h} z_{od}^{h^{\gamma_h}} f_{kd}^{h^{-[\frac{\gamma_h}{\sigma_h - 1} - 1]}} \tag{4}$$

[7]The required assumption to reach the final equation in this memo is in fact weaker than that: we only require that expenditure is a fixed proportion of income such that $X_d = aY_d$ for some constant $0 < a \leq 1$.

Recall that since $\mu_h = \frac{X_d^h}{X_d}$, we can derive the following expression for the expenditure share s_{od}^h:

$$s_{od}^h = \frac{X_{od}^h}{X_d^h} = \frac{Y_o}{Y}\left(\frac{w_o^h \hat{\tau}_{od}^h}{\theta_d^h}\right)^{-\gamma^h} (z_{od}^h)^{\gamma^h} (f_{od}^h)^{-[\frac{\gamma^h}{\sigma^h - 1} - 1]} \tag{5}$$

Immigrants and Trade Flows: We now allow for variables within this model to reflect the fact that immigrants may alter the demand and supply factors shaping trade flows across US counties. There are three candidate variables through which this mechanism might operate: the perceived quality of varieties produced in o, the fixed costs of sourcing varieties from o, and the within-USA iceberg trade costs of shipping varieties from o to d. We assume that the within-USA internal trade costs do not depend on the number of immigrants from o living in d, as these iceberg trade costs simply represent the cost of distance when shipping goods within the US which we assume is driven purely by technology.

The perceived quality of varieties from o and the fixed costs of sourcing those varieties from o, however, are both likely to be influenced by the number of immigrants. That is, immigrants may increase demand for varieties from their origin country but they also may provide links and information which lower the fixed costs of sourcing from these countries. We therefore define a composite measure Z_{od}^h, which we define as the relative "attractiveness" of market d for firms in o selling product h. We define this term explicitly as:

$$Z_{od}^h = [z_{od}^h]^{\gamma_h} [f_{od}^h]^{-[\frac{\gamma_h}{\sigma_h - 1} - 1]}$$

We define this attractiveness measure to be separable between an immigrant-induced attractiveness $Z(M_{od})$ and a residual attractiveness $\tilde{Z}_{od}^h$: $Z_{od}^h = Z(M_{od})\tilde{Z}_{od}^h$ with M_{od} denoting the immigrant stock of households from o living in county d. For ease of exposition, define $\Gamma_h = -[\frac{\gamma_h}{\sigma_h - 1} - 1]$ moving forward. We can now place a functional form on the relationship between immigrants and the attractiveness of county d from producers from o as the following:

$$Z_{od}^h = [M_{od}^{\beta_d/\gamma^h} \tilde{z}_{od}^h]^{\gamma^h} [M_{od}^{\beta_s/\Gamma_h} \tilde{f}_{od}^h]^{\Gamma_h} = [M_{od}]^{\beta_d + \beta_s} [\tilde{z}_{od}^h]^{\gamma^h} [\tilde{f}_{od}^h]^{\Gamma_h} \tag{6}$$

Thus, β_d and β_s represent the two key parameters of interest. They govern, respectively, the effect that immigrants have on price-adjusted demand for varieties from their origin country and the effect of immigrants on the fixed cost of sourcing those same varieties. Notice that when using the aggregated household-level data these two parameters will not be separately identified, but will instead be estimated as a single parameter capturing the effect of immigrants on aggregate trade flows. We then use store-level data to separately identify β_d and therefore provide a decomposition of the effect of immigrants on demand and supply-side factors of trade.

By combining the expression for Z^h_{od} with the expenditure share equation derived earlier and taking the logarithm of both sides, we arrive at the following estimating equation:

$$\ln\left(s^h_{od}\right) = \zeta^h_d + \zeta^h_o + (\beta_d + \beta_s)\ln(M_{od}) + \epsilon^h_{od} \tag{7}$$

where

$$\epsilon^h_{od} = \ln\left([\hat{\tau}^h_{od}]^{-\gamma^h}[\tilde{z}^h_{od}]^{\gamma^h}[\tilde{f}^h_{od}]^{\Gamma_h}\right) \tag{8}$$

Notice that any effect of the aggregate immigrant stock across the US from origin o will be absorbed in ζ^h_o, whereas the relative "remoteness" of county d from all origin countries is captured by ζ^h_d. This is useful in the sense that we are clearly identifying the effect of immigrants on trade flows *across* counties *within* the USA. The error term reflects differences in iceberg trade costs across counties within the US as well as perceived quality and fixed costs of sourcing varieties from o in d that cannot be explained by the immigrant stock.

The identifying assumption for OLS in this case would be $cov[M_{od}, \epsilon^h_{od}] = 0$. This is clearly not the case: so long as the immigrant stock M_{od} and trade costs – whether iceberg or fixed – follow a gravity structure, then the immigrant stock is correlated with the error term and therefore endogenous to market share. We implement an instrument variables strategy and estimate this equation using 2SLS in order to address this endogeneity concern.

3.2 Disentangling Mechanisms Using Retailer Data

Model Overview: Consider the same set-up as above except now all decisions are made at the store level, with stores denoted by s. Each store is in county $d(s)$ and belongs to retail chain $r(s)$. We assume for now that there are no strategic interactions between retailers nor spill-overs in terms of costs. If store s pays a fixed cost of sourcing a variety from origin o in category h, this decision has no effect on the supply or demand considerations of store s', even if s' is in the same market. We therefore explicitly assume that the demand shifters and iceberg trade costs of importing are identical across all stores for sourcing varieties of h from o in d.

Each store therefore must make a decision as to whether they will stock goods from a specific origin country. We assume that there are no "congestion costs" or shelving costs, so each store simply responds to the demand schedule facing each variety as well as iceberg and fixed costs. Similarly, we hold constant the set of stores and do not consider entry or exit of individual stores, or retail chains, across markets.

What is useful about using store-level data to study the effect of immigration on stocking decisions is that we can assume that all stores *within* a given retail chain exhibit identical fixed costs of sourcing from a specific origin country o and category h. This is not to say that immigrants do not affect fixed costs of importing, as retailers may be able to learn from their customers about potential products and firms they did not originally know, but once a given retail chain has formed that sourcing relationship this same fixed cost applies to all stores within the chain, regardless of location. For ease of notation in the equations to follow, we drop the category-level scripts h, but assume that all parameters are category-specific.

We can write our expenditure share equation from above as the following:

$$s_{so} = \frac{Y_o}{Y}\left(\frac{w_o \hat{\tau}_{o,d(s)}}{\theta_{d(s)}}\right)^{-\gamma} z_{o,d(s)}{}^{\gamma} f_{o,d(s)}{}^{\Gamma} \tag{9}$$

We can apply the same decomposition of the last two terms into an immigrant-induced component and an idiosyncratic component, leading to the same expression for $Z(\cdot)$ as described above. Notice that all stores within a given county will face the exact same price-adjusted demand curve for their varieties $z_{o,d(s)}$, as well as the same relationship between the immigrant stock and this price-adjusted demand. The same cannot be said for fixed costs. We therefore decompose fixed costs into a retailer specific component $f_{o,r(s)}$ and an idiosyncratic component $\tilde{f}_{o,d(s)}$ with $f_{o,r(s)}$ replacing the function $M_{od}^{\beta_s/\Gamma}$ from earlier.

Importantly, we have not assumed that immigrants no longer have an effect on the fixed costs of importing. Rather, we have assumed that the fixed costs of sourcing from o for retail chain r are identical across all of their locations. In fact we allow for the possibility that these fixed costs are driven by exposure to immigration via $f_{o,r(s)} = F(\boldsymbol{N^D_{r(s)}}, \boldsymbol{M^D_o})$ with $\boldsymbol{N^D_{r(s)}}$ representing some vector of store counts belonging to retailer $r(s)$ across all destinations D and $\boldsymbol{M^D_o}$ the vector of immigrant stock from o across all D. $F(\cdot)$ then represents some aggregating function mapping retailer presence across locations, and therefore exposure to immigrants and the effect those immigrants have on fixed costs via β_s, into a retailer-specific fixed cost of sourcing from origin o. We leave this mapping unspecified but make the key assumption that regardless of how retail presence in different counties aggregates to a retailer-specific fixed cost, this fixed cost is identical across all stores within that retail chain.

With this assumption made, we can now add the effects of immigrants in a similar manner as to before, except with replacing $M_{od}^{\beta_s/\Gamma} = f_{o,r(s)}$. Taking the logarithm again and linearizing (and adding back in the category h scripts), we arrive at the following estimating equation:

$$\ln s_{so}^h = \zeta_{d(s)}^h + \zeta_{o,r(s)}^h + \beta^d \ln(M_{od}) + \epsilon_{so}^h \tag{10}$$

By recovering an estimate for β^d, we can isolate both parameters β_s and β_d, which govern the effect of immigrants on demand for varieties and on the fixed costs of sourcing varieties from their origin country, respectively.

4 Empirical Strategy

We aim to estimate the causal effect of immigrants on imports and to disentangle the mechanisms driving these imports of consumption goods. To do so, we proceed in three steps. First, we estimate an overall effect of immigrants on imports, which combines both the supply and demand effects of immigration. Second, we leverage variation across retailers to isolate the effect of immigrants on the demand for imports. Third, we use a sample of native-born consumers to show that the demand-side effect of immigrants on imports spills over to natives, who consume substantial quantities of imported goods. In each of these specifications, we use the leave-out push-pull instrumental variables approach of Burchardi et al. (2019) to generate causal estimates.

Estimating the total effect of immigrants on imports. We start by estimating a version of equation 7. The coefficient on log migrants ($\beta_d + \beta_s$) identifies the combined supply and demand effects of immigrants. While we control for county-by-product group fixed effects (ζ_d^h) and origin country-by-product group fixed effects (ζ_o^h), there may still be factors at the bilateral level which drive both migration and imports. For example, distance between o and d will affect both trade and migration costs between o and d, and hence be correlated with $\hat{\tau}_{od}^h$ and $\tilde{f}_{od}^h$ in equation (8). Climatic similarity might also drive both migration and imports if, for example, migrants have a taste for settling a climate similar to their origin country and for food which is best consumed in such climates (e.g., ice cream in hotter regions). This similarity in preferences is captured by the $\tilde{z}_{od}^h$ of error term (8).

We address these endogeneity concerns in three ways. First, we control for the log distance and the difference in latitude between o and d, as well as squared and cubed terms. Second, we add county-by-continent fixed effects and Census division-by-country fixed effects. These fixed effects will absorb any bilateral similarities or differences at the region level. So if migrants in the northeast United States are drawn by the availability of fresh lobster, which is similar to their origin country diet, the Census division-by-country fixed effect will adjust for this potential confounder.

Still, there may be confounders at the country-by-county level which has not been addressed by our rich set of fixed effects and distance controls. For example, two locations may have high migration and imports because of an idiosyncratically high number of flight connections. To deal with such county-country specific confounders, we adopt the instrumental variables approach taken by Burchardi et al. (2019).

Leave-out push-pull instrumental variables. The intuition of the instrument is that a social connection, in this case an immigration decision, between an origin and a destination is likely to occur when the origin is sending many immigrants at the same time the destination is pulling in many immigrants. For example, suppose we want to predict the number of Italians settling in Chicago. To do so, I look at the number of Italians flowing into the United States and the number of immigrants from all origin countries settling in Chicago for the same decade. In particular, the instrument will predict Italians to settle in Chicago if large numbers of immigrants from other countries are also settling there. Similarly, if many immigrants from other origins are settling in Chicago, then an immigrant arriving from Italy will be predicted to settle in Chicago.

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

$$\tilde{IV}^{D}_{o,d} = I^{D}_{o} \times \frac{I^{D}_{d}}{I^{D}},$$

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

Still, there may be threats to the exogeneity of the instrument as defined thus far. One potential exclusion restriction violation occurs when endogenous bilateral immigration is a large share of the instrument's components. For example, if all Italian immigrants coming to the U.S. choose to settle in Chicago due to the large number of flights, then the instrument will pick up the flight connections in its prediction of bilateral immigration. A simple solution then is to leave out bilateral immigration ($I^D_{o,d}$) when computing the instrument.

However, there might also be spatial correlation in confounding variables. For example, both Italian and French immigrants and goods may go to Chicago for the same reason: many flight connections. Then, even leaving out Italy-to-Chicago immigration flows when computing the instrument is not sufficient, because now the French immigration flows to Chicago (used to predict Italy-to-Chicago flows) are contaminated with the confounding flight connections.

To avoid such endogeneity, I again follow Burchardi et al. (2019) and leave out both the continent of origin country o and the Census region of county d to construct the instrumental variable that I use in my baseline estimation:

$$IV^D_{o,d} = I^D_{o,-r(d)} \times \frac{I^D_{-c(o),d}}{I^D_{-c(o)}} \tag{11}$$

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

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

The identification assumption is that any confounding factors that make a given county more attractive for both immigration and importing firms from a given country do not simultaneously affect the interaction of (i) the settlement of immigrants from other continents with (ii) the total number of immigrants arriving from the same country but settling in a different Census regions. A violation may occur if, say, immigrants skilled at importing goods from Italy tend to settle in Chicago and immigrants skilled in importing goods from South Korea settle in Miami in the same decade and for the same reason: a large number of flight connections. This violation is only quantitatively meaningful if Italians are a large fraction of immigrants settling in Chicago, and if South Korean immigrants are a large fraction of the immigrants settling in Miami.

We use equation (11) to predict immigrant inflows into the U.S. decades spanning 1880 to 2000. Burchardi et al. (2019) extensively explore the validity of this instrumental variable and conduct extensive robustness checks for the instrument in the same setting and find that it holds up well.

We show the first-stage results of the leave-out push-pull instruments using our Homescanner data aggregated to the county-by-country-by-product group level in Table 1. We find that the push-pull instrument strongly and positively predicts the contemporary bilateral immigrant population, as well as the population of county residents with ancestry from the origin country as of the 2010 Census.

Estimating the demand effect of immigrants on imports. With an estimate of the total effect of immigrants on imports in hand, we next isolate the demand effect of immigrants by estimate an extended version of equation (10). As before, we use the leave-out push-pull instrument defined in equation (11) to generate exogenous variation in bilateral immigrant populations. Because

Estimating demand externalities on natives. The demand effect of immigrants on imports consists of two components. First, a direct effect, in which immigrants from a given origin country o have a strong preference for goods imported from o. Second, an indirect effect, in which those not from o nevertheless consume increased amounts of goods imported from o. It is this second effect which is novel to the literature on the effects of immigration and which we seek to uncover in this exercise.

The indirect demand effect of immigrants on imports, what we term a demand externality, can arise for various reasons. For example, retailers may face a fixed cost of stocking a given variety, and will only do so once demand for the variety reaches a break-even point. Prior to this break-even point, there is a mass of consumers with unmet demand who would be made better off if the new variety was made available to them. If the good is imported from o and a mass of immigrants from o arrives with a preference for the good, the retailer will begin to stock the good, thus raising the welfare of the natives with previously unmet demand. Alternatively, immigrants may "demonstrate" how to consume novel varieties, which then leads to take up by the native-born. We do not, however, take a stance on the origins of these demand externalities.

To establish the extent of the demand externality from immigrants, we leverage a relatively

Table 1: First-Stage Instrumental Variable Regression Results

	Log ancestry 2010		Log migrants 2010	
	(1)	(2)	(3)	(4)
$I^{1880}_{o,-r(d)} \times \frac{I^{1880}_{-c(o),d}}{I^{1880}_{-c(o)}}$	0.0075	0.014**	0.012***	0.013**
	(0.0052)	(0.0066)	(0.0029)	(0.0059)
$I^{1900}_{o,-r(d)} \times \frac{I^{1900}_{-c(o),d}}{I^{1900}_{-c(o)}}$	0.018*	0.0084	0.015	-0.012
	(0.010)	(0.018)	(0.012)	(0.019)
$I^{1910}_{o,-r(d)} \times \frac{I^{1910}_{-c(o),d}}{I^{1910}_{-c(o)}}$	0.042***	0.033***	0.043**	0.030***
	(0.011)	(0.0081)	(0.017)	(0.011)
$I^{1920}_{o,-r(d)} \times \frac{I^{1920}_{-c(o),d}}{I^{1920}_{-c(o)}}$	0.092***	0.12***	0.076***	0.10***
	(0.011)	(0.018)	(0.010)	(0.018)
$I^{1930}_{o,-r(d)} \times \frac{I^{1930}_{-c(o),d}}{I^{1930}_{-c(o)}}$	0.0066	0.029	0.039***	0.066***
	(0.012)	(0.019)	(0.014)	(0.020)
$I^{1970}_{o,-r(d)} \times \frac{I^{1970}_{-c(o),d}}{I^{1970}_{-c(o)}}$	0.014	0.024**	0.027***	0.038***
	(0.0087)	(0.011)	(0.0088)	(0.012)
$I^{1980}_{o,-r(d)} \times \frac{I^{1980}_{-c(o),d}}{I^{1980}_{-c(o)}}$	0.050***	0.079**	0.052***	0.085**
	(0.016)	(0.030)	(0.018)	(0.033)
$I^{1990}_{o,-r(d)} \times \frac{I^{1990}_{-c(o),d}}{I^{1990}_{-c(o)}}$	0.037*	0.053*	0.036	0.054*
	(0.022)	(0.028)	(0.024)	(0.030)
$I^{2000}_{o,-r(d)} \times \frac{I^{2000}_{-c(o),d}}{I^{2000}_{-c(o)}}$	0.0055	0.021	0.010	0.028*
	(0.0092)	(0.014)	(0.010)	(0.016)
N	946,469	946,469	946,469	946,469
County-Product FE	✓	✓	✓	✓
Country-Product FE	✓	✓	✓	✓
County-Continent FE	✓	✓	✓	✓
Country-Census Division FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
Principal components		✓		✓
1st-stage F-statistic	21.9	36.4	11.1	39.1

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

unused feature of the Nielsen household panel data: that we can observe the country of birth of panelists. Hence, we can distinguish between natives and immigrants in our sample. This variable is only collected in 2008, so we must rely on panelists who survive in the sample through 2016.

We therefore restrict our sample of households to natives and estimate the following equation:

$$\log s_{io}^h = \alpha_o^h + \beta \ln(M_{od}) + \gamma X_{od} + \delta X_i + \varepsilon_{io}^h \tag{12}$$

where $s_i^h o$ is the share of household i's consumption expenditures observed in Nielsen allocated to purchases of goods within product group h that are imported from o. We control for country-by-product group fixed effects, α_o^h. We also control for the full set of distance and latitude difference controls, represented by X_{od}.

With 18,488 native panelists, we lack sufficient variation to include the full set county fixed effects. Instead, we control for a rich set of household demographics in X_i: household income bins, household size, the age and presence of children, as well as dummy variables for the birth year, education status, marital status, race, and Hispanic ethnicity of the head of household. We cluster standard errors both at the origin country-level and at the household-level.

Given the prevalence of zeros in each of our estimating equations (7, 10, and 12), and the fact that our theory suggests logging the expenditure shares, we use pseudo-Poisson maximum likelihood (PPML) estimation (Silva and Tenreyro 2006). Due to the non-linearity of PPML, we take a control function approach to generating exogenous variation in the immigrant population (Petrin and Train 2010; Morten and Oliveira 2018). In particular, we take add the residuals from the first-stage instrumental variable regressions as controls for our main specifications.

5 Preliminary Results

We now estimate each of the above specifications in turn. First, we estimate equation (7) to obtain the total effect of immigrants on exports. We show our results in Table 2.

We find that immigrants increase imports overall. In column 1, we show that–without adjusting for endogenous confounds potentially in log ancestry–more county residents with ancestry from country o corresponds to an increase in imports from o. In parituclar, we estimate a coefficient of $\hat{\beta} = 0.05$ (SE=0.011). We find the same relationship for immigrants, as shown in column 3, with a coefficient size of 0.051 (SE=0.011).

In columns 2 and 4, we control for the residuals from first-stage regression of immigrants on

Table 2: Gravity Estimates: Homescanner Data

	Expenditure share (w/in product category)			
	(1)	(2)	(3)	(4)
Log ancestry 2010	0.050***	0.075***		
	(0.011)	(0.017)		
Log immigrants 2010			0.051***	0.067***
			(0.011)	(0.018)
First-stage residuals		-0.031*		-0.020
		(0.017)		(0.019)
N	948,725	948,725	948,725	948,725
County-Product FE	✓	✓	✓	✓
Country-Product FE	✓	✓	✓	✓
County-Continent FE	✓	✓	✓	✓
Country-Census Division FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		36.4		39.5

Notes: The table presents regression results at the country-county-product group level. Observations weighted using Nielsen county weights. Sample excludes counties with fewer than 10 households surveyed. Standard errors clustered at the origin country level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

the set of leave-out push-pull instrumental variables. We see the more local residents with ancestry from o (or immigrants from o) significantly raises imports from o, estimating $\hat{\beta} = 0.076$ (SE=0.017). This estimate implies that a 10% increase in immigrants from o living in d raises the expenditure share by consumers in d on goods coming from o by 0.5%.[8]

For the number of immigrants, we estimate $\hat{\beta} = 0.068$, which implies that a 10% increase in the number of immigrants from o living in d raises the share of expenditures by consumers in d on goods coming from o by 0.2%.[9]

We next turn to our estimates using store-level data, which allows us to isolate the demand-side effect of immigrants on imports. We show our estimates of equation (10) in Table 3.

Without adjusting for potentially endogenous confounders, we find no statistically significant relationship between immigrants and imports when using the store-level data, as shown in columns 1 and 3 of Table 3. When adding the first-stage residuals, however, we see a strongly positive effect

[8]Using $\hat{\beta} = 0.076$ from column 2 in Table 2 and a mean bilateral ancestry population of 2125, we have: $\exp\left(0.076\left(\ln\left(1+\frac{1.1\times 2125}{1000}\right)-\ln\left(1+\frac{2125}{1000}\right)\right)\right)-1=0.005$.

[9]Using $\hat{\beta} = 0.068$ from column 4 in Table 2 and a weighted mean bilateral immigrant population of 454, we have: $\frac{s^h_{od}[M^{2010}_{o,d}=1.1\times 454]}{s^h_{od}[M^{2010}_{o,d}=454]}-1=\exp\left(0.068\left(\ln\left(1+\frac{1.1\times 454}{1000}\right)-\ln\left(1+\frac{454}{1000}\right)\right)\right)-1=0.0021$.

Table 3: Gravity Estimates: Storescanner Data

	Expenditure share (w/in product category)			
	(1)	(2)	(3)	(4)
Log ancestry 2010	-0.013	0.057**		
	(0.036)	(0.025)		
Log immigrants 2010			0.0024	0.061**
			(0.038)	(0.025)
First-stage residuals		-0.078***		-0.069***
		(0.023)		(0.022)
N	4,209,253	4,209,251	4,209,253	4,209,251
County-Product FE	✓	✓	✓	✓
Country-Retailer-Product FE	✓	✓	✓	✓
County-Continent FE	✓	✓	✓	✓
Country-Census Division FE	✓	✓	✓	✓
Census Division-Retailer FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
1st-stage F-statistic		161.7		158.2

Notes: The table presents regression results at the country-store-product group level. Standard errors clustered three ways at the origin country, retailer, and county levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

of immigrants on imports, as seen in columns 2 and 4.

We find that the number of county residents with ancestry from country o raises the expenditure share on imports from o by 0.2%. Similarly, the effect of a 10% increase in immigrants from o on import expenditures from o is also 0.2%.

For both immigrants and residents with ancestry, we estimate a $\hat{\beta}_d \approx 0.06$. This compares to a total effect of $\beta \approx 0.07$. We conclude that a large fraction of the effect immigrants have on imports is due to the demand channel, with a much smaller part of immigrants' effects on imports resulting from immigrants reducing bilateral trade costs.

Motivated by the fact that the demand channel dominates, we finally seek to uncover whether there exist what we call demand externalities. That is, do natives benefit from the immigrant-induced increase in imports by purchasing more imported goods? We answer this question by estimating equation (12) using the sample of native-born households. We show the results in Table 4.

We find evidence of substantial demand externalities, with natives sigificantly increasing their consumption of imported varieties from a given origin country when there is an influx of immigrants

Table 4: Household-level Estimates

	Expenditure share (w/in product category)			
	(1)	(2)	(3)	(4)
Log immigrants 2010	0.021**	0.021***		
	(0.0095)	(0.0076)		
Log ancestry 2010			0.029**	0.024**
			(0.013)	(0.010)
First-stage residuals		-0.000034		0.0083
		(0.016)		(0.018)
N	7,718,735	7,718,735	7,718,735	7,718,735
Country-Product FE	✓	✓	✓	✓
Country-Census division FE	✓	✓	✓	✓
Distance & latitude difference	✓	✓	✓	✓
Household controls	✓	✓	✓	✓
1st-stage F-statistic		3460.5		8072.6

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

from that origin country. Our results suggest that these demand externalities have nontrivial effects on natives.

6 Discussion

We have shown that immigrants causally increase imports from their origin country. Prior research has shied away from estimating the effect of immigrants on imports, as any import rise may reflect the preferences and consumption patterns of the immigrants themselves. With novel, highly-detailed data, we show that natives consume significantly more imports from an origin when more immigrants from that origin live locally. Our results are suggestive of two plausible mechanisms. First, retailers may only stock a given product if it reaches a critical mass of local demand. Native-born consumers who demand an imported product may not have their demand met until a mass of immigrants from the product's origin arrive locally. Second, natives may follow the consumption of immigrants via a "demonstration effect."

Given our limited data, we cannot claim any kind of welfare effect in a traditional economic model with fixed preferences. Moreover, any welfare gain from what we term demand externalities may be partially mitigated by product displacement. Since shelf space is finite, domestically-produced goods may be displaced by retailers when immigrants arrive. In future additions to this project, we will better and more precisely explore the nature and magnitude of the demand externalities.

References

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

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.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan, "Migrants, ancestors, and foreign investments," *The 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.

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.

Felbermayr, Gabriel, Volker Grossmann, and Wilhelm Kohler, "Migration, international trade, and capital formation: Cause or effect?," in "Handbook of the economics of international migration," Vol. 1, Elsevier, 2015, pp. 913–1025.

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.

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," *NBER Working Paper*, 2018, *22158*, 65.

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," *The Economic Journal*, 2018, *128* (612), F210–F234.

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

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

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