UNDERCOUNTED AND MISUNDERSTOOD: MEASURING HISPANIC IMMIGRATION AND ITS EFFECTS ON HOUSING*

Agustin Hurtado† Francisca Sara-Zaror‡ Pablo Slutzky§

Abstract

The United States has experienced an unprecedented immigration boom since 2021, with net annual inflows averaging 2.5 million, over 75% of whom are Hispanic immigrants. Despite the growing importance of immigration in shaping economic and social outcomes, the availability of precise data for recent years remains limited, constraining researchers' ability to assess its impacts. Official estimates likely undercount the Hispanic immigrant population, particularly in smaller geographic areas, posing challenges for local analyses. To address this gap, we develop a novel framework to measure the local presence of Hispanic immigrants using retail sales data from Hispanic-branded food products. After validating our measure, we examine its implications for housing markets, where immigration's impact is theoretically ambiguous. While immigrant inflows likely increase housing demand, native households may respond through residential sorting, offsetting upward pressure on prices and rents. Across all periods, we find no evidence that Hispanic immigration raises prices or rents, and in 2021-2022, we document a negative association between Hispanic immigration and rent growth. Consistent with this finding, we observe a negative relationship between Hispanic immigration and crime—an important (dis)amenity that can drive native flight and shape housing market dynamics.

*We thank Scott Baker, James Choi, Tony Cookson, Eli Fich, Caitlin Gorback, Harald Hau, Sabrina Howell, Nir Jaimovich, Adriana Kugler, Luis Lopez, Pascal Noel, Max Maksimovic, Adair Morse, Alessandro Previtero, Geoff Tate, Haluk Unal, Liu Yang, and participants of the 2025 FRA Early Ideas Session, Junior Household Finance Seminar, and the University of Maryland Finance Brownbag for their comments and suggestions. We also thank Manuel Molina for research assistance. The views expressed solely those of the authors and do not represent the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System. Researchers own analyses calculated (or derived) 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 those of the researchers 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.

†University of Maryland Smith School of Business.

‡Federal Reserve Board of Governors.

§University of Maryland Smith School of Business.

1 Introduction

Immigration shapes public discourse across societies, sparking intense debates about its economic and social impact. In the United States, this issue has been at the forefront of policy debates, especially concerning its effects on housing. These discussions have gained momentum over the past decade, marked by a notable increase in the number of immigrants entering the country, both authorized and unauthorized.

Despite the critical importance of this issue, the evidence of the impact of immigration on real outcomes is mixed, likely due to two key methodological challenges. The first challenge is that of endogeneity, as immigrants are not randomly distributed across geographies. To address this issue, scholars have employed shift-share instruments, based on the idea that immigrants are more likely to settle in areas where other migrants from the same country reside, given potential network effects. The second challenge, which is the focus of this paper, is the underestimation of these populations in conventional data sources.

One group particularly affected by undercounting is noncitizens, especially individuals with unknown legal status. For example, researchers at the U.S. Census Bureau found that approximately 20% of noncitizens listed in administrative records report addresses that could not be linked to an address in the 2020 Census, compared to only 5.7% for citizens (Brown et al., 2023). This suggests that a significant portion of these individuals were omitted from the 2020 Census. Additionally, Bates and Trejo (2019) found that individuals lacking permanent immigration status feared that Census Bureau data might be used for immigration enforcement, leading to lower response rates among this group. Indeed, we show that these lower response rates are not unique to the Decennial Census data. For example, the response rate for the Census Bureau's American Community Survey declined sharply, from over 97% in early 2010s to under 85% in 2022-2023. Furthermore, states with a high proportion of Hispanic noncitizens have experienced substantially lower response rates in recent years, likely exacerbating immigrant undercounting.

To address these data and measurement challenges, in this paper we develop a novel

proxy for estimating number of immigrants. More specifically, we use retail sales data of food products predominantly consumed by immigrants of Hispanic origin to measure the presence of immigrants in different geographies. This innovative approach allows us to infer the presence and scale of immigrant populations more precisely than traditional methods and at a higher frequency. For illustration, consider the pre-cooked P.A.N. White Corn Meal (or harina P.A.N.), a specialized product consumed predominantly by Venezuelan nationals. While Venezuelan immigrants —and especially those with no permanent immigration status in the country— are probably associated with low response rates in official surveys, mapping the consumption of products such as harina P.A.N. allows us to better proxy for immigration.

Using store-level point-of-sale data for over half a million brands and more than nine million unique product IDs, we identify approximately 50,000 food products associated with Hispanic immigrants sold in over 30,000 stores nationwide. In 2022 alone, these data capture $135 billion in total sales, with approximately $2 billion corresponding to Hispanic products, representing a notable share of retail sales. With this data, we construct a measure of Hispanic immigration for each geography by mapping the consumption of such products across time.

We validate our proxy through multiple approaches. First, we use information from the 2010 Decennial Census, as the most recent Census is estimated to significantly undercount Hispanic individuals. Specifically, a 2022 U.S. Census Bureau report estimates this undercount at 5% of the Hispanic population, a rate more than three times that of 2010 (Khubba et al., 2022). We show that our data closely aligns with Hispanic immigrant estimates from the higher-quality 2010 Decennial Census. Additionally, we use Hispanic student enrollment data from public schools to further validate our proxy.

We additionally validate our measure of Hispanic immigration using data on the removal of unauthorized immigrants. We first exploit the staggered rollout of the Secure Communities program across counties. This program, which facilitated the exchange of information between local, state, and federal law enforcement agencies and the Department of Homeland Security, was initiated in 2008 and resulted in the removal of over 700,000 individuals, mostly

of Hispanic origin. We show that our proxy captures the staggered implementation of the program, with a notable reduction in the sale of Hispanic products in different areas following the deployment.

Last, we show that our proxy captures the effect of the largest workplace raid in U.S. history, Operation Wagon Train. This operation, which took place on Dec. 12, 2006, targeted Swift & Co. meatpacking plants in six States, and resulted in the arrest of nearly 1,300 workers, most of whom were subsequently deported. We show that our proxy identifies a decline in sales of Hispanic products in the six locations where the raids took place, with no significant change in sales of non-Hispanic products. Importantly, official data, such as that in the American Community Survey, do not capture this decline.

After validating our measure and applying it to assess local levels of immigration, we use it to evaluate the impact of Hispanic immigration on key social and economic outcomes, with a particular focus on housing. While the economic literature has extensively examined immigration's labor market effects, its implications for housing markets remain less well understood. In principle, the relationship between immigration and housing is theoretically ambiguous. On the one hand, standard demand-side theory suggest that an influx of immigrants should increase demand for housing, placing upward pressure on rents and home prices. Prior research has largely confirmed this effect at the city level, where immigrant arrivals have historically been associated with rising housing costs (Saiz, 2007; Saiz and Wachter, 2011). However, neighborhood-level dynamics within cities are more complex. Native-born households may respond to immigration inflows through residential sorting, either clustering in different neighborhoods or leaving urban areas altogether. This process of "native flight" resembles the "white flight" phenomenon in Boustan (2010) and may dampen housing demand, partially offsetting broader price pressures. Moreover, post-pandemic urban and social shifts—such as the rise of suburban areas beyond the urban fringe documented by Davis et al. (2024) and anti-immigrant sentiment—may lead to lower within-city housing demand despite immigrant inflows. At the same time, Hispanic immigrants constitute a crucial labor force in the construction industry, potentially mitigating housing supply

constraints (Howard et al., 2024). Across all periods, we find no evidence that Hispanic immigration increases housing rents. Notably, in 2021-2022, a period of extraordinary immigration inflows, we document a negative association between Hispanic immigration and rent growth.

We also examine the effect of Hispanic immigration on crime, an important (dis)amenity that can induce native flight and influence house price dynamics and where empirical evidence remains mixed. Economic theory provides conflicting predictions. If immigration leads to labor market displacement or social dislocation, it could, in principle, increase crime rates. Conversely, if immigrants self-select based on economic ambition and risk aversion—or if increased immigration revitalizes local economies—crime could decline. Using our novel measure, we find a negative relationship between Hispanic immigration and property, violent, and total crime, driven by the negative correlation in 2021-2022.

Additionally, we show that our data can be used to expand available information on the geographic distribution of certain groups of immigrants and assess their impact on other social and economic outcomes. More specifically, we use a subset of products from the retail sales data to create country of origin-specific measures of immigration. Using data on Venezuelan products, for instance, we create a proxy for the presence of Venezuelan immigrants across the country. This group is of particular interest, given that it represents a large fraction of immigrants who arrived to the United States in the most recent period. We show that our proxy correlates strongly with the data in the American Community Survey (2022) on the number of Venezuelan nationals in the 21 counties for which it includes information. We then expand this information to over 700 counties for which we have consumption data and for which there are no official measures.

Overall, our paper contributes to several strands of the literature. First, it contributes to the literature on the measurement and effect of total immigration, including authorized and unauthorized, on several outcomes. This area remains relatively underexplored in financial economics. To the best of our knowledge, only two recent papers address this topic, both primarily focusing on municipal finance outcomes. Zimmerschied (2024) examines the impact

of immigration on local governments' access to finance, measured through municipal bond yields, while Cornaggia et al. (2024) investigates how unauthorized immigration affects the fiscal health of local governments. A central difference between these papers is that Cornaggia et al. (2024) relies on administrative data, whereas Zimmerschied (2024) uses Census Bureau data, which systematically undercounts immigrants.[1] We contribute to this nascent literature by developing a new measure of immigration that improves upon approaches that use static administrative data. By using consumption microdata, we can account for the high mobility rates of this group, allowing us to analyze the effect of immigration on housing markets.

In terms of measurement and labor outcomes, Borjas (2017) develops a method to impute the unauthorized status of individuals in the Current Population Survey and uses these data to measure the labor force participation rate of this group relative to others. Meanwhile, Borjas and Cassidy (2019) uses the same method to estimate the wage gap for unauthorized immigrants. We contribute by developing a novel measure of immigration that takes into account both types of immigrants and that can additionally mitigate issues related to the increasing non-response rates in official data sets.

Second, we add to the literature on the effect of immigration on the housing market. This relationship has been a focal point of economic research, with various studies examining how influxes of immigrants influence housing prices and rents. Within this literature, Saiz (2007) provides a foundational analysis of this dynamic, finding that immigration can significantly affect housing demand, thereby influencing price levels in urban areas. This study highlights the elasticity of housing supply as a critical factor in determining the magnitude of price effects. Building on this work, Saiz and Wachter (2011) further explores the spatial distribution of immigrants and its implications for local housing markets, finding that cities with higher concentrations of immigrants tend to experience more pronounced price effects. In a subsequent study, Sá (2015) investigates the differential impacts of immigration on housing prices across regions with varying economic conditions, underscoring the role of local market characteristics in mediating these effects.

[1] Another important difference is that Cornaggia et al. (2024) examines a range set of municipal finance outcomes than those analyzed by Zimmerschied (2024).

One key aspect of these studies is that they rely on official data on immigration, such as flows of immigrants admitted as permanent residents ("Green card holders") or changes across Decennial Census editions. These numbers likely undercount other groups of immigrants, such as unauthorized or those seeking asylum, which account for a significant fraction of immigrants in the last waves. We contribute to this literature by focusing on an alternative measure of immigration that incorporates a wider set of migrants and that can additionally account for the usually high mobility rates among this group of individuals.

Last, we contribute to the literature on the effect of immigration on crime rates. Within this literature, Bell et al. (2013) explore the UK context, offering insights into how immigration affects crime rates across different regions. Their findings suggest that the type of immigration impacts crime rates differently, with a wave of asylum seekers associated with an increase in property crime and a wave of workers from EU accession countries associated with a reduction in such rate. Meanwhile, Pinotti (2015) investigates the Italian context and finds that the legalization of immigrants can significantly reduce crime rates. This research highlights the role of policy interventions in shaping the relationship between immigration and crime.

In a recent article, Marie and Pinotti (2024) provide a survey of papers in this topic, while Fasani et al. (2019) explore this issue in three settings, Italy, the United Kingdom, and the United States. Overall, the literature finds that immigration does not inherently lead to higher crime rates and, in some cases, can contribute to reductions in certain types of crime. These studies emphasize the importance of considering socio-economic integration and local context when assessing the impact of immigration on crime. We contribute to this body of literature by developing a novel proxy that enhances the measurement of immigrant presence, encompassing both registered and unregistered individuals. This measure facilitates a deeper understanding of the impact of immigration on crime.

The rest of the paper proceeds as follows. In Section 2 we provide information on immigration in the United States and discuss measurement problems in official and non-official data. In Section 3 we describe the data and how we identify Hispanic brands.

In Section 4, we describe how we develop our proxy and how we validate it. In Section 5 we estimate the effect of immigration on housing rents and crime. Section 6 concludes.

2 Institutional Background

The United States is home to the largest number of foreign-born residents globally, with over 45 million immigrants comprising 14% of the population and more than 20% in large states like California, New York, and Florida. This demographic importance has been amplified by an unprecedented surge in immigration between 2021 and 2023, during which net annual inflows averaged 2.5 million—the largest wave in U.S. history, even after adjusting for population size—bringing the foreign-born share to a historic high of 15.2%

Despite the increasing importance of immigration, official statistics have faced two persistent issues that have worsened in recent years. First, survey-based estimates often differ from those estimated from administrative records. According to Table I, Census Bureau surveys capture less than a third of the inflows estimated from administrative data. This difference reflects challenges in survey data collection for immigrants, including high nonresponse rates, inaccurate responses, and differences in the characteristics of respondents vs. non-respondents and accurate vs. inaccurate respondents. As Appendix A.1 illustrates, these challenges reduce the accuracy of survey-based estimates of immigration and likely contribute to its underestimation.

Second, official immigration statistics suffer from limited geographic and time coverage. The American Community Survey (ACS), while providing annual estimates, has limited coverage, particularly in geographies without large populations. The Decennial Census, though more comprehensive, is conducted only once every ten years and fails to capture short-term immigration fluctuations. Additionally, data release lags further constrain usability. For example, granular population estimates from the 2020 Census became available 16 months post-collection. These limitations hinder timely and precise measurement of immigration trends, particularly at subnational levels. For instance, ACS 1-year estimates

often omit immigrant populations in smaller counties, making local-level analysis difficult.

These two persistent data issues are particularly severe for Hispanic immigrants. Survey-based undercounts are exacerbated by nonresponse biases and inaccurate reporting among Hispanic noncitizens who may be reluctant to disclose their status. Fear of immigration enforcement has further discouraged participation, with ACS and Decennial Census data showing systematically lower response rates among Hispanic noncitizens (Bates and Trejo, 2019). These measurement gaps are most acute in states with high Hispanic noncitizen concentrations, where Census coverage has deteriorated in recent years. As a result, official data likely understate both the scale and composition of the recent immigration wave.

Recent immigration trends intensify the data challenges. Immigration is increasingly dominated by Hispanic individuals, with a large majority of new immigrants between 2021 and 2022 being of Hispanic origin. Figure I uses administrative data to illustrate this shift, showing both the absolute number of Hispanic immigrants (bars, left axis) and their share of total immigration (line, right axis) over time. After a period of relative stability from 2014 to 2018, Hispanic immigration surged in 2021 and 2022, reaching nearly 3 million individuals in 2022—the highest level on record. Simultaneously, the Hispanic share of total immigration remained elevated, fluctuating between 70% and 85% in recent years—a sharp increase compared to 2014, when the share was below 50%.

2.1 Definitions and Terminology

Throughout this paper, we use several terms related to immigration that warrant clarification to ensure consistency. Below, we define key terms and describe their alignment with commonly used classifications in official statistics.

Immigrant An immigrant refers to any foreign-born individual residing in the United States. This definition aligns with the U.S. Census Bureau's classification of the foreign-born population, which includes naturalized citizens, lawful permanent residents, temporary visa

holders, refugees, and unauthorized individuals.

Hispanic Immigrant We define a Hispanic immigrant as an individual born in Latin America who identifies as Hispanic or Latino. This classification uses the Census Bureau's definition of Hispanic ethnicity, which is based on self-identification and encompasses individuals of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish-speaking origins.

Immigrant vs. Hispanic Immigrant The foreign-born population consists of all individuals born outside the United States, irrespective of nationality or ethnicity. The Hispanic foreign-born category is a subset of the foreign-born population, those born in Latin America that identify as Hispanic or Latino. Panel A of Figure II illustrates this relationship, showing that the total number of immigrants in the U.S. ranged between 40 and 45 million from 2012 to 2018, while Hispanic immigrants accounted for roughly 18 to 20 million within this group.

Noncitizen A noncitizen is any foreign-born individual who has not acquired U.S. citizenship. While all noncitizens are immigrants under our definition, not all immigrants are noncitizens, as naturalized individuals are considered citizens.

Hispanic Noncitizen We define a Hispanic noncitizen as a noncitizen born in Latin America who identifies as Hispanic or Latino, encompassing noncitizens of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish-speaking origins.

Noncitizen vs. Hispanic Noncitizen The noncitizen population consists of all foreign-born individuals who have not acquired U.S. citizenship, regardless of ethnicity. The Hispanic noncitizen category is a subset of this population, comprising individuals born in Latin America who identify as Hispanic or Latino and who have not acquired U.S. citizenship. Panel B of Figure II illustrates this relationship, showing that the total number of noncitizens in the U.S. remained between 20 and 22 million from 2012 to 2018, while Hispanic noncitizens

accounted for approximately 10 to 12 million within this group.[2]

2.2 Challenges with Official Data

2.2.1 American Community Survey

The American Community Survey (ACS) is a comprehensive, ongoing survey conducted by the U.S. Census Bureau that provides vital information about the American population and housing characteristics. Initiated in 2005, the ACS collects data annually from a sample of approximately 3.5 million households, making it one of the most extensive sources of demographic, social, economic, and housing information in the United States. Unlike the decennial census, which provides a snapshot every ten years, the ACS aims at providing more timely data and covering a wide range of topics, including demographics, income, health, education, and employment.

The ACS employs a stratified sampling technique, which allows for detailed estimates at various geographic levels, from national to state levels down to counties. The ACS uses both mail and online questionnaires, supplemented by telephone and personal interviews for non-responding households, seeking a robust data collection process. The survey's results are released in the form of 1-year and 5-year estimates, with the former providing more timely but limited data and the latter providing more reliable data for smaller geographic areas and population groups.

The ACS has experienced a sharp decline in response rates over the past decade, falling from 97.5% in 2010 to 84.7% in 2023. Figures III and IV document this trend at both national and state levels, while Figure V highlights a key driver: a surge in refusal rates.[3] This issue may pose a challenge to the accuracy of data extrapolation. In particular, while the likelihood of inclusion of a group in the sample and the weights used to correct for heterogeneous representation depend on its relative size in the population, the uncertainty in the underlying structure of the population might add biases to the data. A growing body of

[2] Panel A of Figure VI plots the ratio of these quantities, showing that a large share of noncitizens are Hispanic.

[3] The sharp drop in 2013 reflects the impact of a government shutdown on follow-up operations, while the decline in 2020 reflects pandemic-related disruptions.

evidence suggests that declining response rates and untruthful reporting disproportionately affect immigrant counts, leading to systematic underestimation (Brown et al., 2023; Jasso and Rosenzweig, 2020; Jensen et al., 2015, 2022; Mira and Bollinger, 2024).[4]

The undercounting of immigrants by the ACS may be particularly severe in geographies with large shares of noncitizens, who, as Panel A of Figure VI shows, are predominantly Hispanic. Fear of immigration enforcement has been a major deterrent to ACS participation, as documented by Bates and Trejo (2019), USDA (2025), among others. Many noncitizens have expressed concerns that Census data could be used for enforcement purposes, leading to both lower participation and higher misreporting rates.[5] This phenomenon is evident in Panel B of Figure VI, which shows that states with a high share of Hispanic noncitizens experienced significantly lower response rates, particularly since 2017. Response rate differences remained relatively stable before 2017, hovering around 2 percentage points. However, the gap widened sharply between 2017 and 2021, reaching 6 percentage points in 2018. The decline deepened to nearly 12 percentage points by 2021 and has since fluctuated between 12 and 6 percentage points.

These measurement issues have become even more pronounced since 2021, amid unprecedented Hispanic immigration inflows. In smaller geographies, limited sample sizes render estimates less reliable, particularly for small subgroups. In cases where estimates fall below reliability thresholds, the ACS does not even report 1-year estimates, exacerbating coverage issues for states without large populations. For example, the ACS 1-year estimates only provide Venezuelan immigrant counts for 21 counties out of more than 3,000 nationwide.

[4] Nonresponse issues have also affected other key Census surveys, raising concerns about data reliability. Bee and Rothbaum (2024) documents that widening differences between respondents and nonrespondents in the Current Population Survey (CPS) have led to an upward bias in income estimates (by 2-3%) and a downward bias in poverty rates (by 0.3-0.5%) between 2020 and 2024.

[5] Appendix Figure A.4.1 shows that states with a high share of Hispanic individuals experienced a significant increase in refusal rates between 2012 and 2022, suggesting that immigration enforcement may have affected the response rates of Hispanic individuals more broadly—not just Hispanic noncitizens.

2.2.2 Decennial Census

The U.S. Decennial Census, mandated by the Constitution and conducted every ten years, is the most comprehensive population enumeration in the country. Unlike the ACS, which relies on annual sampling, the Census aims for full coverage.[6] It collects key demographic variables, including age, sex, race, ethnicity, household composition, and housing tenure. Despite its scope, the Census faces many of the same measurement challenges as the ACS, particularly in undercounting certain populations. As these issues are detailed in the ACS subsection, here we briefly highlight those that are particularly salient in the Decennial Census.

Undercounting has been a persistent concern, particularly for foreign-born individuals, noncitizens, and those in limited-English-speaking households (O'Hare et al., 2019; O'Hare, 2019; U.S. Census Bureau, 2020). Structural barriers—ranging from lower response rates to logistical constraints—exacerbate this issue, while policy changes and public perceptions further suppress participation. A major driver of this undercounting is fear of immigration enforcement, particularly among noncitizens without permanent legal status. Bates and Trejo (2019) document that these individuals exhibit lower response rates due to concerns that Census data could be used for enforcement purposes. The 2020 Census, for instance, faced heightened scrutiny over the potential misuse of citizenship data, amplifying these fears and likely deterring participation among immigrant communities.[7] These concerns were not unfounded: the U.S. Census Bureau estimates that the Latino population was undercounted by 5.0% in 2020, a sharp increase from the 1.5% undercount in 2010 (Khubba et al., 2022).

These measurement challenges result in a systematic omission of noncitizens from the Census count. Recent evidence highlights significant difficulties in accurately linking noncitizens to Census records, with approximately 20% of noncitizens listed in administrative records failing to match a Census address, compared to just 5.7% of citizens (Brown et al., 2023). This discrepancy suggests that a substantial share of noncitizens is

[6] For example, the 2020 Census accounted for 99.98% of housing units nationwide U.S. Census Bureau, 2020.

[7] See, for example, www.washingtonpost.com/politics/2019/04/22/new-research-shows-just-how-badly-citizenship-question-would-hurt-census/.

entirely excluded from official population estimates, further exacerbating concerns over their underrepresentation.

While highly comprehensive, the main limitation of the Decennial Census is its infrequency. Conducted every ten years, it fails to capture interim demographic shifts. Additionally, data release is delayed; for example, 2020 Census state-level population results were published nearly a year post-collection, with detailed breakdowns (e.g., race, sex, smaller geographies) released 16 months later. These delays hinder timely analyses.

2.2.3 Department of Homeland Security

Another commonly used source of data on immigration is that from the Office of Homeland Security Statistics, which provides data on the number of new lawful permanent residents ("green card holders"), non-immigrant admissions (such as students), naturalizations, refugees and asylees. However, this data is aggregated at the state level, and in the case of admissions, it does not represent unique individuals but admissions (one individual might travel and be admitted multiple times). In addition, it covers primarily legal immigration and does not provide detailed demographic information.

2.2.4 State and Local Records

While several local governments collect this type of information, there is wide inconsistency in terms of the type and methods of data collection, making the data hard to compare across areas.

3 Data and Brands Identification

To address the data limitations depicted in the previous section, we develop a novel proxy for Hispanic immigration based on retail sales data for Hispanic food products. In this section, we describe the underlying data and the methodology employed to construct this proxy.

3.1 Hispanic Products

Our main source of data is the NielsenIQ Retail Scanner Data for the years 2006-2022. The dataset provides a granular, weekly record of prices, sales volume, and store-specific information across extensive product categories within the retail sector. Generated by point-of-sale systems from over 90 retail chains and roughly 50 thousand stores per year, the data offers comprehensive coverage at national, regional, and local geographic levels.

Our data captures point-of-sale information for over nine million different products corresponding to over half a million different brands. This data covers sales amounting to an average of over $100 billion dollars per year during the 2010-2022 period. NielsenIQ organizes products into one thousand modules based on broader product categories; e.g., module "Carbonated Drinks" in the category "Beverages". Within each module, the data is collected at the Universal Product Code (UPC)—or barcode—level and contains detailed product information, including brand, size, and other attributes, e.g., flavor.

We start by identifying products and brands that target or appeal to the Hispanic or Latino market. Such brands often reflect the community's cultural preferences, tastes, and values. These brands typically exhibit (1) marketing materials and labels in Spanish or bilingual Spanish/English, (2) a Hispanic origin and identity, and (3) culinary and cultural relevance.[8]

Using this definition, we build a list of potential Hispanic products from three main sources. First, we include products labeled "Made in Latin America" or featuring Spanish in their marketing materials or labels in Label Insight, which provides historical label packaging data starting in 2005 (brands, product descriptions, etc.).[9] Second, we consider items classified as "Hispanic" on www.Walmart.com. Third, we add products available through online stores catering to this population such as www.MexGrocer.com and www.saboresmarket.com.

[8] We assess each brand's origin, identity, and relevance using country of origin, countries where the brand is sold, and brand history.

[9] Label Insight covers the following countries in Latin America: Argentina, Chile, Colombia, Costa Rica, Dominican Republic, El Salvador, Honduras, Mexico, Peru, and Uruguay.

With this list, we proceed as follows. First, we retain well-known Hispanic brands and exclude non-Hispanic brands. Second, we remove brands with non-Hispanic origins despite popularity in Hispanic countries, such as Knorr.[10] Third, we classify brands as Hispanic if the labels are in Spanish or Spanish and English, if their origin or distribution countries are Hispanic, or if their histories reveal Hispanic origin, identity, or relevance, as with Arcor, an Argentine brand with Spanish/English labels in their products.

We expand our list of Hispanic brands in two ways. First, we identify manufacturers by extracting the first 6-9 digits of the UPCs from the list of brands already identified as Hispanic and scan Nielsen's product list for other products by these manufacturers. Second, we include all products with a UPC starting with 750, denoting headquarters in Mexico. These additional products are verified as Hispanic through the criteria above.

We restrict our sample to products with a wide range of varieties and that are easily substitutable, so consumers have the option of choosing an Hispanic brand. For this, we focus exclusively on food products in the following categories: dry grocery, frozen foods, dairy, deli, and packaged meat. We exclude products in health and beauty care, non-food grocery, and fresh produce, alcoholic beverages, and general merchandise. Under this restriction, we exclude certain products, such as Mexican tequila or Chilean wines, that might appeal to a broader audience. This process yields approximately 3,000 distinct brands (such as "El Guapo", "La Fe", "El Mexicano", and "La Preferida" and 33,000 unique barcodes. We present summary statistics of shares of Hispanic products (Table II), volume of retail sales of Hispanic products (Table III), and volume of retail sales of non-Hispanic products (Table IV) by county and year.

[10] Knorr was founded in Germany in 1838 by Carl Knorr. Its headquarters are in Heilbronn, Germany.

4 Hispanic Immigration Proxy

4.1 Proxy Construction

Using retail sales data, we construct our proxy for Hispanic immigration: the Hispanic Consumption Share (HCS). To do so, we link our list of Hispanic brands to the NielsenIQ Retail Scanner Data, yielding prices p and quantities q for product i, in product module m, sold in store k, located in geography g, over time t.

To avoid biases from product module or store entry, we construct a balanced panel of modules and stores across geographies over time. Using this multi-dimensional panel, we calculate our proxy for Hispanic immigration in geography g and time t, as:

$$HCS_{g,t} = \frac{\mathcal{S}^H_{g,t}}{\mathcal{S}_{g,t}} = \frac{\sum_i \sum_m \sum_k p^{Hispanic}_{i,m,k,g,t} \times q^{Hispanic}_{i,m,k,g,t} / P^H_{g,t}}{\sum_i \sum_m \sum_k p_{i,m,k,g,t} \times q_{i,m,k,g,t} / P_{g,t}}, \tag{1}$$

where $\mathcal{S}^H_{g,t}$ and $\mathcal{S}_{g,t}$ are deflated nominal sales of Hispanic food products and all food products, respectively. We provide details on the construction of deflators $P^H_{g,t}$ and $P_{g,t}$ in Appendix A.2.

This proxy captures the share of retail sales in geography g and time t that is associated with Hispanic immigrants. We present the estimates of the national measure in Figure VII and a map of the distribution of retail sales in 2010 in Figure VIII. Importantly, given that Hispanic immigrants do not exclusively consume products associated with Hispanic brands, a 1% retail sales share does not map into a 1% population share.

4.2 Hispanic Immigration Proxy Validation

In this section, we describe the steps we take to validate our measure. We start by showing how it coincides with official data in times when response rates and the potential undercounting of immigrants were likely less problematic. We then use information on the removal of unauthorized immigrants to show how our proxy captures the departure of those

individuals.

4.2.1 U.S. Census data

We start by using information collected during the 2010 decennial census to validate our measure. We use that year given that, in 2010, the risk for an individual to truthfully report his country of origin or ethnicity was arguably smaller compared to more recent years. As previously mentioned, a report by the U.S. Census Bureau in 2022 estimated that the Latino population had an undercount estimated at 5% in 2020, compared to 1.54% in 2010. While some of the increase can be attributed to the difficulties that the global pandemic presented, this increase was significantly smaller for other groups, such as the Black population (for which the estimated undercount increased from 2.1% to 3.3%). One factor that was frequently cited was that of a proposed citizenship question, which despite not being included in the final version led to concerns that census data could be used against immigrants and reduced response rates.

We use this 2010 data as a reference point to validate our measure. By drawing on the relatively accurate demographic information from this year, we are able to assess the robustness of our measure in capturing the true composition of populations across different geographic areas. This approach allows us to ensure that our measure is not only theoretically sound but also empirically grounded in reliable data. By aligning our measure with the 2010 data, we can confidently assert its validity in reflecting the demographic composition of different geographical areas.

More specifically, we construct a binned scatterplot that illustrates the correlation between population and consumption shares—as defined in Equation (1)—at the county level. The scatterplot presented in Figure IX reveals a strong positive correlation between Hispanic population and consumption shares, providing compelling evidence of the validity of our measure.

4.2.2 Schooling data

We run a similar analysis using data from the National Center for Education Statistics. In particular, we obtain detailed data on the number of K-12 students enrolled in each public school throughout the country and the number of Hispanic students enrolled. With these data, we construct a measure at the county level on the percentage of Hispanic students for each year.

Consistent with other measures of immigration, we find a consistent increase –from 16% in 2001 to 28% in 2023– in the percentage of students identified as Hispanic in public schools at the national level. As with the data from the 2010 Census, we construct a binned scatterplot that illustrates the correlation between the share of students in public schools identified as Hispanics and consumption shares at the county level. We present the binned scatterplot in Figure X, which reveals a strong correlation. This evidence further suggests that our measure of retail sales properly captures immigration at the local level.

4.2.3 Removals of unauthorized immigrants

In this subsection, we show that our measure captures the departure of unauthorized immigrants, most likely not captured in official data.

Secure Communities Secure Communities is a data-sharing partnership established between the Department of Homeland Security (DHS) and various local, state, and federal law enforcement agencies. This initiative was designed to enhance the identification and removal of unauthorized immigrants who are deemed threats to public safety. By facilitating the exchange of information, Secure Communities enables law enforcement agencies to access and share biometric data, such as fingerprints, to verify the immigration status of individuals who are arrested or detained.

Since its inception in November 2008, Secure Communities has resulted in the removal of over 700,000 individuals. While the initiative has been praised for its ability to enhance

public safety, it has also sparked considerable debate regarding its impact on immigrant communities and the potential for ethnic profiling. We obtain data on the date the program was implemented in each county and take advantage of its staggered implementation to explore whether these counties experience a decline in retail sales of Hispanic food products following the initial implementation.[11] We provide in Figure XI information on the staggered nature of the program. These data have previously been used to explore the impact of immigration enforcement on outcomes such as residential construction and housing affordability (Howard et al., 2024), employment of likely unauthorized immigrants (East et al., 2023), crime Treyger et al. (2014), and participation in federal safety net programs (Alsan and Yang, 2022).

We examine whether our proxy for Hispanic immigration captures changes in consumption that follow the implementation of the program and subsequent removal of unauthorized immigrants. In particular, in a staggered Difference-in-Differences setting, we study the effect of the program on sales of Hispanic products, using the logarithmic transformation. We present the results in Table V. In columns (1) and (2), we present the results of regressing the volume of sales of Hispanic products on an indicator variable, *Post SC*, set to one following the implementation of the program. We find that, following the implementation of the program and the subsequent removal of unauthorized immigrants, sales of Hispanic products declined by approximately 5%, regardless of the set of fixed effects included. In columns (3) and (4) we run a placebo test where we focus on the set of counties where the program was implemented but where there were no effective deportations. We find no effect on sales of Hispanic products, with a coefficient that is statistically insignificant and positive. In columns (5) to (8) we repeat the exercise but focus on sales of non-Hispanic products. We find that the effect is much smaller and statistically insignificant in all cases, which suggest that, while Hispanic immigrants might consume non-Hispanic products, their removal does not significantly affect aggregate non-Hispanic consumption. We plot the quarter-by-quarter estimate for Hispanic products using the estimator developed by Sun and Abraham (2021)

[11] While the program did not especifically target Hispanic immigrants, more than 74% of the removed individuals were Mexican, 6.85% Guatemalan, 6.77% Honduran, and 4.9% Salvadorean, with Hispanics representing over 95% of the removals.

and provide the results in Figure XII.

Overall, these results provide further support for our proxy of Hispanic immigration, showing that it captures the removal of individuals who are likely uncounted in official measures. Unfortunately, a similar exercise with official data cannot be performed, given that the U.S. Census Bureau changed the wording of the question from "Is this person Spanish/Hispanic/Latino?" to "Is this person of Hispanic, Latino, or Spanish origin?" in 2010, which resulted in the more than doubling of positive responses and makes comparisons across time difficult during the period in which the program was implemented.

Raids On December 12, 2006, U.S. Immigration and Customs Enforcement (ICE) conducted coordinated raids on Swift & Company meatpacking plants across six states. These raids were part of an investigation into identity theft and the use of fraudulent Social Security numbers by unauthorized workers and led to the arrest of 1,297 individuals, most of them Hispanic and most of which were subsequently deported. At that time, the operation was the largest immigration enforcement action in the history of the country.

The raided plants were located in the following counties: Hyrum, Utah; Greeley, Colorado; Marshalltown, Iowa; Grand Island, Nebraska; Cactus, Texas; and Worthington, Minnesota. Using this information, we explore whether retail sales data of Hispanic products captures the removal of these individuals by aggregating sales of Hispanic products in these six counties. We plot retail sales of Hispanic products in these counties in Figure XIII, Panel A. For comparison, we also plot sales of non-Hispanic products. The figure shows a significant decline in sales of Hispanic products following the raids and the removal of these immigrants. In contrast, sales of non-Hispanic products do not experience a similar decline.

Importantly, this effect is not reflected in official ACS data, with only two out of the six affected counties providing information on the number of Hispanic immigrants. For the two counties that do report data, there is no significant decline in the number of registered Hispanic immigrants in the ACS data (Panel B in Figure XIII).

5 Applications

In this section, we show how our novel measure can be used to track immigration flows and assess their impact on relevant economic and social outcomes. We begin by explaining how we map retail sales into immigration flows, using two complementary proxies: a shift-share proxy and a model-based proxy. We then use our shift-share proxy to examine whether the surge in immigration during 2021-2022 contributed to the sharp rise in housing rents observed over this period. Furthermore, we use this measure to investigate the relationship between recent immigration flows and crime as a potential mechanism through which immigration affects housing rents. Finally, we illustrate how our measure enhances and extends existing immigration metrics.

5.1 Shift-Share Proxy

Our consumption share proxy developed in section 4 cannot be used directly to measure the effect of immigration on economic and social outcomes for two reasons. First, it does not provide a direct estimate of the number of Hispanic immigrants. Second, changes in the Hispanic consumption share proxy might be mechanically driven by changes in non-Hispanic food consumption. To address these issues, we construct a shift-share proxy by allocating the gross national Hispanic immigration inflow, "the shock" shown in Figure I, across geographies using geographic-level Hispanic food sales growth as "the shares":

$$HII_{g,t} = \left(\frac{\Delta \mathcal{S}^{H}_{g,t}}{\sum_{g} \Delta \mathcal{S}^{H}_{g,t}} \right) \times NHI_{t} \tag{2}$$

where $HII_{g,t}$ is our Hispanic immigration inflow shift-share proxy, $\mathcal{S}^{H}_{g,t}$ is Hispanic food sales, and NHI_t is the gross national Hispanic immigration inflow in geography g and time t. We set $\Delta \mathcal{S}^{H}_{g,t} = 0$ if $\Delta \mathcal{S}^{H}_{g,t} < 0$.

For illustration, consider that retail sales of Hispanic products increased by one billion dollars in a particular year at the national level. If retail sales of Hispanic products in Los

Angeles County increased by 50 million dollars that same year, we allocate 5% of the gross national inflow of Hispanic immigrants to that county.

5.2 Model-Based Proxy

One limitation of our shift-share proxy is its implicit assumption that only Hispanic immigrants consume Hispanic food, or that the share of non-Hispanics who consume Hispanic products is similar across geographies. To address this, we construct a proxy based on a consumption model, outlined in Appendix A.3. The model includes three agent types—Hispanic immigrants, Hispanic natives, and non-Hispanics—and two goods—Hispanic and non-Hispanic food. We derive an equation linking Hispanic food sales to Hispanic immigrant, Hispanic native, and non-Hispanic populations, which we take to the data.

As Figure XIV shows, our model-based proxy, the HSS estimate, is strongly correlated with our shift-share proxy and exceeds ACS estimates in most states in 2021 and 2022, often by a factor of two or three. In Maryland, for example, our 2022 estimate approaches 900,000—nearly three times the ACS figure.

Since this proxy is limited to a subset of states and counties, we rely on our shift-share measure to analyze the local effects of Hispanic immigration.

5.3 Housing Rents

Starting in 2021, immigration inflows surged, with over 75% of the surge attributed to immigrants of Hispanic origin, as depicted in Figure I. Within this setting, we study the relationship between Hispanic immigration inflows and rent prices, using data from the CoreLogic Single-Family Rent Index (SFRI) at the Core Based Statistical Area (CBSA) level from 2014 to 2022. This index is constructed using a repeated pairing method, described in Nothaft (2018).

The primary rent index we use is the Single-Family Combined Index, which includes both

attached and detached single-family rentals. Additionally, we use additional SFRI indices that divide the combined index into tiers based on whether the property is rented for less than 75% of the CBSA's median, 75%-100%, 100%-125%, or above 125% of the median and focus on the lowest and highest tiers.

We estimate regional Hispanic Immigration inflows by year using a *shift-share analysis*. Following the specification in Equation (2), we allocate the annual national Hispanic immigration inflow shown in Figure I across the areas depending on how much the sales of Hispanic branded products increased in each area in each year. We validate this regional immigration proxy by examining its correlation with official measures of MSA-level Hispanic immigration during the 2014-2019 period from the ACS, as shown in Figure XV. Our proxy for Hispanic immigration inflows is positively correlated with both the Hispanic population that moved from abroad in the previous year (Panel A) —i.e. the location of recent Hispanic immigrants—and with the MSA Hispanic population change (Panel B)—i.e. the mobility of the Hispanic population.

We estimate the relationship between the inflow of Hispanic immigrants and rent prices with the following specification:

$$Rent_{g,t} = \beta_h HII_{g,t-1} + \beta_d DII_{g,t-1} + \delta_t + \delta_g + \epsilon_{g,t} \tag{3}$$

where $Rent_{g,t}$ is the logarithmic transformation of CoreLogic Single-Family Rent Index, $HII_{g,t-1}$ is the lagged Hispanic immigration inflow (HII) rate, defined as Hispanic immigration inflow normalized by the total population in the previous period, $DII_{g,t-1}$ is the lagged Domestic immigration inflow (DII) rate, defined as the domestic immigration inflow normalized by the total population in the previous period. We include time and CBSA fixed effects and cluster standard errors at the CBSA level.

We present the results in Table VI for the period between 2014 and 2022. In column (1) we present the results for the main index. The coefficient for Hispanic Immigration suggests a negative relationship between immigration and rent prices and no significant

relationship between domestic migration and rent prices. Similar conclusions can be drawn from separately analyzing relatively inexpensive (tier 1 units, column 2) and relatively expensive units (tier 4 units, column 3).

To test whether the relationship was different in the period with a surge in immigration (2021-2022), we create an indicator variable, *Post*, for this period and interact it with the immigration flow variables. We present the results in Table VII, which show that the negative relationship between Hispanic immigration and rent prices is concentrated in the period with a rapid surge in immigration. During this period, we find that domestic migration is positively correlated with rent prices.

While this exercise does not allow us to make causal statements, our observational analysis suggests that housing rents are not rising faster in places where Hispanic immigrants established during the 2021-2022 immigration surge. Similar conclusions can be drawn from separately analyzing relatively inexpensive (tier 1 units, column 2) and relatively expensive units (tier 4 units, column 3). In contrast, we find that rents do appear to have been rising faster in places with higher inflows of domestic migrants.

5.4 Crime

We explore whether the decrease in housing rents can be explained by an increase in crime in these areas. To this end, we conduct a similar analysis exploring the relationship between the surge in Hispanic immigration and crime, using data from the FBI's Uniform Crime Reporting (UCR) Program. This database includes information on the number of crimes for a series of different types of criminal activities, such as murder, manslaughter, rape, robbery, assault, and theft, and then aggregates the data into property, violent, and total crime indexes.[12]

We repeat the exercise from the previous subsection with a focus on three indexes of crime, property, violent, and total crime. We present the results on property crime in Table VIII. The coefficients in column (1) suggest a negative correlation between Hispanic immigration

[12]We thank Jacob Kaplan for sharing this data, sourced from Kaplan (2024).

and property crime and no significant correlation with domestic immigration. When we allow the coefficients to vary across periods, we find that this effect is concentrated in the latter part of the sample, when immigration increased rapidly. We find similar effects when we explore violent crime (column (2)) and total crime (column (3)). These results suggest that the negative association between Hispanic immigration and rents is not driven by an increase in crime.

5.5 Expansion of Available Data

With the recent surge in immigration from countries such as Venezuela and Haiti, public agencies and numerous non-profit organizations have begun assisting these individuals and seeking data to better understand their settlement patterns and explore their impact on various outcomes. We explore whether our data can capture the location of these immigrants. To this end, we focus on subsamples of products almost certainly associated with Venezuelan consumers and their retail sales data, using the most recent available year (2022) in the data. We first compare our data with those available in the 1-year American Community Survey of 2022, which provides information on the number of Venezuelan nationals for 21 counties.[13] We plot this correlation in Figure XVI, which suggests that our data captures the number of immigrants in these counties properly.

Importantly, the richness of our data allows us to expand this information and assess the presence of immigrants in other regions. In particular, while the ACS provides data for 21 counties, our data allows us to expand this measure to a total of 748 counties.[14] While these 21 counties are some of the largest and correspond to those with the largest numbers of immigrants, they account for less than 50% of the total consumption of products we associate with Venezuelan nationals, suggesting that the exclusive use of the ACS data would miss over half of the potential number of Venezuelan immigrants.

For a visual representation, we provide in Figure XVII the counties included in the ACS

[13]The 5-year version includes more counties, but the data is less informative in setting with rapid changes such as the recent immigration wave where currency of the data is more important than precision.

[14]The ACS data shows "null" for the vast majority of the observations, which means there is no data available for that geography.

data (Panel A), with the circle size proportional to the number of immigrants, and the counties we capture with retail consumption data (Panel B). Overall, this analysis suggests that our data can be used to expand the information on the geographic location of individuals of certain nationalities to explore their impact on social and economic outcomes.

6 Conclusion

This study addresses a critical gap in our understanding of recent immigration trends in the United States and their impact by developing a novel framework to measure the local presence of Hispanic immigrants. Using retail sales data of Hispanic-branded food products, we develop a novel proxy to capture immigrant populations across geographic areas.

We use this measure to explore the relationship between Hispanic immigration and economic and social outcomes, such as housing rents and crime rates. We find a negative association between the inflow of Hispanic immigrants and housing rents in the 2021-2022 period, when immigration surged rapidly. When exploring the effect on crime, we find a negative association between inflows of Hispanic immigrants and both violent and property crime indexes.

Overall, our paper contributes by providing researchers with a tool for timely measuring and assessing the effect of demographic shifts, a key issue in recent years. Future research can build on this framework to explore further dimensions of immigration's impact, fostering a deeper understanding of its role in shaping economic and social outcomes.

References

Alsan, M. and C. S. Yang (2022): "Fear and the safety net: Evidence from secure communities," *Review of Economics and Statistics*, 1–45.

Bates, N. and Y. G. Trejo (2019): "2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Focus Group Final Report," *US Census Bureau*, 1, 176.

Bee, A. and J. Rothbaum (2024): "Using Administrative Data to Evaluate Nonresponse Bias in the 2024 Current Population Survey Annual Social and Economic Supplement," *U.S. Census blog, Research Matters, September 2024*.

Bell, B., F. Fasani, and S. Machin (2013): "Crime and immigration: Evidence from large immigrant waves," *Review of Economics and statistics*, 95, 1278–1290.

Borjas, G. J. (2017): "The labor supply of undocumented immigrants," *Labour Economics*, 46, 1–13.

Borjas, G. J. and H. Cassidy (2019): "The wage penalty to undocumented immigration," *Labour Economics*, 61, 101757.

Boustan, L. P. (2010): "Was Postwar Suburbanization âWhite Flightâ? Evidence from the Black Migration," *Quarterly Journal of Economics*, 125, 417–443.

Brown, J. D., M. L. Heggeness, and M. Murray-Close (2023): "Noncitizen Coverage and Its Effects on US Population Statistics," Tech. rep., US Census Bureau, Center for Economic Studies.

Cornaggia, J., K. Cornaggia, and R. D. Israelsen (2024): "Unauthorized Immigration and Local Government Finances," *Available at SSRN 5026977*.

Davis, M. A., A. C. Ghent, and J. Gregory (2024): "The work-from-home technology boon and its consequences," *Review of Economic Studies*, rdad114.

East, C. N., A. L. Hines, P. Luck, H. Mansour, and A. Velasquez (2023): "The labor market effects of immigration enforcement," *Journal of Labor Economics*, 41, 957–996.

Fasani, F., G. Mastrobuoni, E. G. Owens, and P. Pinotti (2019): *Does Immigration Increase Crime?*, Cambridge University Press.

Howard, T., M. Wang, and D. Zhang (2024): "How Do Labor Shortages Affect Residential Construction and Housing Affordability?" *Available at SSRN*.

Jasso, G. and M. R. Rosenzweig (2020): "What Is the Size and Legal Composition of the US Foreign-Born Population?" *International Migration Review*, 54, 640–673.

Jensen, E., M. Benetsky, and A. Knapp (2022): "A Sensitivity Analysis of the Net Undercount for Young Hispanic Children in the 2010 Census," in *Census Coverage of the Hispanic Population:*

Analyses of Data Quality and Coverage for the 2020 Census, ed. by J. del Pinal, A. R. Cresce, and T. Gryn, Springer, 95–116.

Jensen, E. B., R. Bhaskar, and M. Scopilliti (2015): *Demographic analysis 2010: Estimates of coverage of the foreign-born population in the American Community Survey*, US Census Bureau, Population Division Washington, DC.

Kaplan, J. (2024): "Summary Reporting System (SRS)," .

Khubba, S., K. Heim, and J. Hong (2022): *National Census Coverage Estimates for People in the United States by Demographic Characteristics: 2020 Post-enumeration Survey Estimation Report*, US Department of Commerce, US Census Bureau.

Marie, O. and P. Pinotti (2024): "Immigration and crime: An international perspective," *Journal of Economic Perspectives*, 38, 181–200.

Mira, A. E. and C. Bollinger (2024): "Credible Estimates: An Empirical Guide to Measurement Error and Misclassification," Working paper.

Nothaft, F. (2018): "Single-Family Rental: Measuring Rent Growth and Return," *The Journal of The Center for Real Estate Studies*, 6, 5–14.

O'Hare, W. (2019): "Census coverage of the Hispanic population," *Differential undercounts in the US census: Who is missed?*, 71–82.

O'Hare, W., L. Jacobsen, M. Mather, A. VanOrman, and K. Pollard (2019): "What Factors Are Most Closely Associated With the Net Undercount of Young Children in the US Census," *Population Reference Bureau, Washington, DC. https://www. prb. org/wpcontent/uploads/2019/03/net-undercount-children-acs. pdf.*

Passel, J. S. and J. M. Krogstad (2023): "What we know about unauthorized immigrants living in the US," *Pew Research Center*, 6.

Pinotti, P. (2015): "Immigration enforcement and crime," *American Economic Review: Papers and Proceedings*, 105, 205–209.

Sá, F. (2015): "Immigration and house prices in the UK," *The Economic Journal*, 125, 1393–1424.

Saiz, A. (2007): "Immigration and housing rents in American cities," *Journal of urban Economics*, 61, 345–371.

Saiz, A. and S. Wachter (2011): "Immigration and the neighborhood," *American Economic Journal: Economic Policy*, 3, 169–188.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 225, 175–199.

Treyger, E., A. Chalfin, and C. Loeffler (2014): *Estimating the Effects of Immigration*

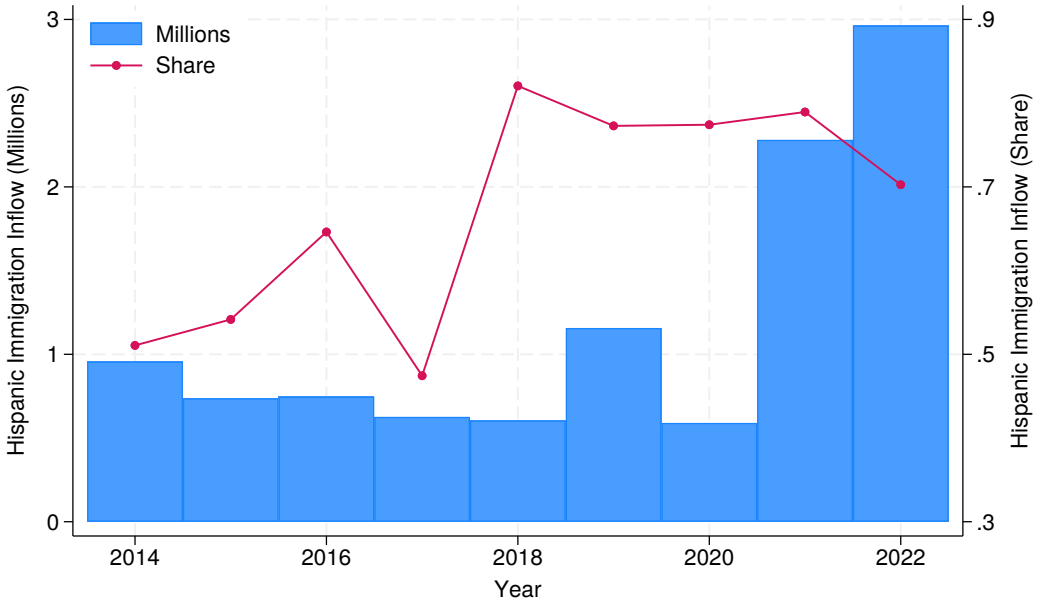
Enforcement on Local Policing and Crime: Evidence from the Secure Communities, George Mason University School of Law.

U.S. Census Bureau (2020): "Counting Young Children in the 2020 Census," Accessed: January 29, 2025.

USDA (2025): "Farm Labor," Available at: https://www.ers.usda.gov/topics/farm-economy/farm-labor (accessed January 23, 2025).

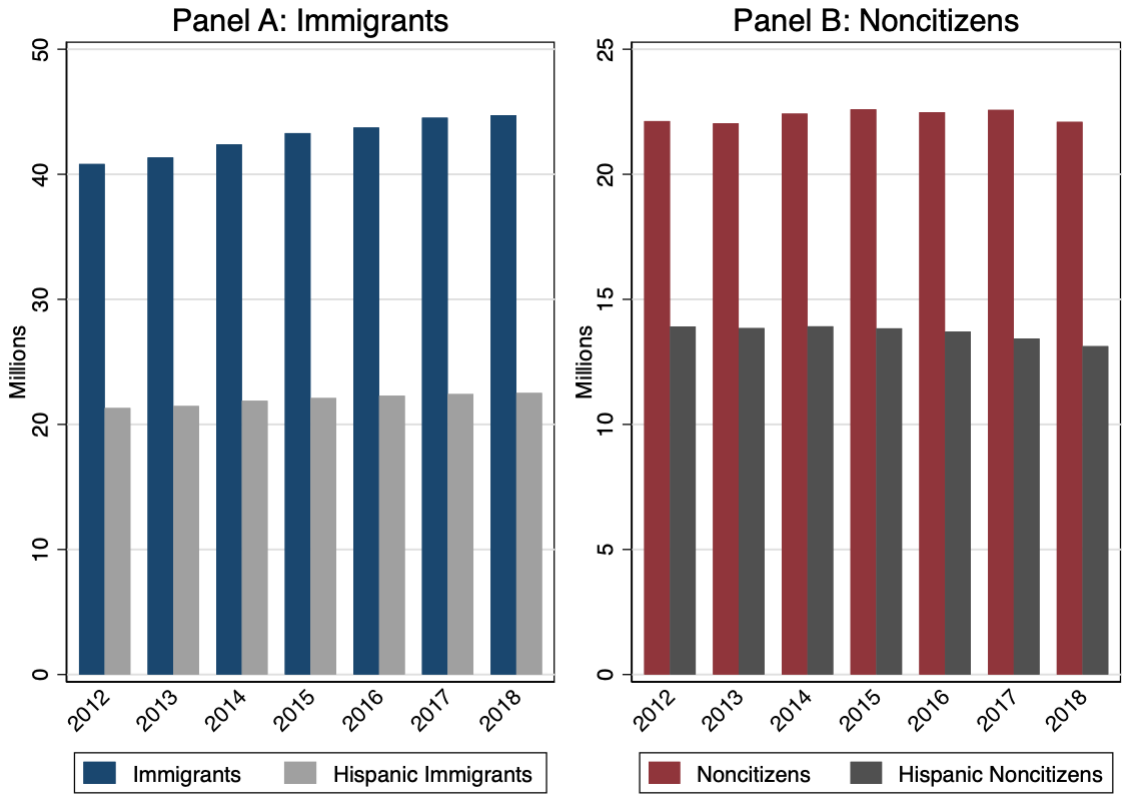
Zimmerschied, K. (2024): "Should They Stay or Should They Go? Immigration and Municipal Bonds," Tech. rep., working Paper.

Figure I: Estimated Hispanic Immigration Flows



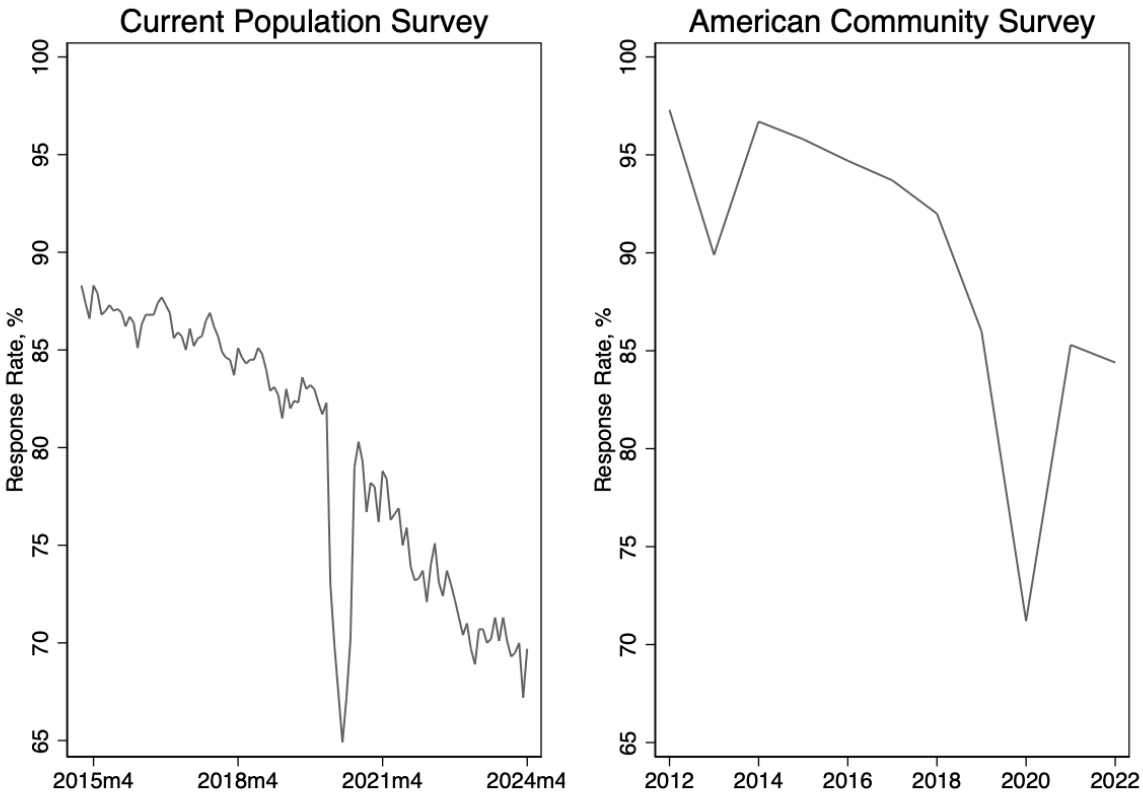
This figure plots estimated Hispanic immigration flows and the Hispanic immigration inflow share. The Hispanic immigration inflow (millions of individuals) is estimated as the sum of legal permanent resident immigrants, temporary immigrants, and Other Foreign Nationals (OFN) from Hispanic countries in Latin America. The OFN category includes individuals admitted with a notice to appear, individuals entering without apprehension, and those granted humanitarian paroles. The Hispanic immigration inflow share corresponds to the Hispanic immigration inflow over total immigration inflow.

Figure II: ACS Estimates of Immigrants and Noncitizens, 2012â2018



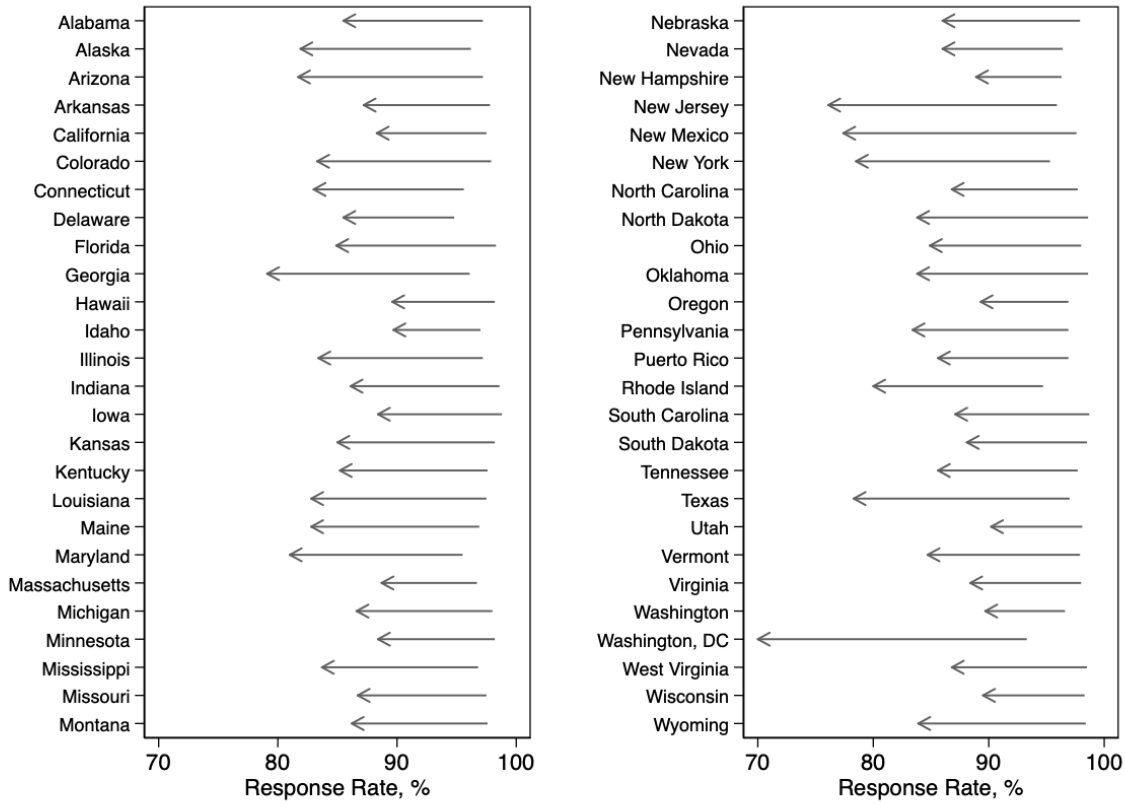
This figure presents ACS 1-year estimates of the number of immigrants and Hispanic immigrants (Panel A) and noncitizens and Hispanic noncitizens (Panel B) from 2012 to 2018. All figures are in millions. Definitions for each category can be found in Section 2.1.

Figure III: Current Population Survey and American Community Survey, Response Rates



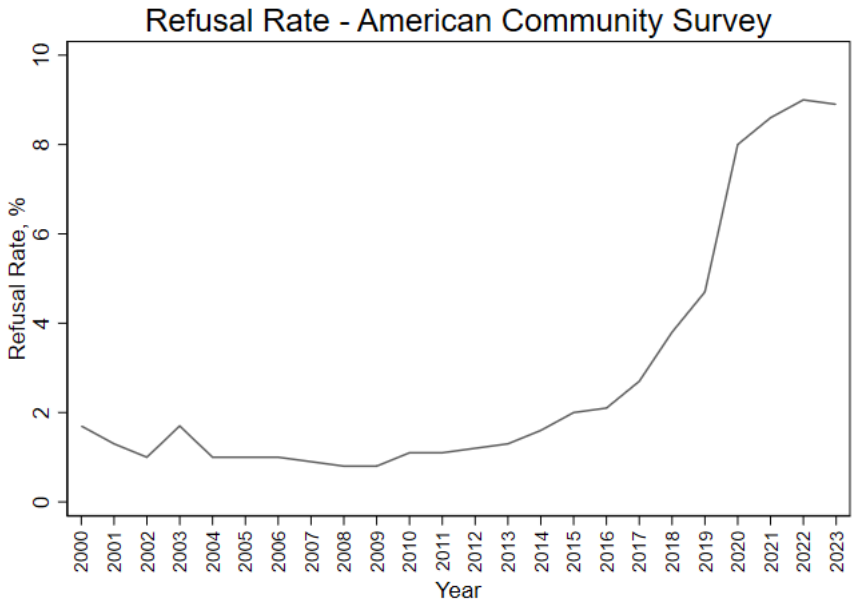
This figure plots the evolution of the response rates for the Current Population Survey (left) and the American Community Survey (right) at the national level. Data are from the U.S. Census Bureau and can be found at www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/index.php. The sharp drop in 2013 reflects the impact of a government shutdown on follow-up operations, while the decline in 2020 reflects pandemic-related disruptions.

Figure IV: American Community Survey, 2012-2022 State-Level Changes in Response Rates



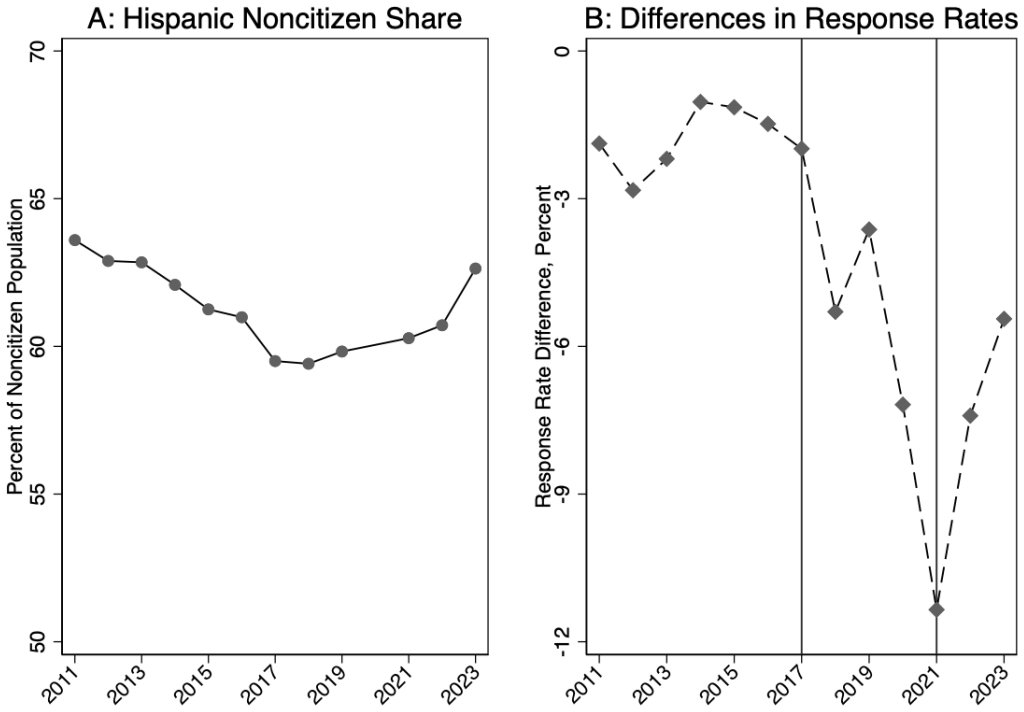
This figure plots the change in the response rates between 2012 and 2022 for the American Community Survey for each State and the District of Columbia. Data are from the U.S. Census Bureau.

Figure V: American Community Survey, 2000-2023 Refusal Rates



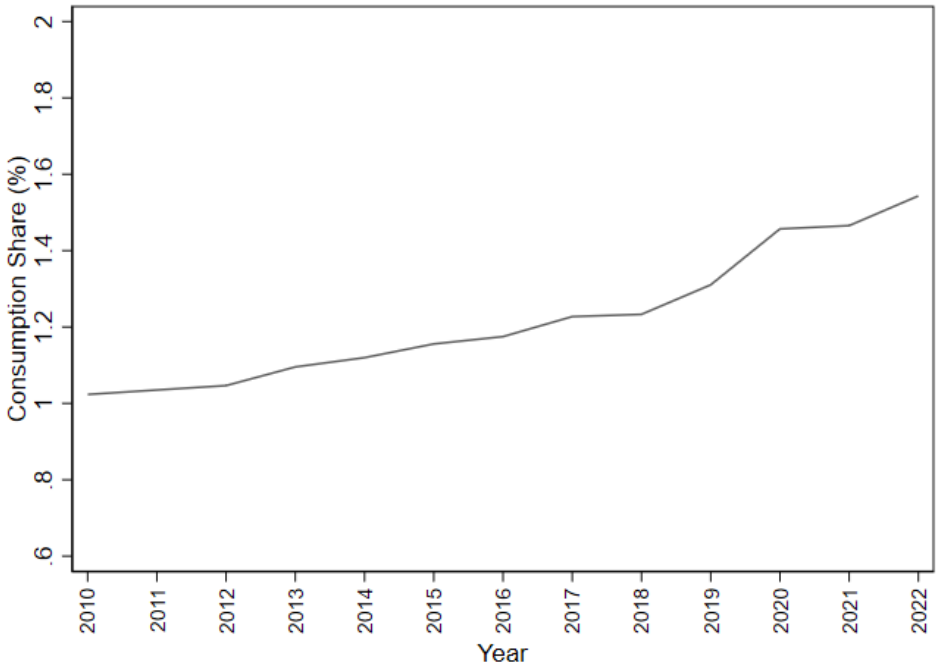
This figure plots the evolution in the refusal rate for the American Community Survey between 2000 and 2023. Data are from the U.S. Census Bureau and can be found at www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/index.php.

Figure VI: Hispanic Noncitizen Share and Response Rate Difference, 2011-2023



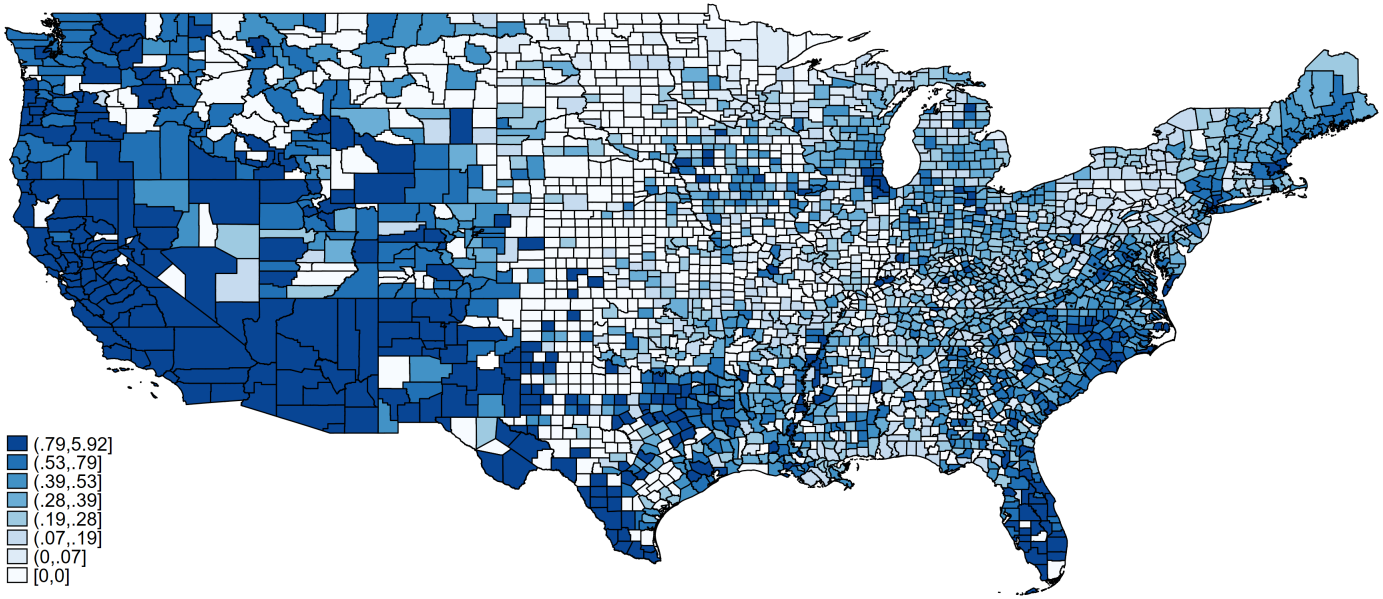
This figure presents ACS 1-year estimates of the share of noncitizens who are Hispanic (Panel A) and the difference in response rates between states with a high versus low share of Hispanic noncitizens (Panel B) from 2011 to 2023. High-share states are those where the percentage of noncitizens who are Hispanic is above the national median, while low-share states fall below this threshold. Definitions for each category can be found in Section 2.1.

Figure VII: Hispanic Products Share



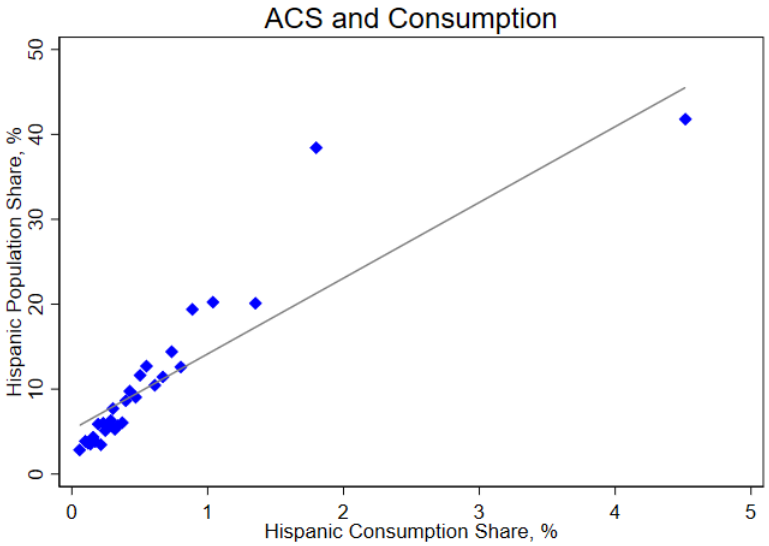
This figure plots the evolution of the share of retail sales corresponding to Hispanic products between 2010 and 2022. Data are from NielsenIQ Retail Scanner Data.

Figure VIII: Retail Sales of Hispanic Products by County



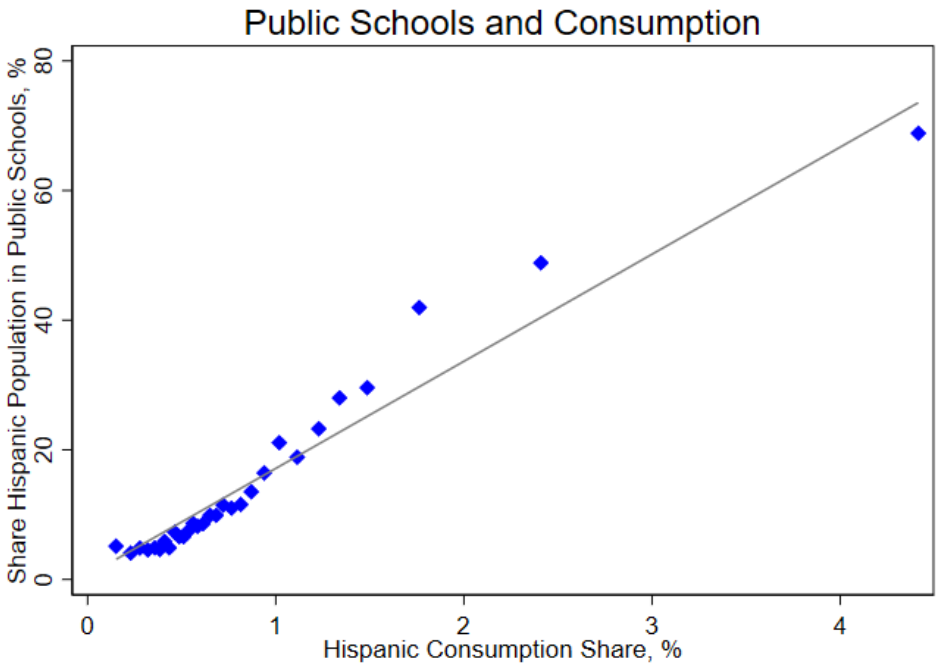
This figure maps the share of retail sales (in %) corresponding to Hispanic products for each county in the 48 contiguous states and the District of Columbia in 2010. Data are from NielsenIQ Retail Scanner Data.

Figure IX: Census Data and Retail Sales



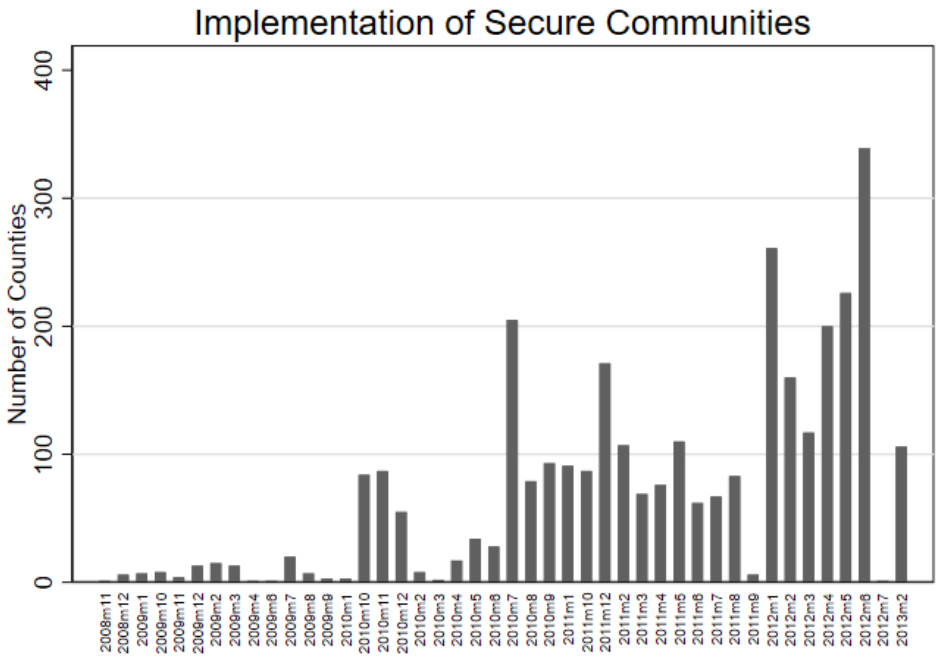
This binned scatter plot presents the relationship between the share of retail sales corresponding to Hispanic products and the share of Hispanic residents in each county in 2010. Data on retail sales are from NielsenIQ Retail Scanner Data and demographic data is from the American Community Survey.

Figure X: School Enrollment Data and Retail Consumption, 2022



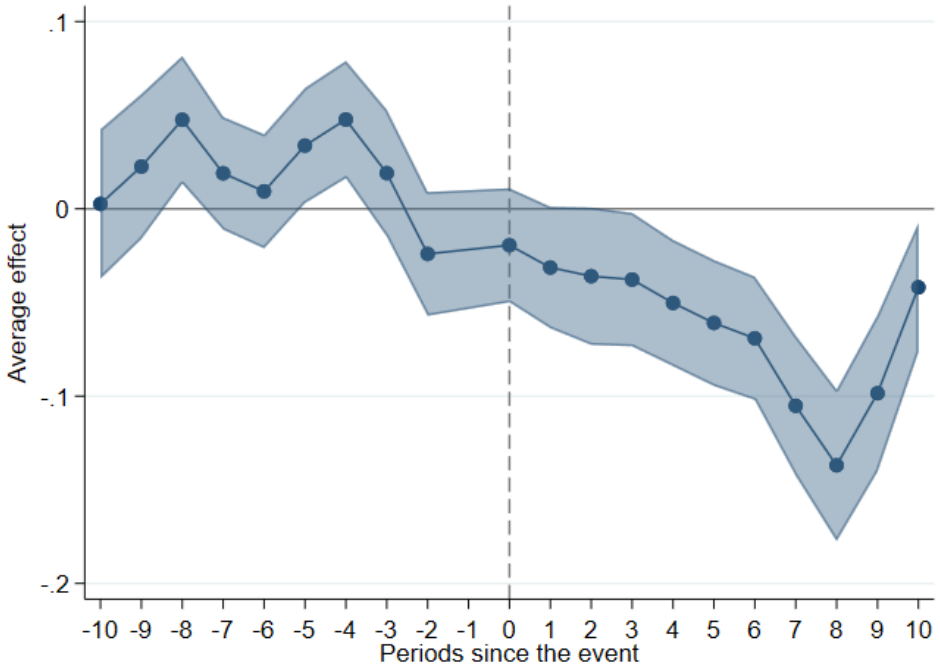
This binned scatter plot presents the relationship between the share of retail sales corresponding to Hispanic products and the share of Hispanic students enrolled in K-12 public schools in each county. Data on retail sales are from NielsenIQ Retail Scanner Data and schooling data are from the National Center for Education Statistics.

Figure XI: Implementation of Secure Communities



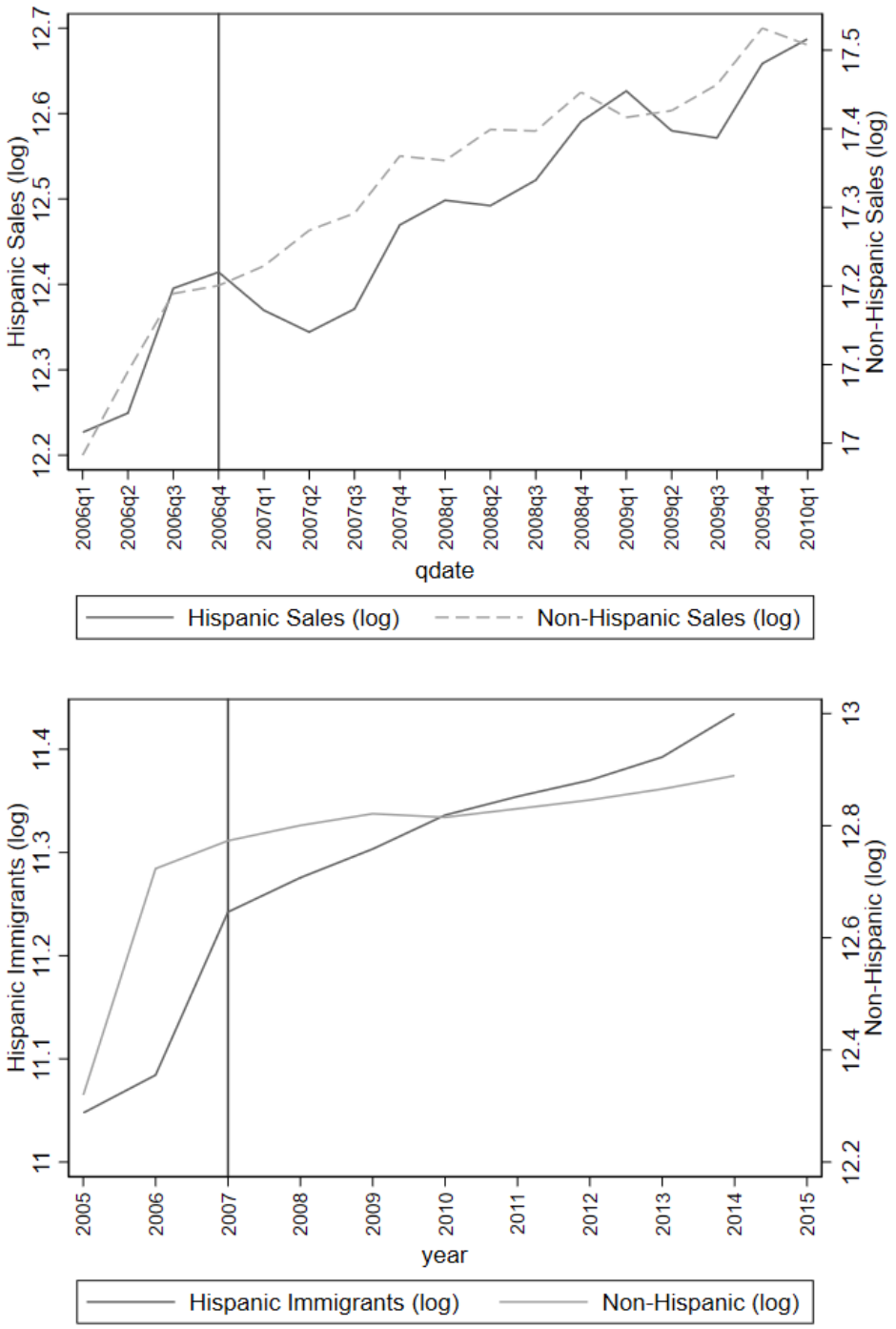
This figure plots the time series of the number of counties where the Secure Communities program was implemented for the first time. By February 2013, all the counties had implemented the program. Data are from the U.S. Immigration and Customs Enforcement Agency.

Figure XII: Secure Communities and Retail Consumption



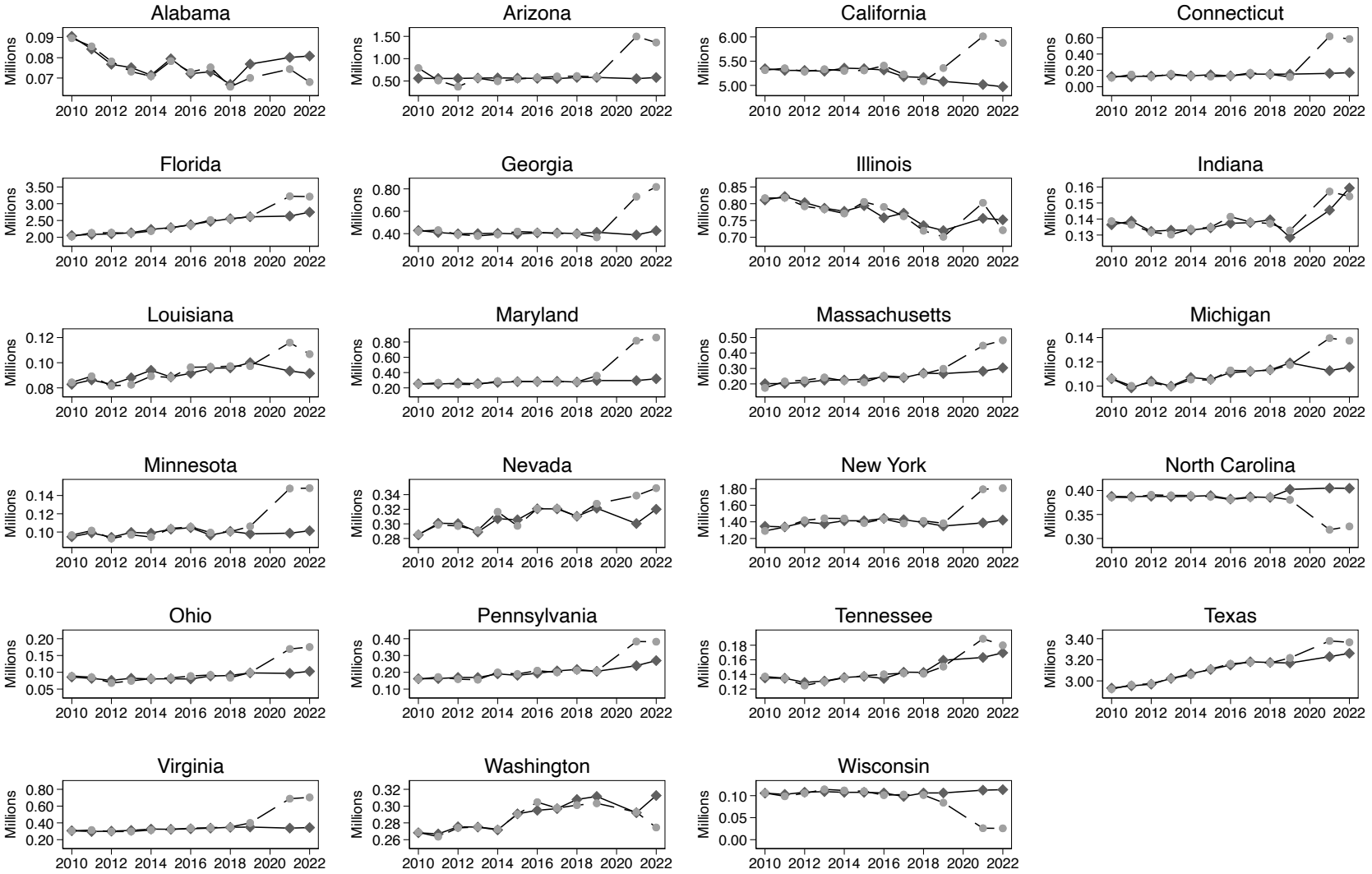
This figure presents the coefficients from the estimator created by Sun and Abraham (2021) for the relationship between the implementation of the Secure Communities program and retail sales corresponding to Hispanic products. Data on retail sales are from NielsenIQ Retail Scanner Data and data on Secure Communities are from the U.S. Immigration and Customs Enforcement Agency.

Figure XIII: Retail Consumption and Raids



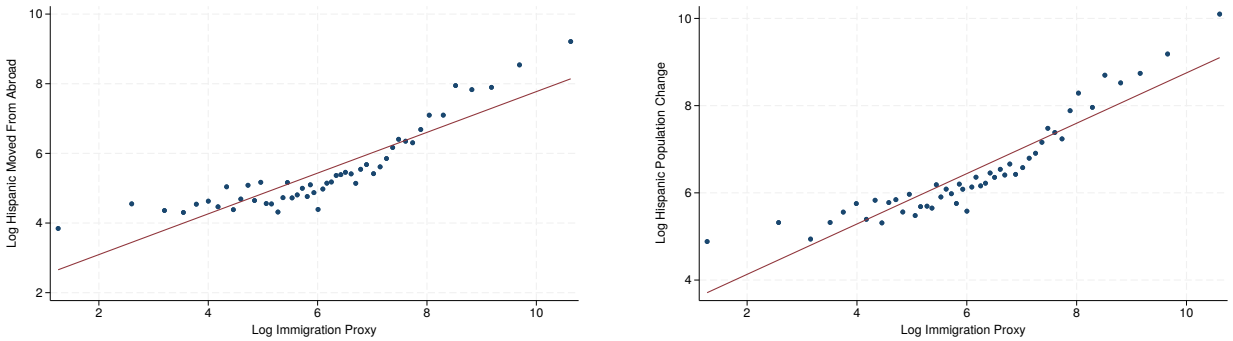
This figure presents the evolution of retail sales corresponding to Hispanic products in counties affected by the Immigration Enforcement Actions on six Swift & Company meat processing plants in 2006 (Panel A). Panel B plots the evolution of the number of Hispanic immigrants in two affected counties for which the American Community Survey provides data. Data on retail sales are from NielsenIQ Retail Scanner Data and demographic data is from the American Community Survey.

Figure XIV: ACS vs. HSS Hispanic Immigration Estimates by State, 2010-2022



This figure presents ACS and HSS estimates of Hispanic immigration by state from 2010 to 2022.

Figure XV: Regional Hispanic Immigration Inflows Proxy, Validation

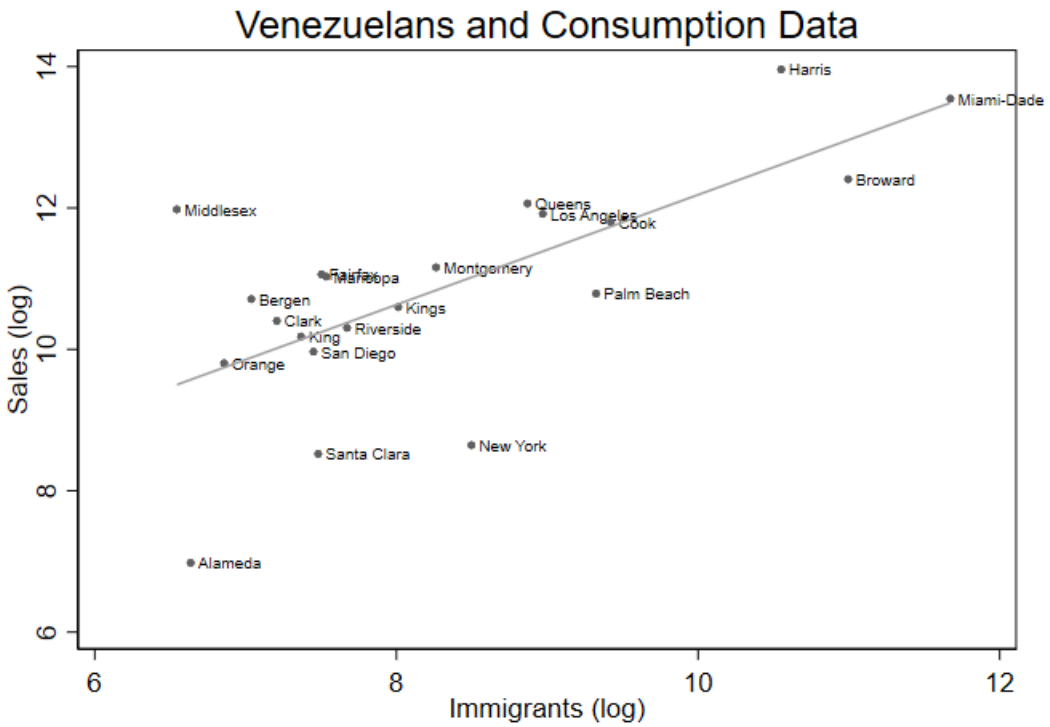


(a) Hispanic Moved from Abroad

(b) Hispanic Population Change

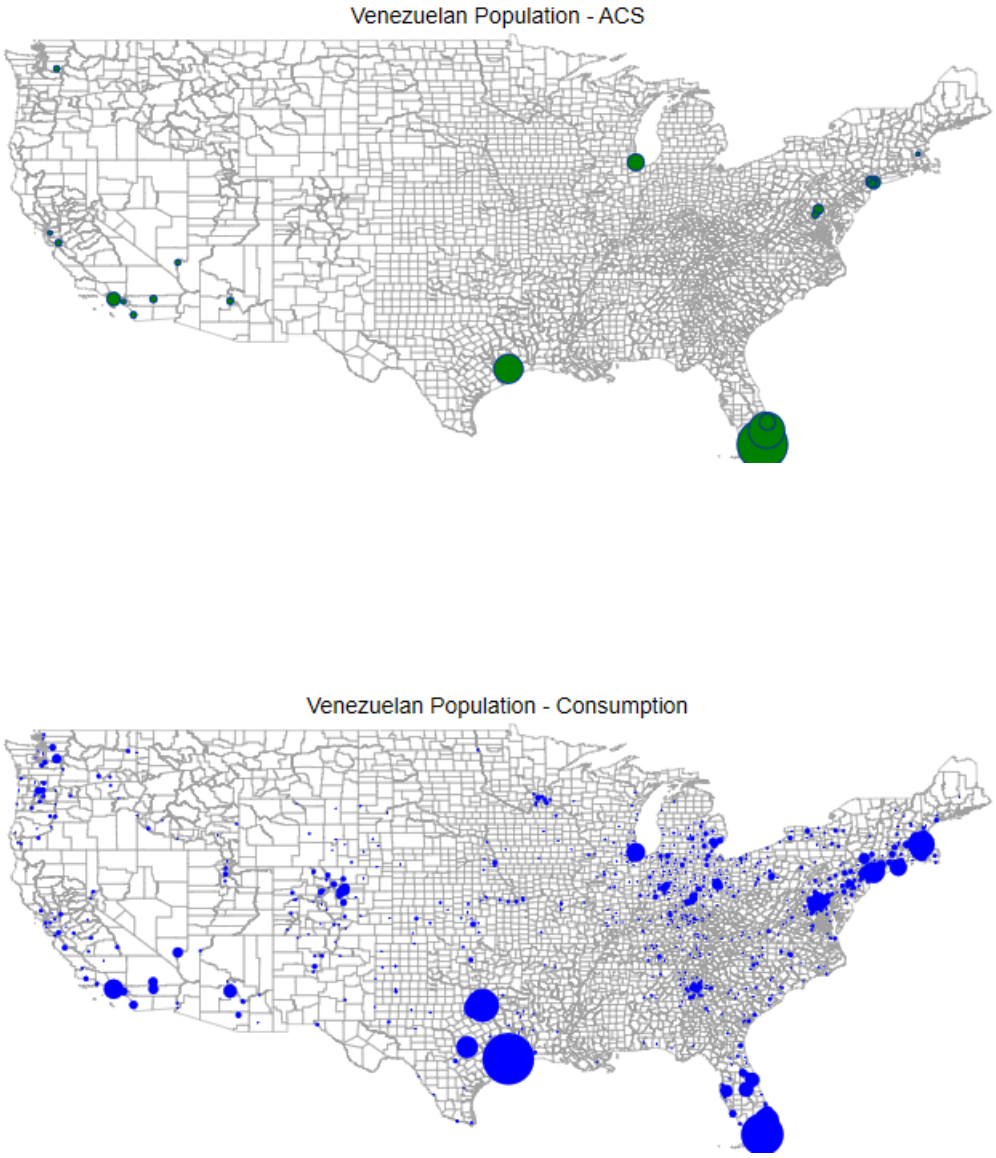
A simple linear regression of "log Hispanic population moved from abroad" on "log Hispanic immigration proxy" has an R^2=0.34, with β=0.59 (left panel). The same regression but with "log Hispanic population change" as the dependent variable has an R^2=0.41, with β=0.58 (right panel).

Figure XVI: Presence of Venezuelan Immigrants, 2022



This figure plots the correlation between the number of Venezuelan nationals in the American Community Survey and the level of sales of products most commonly consumed by Venezuelan at the county level. Data on retail sales are from NielsenIQ Retail Scanner Data and demographic data are from the American Community Survey, both measured in 2022. Data from the American Community Survey is available for 21 counties.

Figure XVII: Presence of Venezuelan Immigrants, 2022



This figure maps information on the number of Venezuelan immigrants reported in the American Community Survey (top panel), with circle sizes proportional to this number, and the volume of retail sales of products commonly consumed by Venezuelan nationals (bottom panel) at the county level. Data on retail sales are from NielsenIQ Retail Scanner Data and demographic data are from the American Community Survey, both measured in 2022.

Table I: Aggregate Estimates of Immigration Flows, 2019-2022

Organization	Instrument Name	Method	Estimated Inflow
Census Bureau	ACS	Survey	866 Thousand
CBO	DE	Administrative	2.43 Million
Goldman Sachs	Report	Administrative	3.1 million
Oxford Economics	Report	Administrative	3.3 million

This table presents summary statistics on aggregate national estimates of immigration inflows between 2019 and 2022 based on survey and administrative data. The first column lists the organizations producing the estimates, the second column specifies the name of the instrument, the third column indicates the methodological approach (survey or administrative), and the fourth column reports the estimated inflow. Data sources include the Census Bureau, the Congressional Budget Office (CBO), Goldman Sachs, and Oxford Economics.

Table II: Summary Statistics - Hispanic Share - County Level

	(1)					
	Mean (%)	Std. Dev.	p.10	p.50	p.90	Obs
2010	0.46	0.66	0.10	0.27	0.91	2,181
2011	0.47	0.65	0.11	0.28	0.96	2,184
2012	0.47	0.64	0.08	0.29	1.00	2,197
2013	0.47	0.67	0.06	0.27	1.05	2,203
2014	0.49	0.67	0.07	0.29	1.09	2,208
2015	0.53	0.67	0.10	0.33	1.10	2,208
2016	0.60	0.66	0.15	0.41	1.20	2,212
2017	0.70	0.74	0.19	0.49	1.35	2,211
2018	0.76	0.81	0.22	0.52	1.42	2,212
2019	0.81	0.86	0.24	0.56	1.51	2,211
2020	0.88	0.85	0.30	0.63	1.59	2,213
2021	0.92	0.82	0.36	0.68	1.58	2,213
2022	0.98	0.83	0.40	0.74	1.70	2,213
N	28,666					

This table presents summary statistics on the share of retail sales corresponding to Hispanic products by county and year between 2010 and 2022. Data are from NielsenIQ Retail Scanner Data.

Table III: Summary Statistics - Retail Sales of Hispanic Products

	(1)					
	Mean (log)	Std. Dev.	p.10	p.50	p.90	Obs
2010	9.57	2.84	6.08	9.93	13.01	2,181
2011	9.73	2.78	6.30	10.05	13.14	2,184
2012	9.73	2.83	6.25	10.06	13.15	2,197
2013	9.66	2.96	5.84	10.15	13.22	2,203
2014	9.80	2.91	5.95	10.23	13.31	2,208
2015	9.94	2.85	6.22	10.36	13.38	2,208
2016	10.20	2.67	6.78	10.48	13.44	2,212
2017	10.37	2.53	7.18	10.56	13.52	2,211
2018	10.44	2.52	7.28	10.63	13.54	2,212
2019	10.56	2.49	7.39	10.73	13.64	2,211
2020	10.83	2.50	7.59	11.05	13.96	2,213
2021	10.98	2.44	7.77	11.22	14.05	2,213
2022	11.13	2.42	7.96	11.36	14.20	2,213
N	28,666					

This table presents summary statistics on the level of retail sales corresponding to Hispanic products by county and year between 2010 and 2022. Data are from NielsenIQ Retail Scanner Data.

Table IV: Summary Statistics - Retail Sales of Non-Hispanic Products

	(1)					
	Mean (log)	Std. Dev.	p.10	p.50	p.90	Obs
2010	15.42	2.22	12.38	15.74	18.26	2,196
2011	15.53	2.17	12.58	15.81	18.35	2,212
2012	15.58	2.14	12.70	15.84	18.37	2,213
2013	15.62	2.13	12.77	15.86	18.39	2,213
2014	15.65	2.14	12.81	15.92	18.42	2,213
2015	15.66	2.15	12.79	15.94	18.44	2,213
2016	15.69	2.16	12.79	15.97	18.49	2,213
2017	15.66	2.17	12.75	15.96	18.49	2,213
2018	15.66	2.20	12.72	15.97	18.50	2,212
2019	15.69	2.18	12.79	15.99	18.51	2,213
2020	15.82	2.20	12.90	16.15	18.66	2,213
2021	15.89	2.20	12.95	16.24	18.72	2,213
2022	15.95	2.21	12.99	16.31	18.78	2,213
N	28,750					

This table presents summary statistics on the level of retail sales corresponding to non-Hispanic products by county and year between 2010 and 2022. Data are from NielsenIQ Retail Scanner Data.

Table V: Secure Communities and Retail Sales

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sales Hispanic				Sales Non-Hispanic			
Post SC	-0.055***	-0.052***			-0.008	-0.006		
	(-3.77)	(-2.77)			(-1.37)	(-0.78)		
Post SC (Placebo)			0.084	0.059			0.024	0.030
			(1.51)	(1.07)			(0.99)	(1.22)
N	83,905	83,865	83,905	83,865	86,461	86,421	86,461	86,421
R-squared	0.950	0.958	0.950	0.958	0.983	0.986	0.983	0.986
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	No	Yes	No	Yes	No	Yes	No
State x Quarter FE	No	Yes	No	Yes	No	Yes	No	Yes

This table presents the results of the effect of the implementation of the Secure Communities program across counties on sales of Hispanic and non-Hispanic products. The dependent variables are the logarithmic transformation of the volume of retail sales of Hispanic products, *Sales Hispanic* and the logarithmic transformation of the volume of retail sales of non-Hispanic products, *Sales Non-Hispanic*. *Post SC* is an indicator variable at the county-year level set to one for counties where the Secure Communities program has already been implemented, while *Post SC (Placebo)* is an indicator variable at the county-year level set to one for counties where the Secure Communities program has already been implemented but where the county experienced no effective removals during the analyzed period. Standard errors are clustered at the county level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are from NielsenIQ Retail Scanner Data and the U.S. Immigration and Customs Enforcement Agency.

Table VI: Rent and Hispanic Immigration Inflows

	(1) Rent Index All	(2) Rent Index Tier 1	(3) Rent Index Tier 4
Hispanic Immigration Inflow	-1.073***	-1.158**	-1.231***
	(-2.90)	(-2.65)	(-3.46)
Domestic Immigration Inflow	0.643	0.413	0.963*
	(1.47)	(1.40)	(2.00)
Year FE	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes
Trends	Yes	Yes	Yes
Observations	344	344	336

This table presents the results of the analysis of the relationship between immigration and rent prices. The dependent variable is the logarithmic transformation of Corelogic's Rent Index at the CBSA level. *Hispanic immigration* measures the inflow of Hispanic immigrants from our shift-share proxy, *Domestic immigration* is the net inflow of domestic migrants from the U.S. Census Bureau. Column 1 uses the overall rent price index. Column 2 uses the price index for tier 1 (relatively inexpensive) units and Column 3 uses the price index for tier 4 (relatively expensive) units. Standard errors are clustered at the county level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are from NielsenIQ Retail Scanner Data, Corelogic, and the U.S. Census Bureau.

Table VII: Rent and Hispanic Immigration Inflows - By Period

	(1) Rent Index All	(2) Rent Index Tier 1	(3) Rent Index Tier 4
Hispanic Immigration Inflow	-0.200	-0.222	-0.258
	(-0.76)	(-0.50)	(-0.97)
Post × Hispanic Immigration Inflow	-1.949***	-2.067***	-2.170***
	(-3.92)	(-3.59)	(-4.57)
Domestic Immigration Inflow	-0.863	-1.047**	-0.311
	(-1.35)	(-2.07)	(-0.52)
Post × Domestic Immigration Inflow	1.577***	1.321***	1.743***
	(2.91)	(3.45)	(3.13)
Year FE	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes
Trend	Yes	Yes	Yes
Observations	344	344	336

This table presents the results of the analysis of the relationship between immigration and rent prices by period. The dependent variable is the logarithmic transformation of Corelogic's Rent Index at the CBSA level. *Hispanic immigration* measures the inflow of Hispanic immigrants from our shift-share proxy, *Domestic immigration* is the net inflow of domestic migrants from the U.S. Census Bureau. *Post* is an indicator variable set to one for the years 2021 and 2022. Column 1 uses the overall rent price index. Column 2 uses the price index for tier 1 (relatively inexpensive) units and Column 3 uses the price index for tier 4 (relatively expensive) units. Standard errors are clustered at the county level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are from NielsenIQ Retail Scanner Data, Corelogic, and the U.S. Census Bureau.

Table VIII: Crime and Hispanic Immigration Inflows

	(1) Property	(2) Violent	(3) Total
Hispanic Immigration Inflow	-1.573**	-1.385*	-1.499**
	(-2.29)	(-1.93)	(-2.28)
Domestic Immigration Inflow	-0.904	-1.525	-1.021
	(-0.92)	(-1.39)	(-1.04)
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Trend	Yes	Yes	Yes
Observations	12180	12143	12180

This table presents the results of the analysis of the relationship between immigration and crime. The dependent variable is the logarithmic transformation of crime indexes from the FBI's Uniform Crime Report. *Hispanic immigration* measures the inflow of Hispanic immigrants from our shift-share proxy, *Domestic immigration* is the net inflow of domestic migrants from the U.S. Census Bureau. Column 1 presents the results for property crime. Column 2 presents the results for violent crime and Column 3 presents the results for total crime. Standard errors are clustered at the county level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are from NielsenIQ Retail Scanner Data, the Federal Bureau of Investigation, and the U.S. Census Bureau.

Table IX: Crime and Hispanic Immigration Inflows - By Period

	(1) Property	(2) Violent	(3) Total
Hispanic Immigration Inflow	1.054	0.882	1.130
	(0.93)	(0.84)	(1.05)
Post × Hispanic Immigration Inflow	-4.212***	-3.688***	-4.138***
	(-4.57)	(-3.72)	(-4.62)
Domestic Immigration Inflow	-1.044	-1.747*	-1.150
	(-1.10)	(-1.78)	(-1.21)
Post × Domestic Immigration Inflow	-0.522	-0.959	-0.665
	(-0.34)	(-0.49)	(-0.43)
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Trend	Yes	Yes	Yes
Observations	12,180	12,143	12,180

This table presents the results of the analysis of the relationship between immigration and crime. The dependent variable is the logarithmic transformation of crime indexes from the FBI's Uniform Crime Report. *Hispanic immigration* measures the inflow of Hispanic immigrants from our shift-share proxy, *Domestic immigration* is the net inflow of domestic migrants from the U.S. Census Bureau. *Post* is an indicator variable set to one for the years 2021 and 2022. Column 1 presents the results for property crime. Column 2 presents the results for violent crime and Column 3 presents the results for total crime. Standard errors are clustered at the county level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. Data are from NielsenIQ Retail Scanner Data, the Federal Bureau of Investigation, and the U.S. Census Bureau.

Online Appendix to

UNDERCOUNTED AND MISUNDERSTOOD: MEASURING HISPANIC IMMIGRATION AND ITS EFFECTS ON HOUSING*

Agustin Hurtado† Fran Sara-Zaror‡ Pablo Slutzky§

March 14, 2025

*For Online Publication. The views expressed in this paper are those of the authors and do not reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

†University of Maryland Smith School of Business.

‡Federal Reserve Board of Governors.

§University of Maryland Smith School of Business.

A.1 Official Data Appendix

A.1.1 Non-Response and Small Sample Issues

In the context of surveys, the Mean Squared Error (MSE) is a key metric used to assess the accuracy of an estimate by incorporating both bias and variance. It measures how close an estimator is to the true population parameter.

Let $\hat{\phi}$ be the sample mean estimate for the national immigration rate ϕ, a population parameter. If we assume $\hat{\phi}$ is unbiased, its accuracy is given by:

$$MSE(\hat{\phi}) = Var(\hat{\phi}) = \frac{\sigma^2}{n}$$

Non-response In a survey, the effective sample size n is the product of the original number of people sampled N and the response rate r (the proportion of respondents who complete the survey). Thus,

$$Var(\hat{\phi}) = \frac{\sigma^2}{r \times N}.$$

The recent immigration surge, the significant number of unauthorized immigrants, and fear-inducing policy changes reduce the response rate r, as shown in Figure VI, worsening the accuracy of the immigration estimate $\hat{\phi}$.[1]

Small-sample Let $\hat{\phi}_i$ be the sample mean estimate for county i's immigration rate ϕ_i, a population parameter. If we assume $\hat{\phi}_i$ is unbiased, its accuracy is given by:

$$MSE(\hat{\phi}_i) = Var(\hat{\phi}_i) = \frac{\sigma_i^2}{n_i}$$

County i's effective sample size n_i is smaller than the nation's effective sample size n, so assuming $\sigma_i \approx \sigma$, the estimate for county i's immigration rate is less accurate than that of the nation.

[1] According to Passel and Krogstad (2023), approximately 11 million unauthorized immigrants lived in the U.S. in 2022, with this number likely increasing in the following years due to events in source countries like Venezuela.

A.2 Deflator Construction Appendix

For each product i, geography g, and month t, we define a sales-weighted price:

$$p_{ig,t} = \frac{\sum_{k \in K_{igt}} (pq)^k_{ig,t}}{\sum_{k \in K_{igt}} q^k_{ig,t}}$$

where k indexes a store in K_{igt}, $(pq)^k_{ig,t}$ is its monthly sales, and $q^k_{ig,t}$ total units sold in a month. Inflation is defined as:

$$\pi^E_{g,t} = \sum_{l \in L^E_{g,y}} \omega_{lg} \sum_{i \in I^E_{lg,y}} \omega_{ig} \Delta \ln p_{ig,t}$$

where the weight ω_{ig} is the annual sales share of product i in category l and geography g; ω_{lg} is the equivalent weight for category l in geography g. The sets of products, $I^E_{lg,y}$, and categories, $L^E_{g,y}$, in year y vary with ethnicity such that $E \in \{H, NH\}$. The corresponding deflator, with January 2010 as the initial period, is:

$$\begin{aligned} P^E_{g,0} &= 1; \\ P^E_{g,t} &= \exp\left(\sum_{\tau=1}^{t} \pi^E_{g,\tau}\right). \end{aligned}$$

To get a state-level deflator, define

$$\pi^E_{s,t} = \sum_{g \in G_s} \omega_g \pi^E_{g,t}$$

where the weight ω_g is the annual sales share of geography g in state s, and G_s is the set of all geographies in state s. The corresponding deflator, with January 2010 as the initial period, is:

$$\begin{aligned} P^E_{s,0} &= 1; \\ P^E_{s,t} &= \exp\left(\sum_{\tau=1}^{t} \pi^E_{s,\tau}\right). \end{aligned}$$

A.3 Model-Based Proxy Appendix

A.3.1 Setup

Population Total population T is the sum of Hispanic immigrants and natives, I and H, and Non-Hispanic individuals N. Hispanic immigrants and natives are born in Latin America and the U.S.

Goods Individuals consume: Hispanic-branded and Non-Hispanic-branded foods, x_h, and x_n.

Preferences Type $i \in \{I, H, N\}$ individuals' utility function is given by:

$$\mathcal{U}^i(x_h, x_n) = x_h^{\phi^i} x_n^{1-\phi^i},$$

Income & prices Individuals have income y, and face prices of Hispanic and non-Hispanic food (p_h, p_n).

A.3.2 Individual Consumption

Each individual j maximizes her/his utility function $\mathcal{U}_j^i(x_{j,h}^i, x_{j,n}^i)$ subject to the budget constraint:

$$p_h x_{j,h}^i + p_n x_{j,n}^i \leq y.$$

Solving this problem yields the following expression for individual j's Hispanic food consumption:

$$x_{j,h}^i = \phi^i \frac{y}{p_h}, i \in \{I, H, N\}.$$

A.3.3 Aggregate Consumption

Aggregate Hispanic food consumption is the sum of Hispanic food consumption across all individual types $i \in \{I, H, N\}$ and all individuals $j = \{1, 2, ..., T\}$:

$$
\begin{aligned}
\mathcal{X}_h &= \sum_{i \in I,H,N} \sum_{j=1}^{T} x_{j,h}^{i}, \\
\mathcal{X}_h &= \sum_{j=1}^{I} x_{j,h}^{I} + \sum_{j=I+1}^{H} x_{j,h}^{H} + \sum_{j=H+1}^{N} x_{j,h}^{N}, \\
\mathcal{X}_h &= \left(\frac{\phi^I \times y}{p_h}\right) I + \left(\frac{\phi^H \times y}{p_h}\right) H + \left(\frac{\phi^N \times y}{p_h}\right) N.
\end{aligned}
$$

Multiplying the last equation by p_h yields:

$$
p_h \times \mathcal{X}_h = (\phi^I \times y) I + (\phi^H \times y) H + (\phi^N \times y) N,
$$

where $\mathcal{S}^H \equiv p_h \times \mathcal{X}_h$ is aggregate Hispanic food sales. Defining $\beta_i \equiv \phi^i \times y$, we can express $\mathcal{S}_h$ as:

$$
\mathcal{S}^H = \beta I + \gamma H + \delta N.
$$

A.3.3.1 Estimation

We construct a state-year panel of $\mathcal{S}_t^H$, I, H, N for 2010-2022 from NielsenIQ Retail Scanner Data and ACS 1-year estimates, respectively. To avoid biases arising from the sudden inclusion of information on the number of agents of each type in the ACS data, we use a balanced panel of states that consistently report the data in the 2010-2022 period.

Because the ACS experienced a sharp decline in its response rate in 2019-2022 and respondents and non-respondents seem to differ, we use these data for 2010-2018 to estimate the following panel regression:

$$
\mathcal{S}_{g,t}^H = \alpha_g + \beta_g I_{g,t} + \gamma_g H_{g,t} + \delta_g N_{g,t} + \epsilon_{g,t},
$$

where g corresponds to a state and $\beta_g, \gamma_g, \delta_g$ are state-specific coefficients. We estimate state-specific coefficients because, in our framework, the relationship above is state-specific, allowing for differences in the consumption basket of similar types of agents across states.[2]

The (out-of-sample) prediction of Hispanic food consumption using information available up to 2018 is:

[2] In the language of Appendix A.3, ϕ^N is most likely larger in Arizona than in DC.

$$\hat{\mathcal{S}}^{H}_{g,t|2018} = \hat{\alpha}_g + \hat{\beta}_g I_{g,t} + \hat{\gamma}_g H_{g,t} + \hat{\delta}_g N_{g,t}, \quad t = 2019, ..., 2022.$$

We define excess Hispanic food sales ES as:

$$ES_{g,t} = \mathcal{S}^{H}_{g,t} - \hat{\mathcal{S}}^{H}_{g,t|2018}, t = 2019, ..., 2022.$$

Excess Hispanic food sales $ES_{g,t}$ is a proxy of uncounted Hispanic immigration in consumption units. We translate it into population units by rewriting the 2019-2022 actual and excess sales as:

$$\begin{aligned}
\mathcal{S}^{H}_{g,t} &= \tilde{\alpha}_g + \tilde{\beta}_g I_{g,t} + \tilde{\gamma}_g H_{g,t} + \tilde{\delta}_g N_{g,t} + \tilde{\theta}_g U_{g,t}, \\
ES_{g,t} &= \tilde{\theta}_g U_{g,t} + (\tilde{\alpha}_g - \hat{\alpha}_g) + (\tilde{\beta}_g - \hat{\beta}_g) I_{g,t} + (\tilde{\gamma}_g - \hat{\gamma}_g) H_{g,t} + (\tilde{\delta}_g - \hat{\delta}_g) N_{g,t},
\end{aligned}$$

where $U_{g,t}$ are uncounted immigrants, i.e., Hispanic immigrants uncounted by the Census Bureau.

Assuming that Hispanic immigrants and uncounted immigrants have similar preferences and that individuals' preferences and income do not change dramatically over time, we obtain our proxy of Hispanic immigration in population units:

$$ES_{g,t} = \tilde{\theta}_g U_{g,t} \implies U_{g,t} = \frac{1}{\tilde{\theta}_g} ES_{g,t} = \frac{1}{\tilde{\beta}_g} ES_{g,t}.$$

Finally, we construct our Hispanic immigration estimates as:

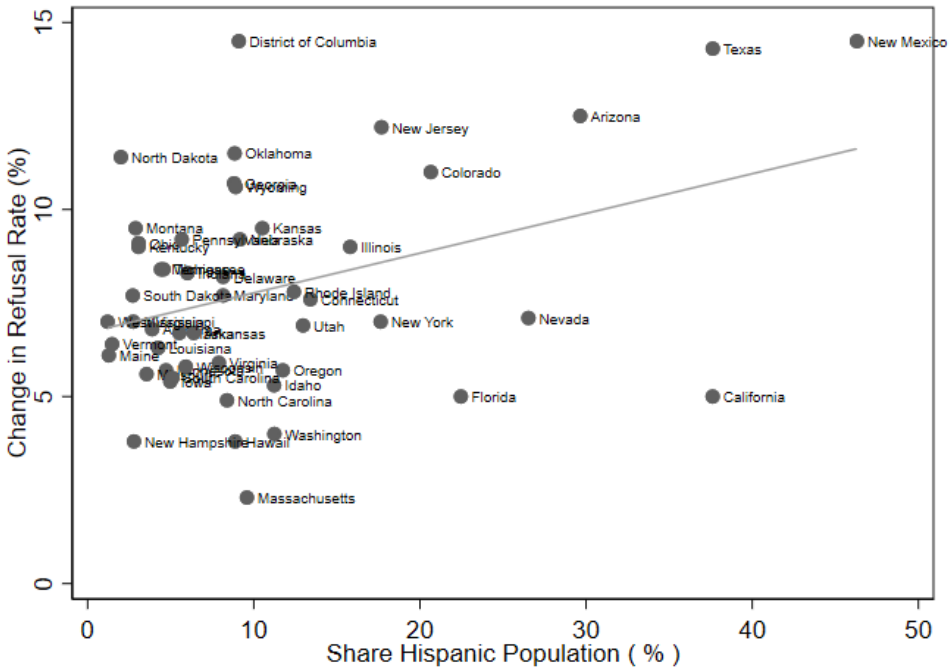
$$HSS_{g,t} = U_{g,t} + ACS_{g,t},$$

where $HSS_{g,t}$ is our estimate and $ACS_{g,t}$ is the ACS estimate of Hispanic immigration, equivalent to $I_{g,t}$ in our framework. We plot our estimates and compare them with ACS estimates in Figure XIV, where we include the 23 states for which ACS provides data on the three types of agents for 2010-2022.

We find that our estimates for Hispanic immigrants are substantially larger than the ACS estimates in most states. For example, according to Figure XIV, our estimate of the number of Hispanic immigrants for the state of Maryland in 2022 is close to 900,000, almost three times the Hispanic immigration estimate from the ACS shown in the Figure. Our estimates are slightly smaller than the ACS estimates only in four (out of 23) states: Alabama, North Carolina, Washington, and Wisconsin.

A.4 Figures Appendix

Figure A.4.1: American Community Survey, Changes in Refusal Rates and Hispanic Share



This figure plots the correlation between the change in the refusal rates between 2012 and 2022 for the American Community Survey and the share of Hispanic population as of 2010 for each State and the District of Columbia. Data are from the U.S. Census Bureau.