# Cross-Border Product Adoption: Individual Imports, Migrant Networks, and Domestic Retailers*


David Argente
Yale University and NBER

Esteban Méndez
Central Bank of Costa Rica

Diana Van Patten
Yale University and NBER


March 8th, 2024


**Abstract**

This paper studies the role of demand externalities in the process of importing foreign products. Using comprehensive data on *individuals'* foreign purchases and on individual-level networks, we propose a new instrument which exploits demand shocks across international migrant networks. We then analyze how the likelihood of an individual importing a product depends on whether her peers recently purchased the same product. We find that a product imported by a close neighbor, a co-worker, or a friend increases an individual's own demand. This demand increase can trigger a response from retailers, as firms increase (decrease) their likelihood of importing a product after it is heavily (lightly) imported by consumers; a response driven by firms' learning about the local demand from individuals' behavior. These externalities suggest that the gains from trade may be larger than previously documented, especially in developing countries or remote markets, and more so as direct-to-consumer online shopping soars.



Keywords: product adoption, local development, social networks.



*Corresponding author: diana.vanpatten@yale.edu.


We thank Harum Alp, Fernando Alvarez, Costas Arkolakis, David Atkin, Lydia Cox, Kevin Donovan, Mayara Felix, Teresa Fort, Amit Khandelwal, Sam Kortum, Theresa Kuchler, Jeremy Majerovitz, Isabela Manelici, Mushfiq Mobarak, Monica Mrazova, David Nagy, Ezra Oberfield, Luigi Pistaferri, Steve Redding, Richard Rogerson, Todd Schoellman, Daniel Xu, and participants at various seminars and conferences for helpful comments and discussion. The views expressed herein are those of the authors and do not necessarily represent the views of the Central Bank of Costa Rica (BCCR). Results have been reviewed by the BCCR to ensure that no confidential information is disclosed.

# 1 Introduction

It is widely accepted that the global consumer landscape has undergone a remarkable transformation with the ascent of online shopping, and that, far from a momentary surge, the rise of e-commerce has had a consistent upward trend, that is expected to continue.[1] It is less well-known that in developing, remote, or small countries, online shopping often takes the form of cross-border shopping. For instance, a consumer might make an online purchase from an overseas retailer, followed by the international shipping and delivery of these items to her local residence. The relevance of cross-border shopping can be sizable; for example, in Latin America, cross-border shopping represents more than 50% of all e-commerce in Bolivia, Costa Rica, Ecuador, the Dominican Republic, El Salvador, Panama, and Paraguay; and on average represents 39% of all online shopping (EBANX, 2023).

Beyond convenience, in developing countries, remote areas, or small markets, online shopping has expanded the frontier of possibilities for shoppers, providing access to goods and varieties that are not available at local stores. This situation creates a context where information frictions can flourish. For instance, online purchasing makes it harder for a buyer to assess a product's appeal compared to seeing it in a physical store and the vast array of options available on the internet can make it difficult to grasp the full extent of product varieties. Moreover, online purchases from abroad are often hard and expensive to return after delivery, thereby increasing the risk associated with purchasing given information frictions. Similarly, for local retailers, searching online to source products for local markets is both costly and risky, especially when the local consumer preferences for these new products are unknown.

This paper studies the role of direct and indirect demand externalities in the process of importing final goods. At the individual level, the role of externalities is direct and intuitive; when an individual buys a foreign product, others around her might learn about the product's appeal or characteristics to then decide whether to import it themselves. In this context, learning may come from communication with a buyer in one's network or from the observation of her purchasing decisions, which can provide valuable information not internalized by the original buyer. A less obvious externality, that might be particularly relevant for remote markets or developing

[1] In 2023, the number of digital buyers is at 2.6 billion globally, so that almost one in three people is an online shopper; a number that is expected to increase (Oberlo, 2023).

countries, is how once individuals in a location verify the quality or appeal of a product, *retail firms* might gain information about the local demand for this good, to then import it and sell it domestically. In this sense, this *indirect* externality leads to an effect of demand shocks on supply responses. Both of these externalities can lead to a multiplier effect after an import takes place, which can impact consumer welfare and profits. As a result, the gains from trade would be larger than previously measured, and network effects may potentially affect the impact of policies attempting to stimulate aggregate demand via, for example, tariff reductions.

The impact of an individual's adoption of a foreign product on her network and on firms' decisions is hard to measure. First, data on purchases of specific products by individual consumers is difficult to obtain. Second, data on an individual's networks is scant. Third, because preferences are likely to be correlated among peers, observing that individuals in the same group make similar consumption decisions may simply reflect shared preferences or common shocks, not demand externalities. Finally, even if identifying the peer effect was possible, linking individual decisions to firm behavior is non-trivial. Given these challenges, there is limited evidence on this topic.

Our paper contributes by analyzing how the likelihood of an individual acquiring a *specific* foreign product depends on whether peers in her relevant network had previously purchased the same product. We also study how the probability that a *retailer* imports a good reacts to its customers importing this product. Our context is the adoption of foreign goods by Costa Rican individuals and firms. Costa Rica is a small open economy where many goods, which are available online, cannot be bought domestically and international returns are rare. Our analysis employs panel data with detailed information on Costa Ricans' foreign purchases and domestic networks.

We begin by developing a simple conceptual framework to guide our empirical examination of how an individual's imports affect the behavior of her peers in a network, and whether a mass of individual imports can trigger a response from local firms. In the model, individuals learn about new foreign products from each other. Only after someone imports a product, its true appeal becomes common knowledge to everyone in the network. Individuals can then decide whether to import the product or not, and retailers can observe the residual local demand for the product and may choose to import and sell it domestically to maximize profits. Thus, externalities play an important role affecting both individuals' and retailers' demand for foreign goods. The model provides a framework to think about this new channel by which individuals can learn about products from each other and retailers can learn about the level of the domestic demand for foreign goods. The main goal of the paper is to explore these forces in the data.

Identifying demand externalities empirically entails overcoming important hurdles. First, it involves data constraints, as it requires knowledge on both an individual's consumption and on an individual's network. These constraints have led most of the previous research studying this topic to identify networks using characteristics that are common among individuals, such as race, cohort, nationality, location, among others.[2] We largely overcome this challenge by leveraging Costa Rican data on imports of specific products by *individuals*, for example, a purchase at a U.S. online retailer that was then delivered by mail to Costa Rica. Each import includes details like date, price, weight, product code up to the HS-10 level, and country of origin. Moreover, imports data can be linked to information on the importer's networks of neighbors, family, coworkers, and friends. These networks are computed for the universe of adults in Costa Rica, and leverage information on family trees, neighborhood composition, and employer-employee records, along with a new measure of friendships that we develop for this paper and leverages big data on peer-to-peer money transfers. To speak to retailers' responses, we complement our study with data for all formal firms in the country, which includes imports, sales, and location.

Second, consumption externalities usually face identification problems (Brock and Durlauf, 2001; Manski, 1993; Moffitt, 2000). We want to understand if, once an individual imports a product, the probability of importing the same product for people in her relevant network increases. This poses endogeneity and econometric challenges, as it is hard to distinguish the true network effect from correlated shocks and common characteristics, which might make individuals import a product for reasons which are independent from the influence of their connections. To overcome this challenge, we propose a new identification strategy which leverages several aspects of our context and our data. In particular, we construct an instrument using the following idea:

Suppose individuals $L$ and $N$ live in Costa Rica, and $L$ has a sister living in Los

[2] A few exceptions with more detailed data include work which surveys respondents to obtain data on their relevant network, like Bandiera et al. (2009); Banerjee et al. (2013), who use details on the network of microfinance clients; and De Giorgi et al. (2019), who rely on co-workers as the relevant reference group.

Angeles while $N$ has a sister living in New York City. *If product $i$ becomes more popular in LA as compared with NYC in period $t$, then $L$ is more likely than $N$ to import product $i$ in period $t+1$.*

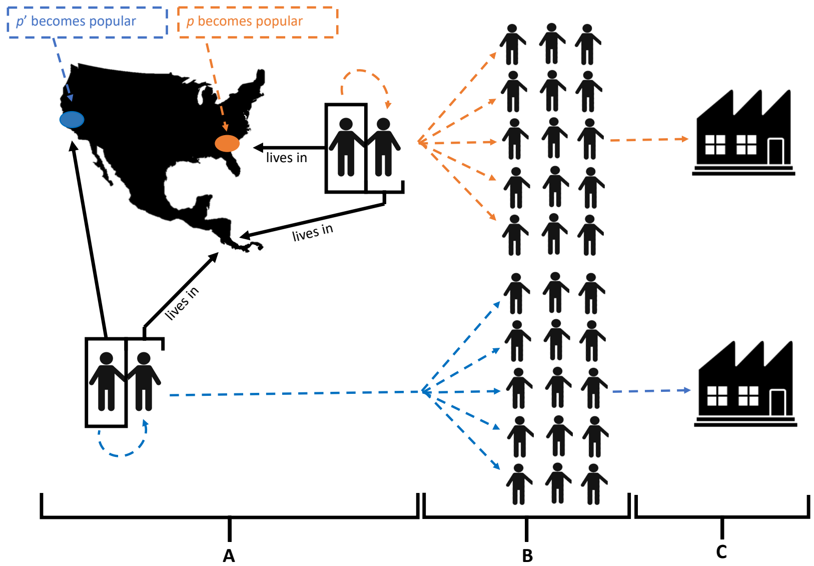
The spirit behind this instrument is how, anecdotally but also intuitively, information on product trends is transmitted to developing countries after relatives migrate to developed countries, where more products are available. This instrument exploits that (i) we can identify Costa Ricans living abroad and where they reside in the U.S. (1% of the Costa Rican population), (ii) we can link these immigrants to their family network still living in Costa Rica (5% of the population), (iii) we collected regional data from several sources to follow product-specific trends across the U.S., (iv) consumer trends in the U.S. do not respond to local conditions in Costa Rica, and (v) we can track Costa Rican product-specific purchases at a daily frequency. This strategy also has the large advantage that the analysis can be run using product-level consumption, as opposed to total consumption by an individual, which aids in separating the consumption network effect from income shocks. The instrument, summarized in Panel A of Figure 1, can strongly predict the product-specific imports of individuals with close relatives in the U.S. across time.[3] Thus, this instrument allows us to document how demand shocks can propagate across international migrant networks; which is, in itself, a contribution.

We then study if, after a Costa Rican with a relative living abroad increases her exposure to a specific product, other individuals in her network *without* relatives residing abroad become more likely to import *the same* product within one quarter (Panel B of Figure 1). We document that this pattern holds regardless of the definition of network that we consider. After someone experiences an exogenous increase in exposure to a product, individuals in her networks of neighbors, co-workers, and friends all become more likely to import the product. In fact, the effect is similar across types of networks.[4] An increase of one standard deviation in the probability of importing a product for those with relatives abroad leads to a 21% higher probability

[3]Our main specification defines "close relative" as including parents, siblings, own children, partner, partner's parents, siblings, and children.

[4]This similarity is partly due to us controlling for network-time, network-product, and product-time fixed effects; so that, for instance, differences in network size will not affect our coefficient. Note that we do not study the impact of the instrument on family networks, as the instrument is constructed based precisely on this network type.

Figure 1: Instrument Relying on International Family Networks and Exogenous Product Trends




*Notes*: The figure summarizes the idea behind our main instrument and analysis. The instrument (Panel A) leverages information on the family networks of migrants to different U.S. cities, along with variation on product trends across these cities. Panel B represents our second stage, in which we measure the effect of exogenous exposure to a product on the probability of importing the same product for people who share a network. Panel C represents our study of the supply response after individuals import a product.

of importing this specific product for people in the same neighborhood—but without foreign connections—within one quarter, as compared to the mean neighborhood. The corresponding increase in probability when considering people in the same firm is 17%, while for people in the same friends network the probability of importing the same product is 20% higher. As a robustness exercise, we push our data further and construct a distance-3 nodes instrument in the spirit of De Giorgi et al. (2019); results hold and remain statistically equal to the ones using our baseline approach.

While we document that an increase in exposure to a foreign good increases the probability of importing it within the network, not all products diffuse with the same intensity. For a given level of exposure, some products exhibit a strong propagation, i.e., many people within the network import them within a short period of time, while for others the propagation is relatively weak. We document that foreign products diffuse more strongly for goods in dynamic product categories, as opposed to established ones. These results are consistent with information frictions being key. Moreover, our results suggest that goods propagate more within a network if their initial importer is well-connected with others, and opposed to having low centrality.

After exploring the process of diffusion across individuals in a network, we document a new channel by which retailers learn about the local demand for foreign products. We find that retail firms in neighborhoods with a larger exposure to a foreign product are more likely to import this product within a short period of time (Panel C of Figure 1). The magnitude of the effect is quite large, for instance, a one standard deviation increase in the probability of importing a product for those with relatives leads to an 11% higher probability of the retailer importing the product within a quarter.

We then study the mechanism behind the response of retailers. We show that this effect is driven by products which exhibit high propagation among networks of individuals after they are first imported. In other words, if networks of people reflect a *strong* local demand for a foreign good—so that after an increase in exogenous exposure, many people import the product—then retailers respond to this observed high demand by importing the same product themselves. However, if instead the product has *low* propagation after a first import—which would be the case if local demand for the product is low—the retailers become *less likely* to import this product. These supply responses are mainly driven by small retailers, who face greater search frictions and may benefit more from information from individuals, and who might be more receptive to local consumers' needs. Taken together, our findings point to an indirect externality such that retailers—particularly small ones—are learning about the *level* of the local demand for foreign goods by observing the degree of propagation of imports across people in their catchment area after a foreign product is imported by individuals.

Finally, we combine the documented effects to grasp their overall impact. Our estimates imply that an additional 10 million USD in U.S. spending on a product increase *total trade in final goods* by 2.2%, as both Costa Rican individuals and retailers would import more from the U.S. via our channel, thereby creating a multiplier effect through an additional demand response. This is a sizable effect, which can also be expressed as an elasticity in dollars; we find that $1 of spending on a product in the U.S. leads to an additional 5 cents in imports from Costa Rica, due to both the direct and the indirect externalities that stimulate local demand.

**Related Literature** Our paper speaks to several strands of the existing literature. First, it contributes to a longstanding literature studying how social interactions can impact behavior (Bailey et al., 2022; Bandiera et al., 2009; Bandiera and Rasul, 2006; Bertrand et al., 2000; De Giorgi et al., 2010; Duesenberry, 1948; Duflo and Saez, 2003; Mas and Moretti, 2009; Veblen, 1899).

With respect to *consumption* network effects, in particular, the literature has been constrained by the availability of consumption data and has focused on direct externalities. We make progress by relying on (i) a battery of definitions of networks for the entire population and (ii) a product-specific consumption analysis, as well as by (iii) documenting indirect externalities from individuals to firms. First, the richness of the networks data allows us to estimate economy-wide effects and to document heterogeneous effects depending on, for instance, the centrality, demographics, or degree of homophily of the agents involved. Second, we are able to observe a vast array of products; the product-specific analysis delivers variation across products, which a key component of our identification strategy, but beyond that, allows us to explore heterogeneous effects across product types, for instance, depending on the dynamism of the product category or on its degree of visibility. These first two points push the current frontier as consumption data has typically come from relatively small surveys which rarely have a longitudinal component. De Giorgi et al. (2010) make progress in relying on consumption measures based on household income from tax records, but observe only *aggregate consumption*. Bailey et al. (2022), on their part, also make progress by using Facebook data to define networks, but are constrained to study a single product, cell-phones, to identify peer effects. These papers are therefore constrained in terms of studying heterogeneous effects across products and network types and estimating economy-wide effects. Finally, a third key contribution of this paper is to explore the interplay between direct externalities via peers and indirect externalities on retailers. While hard to measure, this is, to the best of our knowledge, the first direct evidence on how retailers can learn from consumers and use that knowledge to inform their sourcing choices. In other words, by analyzing if retailers respond to the observed local demand for a product, the paper explores new ground and investigates the indirect impact of network effects on supply-side responses.

This paper focuses on cross-border purchasing; a new avenue by which social interactions can have a significant impact on consumption, and that only promises to become more relevant as direct-to-consumer trade continues to grow. As such, the paper also contributes to the literature on international trade. First, to the best of our knowledge, this is the first paper to leverage and explore data on imports by individuals. While, historically, imports by individuals have been rare, the rapid expansion of the direct-to-consumer market is only expected to accelerate due to factors like more global internet penetration, better transport and logistics infrastructure, and overall more globalization. Therefore, this paper aims to be a starting point for what promises to be a rich area for future research. Second, the literature on foreign trade has studied the decisions of firms to "enter" (export to) a foreign market.[5] We document the presence of demand externalities across consumers, which should be taken into account when assessing the profitability of firms to enter a market. Further, we also document that, even when a firm does not export to a country, individuals can decide to import products, and the revealed information on local demand after this event may trigger *domestic* firms' import decisions, which is a new mechanism that had not been proposed by previous work. Our paper also speaks to the role of consumers as drivers of international trade, both directly via network effects and indirectly via their effect on retailers, which has been largely understudied. Instead, trade studies on the role of expenditures have largely focused on the effects of trade on inequality, both using microdata and exploiting major reforms in individual countries (Atkin et al., 2018; Faber, 2014; Porto, 2008), and leveraging theoretical frameworks to measure inequalities in gains from trade across consumers as in Fajgelbaum and Khandelwal (2016). Lastly, Akerman et al. (2022) document that internet access makes international trade patterns more sensitive to distance and economic size in a small open economy, an even more so for differentiated goods. While this paper focuses of trade between firms, this finding is consistent with our results, which constitute a micro-foundation for these aggregate patterns.

The paper also contributes to studies on the diffusion of ideas from developed to developing countries, which has been modeled by several recent papers, most of which assume that ideas are embodied in products, and can thus travel across countries via trade flows and increase productivity and growth (Alvarez et al., 2013; Buera

[5]For instance, Arkolakis (2010) models how a firm would export to a market if it is profitable to incur the marginal cost to reach a single consumer, and then faces an increasing marginal penetration cost to access more consumers.

and Oberfield, 2020; Perla and Tonetti, 2014; Perla et al., 2021; Sampson, 2016; Van Patten, 2021). These papers propose a causal channel though which trade flows lead to knowledge. Instead, our paper proposes a new mechanism in which knowledge flows through social networks can lead to increases in trade flows, both via individual- and firm-level imports. In terms of cross-country aggregate flows, our mechanism would then imply that stronger social networks across countries can lead to more trade between them. Indeed, Bailey et al. (2021) document that two countries trade more when they are more socially connected via Facebook, especially for goods where information frictions may be large. Our paper then provides a micro-foundation for these aggregate cross-country patterns—and, in particular, for the role of information frictions—using a range of definitions of national networks, retailer responses, and a research design that allow us to identify product-specific externalities.

The rest of the paper is organized as follows. Section 2 includes details on the data used in our analysis. We describe our estimation framework and empirical results for individuals' networks in Section 3, including a study of heterogenous effects across networks, demographics, and products. Section 4 describes our results for retail firms, and Section 5 concludes.

# 2 Data

We now describe a series of administrative datasets which we use in our analysis. Importantly, while the the data that we used is anonymized, the variables *across* data-sets can be linked via unique (pseudonymous) identifiers, so that the information described below can be combined at the individual level.

**Customs Data** Customs records are available for the period 2005-2019. Each import includes up to a 10-digit HS code, along with information on the amount transacted, the quantity traded, and the country of origin.[6] As in other countries, customs records are available for firm-level imports. In addition, if an *individual*

[6] The HS-10 classification exists only for a subset of goods. For some categories, the HS-10 classification does not exist, so the HS-8 or HS-6 code is the narrowest way to classify goods. In all cases, we use the most disaggregated category available. While HS codes are not barcodes, this can be seen as an advantage in our setting; a person might learn about a new type of flask bottle from a peer, but order a blue one instead of a green one, which would typically be in the same HS code but would not have the same barcode.

imports a good or service on her own (for instance, if she bought an item from an online retailer in the U.S. and shipped it to Costa Rica), this transaction is also recorded.[7] To the best of our knowledge, this is the first paper to leverage this type of information despite the fact that, far from an unusual practice, online shopping in general, and cross-border shopping in particular, has become prevalent. In fact, about 53% of total e-commerce in Costa Rica is cross-border, while the Latin American average is 39%; further, the annual growth rate for cross-border e-commerce in Latin America in 2022 was 44% (EBANX, 2023).

**Family Networks** We reconstruct nationwide family networks based on information from Costa Rica's National Registry. This data includes official information to reconstruct each person's family tree based on existing records and without relying on name-matching. The data is dynamic and at a monthly frequency, meaning that family networks can change over time between 2015 and 2019.

**Networks of Neighbors** Networks of neighbors for adult citizens can be constructed from official records maintained by the National Registry. While records include district of residence, and there are 488 districts across the country, they also include the voting center which is closest to each citizen's residence *for each adult citizen*. With 2,028 voting centers in total, the median number of adults assigned to each voting center is 586.[8] Thus, we leverage the latter to get a more precise measure of close-by neighbors.

**Networks of Coworkers** Matched employer-employee data was obtained from the Registry of Economic Variables of the Central Bank of Costa Rica, which tracks the

[7]While individual imports could potentially also include imports from informal sellers, it will become clear that this would only make our estimates a lower bound: if a person imports a product and informally sells it domestically, then the incentives for others to import it decrease, which would attenuate our coefficient of interest. It is worth noting that informal workers are a relatively small share of all workers in Costa Rica (27.4%), and significantly below Latin America's average of 53.1% (ILO, 2002). Moreover, nearly all individuals import less than 20 times in total within our sample period (2015-2019), , which suggests these are not informal sellers who use this method to stock, and dropping those who do import more regularly does not meaningfully change our results.

[8]For more details on voting centers, see Méndez and Van Patten (2022).

universe of formal employment at a monthly frequency between 2015 and 2019.[9] These data allows us to recover networks of coworkers which change at a monthly frequency, as people change their employers.

**Networks of "Friends"** We provide a novel measure of networks of friends which relies on connecting pairs of individuals who have sent money to each other *bilaterally*.[10] To do this, we use data on comprehensive transactions on Sinpe Móvil, an application that allows Costa Ricans to make peer-to-peer money transfers using their mobile phones, which has been adopted by over 60% of all adults in the country (Alvarez et al., 2023), and which processes the equivalent of 15% of GDP in transactions each year. The measure we construct is time-invariant, and its logic is the following: we start at the *end* of the sample period and retrospectively ask: which pairs of peers have made transfers to each other bilaterally? These pairs are considered friendships, which has the advantage of eliminating transfers to, for instance, a nanny or a housekeeper. This method allows us to proxy for networks of friends which are usually impossible to recover.

**Firm-Level Data** We leverage data on corporate income tax returns, which cover the universe of formal firms in the country and contain typical balance sheet variables, including sales. The data span 2015 to 2019, and includes details on each firm's sector and location.

The instrumental variables strategy proposed in this paper requires two additional data sources, which are described below.

**Networks of International Migrants** In Costa Rica, voting is mandatory, and it is one of the countries where more immigrants residing abroad register to vote in their corresponding Costa Rican consulate (approximately 51%).[11] The latter leads

[9]As previously mentioned, it is advantageous in our setting that informal workers are a minor share of all workers in Costa Rica (see footnote 12).

[10]For instance, if user A has only sent money to user B, we would not record this relationship as a friendship. If, however, both A and B have sent money to each other, then their relationship is classified as a friendship.

[11]For instance, the equivalent share of migrants residing in the U.S. and registered to vote in their home country in Mexico is 1.5%, and the median in Latin America is 17.6%.

to a registry of migrants' residence abroad, which is recorded by the Supreme Court of Elections. The data is updated monthly, and maps registered voters residing in foreign land to the consulate closest to their residence from 2014 to 2022.[12] Large countries, like the U.S., have several consulates, which are located in the cities with the largest mass of Costa Rican residents.[13] While this information is available in other countries, to the best of our knowledge this is the first paper to leverage it to recover international networks of migrants.

**Consumer Trends in the U.S.** We obtain U.S. consumer trends by product from two alternative sources, which complement each other. First, we rely on the Consumer Expenditure Survey (CEX), which includes quarterly data by Metropolitan Statistical Area (MSA) for 700 categories of products between 2015 and 2022.[14] While the MSAs for which estimates are produced do not span the entirety of the U.S.'s territory, they do include every city where a Costa Rican consulate is located, which correspond with the main cities where Costa Ricans reside abroad. In fact, according to the American Community Survey, over 82% of Costa Ricans living in the U.S. reside in one of these cities during our sample period. The CEX data (UCC codes) can be then mapped to HS codes using the concordance developed by Furman et al. (2017); also used by Hottman and Monarch (2020) and Borusyak and Jaravel (2021). The variation from this mapping is mainly at the HS-4 or HS-6 level, as CEX categories are often more aggregated than customs' HS codes.

While our main results are based on the CEX, we leverage a second source of data on consumer trends by product which aims to complement the CEX, precisely by providing variation for narrower product codes. The logic behind this second source is the following: In the U.S., many tradable products are imported. Thus, U.S. expenditures on these products by region should co-move with the imports of these products in those areas.[15] Following this idea, we use HS-10 level quarterly imports by customs districts in the U.S. from the Census Bureau, which include over

[12]The area serviced by each consulate is well-defined and public information.

[13]The cities with Costa Rican consulates with ratified voting centers are: Atlanta, Chicago, Houston, Los Angeles, Miami, New York, and Washington D.C. There are also honorary consulates in Minneapolis, Puerto Rico, and Tucson.

[14]Details on geographic coverage are available in the BLS website (link).

[15]Note how it is helpful that we will ultimately rely only on variation in (internationally) tradable products, not on changes in expenditures on non-tradables.

20 thousand product codes, to obtain variation at the HS-10 level. Conveniently, while HS-10 codes do not necessarily coincide across different nations, U.S. being Costa Rica's main trading partner, they do align for these two countries.[16]

The U.S. has 47 customs districts; instead of assuming a product is consumed in the same customs district it is imported into, we follow Acosta and Cox (2019) and allow for movements of imports within the U.S. using data from the Department of Transportation's Freight Analysis Framework (FAF), which provides estimates of where imported goods travel once they enter into U.S. borders across 132 FAF zones.[17]

Reassuringly, Appendix B.2 presents evidence—both in levels and in *changes*—of a strong correlation between expenditures in the CEX and the one-quarter lagged value of imports by product code and by region. This lag is intuitive, as it takes time both for goods which cross the U.S. border to arrive to retailers and for them to be consumed by households and show up in the CEX; thus, throughout the paper, we use one-quarter lagged U.S. imports as a proxy of contemporary expenditures on those products. Reassuringly, and in line with this strong correlation, we will show that all our main results are statistically equal regardless of whether we measure U.S. expenditures via the CEX or via the U.S. imports data.

To further validate the CEX, Appendix B.3 leverages transaction-level data on debit card expenditures by region and by Merchant Category Codes (MCCs) in the U.S., with over 10 million cards between 2017 and 2020.[18] While imperfect, as MCCs tend to be specialized codes designed for financial transaction tracking, it is reassuring that—just as in the case of customs data—the correlation between CEX and card expenditures by region and product code is strong both in levels and in changes.

# 3 Direct Externalities in Individual Imports

**Conceptual Framework** We propose a simple theoretical framework in Appendix A to guide our empirical analysis. We extend insights from Caplin and Leahy (1998) to

[16]More precisely, we manually check that the definition of all the HS-10 codes which are imported by individuals in Costa Rica is the same as in the U.S.'s Harmonized Tariff Schedule (HTS).

[17]While FAF zones are of moderate size, it will become clear later that, for our purposes, it is not crucial to pinpoint the precise location where a good was consumed; instead, we are interested in the broad area within the U.S. where consumption took place.

[18]These data come from Facteus, a provider of financial data for business analytics.

the process of importing final goods and aims to highlight three features: i) the initial delay in the adoption of particular foreign products, ii) the rapid adoption of these products in the network after someone first imports them, and iii) the subsequent adoption of these products by local firms. In the model, the initial delay in the adoption of a foreign product variety is because its quality or appeal is uncertain. As a result, agents wait until they have better information about a variety before importing it.[19] Importantly, the delay is not optimal and it is a consequence of a direct demand externality; individuals do not internalize that information is revealed to others in their network once they import a product variety. Once someone in their network imports this variety, agents gain information about its type and can use this information to decide whether they want to import it or not. Similarly, there is also an indirect externality: firms respond to the revealed information by importing a product, but only when the expected gains are sufficiently large. In fact, firms only import varieties with strong enough propagation among consumers after they are imported, as only in this case there are profits to be made from selling the foreign good domestically.

**Empirical Approach** In general, pinning down consumption externalities involves both data challenges, as it requires information on an individual's network, and identification problems, as the decisions of peers can be endogenous. Regarding network data, data constraints are key (De Paula, 2017) and have forced previous research to identify networks using characteristics which are common between individuals, like race, cohort, location, among others.[20] This paper undertakes an effort to combine several reference groups and create a relatively complete picture of the network of each individual in Costa Rica. In particular, we define networks in three different ways. First, we assume an individual's network is composed of those living in close proximity to her. This definition of neighbors is relatively precise, as our data disaggregates the country into small locations of approximately 600 people each.

Second, we assume that the relevant network is composed of co-workers. The latter

[19] For example, many brands of perfume available in the U.S. are not available in Costa Rica; as a result, Costa Ricans might need to find out more information about a specific fragrance before importing it.

[20] A few exceptions with more detailed data include Banerjee et al. (2013), who use details on the network of microfinance clients, and De Giorgi et al. (2019), who rely on co-workers as the relevant reference group.

is in line with De Giorgi et al. (2019), who identify co-workers as a good reference group given the large share of the day spent with them, among other reasons. Third, we create a proxy of friendship to generate networks. This proxy relies on data from *Sinpe Móvil*, an app that allows for peer-to-peer money transfers. As more than 60% of the adult population in Costa Rica is a user of the app and annual transactions in the app account for over 15% of the country's GDP (Alvarez et al., 2023), it is possible to identify close connections from the app's data under some assumptions, which will be discussed in detail later in this section. To the best of our knowledge, the breadth of our networks spans more ground than has been previously available, and allows us to compare the impact of demand externalities across different domestic networks.

Panel (a) of Table B.1 displays summary statistics describing each network. Compared to networks of coworkers and friends, networks of neighbors are fewer and larger in size. Networks of friends are more numerous and exhibit the lowest median number of members. Panel (b) of Table B.1 describes the number of product codes, both in the CEX and in the U.S. imports data, *which are imported by individuals* in Costa Rica, along with the top codes by import volume according to each data source.

Leveraging these data, we want to understand if, once an individual imports a product, the probability of importing the same product for people in her relevant network meaningfully changes. This poses endogeneity and econometric challenges, which have been pointed out by the previous literature, starting from Manski (1993). In particular, it is hard to distinguish the true network effect from common shocks and common personal characteristics, which might make individuals import a product for reasons which are independent from the influence of their connections.

## 3.1 Identification Strategy

We propose an identification strategy that leverages several aspects of our data. In particular, we construct an instrument based on the following idea: *Suppose both $L$ and $N$ live in Costa Rica, and $L$ has a sister living in Los Angeles (LA) while $N$ has a sister living in New York City (NYC). If product $i$ becomes more popular in LA as compared with NYC in period $t$, then $L$ is more likely than $N$ to import product $i$ in period $t+1$.*

Panel A of Figure 1 shows a diagram summarizing the instrument in a more general fashion. Following the figure's notation, if a family in Costa Rica has a member who migrated to a U.S. city on the West Coast, in blue, while another family has a member who migrated to a U.S. city on the East Coast, in orange, and a specific product $p'$ ($p$) becomes more popular on the West (East) Coast city as compared with other cities in the U.S., then relatives of each migrant in Costa Rica become more exposed to each of these products and are more likely to import it.

The spirit behind this instrument is how, anecdotally but also intuitively, information on product dynamics is transmitted to developing countries once relatives migrate to developed countries, where more products are available. The instrument exploits (i) that we can identify Costa Rican citizens who are living abroad along with the location where they reside in the U.S. (1% of the total population), (ii) that we are able to link these immigrants to their relatives who still reside in Costa Rica (5% of the total population is a close relative), (iii) that we collected data at the MSA-level and customs district-level to follow product-specific dynamics across the U.S. over time, (iv) that product-specific consumer trends in the U.S. do not respond to local conditions in Costa Rica, and (v) the availability of individuals' product-specific purchases at a daily frequency. The instrument also has the significant advantage that the analysis can be run at the product-level, as opposed to considering total consumption by an individual, which aids in identifying the true consumption network effect.

Our starting point is to construct a measure of consumer trends in the U.S., which can vary across time, cities, and products, but that we can purge from the impact of business cycles in the U.S., national product trends, and differential level effects. More rigorously, let $s$ denote a U.S. city, $p$ a product, $t$ a quarter, and $c$ a Costa Rican consulate in the U.S. Consider the following specification:

$$\ln E_{spt} = \alpha + \underbrace{\lambda_{sp}}_{\text{level}} + \underbrace{\mu_{st}}_{\text{local business cycles}} + \underbrace{\phi_{pt}}_{\text{national product trends}} + \epsilon_{spt}, \tag{1}$$

where $E_{spt}$ are expenditures in city $s$ on product $p$ at time $t$. Let $\ln \widetilde{E}_{spt}$ be residuals of this regression; these residuals would then capture the differential changes in product taste across U.S. cities. Note that equation (1) includes fixed effects that would prevent $\ln \widetilde{E}_{spt}$ from varying (i) because people in a city are more prone to buy a certain product, for example, people in Chicago buying more winter coats (level effect); (ii) because a particular region had a positive or negative income shock (local cycles); or (iii) because a product became more or less popular in general (national product trend). As Costa Rican consulates sometimes span several cities, we proceed by constructing a *consulate-product-time* specific exposure, namely:

$$\ln \widetilde{E}_{cpt} = \sum_{s \in c} \left( \frac{CR_s}{CR_c} \right) \ln \widetilde{E}_{spt}, \tag{2}$$

where $CR_s/CR_c$ is the share of Costa Ricans living in consulate $c$ who reside in city $s$. This share is time invariant. That is, we aggregate our expenditures measure to the consulate-level weighting by population shares in each city.[21]

The expenditure shares in this regression can be measured in two alternative and complementary ways: through the CEX by MSA and through U.S. imports by customs district, as explained in Section 2. In both cases, regions (MSAs and customs districts), are aggregated following equation (2). These measures complement each other: the CEX is more representative of overall consumption than imports, but its relatively aggregated categories provide variation which is mostly at the HS-4 level, as shown in Panel (a) of Table B.2. The imports by customs district proxy consumption via HS codes which have up to 10 digits and provide variation at a more precise level, as shown by Panel (b) of Table B.2, but do not capture changes in domestic production. It is worth noting, however, that since our focus lies solely on internationally traded products, the omission of non-tradables is not a pronounced restriction.[22]

**Network-Level Exposure** We proceed by constructing measures of exposure at the network-level to set up the first stage of our instrumental variables strategy.[23]

[21] Costa Rican residents by city are obtained from the American Community Survey. The share is time invariant as the average Costa Rican residing in the U.S. by 2019 migrated in 1994; thus, movements abroad are rare, and could lead to selection which we prefer to shut down. Therefore, we fix these shares in 2014, one year before our sample period starts.

[22] As shown in panel (b) of Table B.1, top codes imported by Costa Ricans are manufactured goods (clothing, toys, etc.), which are likely to have been imported into the U.S.

[23] In Section 3.3, we run first-stage regressions at the individual level, which deliver statistically equal results to those at the network-level. This subsection also explains why it is not feasible to conduct a product-level analysis at the individual level.

The exposure of Costa Rican network $b$ to product $p$ via imports is then given by:

$$\ln \widetilde{E}_{bpt} = \sum_c s_{bct} \ln \widetilde{E}_{cpt},$$

where $s_{bct}$ is the share of people in network $b$ who have a relative living abroad in $c$ at time $t$.[24] It is worth noting that $s_{bct}$ only varies across time due to movements *within* networks in Costa Rica, but Costa Ricans residing in a U.S. consulate are time invariant, so that $s_{bct} = N^c \frac{s^c_{bt}}{N_{bt}}$, where $N^c$ is the total of Costa Ricans who have relatives in $c$ (and does not depend on time), $s^c_{bt}$ is the share of Costa Ricans with relatives in $c$ who belong to network $b$ at time $t$, and $N_{bt}$ is the total members of $b$ at time $t$.[25] Note that, as the network of friends is time-invariant, $s_{bct}$ would not be changing across time for this network. We then consider the following linear probability model for network $b$:

$$\underbrace{\text{Import}^{US\,exposure}_{bpt}}_{\substack{\text{Prob. importing } p \text{ for those} \\ \text{with relatives in the U.S.}}} = \alpha + \beta \underbrace{\ln \widetilde{E}_{bp,t-1}}_{\text{Network exposure abroad}} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt}, \qquad (3)$$

where $\text{Import}^{US\,exposure}_{bpt}$ equals one if an individual in $b$ imports product $p$ *for the first time* at time $t$ and is zero otherwise. $\gamma_{np}$, $\gamma_{nt}$, and $\gamma_{pt}$ represent network-product, network-time, and product-time fixed-effects, respectively. The regression is run separately for each type of network, so that $b \in B$ and $B$ is either a neighborhood, a firm, or a friends network; moreover, standard errors are clustered by network-product.[26]

A few remarks are in order. First, this regression includes *only* imports and exposure of people who reside in Costa Rica *and* have a relative living in the U.S. Second, note the left-hand-side variable is quite conservative: if *anyone* in the network

[24]For most people who have relatives in the U.S., all relatives live in the same consulate. However, there are a few individuals who have relatives living abroad in different consulates. For them, this regression considers a weighted sum of relatives as the main regressor, namely, $\sum_c s_{nc,t-1} \ln \widetilde{E}_{cp,t-1}$, where $s_{nc,t-1}$ denotes $n$'s relatives who reside in consulate $c$ as a share of all her relatives who live in the U.S. Controls at the product-level imply HS-4 codes; for the case of the CEX, in particular, this is extremely demanding, as most of the variation is at the HS-4 level.

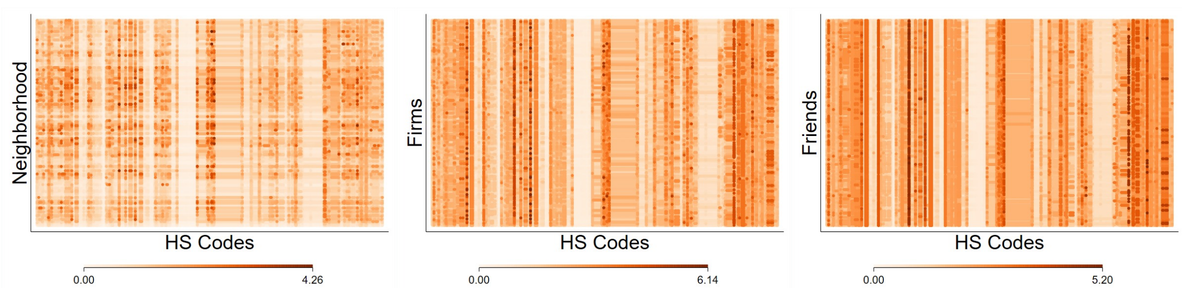
[25]Just as before, Costa Ricans living abroad are fixed in 2014 to avoid selection issues and as migrations abroad tend to be infrequent and permanent.

[26]Appendix D.1 explains why, in our particular setting clustering by network-product is sufficient and even on the conservative side.

was importing the product, its value is one. Thus, changes in this variable should arise from unusual first-time imports.[27]

Third, our battery of fixed-effects is an additional aid in identifying the effect of relatives' exposure abroad, as opposed to shocks faced at the network-product, network-time, or product-time level. For example, $\gamma_{bp}$ would address if a neighborhood is rich and tends to import some products more, $\gamma_{bt}$ would capture if a network was exposed to a shock and prevents biases which are constant across products, and $\gamma_{pt}$ would capture if a product becomes more popular in Costa Rica in a given period.[28] While we saturate the regression with these fixed effects, significant variation remains; to visualize it, we compute the term $\ln \widetilde{E}_{bp,t}$ *netted of all the fixed-effects* in equation (3), and then calculate the variance of this term for each network-product pair. Figure 2 shows that, regardless of how we define a network, there are significant differences in this variance both across networks and product codes.

Figure 2: Identifying Variation by Network-Product Pair




(a) Neighbors (b) Coworkers (c) Friends

| **Top codes:** | **Bottom codes:** |
|---|---|
| Women's overcoats...not knitted or crocheted (62021100) | Footwear for sports (640391, 640399) |
| Gym equipment (95069100); surfboards & water-sport equipment (95062900) | Printed books (49019900) |
| Medical, surgical, or laboratory sterilizers (841920) | Waterproof footwear (6401100000) |

*Notes*: The figure summarizes the differences in the identifying variation across network-product pairs. Namely, we compute the term $\ln \widetilde{E}_{bp,t}$ netted of fixed-effects, and then calculate the variance of this term for each network-product pair. Each panel shows this variance for a network type.

[27]Note that measurement error on the left-hand-side variable would in general not bias this coefficient. This result holds as long as the exposure's residual is uncorrelated with the measurement error, which in our case is likely to occur. Figure B.1 shows the frequency of importing events among individual importers with a relative abroad. An event is defined as importing a particular product; as shown, almost all individual importers import a particular product only once.

[28]Controls at the product-level imply HS-4 codes for the case of the CEX—which this is extremely demanding, as most of the variation is at the HS-4 level—and HS-6 codes, for the case of U.S. imports data.

Fourth, equation (3)'s timing is based on an Akaike information criterion, which also guides the timing of other regressions in the paper. Appendix D.6 conducts local projection exercises (Jordà et al., 2020; Montiel Olea and Plagborg-Møller, 2021), which support the choices of lag structure for all regressions in the paper. Finally, equation (3) describes a linear probability model. Angrist and Pischke (2009) explain that, while a nonlinear model may fit the conditional expectation function for limited dependent variables more closely than a linear model, this probably matters little for marginal effects. Moreover, non-linear models involve many decisions (e.g., weighting scheme, derivatives versus finite differences), while OLS is standardized. Further—and particularly relevant for our case—non-linear models involve important complications when dealing with IVs and panel data, some of which are yet to be resolved by the literature.[29]

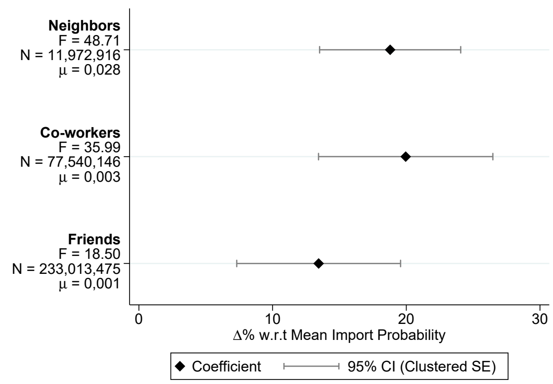
The results of this first-stage regression are shown in Figure 3.[30] Instruments are strong for every network, as reflected by the F-statistics to the left of each panel. Moreover, results are remarkably similar across networks of neighbors, co-workers, and friends: a one standard deviation increase in the exposure to a product of people *with relatives abroad* leads to a 13-20% higher probability of importing this product next quarter, as compared to people *with relatives abroad* in the mean network.[31] Figure D.2 displays analogous results relying on U.S. imports by customs districts to construct our instrument; reassuringly, these results are statistically equal to the baseline results based on CEX data.

[29] Angrist and Pischke (2009) also mention challenges with inference and correct calculation of standard errors in non-linear models, and quote Deaton (1997), who mentions "Absent knowledge of $F$ [the distribution of the errors], this regression function does not even identify the $\beta$'s [Tobit coefficients]—see Powell (1989)—but more fundamentally, we should ask how it has come about that we have to deal with such an awkward, difficult, and non-robust object." Deaton, 1997, p. 230.

[30] Tables corresponding with these figures are reported in Appendix D.3. Appendix D.5 provides details on the samples of products used in each regression.

[31] A one standard deviation is equivalent to an increase in spending of 69 million USD. The average spending per product in a consulate in a given quarter is 84 million USD. Note that regressions control for network-time fixed effects, thus, for instance, network size would not affect our coefficients.

Figure 3: First Stage: Imports by Costa Ricans with Relatives in the U.S.




*Notes*: The figure describes our first stage results by network when constructing exposure measures based on data on expenditures by product and by MSA from the CEX. Figure D.2 shows analogous results relying on expenditures data from imports by U.S. customs districts. The horizontal axis describes the effects as a percentage change with respect to the mean import probability in each network. Gray horizontal bars denote 95% confidence intervals, clustered by network-product. Mean import probabilities are reported to the left of each panel. The same left panel also reports the F-statistics of this first stage. All regressions include network×product, network×time, and product×time fixed-effects. Data is quarterly and spans 2015-2019. Tables corresponding with these figures are reported in Appendix D.3. Appendix D.5 presents details on the sample of products used in each regression.

## 3.2 Propagation of Importing Probabilities Across Networks

Using our network-product-time measure of exposure, we want to understand if people in a network *who are unrelated to migrants in the U.S.* increase their probability to import a particular product after being exposed to it through their peers who do have relatives abroad. Thus, from our IV's first stage, we will use the predicted values for the probability of importing a product for those in the network *with relatives in the U.S.* as explanatory variable. Our dependent variable would instead depend on the probability of importing a particular product for people in the network without relatives in the U.S., as follows:

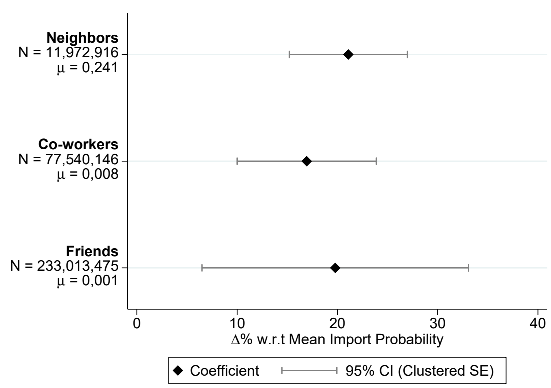
$$\text{Import}_{bpt} = \alpha + \beta \widehat{\text{Import}}^{US\,exposure}_{bp,t-1} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt}, \tag{4}$$

where $\text{Import}_{bpt}$ equals one if people in network $b$ *without relatives in the U.S.* import product $p$ at time $t$, and where we again include a battery of fixed-effects so that we only exploit $bpt$-level variation.[32] Just as for the first stage, we consider three

[32] Just as in the first stage, note that the network-time fixed-effect would prevent estimates from varying depending on network size.

different types of networks: neighbors, coworkers, and friends, and run independent regresions for each of them. Why so many networks? Each of them has different, and complementary, strengths and limitations. The network of neighbors spans all the population and will allow us to study the role of indirect externalities in triggering a supply-side response; however, it can be argued that there is less connectivity among groups of neighbors who are composed of hundreds of people than in other smaller networks. The networks of coworkers provide strong identification given our empirical strategy, however, they only span the formally employed; 41% of the population. The networks of friends are a novel way of measuring connections beyond observables, but the analysis is limited to those people who have adopted the mobile payment app; 60% of the population. Details on how we define this last network are presented in Appendix C. Overall, we believe that utilizing all networks paints a better and more robust picture of the role of externalities in product adoption.

Figure 4: 2SLS: Propagation Within Network




*Notes*: The figure shows the two-stage least squares regression corresponding with equation (4) when constructing exposure measures based on data on expenditures by product and by MSA from the CEX. Figure D.3 shows analogous results relying on expenditures from imports by U.S. customs districts. The horizontal axis describes the effects as a percentage change with respect to the mean import probability in each network. Gray horizontal bars denote 95% confidence intervals, clustered by network-product. Mean import probabilities are reported to the left of each panel. All regressions include network×product, network×time, and product×time fixed-effects. Data is quarterly and spans 2015-2019. Tables corresponding with these figures are reported in Appendix D.4. Appendix D.5 presents details on the sample of products used in each regression.

The baseline results of the two-stage least squares (2SLS) estimations are shown in Figure 4.[33] The horizontal axis describes the effects as a percentage change with

[33]Tables corresponding with these figures are reported in Appendix D.4. Appendix D.5 provides details on the samples of products used in each regression.

respect to the mean import probability in each network, and the means of each independent variable are reported to the left of the figure.[34]

The magnitudes of the 2SLS coefficients are similar across networks; we find that a one standard deviation increase in the probability of importing a product for people with relatives in the U.S. leads to an increase of between 17% and 21% in the probability to import this specific product next quarter for people in the network without relatives in the U.S., as compared with the mean network.[35] Reassuringly, results when relying on U.S. imports data to construct our instrument are very similar (and statistically equal) to the baseline results based on the CEX, as shown in Figure D.3. Further, as discussed in Appendix D.6, this effect is persistent: after an exogenous import, at least one person in the network without relatives abroad keeps importing the product each quarter. Appendix D.2 discusses the ordinary least squares (OLS) and the reduced form results.

Our baseline regressions' dependent variables are quite conservative: if *anyone* in the network was importing the product, its value is one. Thus, changes in this variable should arise from unusual imports. It is possible to also conduct an intensive margin analysis, which considers changes in the quantity of importing events. Appendix D.7 reports these intensive margin effects, which are in line with our baseline results. Taking networks of neighbors as an example, the estimates imply that a one standard deviation increase in exposure leads to a 19% increase in imports of people with relatives abroad, with respect to the mean import quantity. This effect is followed by a 23% increase in imports for those without relatives abroad, with respect to their mean import quantity. A one standard deviation equals 69 million USD in additional spending. Thus, to put this differently, if U.S. spending on a product increases by 10 million USD, then total cross-border *individual* imports would increase 3.45%.

[34]Note that mean import probabilities differ substantially by network, which is why normalizing the effects with respect to the mean import probabilities is a helpful way to compare them. The main reason why mean import probabilities differ is intuitive: some networks, like neighborhoods, are large (hundreds of people) compared with coworkers who have an average of less than ten people. The smallest probability is for friends, as the left-hand-side variable in this case is individual-specific, as explained in detail in Appendix C. Similarly, import probabilities are smaller when relying on U.S. imports than on the CEX, as the former has more narrowly defined product categories.

[35]One standard deviation in the probability of importing a product $p$ by people with relatives in the U.S. is 2% for neighbors, 1% for coworkers, and 0.4% for friends.

## 3.3 More Demanding Specifications and Robustness Exercises

It is worth spelling the exclusion restriction of our instrument. Our identification strategy requires that the likelihood of buying product $p$ of a Costa Rican—without relatives abroad—in a network connected to a U.S. city via family ties co-moves with *changes* in expenditures on $p$ in this U.S. city only through the relatives' influence.[36] Arguably, our main specification, which is saturated with a battery of fixed effects, takes care of most first-order concerns related to this statement. To complement it, we now conduct a series of robustness exercises with yet more demanding specifications.

**Instrument Using Distance-3 Nodes** We can push our data to better understand the robustness of the effects that we documented. In particular, we construct an alternative instrument which can rule out several alternative hypotheses. For instance, it allows us to verify that our results are not driven by the co-movement of product-specific tastes within a network.[37] The instrument exploits that we have information on both co-workers and spouses, and that spouses who work at different firms can be seen as a bridge between sets of co-workers that are otherwise disjoint; an observation that is present in De Giorgi et al. (2019). This approach leverages the existence of intransitive triads, and in our case, would rely on the notion that *if the co-worker of the spouse of my co-worker* has a relative in the U.S. and becomes exogenously more likely to import a product—controlling for common shocks experienced at my firm—this should not influence my probability of importing this exact same product directly, only indirectly though peer effects.

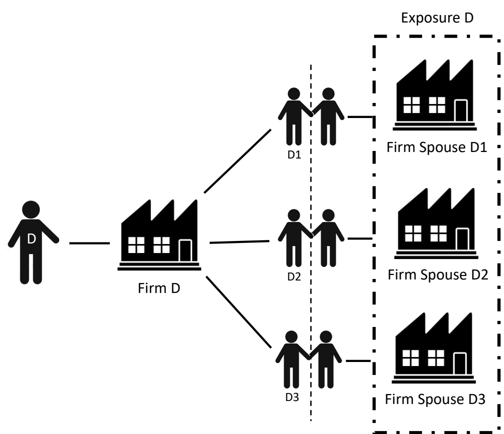
Figure 5 presents a diagram to make this notion more clear; it considers an individual $d$ working at firm $D$. The individual's exposure to a particular product $p$ depends on the exposure to $p$ faced by the spouses of her co-workers (D1, D2, D3)

[36]While, given our specification and fixed effects, this statement only has to hold in changes, Appendix B.5 also presents evidence in support of this statement in levels. Namely, we find balanced observables (age, gender, and wage) among Costa Ricans who migrate to different U.S. cities.

[37]Note that this threat already makes a hard conjecture: to represent a problem for our estimation, it would have to be the case that people reside in U.S. cities where product-specific trends *change* in synchrony with theirs over time, and that for some reason *unrelated* to their influence, their relatives who live in Costa Rica have product-specific tastes that also co-move with those of the U.S. city where they live.

at their firms, which in turn depends on the family ties that employees of those firms have with people residing in different U.S. cities, and on how expenditures of product $p$ evolve in those cities. Note that, by nature of the leave-out structure of the measure, each employee at a firm will have a different exposure.

Figure 5: Diagram of Instrument Using Distance-3 Nodes



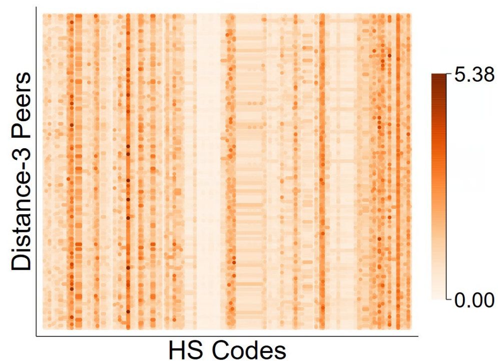

*Notes*: The figure shows the idea behind our instrument, where the relevant exposure is product-specific, time-varying, and depends on exogenous consumer trends, as described in Section 3.1.

We then consider the following regression for individual $d$, which depends on product $p$ at time $t$:

$$\underbrace{Import_{dpt}}_{\text{individual } d} = \delta_0 + \underbrace{\theta_3 \widehat{Import^d_{ip,t-1}}}_{\text{average exposure}} + \underbrace{\delta^d_{xpt}}_{\text{firm-product-time FE}} + \underbrace{\delta_d}_{\text{individual FE}} + \varepsilon_{dpt}, \qquad (5)$$

where the dependent variable, $Import_{dpt}$, is the probability of consuming (importing) foreign good $p$ for individual $d$ at time $t$. On the right-hand side of the regression, $\widehat{Import^d_{ip,t-1}}$ is the average exposure of individual $d$, which is instrumented by the mean residuals of firms employing the coworkers' spouses, $\widetilde{E}_{dp,t-1}$. $\delta^d_{xpt}$ are own-firm×product×time fixed-effects; *these fixed-effects are key, as they force the identifying variation to come from differences between the coworkers' spouses firms and individual d's employer* (see Figure 5 for reference). Finally, $\delta_d$ are individual fixed-effects. Of course, this specification in very demanding, however, there is still significant variation left across individual-product pairs; this identifying variation is summarized in Figure 6. We also list the product codes which were both among the top 10 codes of this sample and were not included in the top codes in Figure 2; remarkably, all of these have to do with work-related products. More details on the instrument and data construction are available in Appendix E.1.

Figure 6: Distance-3 Nodes: Identifying Variation by Individual-Product Pair




| **Distinct top codes**: |
|---|
| Wooden furniture of a kind used in offices (940330) |
| Pens and marker parts (960899) |
| Women's suits of synthetic fibers (610413) |

*Notes*: The figure summarizes the differences in the identifying variation across individual-product pairs. Namely, we compute the term $\widehat{Import}^{d}_{ip,t}$, netted of the fixed-effects in equation (5), and calculate the variance of this term for each individual-product pair. We also display the product codes which were both among the top 10 codes of this sample and were not included in the top codes in Figure 2.

<u>Key advantages of this instrument</u> The distance-3 nodes instrument above is immune to several identification concerns. As an example, consider correlated preferences among people in a network and their relatives abroad. For instance, people with relatives in NYC have different product-time demands than those with relatives in Houston, and their friends in Costa Rica show the same differential demand patterns. The concern would than be that there might be into migration locations, and so people go to NYC given that city has preferences correlated with them and also have Costa Rican friends/colleagues who are similar to New Yorkers.[38] Given this scenario—potentially the worst possible for our baseline instrument–there are two possibilities: (a) there is *not* assortative matching in the marriage market along lines which influence product demands, in which case the instrument constructed based on Figure 5 would deliver a correct estimate; or (b) there is *is* assortative matching in the marriage market along lines which influence product demands, in which case, the own-firm-product-time fixed effect in equation (5) would co-move with our instrument and would prevent $\theta_3$ from being identified from such assortative matching; again,

[38]Note this would have to happen while maintaining balanced observables of migrants across locations (see Appendix B.5).

this approach would deliver a correct estimate.

We find evidence supporting the existence of peer effects in our context, even under this more demanding specification. As shown in column (3) of Table 1, the effect, statistically significant at the 1% level, is an increase of 20% in the probability of the individual importing the specific product within one quarter, with respect to the mean probability of importing. This magnitude is roughly identical to the one documented in column (2), which relies on our baseline IV and the same sample.

Table 1: 2SLS: Individual Imports and Distance-3 Exposure

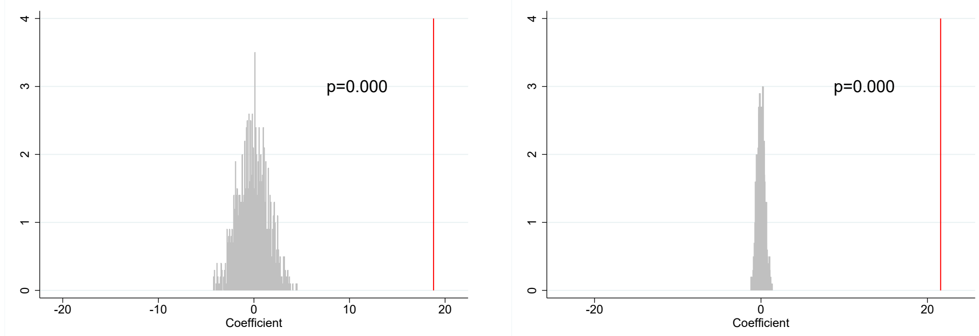
*Dependent variable: $Import_{dpt}$*
*(Prob. of individual d of importing product p at time t)*

| | %Δ w.r.t. mean import probability | | |
|---|---|---|---|
| | OLS | Baseline IV | Distance-3 IV |
| | (1) | (2) | (3) |
| $\widehat{Import}^{d}_{ip,t-1}$ | 29.243 | 21.056 | 20.289 |
| | (11.559)*** | (7.635)*** | (6.417)*** |
| Mean dep. variable | .0002 | .0002 | .0002 |
| F-stat first stage | – | 23,694 | 261,143 |
| *xpt* and *d* FE | Yes | No | Yes |
| *dp*, *dt*, and *pt* FE | No | Yes | No |
| Observations | 348,983,304 | 348,983,304 | 348,983,304 |

*Notes*: All estimations in this table are constructed based on the same sample and run at the individual level. Robust standard errors, adjusted for clustering by individual-product, are in parentheses. Regressions in columns (1) and (3) control for own-firm×product×time and individual fixed effects, while column (2) controls for individual-product, individual-time, and product-time fixed-effects. Mean import probabilities are reported. Appendix D.5 presents details on the sample of products used in each regression.

**Placebo Exposure Measures** We now conduct a randomization exercise in the spirit of Dell et al. (2019) to show that our results do not arise by chance. For each product-network, we randomly re-assign exposure measures $\tilde{E}_{bpt}$ *within HS-4 code category*. For instance, the shock for women's trousers made of wool is replaced by a random shock from another product within the women's clothing category. We then conduct both our first-stage and the reduced form regressions using the reassigned exposure, and repeat this exercise 1,000 times—we focus on the reduced form since there will no longer be a first stage for the IV. Figure 7 plots the distribution of placebo coefficients and depicts the actual coefficient based on the "true" exposure with a vertical red line. The actual coefficients are far in the tails of the placebo distributions, indicating that our effects are unlikely to arise by chance. These results define networks as neighborhoods; other networks are reported in Figure E.1.

Figure 7: Placebo vs. Actual Coefficients




(a) First Stage (b) Reduced Form

*Notes:* The figures plot the distribution of placebo coefficients obtained based on placebo exposure measures defining networks as neighborhoods. Results for other networks are in Figure E.1 The red vertical lines plots the actual first stage (panel a) and reduced form coefficient (panel b). The p-values represent the share of 1,000 placebo coefficients that are larger in magnitude than the absolute coefficient for the actual first stage or reduced form.

**Individual-Level Exposure** While our main empirical strategy involved measures at the network-level, we can explore if our proposed exposure measure can predict relative's imports *at the individual-level*. To do so, consider an individual $n$ with a relative abroad living in a U.S. consulate $c$, and the following specification:

$$\text{Import}_{npt}^{US\,exposure} = \alpha + \beta \underbrace{\ln \widetilde{E}_{cp,t-1}}_{\text{Relatives' exposure abroad}} + \gamma_{np} + \gamma_{nt} + \gamma_{pt} + \varepsilon_{npt}, \qquad (6)$$

where $\text{Import}_{npt}^{US\,exposure}$ equals one if individual $n$ imports product $p$ *for the first time* at time $t$ and is zero otherwise, $\gamma_{np}$, $\gamma_{nt}$, and $\gamma_{pt}$ represent individual-product, individual-time, and product-time fixed-effects, respectively, and standard errors will be clustered by individual-product.

Results are shown in Table E.1.[39] We find that an increase of one standard deviation in the exposure to a product coming from relatives abroad leads to a 13%-18% higher probability of importing this specific product with respect to the mean Costa Rican with a relative abroad. Remarkably, these effects are statistically equal to those when conducting the analysis at the network-level.[40]

Why not run all the analysis at the individual level? This first-stage regression

[39] Appendix D.5 provides details on the samples of products used in each regression.
[40] We also report results with an instrument relying on U.S. customs data.

considers *only* Costa Ricans with a close relative in the U.S. (parent, sibling, children, partner), but it already has over 500 million observations, even though those with a relative are roughly 5% of the population. The latter magnitude should make it clear that, for our second stage, which considers all Costa Ricans *without* close relatives in the U.S. (i.e., the remaining 95% of the population), we cannot run regressions at the individual level; we can, however, run an analysis at the network level, as we did in our main specifications.[41]

**More Demanding Controls** Adding certain fixed-effects to our specification can be a powerful tool to rule out alternative hypotheses. We start by considering a *district*×*product*×*time* fixed-effect, and re-run the analysis defining networks as neighborhoods. Recall that our variation is at the *neighborhood*×*product*×*time* level, so including this fixed-effect limits us to consider variation *within* small areas.[42] Results remain largely unchanged, as shown in column (1) of Table E.2. This is useful, for example, to rule out a story where a seller is targeting an area of the country with advertising about a product.

In a similar spirit, we can add a *network*×*HS-2 product code*×*time* fixed-effect to our analysis. Thus, we would only be exploiting variation *within* relatively narrow product categories. As column (2) of Table E.2 shows, effects are again largely unchanged. Like the exercise using distance-3 nodes, this result speaks against people from a certain network having a preference for a product category, and thus moving to cities where this category is trendy.[43] Also like the analysis with distance-3 nodes, this control would take care of sector-level trends in particular cities.

### 3.3.1 A Remark on Identification

It is not possible to observe all the connections that each Costa Rican has with the U.S. We cannot observe Costa Ricans who are not registered at a U.S. consulate, but beyond this, people might know U.S. residents and communicate with them, even if they are not relatives. Nonetheless, we believe we have enough evidence to show that

[41] In fact, the 2SLS could have, at most, a handful of products if run at the individual-level.

[42] Each district has four neighborhoods on average.

[43] The logic behind the fixed-effect is that, for example, a person might move to NYC because she likes fashion (HS-2), but is unlikely to move because she likes female trousers made of wool.

this is not a major concern. We now briefly summarize it, discussing the evidence in ascending order of strength. First, recall that our second stage includes network-time fixed-effects, which would prevent biases which are constant across products; for instance, suppose there are Costa Ricans living in the U.S. and who we cannot link to a network, this would *not* bias the coefficient as long as their information transmission is symmetric across products. Second, we can relax this condition further: column (2) of Table E.2 shows that results hold controlling for *network*×*HS-2 product code*×*time* fixed-effects, which means that the missed links are not creating a bias as long as the information transmission is symmetric *within* product categories (i.e., people are as likely to transmit information on women's trousers made of wool vs. other clothing items). Third, while the previous conditions might be harder to be satisfied for neighborhoods (for instance, because people from the same neighborhood might be more likely to know people in the same U.S. city who we are not directly linking), this is much more likely to hold for other networks like coworkers. Finally, the distance-3 nodes instrument includes own-*network*×*product*×*time* fixed effects—thus it is immune to this concern—and still delivers results which are statistically equal to the ones in our main specification.

## 3.4 All Products Are Not Imported Equal

We presented causal evidence that importing a product strongly depends on the imports of people in one's network. Of course, while this statement is true on average, some products can exhibit stronger propagation than others. For example, the response of a network might be asymmetric depending on whether the product became more vs. less popular, i.e., whether the change in exposure was positive or negative. In the same spirit, if a product category is particularly innovative, one might expect the propagation to be stronger; while if the product category is not very dynamic, then the observed propagation might be weaker. Propagation might also be stronger if the product is initially imported by a well-connected individual and high centrality, as opposed to a relatively isolated person. We now explore the determinants of the strength of propagation of a product within a network after it is imported.

**Positive vs. Negative Changes in Exposure** Our instrument depends on residuals, as described by equation (1), and therefore has mean zero by construction. Our analysis has so far assumed that the impact of positive and negative changes of this residual is symmetric, however, this does not need to be the case. For instance, intuitively, people might be more likely to transmit information about novelties than about products for which individuals lose interest; in such a case, one would expect the impact of positive changes in exposure to be stronger than the one of negative changes. Indeed, Table F.1 documents that positive values of exposure have an impact approximately *20 times* larger than negative values (in the opposite direction) in our first stage.

**Dynamic Product Categories** In line with the previous exercise, we now compare dynamic product categories with established ones; one would expect the information channel to be particularly relevant for product categories in which there is more dynamism. To explore this, we use Business Dynamics Statistics (BDS) data. The BDS tracks changes in establishments with paid employees over time, providing annual measures of establishment openings and closings and job creation and destruction. These measures are available for the entire economy, and by industrial sector, 3-digit and 4-digit NAICS, state, and MSA.[44] Namely, we use data on the creation of jobs by new establishments and on the entry of new establishments by product category to classify a product as "dynamic" ("established") if its creation of jobs by new establishments and entry of new establishments is above (below) the median within our sample (2015-2019).[45] Table F.2 shows our results, which are consistent regardless of how we define expenditures by product for the instrument and of our definition of dynamic products. In particular for networks of neighbors, we document a stronger propagation of products in more dynamic categories, as shown by the positive coefficients in the interaction terms, which aligns with our narrative and sheds light on the variation's driving forces. These findings align with the results on positive vs. negative changes in exposure presented earlier in this section.

[44]The BDS is created from the Longitudinal Business Database (LBD), a confidential database used by qualified researchers via secure Federal Statistical Research Data Centers.

[45]In particular, the two variables we construct are: (i) entry of establishments, which is the share of new establishments over total establishments in a product category; and (ii) employment gains from new establishments, which equals the share of jobs created by new establishments to total employment in the product category. These elements are then used to define the variable $Dynamic_p$ used in columns (1)-(3)—definition (i)—and columns (4)-(6)—definition (ii)—of Table F.2.

**Centrality of the Importer** Products might propagate more if they are initially imported by someone more connected to others. To explore this notion, we create a measure of degree centrality, which depends on how many friends a person has, using our app-based definition of friendship (i.e., bilateral transactions).[46] We then interact the average centrality of people *with relatives abroad* by network with their probability of importing a product and run our 2SLS. Results in Table F.3 suggest that the more central the importers in the first stage, the stronger the propagation across the network in the second stage. While these results are indicative, note that the interaction terms are noisy; this aligns with recent findings from Akbarpour et al. (2023), who document that choosing optimal seeds can have limited impact on diffusion within a social network.

# 4 Indirect Externalities in Retailers' Imports

The previous section documented the presence of direct demand externalities in individual imports; after an individual imports a foreign product, other members of her network become more likely to import the same product.

Once individuals in a network decide to import a product, there might be useful information about the local demand for this particular product which becomes available to domestic retail firms. In particular, retailers should be eager to import a new product the more locals are willing to acquire it over options available domestically, i.e., the stronger the observed propagation after a first import.

We now test these forces by considering the following regression:

$$\text{Import}^F_{bpt} = \alpha + \psi \widehat{\text{Import}}^{US\,exposure}_{bp,t-2} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt}, \tag{7}$$

where $\text{Import}^F_{bpt} = 1$ if a retail firm in $b$ imports product $p$ at time $t$. Throughout this section, $b$ is always defined as a neighborhood, as retailers' decisions are likely to be influenced by their clients, who in the case of retailers often coincide with people living close to the firm. For example, if a foreign product becomes popular, it might be the

[46] Degree centrality is one of the simplest centrality measures; a node's degree is a count of how many friend connections it has, and the degree centrality for a node is just its degree. For instance, a node with 4 friends would have a degree centrality of 4. Recall that the network of friends is time-invariant (details in Appendix C), so the centrality measure is also fixed across time.

case that people in the firm's catchment area go to stores in their neighborhood to ask if they have the product in stock, which might lead to firms ordering that product in the future given the perceived local demand for it.

We also study if retailers' response depends on likelihood of a product to propagate. To do so, we classify products depending on networks' response after an exogenous import. First, we run *product-specific* regressions in our second-stage, which allows us to recover one coefficient $\beta_p$ per product, and rank products according to the magnitude of their propagation. Then, for instance, the highest $\beta_p$ correspond with products like telephone sets, sports footwear, and knitted women jackets; while the lowest $\beta_p$ correspond with blank compact discs, air conditioning machines, and seats made of wood.[47] Second, we construct two indicator variables that identify the strength of product propagation. Namely, we consider products in the top 75th and bottom 25th percentiles, such that $High_p = 1$ if $\beta_p > \beta_{75}$ and $Low_p = 1$ if $\beta_p < \beta_{25}$. Third, we study the following specification:

$$\begin{aligned} \text{Import}^F_{bpt} = & \alpha + \psi \, High_p \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-2} + \lambda \, Low_p \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-2} + \\ & \zeta \, \widehat{\text{Import}}^{US\,exposure}_{bp,t-2} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt}. \end{aligned} \tag{8}$$

The results of the analysis based on equations (7) and (8) are presented in Table 2. First, from column (1), we find that retailers do respond to an increase exposure to a product. A one standard deviation increase in the (instrumented) probability that people with relatives abroad import a product leads to a 12% higher likelihood that retailers in their neighborhood import the same product, as compared with retailers in the mean neighborhood. Thus, we can document that retailers respond to the observed local demand for foreign goods by importing those products. Appendix D.7 reports intensive margin effects, which are consistent with the baseline results. Based on these results, a 10 million USD increase in U.S. spending on a product translates into an increase of 1.98% in imports of this product by Costa Rican retailers.[48]

[47] Of course, this is an illustration meant to be brief, and product categories corresponding with HS codes are longer and more detailed descriptions. For instance, what we described as knitted women jackets stands for *women's or girls' jackets and blazers of synthetic fibers, knitted or crocheted, excluding suit jackets or blazers.*

[48] Table D.5 reports results based on an increase of one standard deviation in exposure. One standard deviation equals 69 million USD; for exposition, here we interpret results based on a 10 million USD change.

Table 2: Supply Response from Retailers

*Dependent variable: Prob. of retailers importing product $p$ in neighborhood $b$ at time $t$*

| | %Δ w.r.t. mean import probability | | | |
|---|---|---|---|---|
| | All Retailers (1) | All Retailers (2) | Small Retailers (3) | Large Retailers (4) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-2}$ | 11.863 | 10.755 | 25.599 | 9.264 |
| | $(1.679)^{***}$ | $(1.817)^{***}$ | $(3.665)^{***}$ | $(1.336)^{***}$ |
| $High_p \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-2}$ | | 8.759 | | |
| | | $(4.292)^{**}$ | | |
| $Low_p \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-2}$ | | -6.046 | | |
| | | $(1.988)^{***}$ | | |
| F-stat first stage | 49.29 | 31.04 | 48.78 | 48.08 |
| Mean dependent variable | 0.397 | 0.309 | 0.232 | 0.345 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes | Yes |
| Observations | 11,299,497 | 2,679,856 | 11,299,497 | 11,299,497 |

*Notes*: Robust standard errors, adjusted for clustering by neighborhood-product, are in parentheses. Mean import probabilities are reported. All regressions control for neighborhood×product, neighborhood×time, and product×time fixed-effects. Appendix D.5 presents details on the sample of products used in each regression. Data is quarterly and spans 2015-2019.

We now delve deeper into the mechanism behind this result. Column (2) of Table 2 is informative about this mechanism. This column shows that the effect of individual imports is entirely driven by high propagation goods, which are the ones that tend to trigger a supply response from retailers in the form of a higher likelihood of importing specific products.[49] This result suggests that firms are learning about the level of the local demand for those products, as opposed to just a product-discovery story. Appendix G.2 provides an additional test in support of firms learning about the level of local demand by leveraging the imperfect overlap between employer-employee networks and the residential location of employees. The idea behind this exercise is that employees can be exposed to foreign products in their neighborhoods and transmit information about the existence of these products to their employers, which would be relevant under a product-discovery story. However, if employees live in areas which are far away from the retailer, they should be less informative about the particular level of the local demand that their employer will face. Indeed, we document that retail firms are unresponsive to the exposure faced by their employees who live far away from the retailer's catchment area, underscoring the importance of local demand knowledge over mere product awareness.

[49] Appendix G.1 shows that this result remains statistically equal when we allow retailers to import from any origin country, as opposed to considering only imports from the U.S.

Columns (3) and (4) of Table 2 aim to further investigate the mechanism behind the result in column (1), by exploring which retailers are more likely to gain information via individuals. We find that supply responses are mainly driven by small retailers.[50] This result aligns with the idea that small retailers face higher search costs than large ones, and therefore they are more likely to take advantage of information from individuals when choosing which products to import. Further, small retailers have a more direct connection to local consumers and might be more responsive to their requests and needs when choosing which products to source, as compared with large retailers.[51]

**Total Effect** Finally, we bring together the effects we document to understand their total impact. Consider an increase of 10 million USD in U.S. spending on a product. Our results imply that this additional spending would increase *total trade in final goods* by 2.2%, as both Costa Rican individuals and retailers would import more from the U.S. via our channel.[52] This is a sizable effect, that might be useful to understand as an elasticity in dollars: $1 of spending on a product in the U.S. leads to an additional 5 cents in imports from Costa Rica, thereby creating a multiplier effect through an additional demand response.[53]

# 5 Concluding Remarks

This paper documents the role of direct and indirect externalities in the adoption of products, both among individuals and from individuals to retailers. These externalities can have large impacts, especially in developing countries or remote markets where many goods are only available as online foreign purchases and in which information frictions are pervasive. Moreover, as economies become more globalized,

[50] We define a retailer as small if its total employees is below the median at the beginning of our sample period, conditioning the sample on retailers that import at least once.

[51] For instance, if consumers frequent a retail store inquiring about a product, a small retailer whose manager is at the local shop and is closer to the final consumer might be more likely to react to these inquiries.

[52] Note the difference with the effect reported at the end of Section 3.2; there, we reported how total cross-border individual imports would increase due to our channel, while now we are reporting the effect on total trade in final goods, which includes imports by firms.

[53] To calculate this elasticity, we use our intensive margin estimates along with the median import price per product (calculated separately for individuals and for retailers). Prices are available from customs data for each shipment.

with consumers relying on online shopping more every day and with the number of product varieties available online booming, the relevance of these externalities is only expected to increase. Nevertheless, data and identification challenges have led to this topic being understudied.

We make progress by developing a simple conceptual framework and leveraging a battery of novel datasets, including information on networks of neighbors, relatives, coworkers, and friends *for each* adult in Costa Rica. We also use individual-level and firm-level data on imports, which allows us to speak directly to the adoption of specific products across time. Moreover, the paper develops a new instrument to identify the externalities diffusing across networks, which relies on our ability to connect citizens who are living in different countries to their relatives still residing in Costa Rica. This instrument uses variation from product-specific consumer trends across the U.S., and allows us to identify exogenous importing events; the idea is that, when considering two Costa Ricans with relatives living in different U.S. cities, each of them will be influenced by the product-specific trends of their relative's city, which should increase their probability of importing the goods that become more popular in this location relative to others. This new way of identifying peer effects at the product-level is one of the paper's contributions.

Relying on our instrument, we find that after someone in one's network imports a product, the probability of importing *the same product* within a quarter increases for other network members. The latter holds regardless of the network definition that we use. In fact, whether we consider neighbors, co-workers, or friends, the standardized effect of an increase in exogenous exposure is remarkably similar. Further, not all products diffuse across networks with the same intensity. We find that foreign products propagate more strongly across networks the more they belong to dynamic product categories and the more connected their initial importer is; these determinants are in line with a mechanism in which information frictions are key.

The paper also documents a previously unexplored channel by which retailers learn about the local demand for foreign products. Firms in neighborhoods with a larger exposure to foreign products become more likely to import them, and more so the stronger the propagation of the product across networks of individuals. In other words, if networks of people reflect a strong local demand for a foreign good, retailers seem to respond to this observed high demand by importing the product themselves.

This result is mainly driven by small retailers. Further, firms also seem to learn about which products *not* to import, and decrease their probability of importing a product if its observed demand among individuals is low after an exogenous import. The latter points to retailers learning about the level of local demand faced by each product, as opposed to a mere product-discovery story.

The externalities we document are large. We find a multiplier effect such that one dollar of spending in the U.S. leads to an additional five cents in imports from Costa Rica, due to the additional demand response fueled by the externalities on individuals and retailers. The presence of these externalities implies that the gains from trade might be larger than previously documented, and would lead to a multiplier effect for policies that stimulate the demand for foreign products, such as lower tariffs or less import requisites. The latter would be particularly true for developing countries and markets where catalyzers of information frictions are present. Finally, to the best of our knowledge, this is the first paper to study imports by individuals. While imports by individuals have been uncommon in the past, the rapid expansion of the direct-to-consumer market is only expected to accelerate due to factors such as increased internet penetration, improved transport and logistics infrastructure, and overall globalization.[54] Thus, we hope the paper sets as an initial contribution to what promises to be fertile ground for future research.

[54]For instance, Temu, a Chinese app which allows for cross-border direct-to-consumer purchases, was Apple's most downloaded free app in the U.S. for 2023, and low value (*de minimis*) imports represented about 15% of the value of all imports from China in 2021 according to the U.S. Customs Border Protection.

# References


Acosta, M., Cox, L., 2019. The Regressive Nature of the U.S. Tariff Code: Origins and Implications. Technical Report. Working Paper, Columbia University.

Akbarpour, M., Malladi, S., Saberi, A., 2023. Just a Few Seeds More: The Inflated Value of Network Data for Diffusion. Working Paper.

Akerman, A., Leuven, E., Mogstad, M., 2022. Information Frictions, Internet, and the Relationship between Distance and Trade. American Economic Journal: Applied Economics 14, 133–163.

Alvarez, F., Argente, D., Lippi, F., Méndez, E., Van Patten, D., 2023. Strategic Complementarities in a Dynamic Model of Technology Adoption: P2P Digital Payments.

Alvarez, F.E., Buera, F.J., Robert E. Lucas, J., 2013. Idea Flows, Economic Growth, and Trade. Working Paper 19667. National Bureau of Economic Research.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Arkolakis, C., 2010. Market Penetration Costs and the New Consumers Margin in International Trade. Journal of Political Economy 118, 1151–1199.

Atkin, D., Faber, B., Gonzalez-Navarro, M., 2018. Retail Globalization and Household Welfare: Evidence from Mexico. Journal of Political Economy 126, 1–73.

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., Stroebel, J., 2021. International trade and social connectedness. Journal of International Economics 129, 103418.

Bailey, M., Johnston, D.M., Kuchler, T., Stroebel, J., Wong, A., 2022. Peer Effects in Product Adoption. American Economic Journal: Applied Economics .

Bandiera, O., Barankay, I., Rasul, I., 2009. Social Connections and Incentives in the Workplace: Evidence From Personnel Data. Econometrica 77, 1047–1094.

Bandiera, O., Rasul, I., 2006. Social Networks and Technology Adoption in Northern Mozambique. Economic Journal 116, 869–902.

Banerjee, A., Chandrasekhar, A.G., Duflo, E., Jackson, M.O., 2013. The Diffusion of Microfinance. Science 341, 1236498.

Bertrand, M., Luttmer, E.F., Mullainathan, S., 2000. Network Effects and Welfare Cultures. The Quarterly Journal of Economics 115, 1019–1055.

Borusyak, K., Jaravel, X., 2021. The Distributional Effects of Trade: Theory and Evidence from the United States. Technical Report. National Bureau of Economic Research.

Brock, W.A., Durlauf, S.N., 2001. Discrete Choice with Social Interactions. The Review of Economic Studies 68, 235–260.

Buera, F.J., Oberfield, E., 2020. The Global Diffusion of Ideas. Econometrica 88, 83–114.

Caplin, A., Leahy, J., 1998. Miracle on Sixth Avenue: Information Externalities and Search. The Economic Journal 108, 60–74.

De Giorgi, G., Frederiksen, A., Pistaferri, L., 2019. Consumption Network Effects. The Review of Economic Studies 87, 130–163. doi:10.1093/restud/rdz026.

De Giorgi, G., Pellizzari, M., Redaelli, S., 2010. Identification of Social Interactions through Partially Overlapping Peer Groups. American Economic Journal: Applied Economics 2, 241–75.

De Paula, A., 2017. Econometrics of Network Models, in: Advances in Economics and Econometrics: Eleventh World Congress, Cambridge University Press Cambridge. pp. 268–323.

Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. World Bank Publications.

Dell, M., Feigenberg, B., Teshima, K., 2019. The Violent Consequences of Trade-Induced Worker Displacement in Mexico. American Economic Review: Insights 1, 43–58.

Duesenberry, J.S., 1948. The consumption function: A study of relations between income and consumer expenditures. University of Michigan.

Duflo, E., Saez, E., 2003. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. The Quarterly Journal of Economics 118, 815–842.

EBANX, B.B..., 2023. Digital payments connecting businesses and people in rising economies: an overview of online commerce in Latin America and Africa, 2023. Online report. Accessed: 2023-09-13.

Faber, B., 2014. Trade Liberalization, the Price of Quality, and Inequality: Evidence from Mexican Store Prices. Working Paper.

Fajgelbaum, P.D., Khandelwal, A.K., 2016. Measuring the Unequal Gains from Trade. The Quarterly Journal of Economics 131, 1113–1180.

Furman, J., Russ, K., Shambaugh, J., 2017. US tariffs are an arbitrary and regressive tax. VoxEU blog .

Hottman, C.J., Monarch, R., 2020. A matter of taste: Estimating import price inflation across U.S. income groups. Journal of International Economics 127, 103382.

ILO, 2002. Women and men in the informal economy: A statistical picture.

Jordà, Ò., Schularick, M., Taylor, A.M., 2020. The effects of quasi-random monetary experiments. Journal of Monetary Economics 112, 22–40.

Jordà, , 2005. Estimation and Inference of Impulse Responses by Local Projections. American Economic Review 95, 161–182.

Manski, C.F., 1993. Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies 60, 531–542.

Mas, A., Moretti, E., 2009. Peers at Work. American Economic Review 99, 112–45.

Méndez, E., Van Patten, D., 2022. Voting on a Trade Agreement: Firm Networks and Attitudes Toward Openness. Technical Report. National Bureau of Economic Research.

Moffitt, R.A., 2000. Welfare Benefits and Female Headship in U.S. Time Series. American Economic Review 90, 373–377.

Montiel Olea, J.L., Plagborg-Møller, M., 2021. Local Projection Inference Is Simpler and More Robust Than You Think. Econometrica 89, 1789–1823.

Oberlo, 2023. How Many People Shop Online? `https://www.oberlo.com/statistics/how-many-people-shop-online`. Accessed: 2023-09-23.

Perla, J., Tonetti, C., 2014. Equilibrium Imitation and Growth. Journal of Political Economy 122, 52–76.

Perla, J., Tonetti, C., Waugh, M.E., 2021. Equilibrium Technology Diffusion, Trade, and Growth. American Economic Review 111, 73–128.

Porto, G.G., 2008. Agro-Manufactured Export Prices, Wages and Unemployment. American Journal of Agricultural Economics 90, 748–764.

Sampson, T., 2016. Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth. The Quarterly Journal of Economics 131, 315–380.

Van Patten, D., 2021. International Diffusion of Technology: Accounting for Heterogeneous Learning Abilities. Technical Report.

Veblen, T., 1899. The Preconceptions of Economic Science. The Quarterly Journal of Economics 13, 121–150.

# Online Appendix for

## Cross-Border Product Adoption: Individual Imports, Migrant Networks, and Domestic Retailers

March 8th, 2024


David Argente
*Yale University and NBER*

Esteban Méndez
*Central Bank of Costa* Rica

Diana Van Patten
*Yale University and NBER*

# Table of Contents


**Appendix A. Conceptual Framework** **1**

A.1 Solution . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

A.2 Proofs . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

**Appendix B. Setup: Additional Results** **9**

B.1 Supplementary Figures . . . . . . . . . . . . . . . . . . . . 9

B.2 Relationship Between CEX and Imports Data . . . . . . . . 9

B.3 Relationship Between CEX and Debit Card Data . . . . . . 10

B.4 Network Descriptive Statistics . . . . . . . . . . . . . . . . 12

B.5 Balance Tests for Migrants . . . . . . . . . . . . . . . . . . 13

**Appendix C. Details on Networks of Friends** **13**

**Appendix D. Main Analysis: Additional Results** **15**

D.1 Note on Clustering . . . . . . . . . . . . . . . . . . . . . . 15

D.2 Reduced Form and OLS Results . . . . . . . . . . . . . . . 16

D.3 First-Stage Regressions . . . . . . . . . . . . . . . . . . . . 18

D.4 Second-Stage Regressions . . . . . . . . . . . . . . . . . . . 19

D.5 Random Sample vs. Entire Sample of Products . . . . . . . 21

D.6 Timing of the Specifications: Local Projections . . . . . . . 22

D.7 Intensive Margin Results . . . . . . . . . . . . . . . . . . . 23

**Appendix E. Robustness Exercises** **24**

E.1 Instrument Using Distance-3 Nodes . . . . . . . . . . . . . 24

E.2 Placebos: Additional Network Definitions . . . . . . . . . . 25

E.3 Results at the Individual-Level . . . . . . . . . . . . . . . . 25

E.4 Results with Additional Fixed-Effects . . . . . . . . . . . . 26

**Appendix F. Determinants of Product Propagation** **26**

F.1 Asymmetric Response of Positive and Negative Shocks . . . 26

F.2 Dynamic vs. Established Products . . . . . . . . . . . . . . 27

F.3 Centrality of the Importer . . . . . . . . . . . . . . . . . . 27

**Appendix G. Retailers' Response: Additional Results** **28**

G.1 Results Based on Imports from Any Country . . . . . . . . . 28

G.2 Mechanism Based on Employer-Employee Data . . . . . . . 28

# A Conceptual Framework

In what follows, we describe a simple framework to think about demand externalities in the adoption of foreign products, both from an individual's perspective and from the point of view of the firm.

**Setup** $N$ consumers in a network (e.g., neighborhood) want to buy a variety in a product category; they can either buy a domestic variety with a known payoff $D > 0$ or buy a foreign variety abroad with an uncertain (potentially greater) payoff. In a given period, each consumer decides whether to buy domestically and collect $D$ or to search (online) for a product variety abroad. Each consumer who searches finds a variety abroad and decides whether to import it or not. Imported varieties cannot be returned; consumers would rather wait than purchase the wrong variety.[55]

We assume that consumers are risk neutral and maximize utility discounting future periods by $\rho \in (0, 1)$. The utility generated by an imported variety depends on $x\theta$, where $x$ is consumer-specific and $\theta$ is common to all consumers. Consumers are ex-ante identical and have identical priors concerning the distributions of $x$ and $\theta$. We assume that $x$ is drawn independently from a uniform distribution on $[0, 1]$ and is revealed to consumers when they find a variety (before they purchase it). All consumers know that $\theta$ is distributed uniformly on $[0, 2]$, but information on $\theta$ is revealed only after a consumer has imported the product, when both $x$ and $\theta$ become public, and before the next period begins. This is, the first buyers do not observe $\theta$, so their expected utility, if they decide to import a product, is $x\mathbb{E}(\theta) = x$. We assume that, after a variety is imported for the first time, $\theta$ becomes public and subsequent buyers get $x\theta$. Thus, consumers want to learn about $\theta$ from others.

There is a single firm in the network, which sells to all consumers who buy the variety domestically. The firm also learns $\theta$ from the initial importers. If consumers decide to search for a variety abroad, the firm can choose to pay a fixed cost to make a once-and-for all decision to become an importer and sell the product variety domestically. Our information assumptions have two phases: uninformed and informed.

**Uninformed Phase** The uninformed phase takes place before any consumer has imported the foreign variety (i.e., before there is public information on $\theta$).[56] In this

[55] This assumption simplifies the analysis and it is reasonable in the case of Costa Rica, where the costs of returning an item are often too high; anecdotally, consumers often absorb the cost of internationally shipping back the item, the cost of processing the return, and suffer the delays of international shipping. In fact, in the data, *only 0.01% of individual imports are returned.*

[56] We assume that $D$ is small enough so that consumers decide to search for imported products when $\theta$ is unknown. Thus, there are no domestic sales in the uninformed phase. In Section A.1, we show the upper bound of $D$ that is a sufficient condition for search.

phase, consumers search online and simply decide whether to import or not based on their draw of $x$. Intuitively, if $x$ is sufficiently large, consumers decide to import the product variety. Formally, the uninformed agent maximizes

$$V_U(x) = \max\{x, \rho\,[p\,\mathbb{E}V_U + (1-p)\,\mathbb{E}V_I]\}, \tag{9}$$

where $p$ is the endogenous probability that an agent remains uninformed (i.e., all other agents do not import) and $\mathbb{E}V_U$ and $\mathbb{E}V_I$ denote the expected value of being uninformed and informed, respectively. Thus, the consumer imports if $x \geq \widehat{x}$, where $\widehat{x} = \rho\,[p\mathbb{E}V_U + (1-p)\mathbb{E}V_I]$.

**Informed Phase** The informed phase begins after the first cohort has imported the product variety and $\theta$ becomes public. Consumers make two decisions. First, given $\theta$, they decide whether to buy domestically (collect $D$) or continue searching online. Second, if they search, they must decide whether to import or not. Thus, their strategies determine a set of values $\theta \geq \bar{\theta}$ that warrant continued search, and a set of qualities $x \geq \bar{x}$ that determine whether the consumers import the variety or not. The value of an optimal strategy for an informed agent who decides to search is

$$V_I(x,\theta) = \max\{x\theta, \rho\mathbb{E}[V_I(x',\theta)]\}, \tag{10}$$

where the first term is the value of importing the product variety and the second term is the discounted value of keep searching. The cutoff $\bar{\theta}$ is pinned down where the value of searching equals the payoff of buying the variety domestically (i.e., $\mathbb{E}[V_I(x,\bar{\theta})] = D$). In turn, the cutoff $\bar{x}(\theta)$ is the value of $x$ that makes consumers indifferent between importing or keep searching (i.e., $\bar{x}(\theta) = \frac{1}{\theta}\rho\mathbb{E}[V_I(x',\theta)]$).

Also in this phase, the firm learns $\theta$. If consumers buy domestically (i.e., $\theta < \bar{\theta}$), the firm sells them the domestic variety. If consumers search (i.e., $\theta \geq \bar{\theta}$), the firm can decide to pay a fixed cost $c$ and import the good. The firm's optimal strategy is

$$W_I(\theta) = \max\{0, -c + \rho\mathbb{E}[V_I(x',\theta)]\widetilde{N}\}, \tag{11}$$

where, without loss of generality, we normalize firm's profits to zero when it chooses not to import. If the firm decides to import the product, consumers' outside option is to buy abroad. Thus, under Bertrand competition, the firm sets the highest possible price that still prevents consumers from buying abroad. As a result, the firm sells to all other consumers who have not yet bought the good abroad $\widetilde{N} \equiv N\bar{x}(\theta)\widehat{x}$ and obtains all surplus. The firm imports if $\theta \geq \widetilde{\theta}$; the cutoff is pinned down at the point where the cost and the expected gains of importing are equal, $c = \rho\mathbb{E}[V_I(x,\widetilde{\theta})]N\bar{x}(\theta)\widehat{x}$.

**Equilibrium** We focus on equilibria in which the decision rules depend only on information that is payoff-relevant. Further, since all consumers are ex-ante identical, we look for a symmetric Nash Equilibrium. The stationary equilibrium involves a set of cutoff rules, summarized below.

DEFINITION 1. An equilibrium consists of cutoffs $\bar{\theta} \in [0,2]$, $\widetilde{\theta} \in [0,2]$, $\widehat{x} \in [0,1]$, and a function $\bar{x}(\bar{\theta}) : [\bar{\theta}, 1] \to [0,1]$ such that the following strategy is optimal: (a) in the uninformed phase, only varieties with $x \geq \widehat{x}$ are imported, (b) in the informed phase, consumers search happens only if $\theta \geq \bar{\theta}$ and consumers import varieties with $x \geq \bar{x}$, and (c) in the informed phase, the firm imports the good to sell domestically if $\theta \geq \widetilde{\theta}$.

**Properties of the Equilibrium** Proposition 4 in Appendix A.1.3 establishes both existence and uniqueness of an equilibrium of the form given by Definition 1. The equilibrium has intuitive properties, which we describe in a set of propositions, each followed by its intuition. The proofs of all propositions can be found in Appendix A.2.

PROPOSITION 1. The equilibrium level of $\widehat{x}$ is increasing in $D$.

In the uninformed phase, consumers are less willing to search online (or import) a variety if the value of the domestic option, $D$, is high. This result implies that a consumer would never choose to search for a product that is already available domestically with high enough quality or appeal.

PROPOSITION 2. If $p > 0$, then $\widehat{x} > \bar{x}$ and $\mathbb{E}V_U < \mathbb{E}V_I$.

Consumers demand a higher $x$ to import in the uninformed stage (i.e., $\widehat{x} > \bar{x}$), and the possibility of waiting for another consumer to import the good gives rise to a free rider problem. This result derives from an demand externality *that the model generates endogenously*: consumers do not internalize that importing a variety provides valuable information to other consumers in their network. Thus, the equilibrium is inefficient and there is a delay in the adoption of imported varieties, since the expected payoff in the informed phase is greater than that in the uninformed phase.

PROPOSITION 3. $\bar{\theta} \in (0,1)$ is increasing in $D$ and $\widetilde{\theta} \in [\bar{\theta}, 2)$ is decreasing in $N$.

Lastly, in the informed phase, only varieties with high enough $\theta$ relative to the domestic option are imported, either by consumers or the firm. Note $\bar{\theta}$ is strictly less than $\mathbb{E}(\theta) = 1$, reflecting the value of information for consumers and for the firm. Further, the firm is more likely to import a variety and sell it domestically if its market size is large. Figure A.1 summarizes the model's solution for different values of $x$ and $\theta$.

**From Model to Data** Our empirical analysis is guided by this framework. In the uninformed phase, individuals decide whether to import independently from others; there are no common shocks or shared characteristics among peers. As the latter is unlikely to hold empirically, Section 3.1 proposes a strategy to leverage plausibly exogenous demand shocks to the likelihood of importing. Through the lens of the model, these shocks can be understood as shifters of $x$ and can lead to importing varieties without good domestic alternatives (Proposition 1). In the informed phase, the model features direct externalities: once a person imports a variety, peers in her network may become more likely to import it as well (Proposition 2). Whether these externalities exist is an empirical question, which Section 3.2 explores for different networks. Finally, there is also an indirect externality: firms respond to the revealed information by importing, but only when the expected gains are sufficiently large (Proposition 3). In fact, firms only import varieties with strong enough propagation among consumers after they are imported. Section 4 analyzes retailers' responses after individuals import, differentiating between goods with strong and weak propagation.

## A.1 Solution

The solution of the model follows closely Caplin and Leahy (1998). We solve the model backwards beginning with the informed phase. Given that the actual value of $\theta$ is known in this phase, consumers can compute the reservation value $\bar{x}$ by comparing the value of searching online and the value of importing the product variety. Similarly, consumers can compute $\bar{\theta}$ by comparing the value of buying domestically and the value of searching. Finally, knowing $\bar{x}$ and $\bar{\theta}$, the decision of whether to import the product variety or keep searching online in the uninformed phase pins down $\widehat{x}$. For the firm, the solution is simpler. Since there are no domestic sales in the uninformed phase, in the informed phase the firm decides whether to import the variety or not given demand and $\theta$.

### A.1.1 Informed Phase

Recall that in this phase consumers' maximize

$$V_I(x, \theta) = \max\{x\theta, \rho\mathbb{E}[V_I(x', \theta)]\}.$$

where it follows that the reservation level of $x$ is pinned down by:

$$\begin{aligned}\bar{x}(\theta) &= \frac{1}{\theta}\rho\mathbb{E}[V_I(x', \theta)] \\ &= \rho\left(\frac{1+\bar{x}(\theta)^2}{2}\right) \qquad (12)\end{aligned}$$

which shows that $\bar{x}(\theta) \equiv \bar{x}$.[57] Solving equation (12) focusing on solutions in the domain of $\bar{x}$, we find $\bar{x} = \frac{1-\sqrt{1-\rho^2}}{\rho}$.

[57]The second equality in equation (12) follows from $\mathbb{E}[\max\{x, x'|x' = \bar{x}\}] = \frac{1+\bar{x}^2}{2}$ when $x$ and $x'$ are two independent random variables uniformly distributed on $[0, 1]$.

Similarly, $\bar{\theta}$ is pinned down by:

$$\begin{aligned} D &= \mathbb{E}[V_I(x,\bar{\theta})] \\ &= \bar{\theta}\left(\frac{1+\bar{x}^2}{2}\right). \end{aligned} \tag{13}$$

Note that $\bar{\theta} < 1$ since this requires $D < \frac{1+\bar{x}^2}{2}$ and we know the upper bound $D < \frac{1+\widehat{x}^2}{2}$ from our assumption that initial search is more valuable than a domestic purchase in the uninformed phase. Letting $d \equiv \frac{D}{1+\bar{x}^2}$, we can write $\bar{\theta} = 2d$ where $d \in (0, \frac{1}{2})$.

Lastly, $\widetilde{\theta}$ is pinned down by:

$$\begin{aligned} c &= \rho\mathbb{E}[V_I(x,\theta)]\widetilde{N}\} \\ &= \rho\widetilde{\theta}\left(\frac{1+\bar{x}^2}{2}\right)\widetilde{N}, \end{aligned} \tag{14}$$

where $\widetilde{N} \equiv N\bar{x}(\theta)\widehat{x}$. Using the definition of $d$, we can write $\widetilde{\theta} = \frac{2cd}{\rho\widetilde{N}D}$.

### A.1.2 Uninformed Phase

In this phase, an uninformed consumer maximizes

$$V_U(x) = \max\{x, \rho\,[p\mathbb{E}V_U + (1-p)\mathbb{E}V_I]\},$$

where $\widehat{x}$ satisfies the indifference condition between importing or searching and $p = \widehat{x}^{N-1}$ is the probability that an agent remains uninformed. Note that $\mathbb{E}V_U = \frac{1+\widehat{x}^2}{2}$ so that assuming $D < \frac{1+\widehat{x}^2}{2}$ is sufficient so that consumers decide to search for imported varieties when $\theta$ is unknown. The expected value of being informed in this phase is:

$$\begin{aligned} \mathbb{E}V_I &= \int\left[\max\left\{D, \int V_I(x,\theta)dx\right\}\right]d\theta \\ &= (1+d^2)\left(\frac{1+\bar{x}^2}{2}\right) \\ &= (1+d^2)\frac{\bar{x}}{\rho}, \end{aligned} \tag{15}$$

where we use equation (12) in the last line to simplify equation (15). The reservation acceptance level $\widehat{x}$ can be found using the indifference condition

$$\begin{aligned} \widehat{x} &= \rho\,[p\mathbb{E}V_U + (1-p)\mathbb{E}V_I] \\ &= \rho\left[\widehat{x}^{N-1}\frac{1+\widehat{x}^2}{2} + (1-\widehat{x}^{N-1})(1+d^2)\left(\frac{1+\bar{x}^2}{2}\right)\right] \end{aligned}$$

and, using equation (12) to eliminate $\rho$, we find

$$\widehat{x} = \widehat{x}^{N-1}\bar{x}\left(\frac{1+\widehat{x}^2}{1+\bar{x}^2}\right) + (1-\widehat{x}^{N-1})\bar{x}(1+d^2). \tag{16}$$

The value of $\widehat{x}(N)$ as $N$ increase is relevant since it determines the severity of the free rider problem. In particular, the limit of $\widehat{x}(N)$ as $N$ increases is

$$\lim_{N\to\infty} \widehat{x}(N) = \begin{cases} \rho EV_I & \text{if } \rho EV_I \leq 1 \\ 1 & \text{if } \rho EV_I > 1 \end{cases} \tag{17}$$

where the value of $\rho EV_I$ is given in equation (15). Intuitively, if the expected value of being informed is very large (i.e., $\rho EV_I > 1$), the free rider problem becomes very serious and the wait for the first import of a product variety can become arbitrarily long (i.e., $\widehat{x}(N) = 1$).[58]

### A.1.3 Equilibrium

Proposition 4. Let $\widetilde{N} \equiv N\bar{x}\widehat{x}$ and $d \equiv \frac{D}{1+\bar{x}^2}$. Then for $N > 1$, $\widetilde{N} > 1$, $\rho \in (0,1)$, $D \in (0, \frac{D}{2d})$, $d \in (0, \frac{1}{2})$ and $c \in [\frac{\rho\bar{\theta}\widetilde{N}D}{2d}, \frac{\rho\widetilde{N}D}{d}]$, there exists a unique equilibrium of the form given in Definition 1 with $\bar{x} = \frac{1-\sqrt{1-\rho^2}}{\rho}$, $\bar{\theta} = 2d$, $\widetilde{\theta} = \frac{2cd}{\rho\widetilde{N}D}$, and $\widehat{x} \in (\bar{x}, \min[1, \bar{x}(1+d^2)])$ which is uniquely determined.

**Proof.** (of Proposition 4) To establish existence and uniqueness, we need to show that equation (16) provides a unique solution for $\widehat{x}$. Our assumption that initial search is more valuable than a domestic purchase in the uninformed phase implies that $\frac{1+\widehat{x}^2}{2} > D = d(1+\bar{x}^2)$. Since $d \in (0, \frac{1}{2})$, then $\widehat{x} > \bar{x}$. Thus, we need to show that equation (16) has a unique solution $\widehat{x} \in (\bar{x}, 1)$. We begin rewriting equation (16) as

$$H(\widehat{x}) = (1+X_1^2)\widehat{x} + (X_0 - X_1)\widehat{x}^{N-1} - X_0 - X_1\widehat{x}^{N+1} \tag{18}$$

where $X_1 = \bar{x}$, $X_0 = (1+d^2)\bar{x}(1+\bar{x}^2)$ and $X_0 > X_1$. Note that $H(\widehat{x}) < 0$ if $\widehat{x} \to \bar{x}$ and $H(\widehat{x}) > 0$ if $\widehat{x} \to 1$. Thus, there exists a solution $\widehat{x} \in (\bar{x}, 1)$. Since $H(\widehat{x})$ starts below zero and ends above zero, we can show uniqueness by ruling out multiple zeros. This can be done by showing that the function is locally concave at any critical point. To do so, we first find the critical points

$$H'(\widehat{x}^*) = (1+X_1^2) + (N-1)(X_0 - X_1)(\widehat{x}^*)^{N-2} - (N+1)X_1(\widehat{x}^*)^N = 0$$

[58] If $\rho EV_I \leq 1$ then $\lim_{N\to\infty} \widehat{x}(N)^N = 0$. If $\rho EV_I > 1$ then $\lim_{N\to\infty} \widehat{x}(N)^N = \frac{[\rho EV_I - 1]}{\rho[EV_I - 1]} \in (0,1)$.

and then we show that at any $\widehat{x}^* \in [0,1]$, $H''(\widehat{x}^*) < 0$

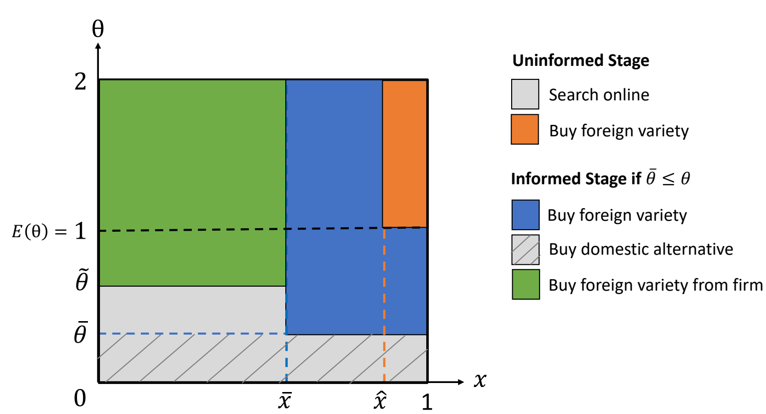
$$\begin{aligned} H''(\widehat{x}^*) &= (N-2)(N-1)(X_0 - X_1)(\widehat{x}^*)^{N-3} - N(N+1)X_1(\widehat{x}^*)^{N-1} \\ &= \frac{1}{\widehat{x}^*}(N-2)(N-1)(X_0 - X_1)(\widehat{x}^*)^{N-2} - N(N+1)X_1(\widehat{x}^*)^{N} \\ &< \frac{N}{\widehat{x}^*}H'(\widehat{x}^*) = 0 \end{aligned}$$

Thus, $H(\widehat{x})$ only has one critical point. □

### A.1.4 State Space

Figure A.1 summarizes the model's solution for different values of $x$ and $\theta$. The orange rectangle is the area of the state space for which it is optimal to import a variety in the uninformed phase. The size of the blue rectangle relative to the orange indicates that, in the informed phase, there is a wider range of values for which importing is optimal. This difference explains the initial delay in the adoption of a foreign variety and its subsequent adoption by the network after someone imports it. Finally, the green rectangle shows the values for which it is optimal for the firm to import the variety and sell it domestically. The importing decision for the firm depends on the common quality or appeal of the foreign variety and on the amount of consumers in the network willing to buy it, but who have not already imported it on their own.

Figure A.1: Model Solution and Properties of the Equilibrium




*Notes:* The figure shows the state space for $\theta$ and $x$, along with the equilibrium thresholds.

## A.2 Proofs

**Proof.** (of Proposition 1) Let $D' > D$. Using equation (18), it can be shown that $H(\widehat{x}; D') < H(\widehat{x}; D)$. From Proposition 4 we know that $H(\widehat{x})$ is increasing in $\widehat{x}$ so an increase in $D$ requires an increase in $\widehat{x}$ to restore equilibrium. Similarly, if $N' > N$ then $H(\widehat{x}; N') < H(\widehat{x}; N)$ if $\widehat{x} \leq (1 + d^2)\bar{x}$. This condition can verified using equation (18) and evaluating it at $\widehat{x} = (1 + d^2)\bar{x}$; in this case, $H(\widehat{x}) > 0$ which implies $\widehat{x} \leq (1 + d^2)\bar{x}$. $\square$

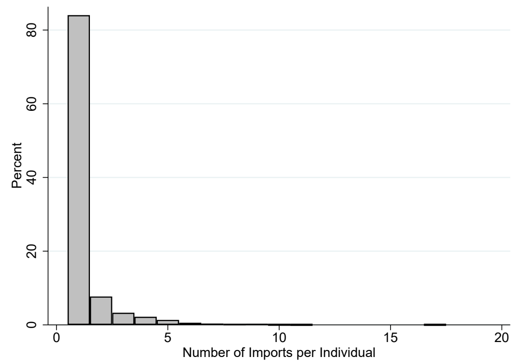
**Proof.** (of Proposition 2) Our assumption that initial search is more valuable than a domestic purchase in the uninformed phase implies that $\frac{1+\widehat{x}^2}{2} > D = d(1 + \bar{x}^2)$. Thus, $\widehat{x} > \bar{x}$ follows from $d \in (0, \frac{1}{2})$. Moreover, using equation (18), we can verify that $H(\widehat{x}) > 0$ at $\widehat{x} = (1 + d^2)\bar{x}$, so that $\widehat{x} \leq (1 + d^2)\bar{x} = \rho \mathbb{E}V_I$. For $p > 0$ and $\widehat{x} > \bar{x}$, $\widehat{x} < \rho \mathbb{E}V_I$ since $\widehat{x} = \rho\,[p\mathbb{E}V_U + (1 - p)\mathbb{E}V_I]$. Thus, $\mathbb{E}V_U < \mathbb{E}V_I$ $\square$

**Proof.** (of Proposition 3) From equation (13) we know that $\bar{\theta} = \frac{2D}{1+\bar{x}^2}$. Thus, $\bar{\theta} < 1$ since this requires $D < \frac{1+\bar{x}^2}{2}$ and we know $D < \frac{1+\widehat{x}^2}{2}$ from our assumption that initial search is more valuable than a domestic purchase in the uninformed phase. This implies that $d \in (0, \frac{1}{2})$ since $\bar{\theta} = 2d$ and using equation (13) it is straightforward to show that $\frac{\partial \bar{\theta}}{\partial D} > 0$. Moreover, note that $\widetilde{\theta}$ is bounded from below by $\bar{\theta}$, since for values of $\theta$ below $\bar{\theta}$ consumers prefer to purchase products available domestically. Using equation (14) it is easy to verify that $\frac{\partial \widetilde{\theta}}{\partial N} < 0$. $\square$

# B Setup: Additional Results

## B.1 Supplementary Figures

Figure B.1: Frequency of Imports Per Product by Individual



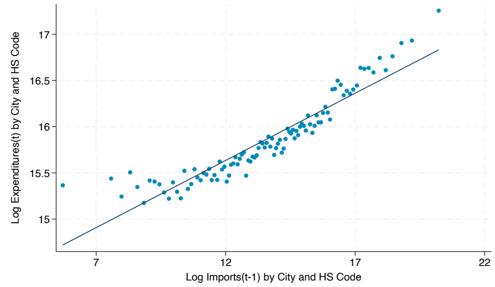

*Notes*: The figure shows a histogram describing the frequency of events by individual importers during the period 2015-2019. An event is defined as an import of a particular HS-code. The sample considers only people who have a relative in the U.S., i.e., the sample which conforms our first-stage regressions.

## B.2 Relationship Between CEX and Imports Data

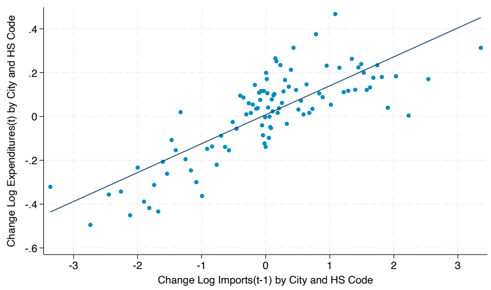
In the U.S., many tradable products are imported. Thus, expenditure shares for these products in the CEX by region should co-move with the imports of these products in these areas. Following this idea, we use data on imports by customs districts in the U.S., adjusted using FAF data from the Department of Transportation as explained in Section 2, to assess the representativeness of the CEX at narrowly-defined categories and geographic areas. This notion follows Acosta and Cox (2019), who show that these customs districts data closely matches aggregate patterns in the CEX. Figure B.2 shows a strong correlation between expenditures in the CEX and expenditures based on customs districts data, when defining products as 4- or 6-digit HS codes, regions as Primary Sampling Units (PSUs), and time as quarters between 2015 and 2019. The correlation is strong both in levels as shown by Panels (A)-(B) and in changes as shown in Panels (C)-(D).

Figure B.2: Expenditure Shares in the CEX vs. Customs Districts

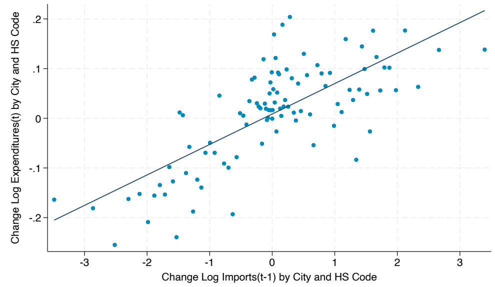

(A) Levels: HS-4




(B) Levels: HS-6




(C) Changes: HS-4




(D) Changes: HS-6

*Notes*: The figures shows the relation between expenditures in the CEX (vertical axis) and expenditure based on customs districts data (horizontal axis), when defining products as HS-4 or HS-6 product codes, regions as PSUs, and time as quarters for the period 2015-2019. Panels (A) and (B) show the correlation in levels for products as HS-4 and HS-6 codes, respectively. Panels (C) and (D) show the correlation for the same definition of products in changes, we trim the top and bottom one percent.

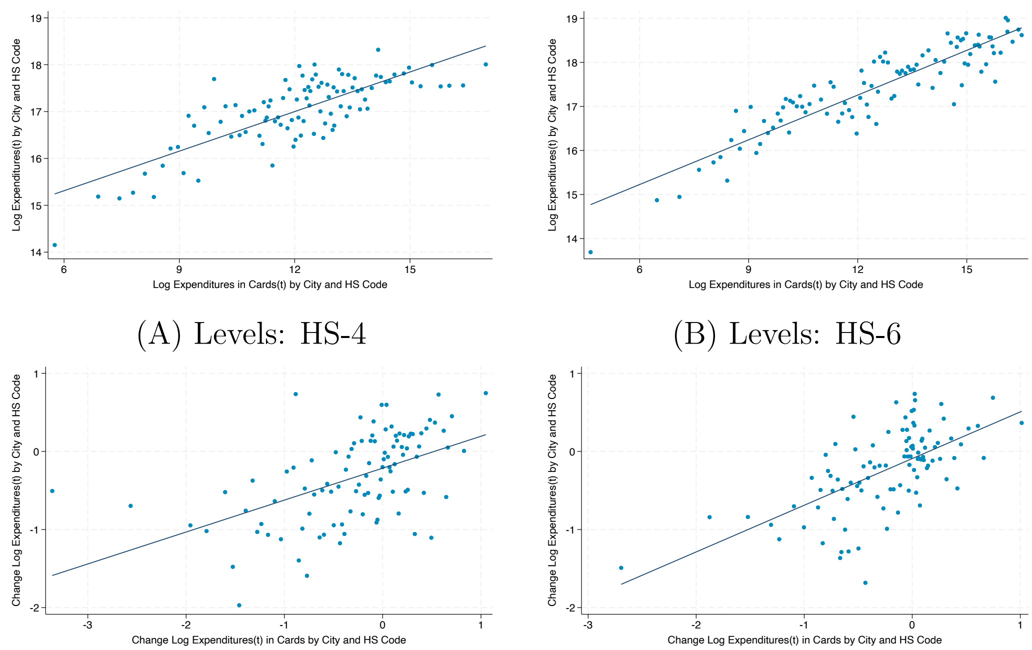
## B.3 Relationship Between CEX and Debit Card Data

Similarly, it is possible to use data on debit card transactions by region and by type to validate the CEX. This data comes from Facteus, a provider of financial data for business analytics. The data set contains information on the total expenditures, total number of transactions, and total number of cards, at the zip-code level and with daily frequency. Approximately 10 million debit cards are included. The data set begins in 2017 and ends in the first week of July 2020. The debit cards in the Facteus panel are issued by "challenger banks." The data set includes information of more than 200 Merchant Category Codes (MCCs), which correspond to the MCC standard as maintained by Visa and Mastercard. Every transaction processed by the card networks is assigned an MCC, which is a four-digit number that denotes the type of business providing a service or selling merchandise. We manually create a bridge between MCC and Standard Industrial Classification (SIC) codes.[59] MCCs were

[59]This bridge was created in parallel by two independent teams of RAs, then cross-checked, and finally revised by a coauthor.

derived from SIC codes, however, MCCs and SIC codes do not always correspond; in some cases, several SIC codes are consolidated into one MCC, while in other cases, such as for "T&E and direct marketing merchants," MCCs do not have corresponding SIC code. Figure B.3 shows a strong correlation between expenditures in the CEX and expenditures based on card transactions data, when defining products as HS-4 or HS-6 product codes, regions as PSUs, and time as years between 2017 and 2019. As in the case of customs data, the correlation is strong both in levels and in changes.

Figure B.3: Expenditure Shares in the CEX vs. Card Transactions




(A) Levels: HS-4

(B) Levels: HS-6

(C) Changes: HS-4

(D) Changes: HS-6

*Notes*: The figure shows the relation between expenditures in the CEX (vertical axis) and expenditures based on card transactions data (horizontal axis), when defining products as HS-4 or HS-6 product codes, regions as PSUs, and time as years for the period 2017-2019. Panels (A) and (B) show the correlation in levels for products as HS-4 and HS-6 codes, respectively. Panels (C) and (D) show the correlation for the same definition of products in changes, we trim the top and bottom one percent..

## B.4 Network Descriptive Statistics

Table B.1: Network-Level and Products Summary Statistics

*Panel (a): Network Summary Statistics*

| Network type | Total networks (1) | Median people per network (2) |
|---|---|---|
| Neighbors | 1,748 | 839 |
| Co-workers | 23,386 | 22 |
| Friends | 49,178 | 3 |

*Panel (b): Products Summary Statistics*

| | Total # products imported by individuals | | |
|---|---|---|---|
| | | Top codes: (by import volume) | Women blouses, knitted/crocheted<br>Vehicle toys, incorporating motor, battery, not metal<br>Women trousers made of cotton<br>Beauty or make-up for skin care (no medicaments)<br>Construction sets, puzzles<br>Toys representing human beings |
| CEX | 430 | | |
| U.S. imports | 2,443 | Top codes: | Women trousers of cotton; women trousers of artificial fibers<br>Tricycles, scooters, pedal cars, and similar wheeled toys<br>Dolls' carriages; dolls; puzzles of all kinds<br>Women jackets and blazers of wool or fine animal hair<br>Printed books, brochures, leaflets, and similar in single sheets<br>Sports footwear, with rubber outer soles designed for basketball |

*Notes*: Panel (a) shows the total number of distinct networks per network type, along with the median number of people who compose each network. These are only networks which have at least one person with a relative abroad. See Appendix C for details on the network of friends. Panel (b) displays the total number of products which are imported more than once by individuals in our data, according to both the CEX and the U.S. imports data by customs districts. Panel (b) also displays (in gray) the top product codes by import volume.

Table B.2: Variation for Exposure Measures
(% of products with underlying variation at each HS-code level)

| *Panel (a): CEX* | | | | *Panel (b): U.S. imports* | | | |
|---|---|---|---|---|---|---|---|
| HS-4 | HS-6 | HS-8 | HS-10 | HS-4 | HS-6 | HS-8 | HS-10 |
| **91.01** | 6.35 | 2.12 | 0.53 | 0 | **59.95** | 21.63 | 18.42 |

*Notes*: The table shows the percentage of products in our sample whose underlying variation is at each HS-code level, which tells us the level at which the exposure measures for our IV strategy are varying, depending on the source from which we obtain expenditures on each product code by region and time in the U.S. Panel (a) shows that most of the variation is at the HS-4 level when using the CEX. Panel (b) shows most of the variation is at the HS-6 level when relying on U.S. imports data by customs districts.

## B.5 Balance Test for Migrants to Different U.S. Consulates

Our instrument exploits variation in consumer trends for specific products across the U.S., and links it to people in Costa Rica based on relatives across different U.S. consulates. While we remove the levels from the relevant variation that we use to construct our instrument in equation (1), we want to verify that the observable characteristics of Costa Rican migrants to different consulates across the U.S. balance. To do so, we calculate normalized differences for different characteristics following Imbens and Wooldridge (2009), namely, for individuals in consulate $c$, one would calculate the following for observable characteristic $X$:

$$\frac{\bar{X}_c - \bar{\mu}_{-i}}{\sqrt{S_i^2 + S_{-i}^2}},$$

where $\bar{X}_c$ ($S_i$) is the mean value (standard deviation) of $X$ for people migrating to $c$ and $\bar{\mu}_{-i}$ ($S_{\mu,-i}$) is the mean value (standard deviation) of $X$ for people migrating to a consulate other than $c$. The rule of thumb is that an absolute value of the normalized difference exceeding 0.25 indicates strong imbalances.

Table B.3: Characteristics of Migrants and Normalized Differences

| Main consulate | Total | Age (years) | | Female (=1) | | Wages | |
|---|---|---|---|---|---|---|---|
| in the U.S. | N | Mean | Norm. diff. | Mean | Norm. diff. | Mean | Norm. diff. |
| Atlanta | 2,605 | 39.14 | -0.07 | 0.47 | 0.05 | 456 | -0.01 |
| Houston | 1,771 | 40.11 | -0.00 | 0.48 | 0.06 | 566 | 0.11 |
| Los Angeles | 3,080 | 42.11 | 0.12 | 0.53 | 0.14 | 575 | 0.12 |
| Miami | 3,458 | 41.59 | 0.09 | 0.51 | 0.10 | 526 | 0.07 |
| New York | 9,785 | 39.77 | -0.04 | 0.38 | -0.16 | 343 | -0.22 |
| Washington | 1,860 | 39.05 | -0.07 | 0.44 | -0.01 | 604 | 0.14 |
| Chicago | 883 | 40.26 | -0.12 | 0.46 | 0.02 | 662 | 0.14 |

*Notes*: Mean wage is in thousands of Costa Rican currency (real terms). Data is monthly and spans 2015-2019.

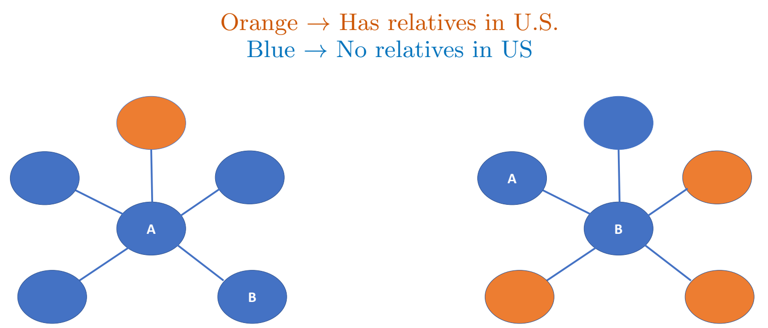
Table B.3 shows these normalized differences for age (in years), gender, and wages for the main Costa Rican consulates in the U.S. While the first two observables are available for all migrants from National Registry data, the last one is only available for migrants who were formally employed before migrating, and whose employment took place at least during one month between 2006 and 2019. As shown, the balance in characteristics of migrants to different U.S. consulates is remarkable; *all of the normalized differences are close to zero and well below 0.25* (in absolute value).

# C Details on Networks of Friends

As briefly explained in Section 2, we use data on comprehensive transactions on *Sinpe Móvil*, an application that allows Costa Ricans to make peer-to-peer money transfers using their mobile phones, to construct networks of friends. Over 60% of all adults in the country are users of this technology to send money to their peers (Alvarez et al., 2023). First, we leverage information on bilateral transactions across users, and their unique identifiers, to identify which pairs of people have sent money to each other in the past. Second, we want to clear this mapping from people who used to app to make a payment (for instance, a parent transferring money to a nanny). Thus, we focus only on pairs of individuals who have sent money to each other *bilaterally*, and use this to construct our proxy of "friends." For instance, if user A has only sent money to user B, we would not record this relationship as a friendship. If, however, both A and B have sent money to each other at some point in time, then their relationship is classified as a friendship. While imperfect, this allows us to proxy for networks of friends which are usually impossible to recover.

Our first stage is relatively straightforward, and works similarly for all networks. Thus, for the set of people with at least one friendship, we consider the share of individuals in a network who have a relative living in the U.S., and examine if their probability of importing a product depends on the exposure of their relative to this product in the U.S. city where they live.

Figure C.1: Networks of Friends: Example




(a) A's Friends (b) B's Friends

Analyzing the second stage presents a greater level of complexity. Figure C.1 shows an example. Suppose A and B are friends. Panel (a) is a diagram showing A's friends, and panel (b) depicts B's friends. Moreover, orange circles represent friends who have relatives living in the U.S. (i.e., they are directly exposed), while blue circles denote friends who are not directly exposed. Focus on panel (a): A only has one exposed friend. Now, is B's exposure coming from this one friend only? Just observing panel (a), it might be tempting to answer positively, however, as shown in panel (b), this is not necessarily the case. Note that this is not an issue for networks of neighbors or coworkers, because they are partitions.

This example illustrates the rationale behind our decision to define networks of friends on an individual-specific basis (i.e., A has three friends, each friend has her friends...). The example also shows why there are as many networks as users of the app with at least one friend, and why for the second stage involving the *friends* network, our dependent variable includes only imports of the *centroid* of the network (i.e., A's imports when considering A's network).

**Sample** Finally, some people have a very large number of friendships (i.e., people with over 50 friends). This goes against the spirit of our measure: we ideally want to capture close relationships with our friendship measure to complement those potentially less close as neighborhoods or coworkers in large firms. Thus, we constraint the sample used in the paper to people with relatively close connections. In particular, we remove close to 30% of the sample with more friends, by restricting the sample to those people with only five friends or less (five being the mean number of friends in the full sample). The latter also aids in making computations manageable, as individuals with a large number of friends pose a challenge in this regard.

# D Main Analysis: Additional Results

## D.1 Note on Clustering

This section explains why, in our particular setting, it is sufficient to cluster standard errors by product-network, and it is *not* necessary (nor computationally feasible) to adjust our standard errors à la Adão et al. (2019) (AKM) or employ randomization inference. We expand on this result below, first intuitively and then more rigorously.

The AKM thought experiment for shift-shares is that, instead of quasi-random assignment happening at the level of the shares, there is quasi-random assignment of shocks. The authors then propose how to do inference if assignment happens at the level of the shocks. For a typical shift-share, everyone in the economy is exposed to each industry-level shock, so clustering is insufficient.

In our setting, however, the observations are at the network-product-time level, thus, a given network-product-time observation will have zero exposure to another product's shocks. Pairing this fact with how, by construction, our shocks are drawn independently across products, then we can justify clustering at the product level; the regressor is drawn independently across products. This can be made more robust by clustering at the network-product level, which (unsurprisingly given the design of our shocks) does not change results much as compared with the product-level clustering. In fact, in our particular setting, this method is strictly more robust than employing AKM: in our case, AKM would have treated shocks for product $p$ in California as independent from shocks to product $p$ in New York, while clustering at the product level allows for arbitrary correlation within product.

More rigorously, in the formula for the standard error, the inverse of the covariance of the instrument and the regressor is usually straightforward to estimate, while the critical question is how to estimate the following object:

$$\Omega := \frac{1}{BPT} \sum_{b,p,t} \sum_{j,q,s} E[Z_{bpt}\varepsilon_{bpt}Z_{jqs}\varepsilon_{jqs}],$$

where $BPT$ is the number of observations, which depends on the number of networks ($B$), products ($P$), and periods ($T$).[60] If we cluster by product, then the estimator is given by: $\frac{1}{BPT} \sum_{b,p,t} \sum_{j,q,s} 1(p = q) \cdot [Z_{bpt}\varepsilon_{bpt}Z_{jqs}\varepsilon_{jqs}]$. This object estimates the within-product terms (the terms where $p = q$), but sets the "across-product" terms ($p \neq q$) to zero. This estimator will converge to the true $\Omega$ under mild conditions, and the main condition to be satisfied is that $E[Z_{bpt}\varepsilon_{bpt}Z_{jqs}\varepsilon_{jqs}] = 0$ when $p \neq q$.

Suppose (as implied by the AKM thought experiment) that the product-level demand shocks (used to construct $Z$) are drawn independently across products, and that this holds when conditioning on $(\varepsilon_{bpt}, \varepsilon_{jqs})$. Then, for $p \neq q$, $E[Z_{bpt}Z_{jqs}|\varepsilon_{bpt}, \varepsilon_{jqs}] = 0$, which implies that $E[Z_{bpt}\varepsilon_{pit}Z_{jqs}\varepsilon_{jqs}] = 0$. Thus, under weaker assumptions than those in AKM, in our setting setting it is appropriate to cluster by product. Further, to be even more conservative, in all our estimations we opt for a two-way cluster by product, $p$, (which encompasses the AKM thought experiment) and by network, $b$, (which would cover the case with quasi-randomly and independently drawn shares).

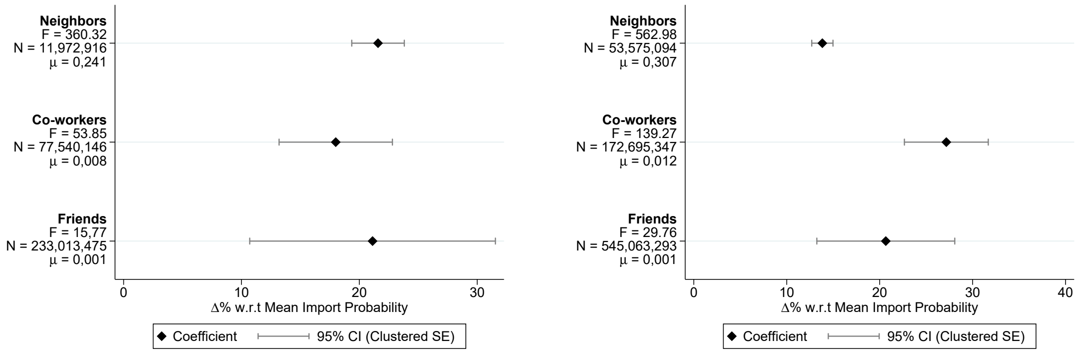
## D.2 Reduced Form and OLS Results

Figure D.1 and Table D.1 report the reduced form and the OLS results, respectively. The OLS estimates are smaller than the IV estimates. This is not entirely surprising. While a pure endogeneity bias would inflate OLS estimates, measurement error in peers' imports would induce a bias in the opposite direction, and may outweigh the endogeneity bias. In our setting, this is likely to occur, as the battery of fixed effects in our saturated specification is precisely aiming to eliminate the endogeneity bias. Therefore, OLS estimates are more likely to reflect the downward bias of measurement error than the upward endogeneity bias. Indeed, if we re-estimate the OLS model without fixed effects, most estimated coefficients more than double in size and become larger than their IV counterparts. Just as in our case, De Giorgi et al. (2019) find smaller peer effects in their OLS than in their IV, and point out that OLS in peer-effect estimation is likely to be downward biased also due to exclusion bias as studied by Caeyers and Fafchamps (2016).[61]

[60] For simplicity, this explanation abstracts from details on partialing-out fixed effects.

[61] Note standard errors are (two-way) clustered, thus, it is possible for them to be smaller in the IV estimation than in the OLS.

Figure D.1: Reduced Form Results




(a) Instrument based on CEX (b) Instrument based on U.S. imports

*Notes*: Panels (a) and (b) display reduced form results. Horizontal axes show effects as a percentage change with respect to the mean import probabilities per product. Mean import probabilities are reported to the left of each panel. Gray horizontal bars denote 95% confidence intervals, clustered by network-product. Regressions include network×product, network×time, and product×time fixed-effects. Appendix D.5 presents details on the sample of products used in each regression. Quarterly data spans 2015-2019.

Table D.1: OLS Regressions

*Dep. var: Prob. importing p in network b at t for those with relatives in the U.S.*

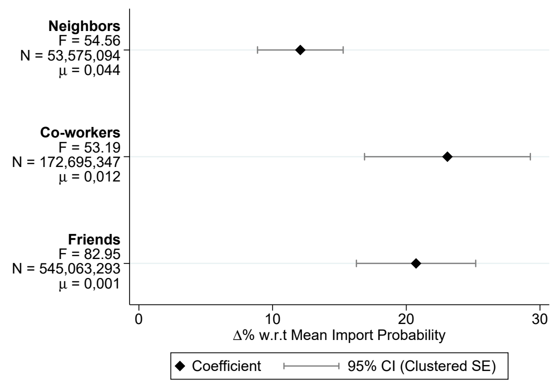
| *Panel (a): Products defined as in CEX* | | | |
|---|---|---|---|
| | %Δ w.r.t. mean import probability | | |
| | Neighbors | Co-workers | Friends |
| | (1) | (2) | (3) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 18.747 | 3.068 | -4.202 |
| | (2.244)*** | (8.312) | (2.318)* |
| Adjusted-$R^2$ | 0.132 | 0.078 | 0.060 |
| Observations | 11,972,916 | 77,540,146 | 233,013,475 |
| Mean dep. variable | 0.241 | 0.008 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |
| *Panel (b): Products defined as in U.S. Imports* | | | |
| | %Δ w.r.t. mean import probability | | |
| | Neighbors | Co-workers | Friends |
| | (4) | (5) | (6) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 24.431 | 26.116 | -7.510 |
| | (1.945)*** | (13.267)** | (3.544) |
| Adjusted $R^2$ | 0.094 | 0.047 | 0.038 |
| Observations | 53,575,094 | 172,695,347 | 545,063,293 |
| Mean dep. variable | 0.307 | 0.012 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |

*Notes*: This table shows our first stage results. Panel (a) relies on exposure measures based on the CEX, while Panel (b) relies on imports by U.S. customs districts. Robust standard errors, adjusted for clustering by network-product, are in parentheses. We include network×product, network×time, and product×time fixed-effects. Mean import probabilities are reported. Appendix D.5 presents details on the sample of products used in each regression. Data is quarterly and spans 2015-2019.

## D.3 First-Stage Regressions

Figure D.2 presents results analogous to those of Figure 3, but relying on U.S. imports data—as opposed to CEX data—to construct our instrument. As shown, the source of data on U.S. expenditures does not change the message, and only slightly changes the magnitude, of our estimates.

Figure D.2: First Stage: Imports by Costa Ricans with Relatives in the U.S. (Instrument based on U.S. imports data)




*Notes:* The figure describes our first stage results by network when constructing exposure measures based on expenditures data from imports by U.S. customs districts. The horizontal axis describes the effects as a percentage change with respect to the mean import probability in each network. Gray horizontal bars denote 95% confidence intervals, clustered by network-product. Mean import probabilities are reported to the left of each panel. The same left panel also report the F-statistics of this first stage. All regressions include network×product, network×time, and product×time fixed-effects. Tables corresponding with these figures are reported in Appendix D.3. Appendix D.5 presents details on the sample of products used in each regression.

For completeness, Table D.2 shows the tabular results of our first stage, which is run independently for three different networks and was summarized by Figures 3 and D.2. Panel (a) shows results when constructing exposure measures based on the CEX, while Panel (b) relies on imports by U.S. customs districts. Means of import probabilities by product across networks are reported; these are smaller for Panel (b), as product categories are much more narrowly defined in Panel (b), which in turn decreases the probability that a specific product will be imported by a particular network on a given quarter. Means are largest for neighbors, followed by coworkers and friends, as neighborhoods are larger, which increases the chances that someone will import a product on a given quarter.

Table D.2: First-Stage Regressions

*Dependent variable: Prob. importing product p in network b at time t for those with relatives in the U.S.*

*Panel (a): Instrument based on CEX*

| | %Δ w.r.t. mean import probability | | |
|---|---|---|---|
| | Neighbors | Co-workers | Friends |
| | (1) | (2) | (3) |
| $\ln \widetilde{E}_{bp,t-1}$ | 18.793 | 19.955 | 13.445 |
| | (2.693)*** | (3.327)*** | (3.126)*** |
| F-statistic | 48.71 | 35.99 | 18.50 |
| Observations | 11,972,916 | 77,540,146 | 233,013,475 |
| Mean dep. variable | 0.028 | 0.003 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |

*Panel (b): Instrument based on U.S. Imports*

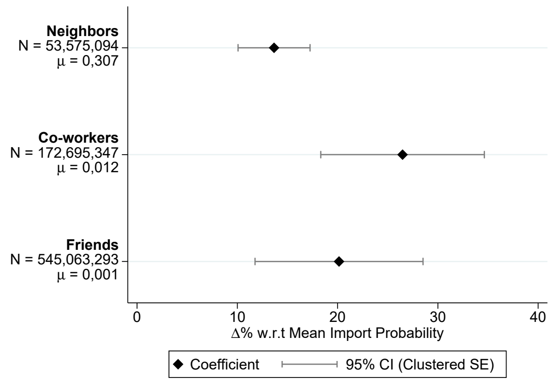
| | %Δ w.r.t. mean import probability | | |
|---|---|---|---|
| | Neighbors | Co-workers | Friends |
| | (4) | (5) | (6) |
| $\ln \widetilde{E}_{bp,t-1}$ | 12.076 | 23.078 | 20.731 |
| | (1.635)*** | (3.164)*** | (2.276)*** |
| F-statistic | 54.56 | 53.19 | 82.95 |
| Observations | 53,575,094 | 172,695,347 | 545,063,293 |
| Mean dep. variable | 0.044 | 0.012 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |

*Notes*: This table shows our first stage results. Panel (a) relies on exposure measures based on the CEX, while Panel (b) relies on imports by U.S. customs districts. Robust standard errors, adjusted for clustering by network-product, are in parentheses. We include network×product, network×time, and product×time fixed-effects. Mean import probabilities are reported. Appendix D.5 presents details on the sample of products used in each regression.

## D.4 Second-Stage Regressions

Figure D.3 presents results analogous to those of Figure 4, but relying on U.S. imports data—as opposed to CEX data—to construct our instrument. Just as for the first stage, the data source for U.S. expenditures does not change 2SLS estimates. Table D.3 shows the tabular results of our 2SLS. Means of dependent variables are reported; these are smaller for Panel (b), as compared with Panel (a), this is to be expected, as product categories are much more narrowly defined in Panel (b), which in turn decreases the probability that a specific product will be imported by a particular network on a given quarter. Means are larger for neighbors, followed by coworkers and friends; this is expected as networks of neighbors are much larger, which increases the chances that someone will import a particular product on a given quarter.

Figure D.3: 2SLS: Propagation Within Network
(Instrument based on U.S. imports data)




*Notes*: The figure shows the two-stage least squares regression corresponding with equation (4) constructing exposure based on imports by U.S. customs districts. The horizontal axis describes the effects as a percentage change with respect to the mean import probability in each network. Gray horizontal bars denote 95% confidence intervals, clustered by network-product. Mean import probabilities are reported to the left of each panel. Regressions have network×product, network×time, and product×time fixed-effects. Tables are reported in Appendix D.4. Appendix D.5 has details on the sample of products per regression.

Table D.3: 2SLS: Propagation Within Network

*Dependent variable: Prob. importing product p in network b at time t for those without relatives in the U.S.*

| *Panel (a): Instrument based on CEX* | | | |
|---|---|---|---|
| | %Δ w.r.t. mean import probability | | |
| | Neighbors | Co-workers | Friends |
| | (1) | (2) | (3) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 21.088 | 16.936 | 19.788 |
| | (3.002)*** | (3.540)*** | (6.782)*** |
| F-stat first stage | 48.71 | 35.99 | 18.50 |
| Observations | 11,972,916 | 77,540,146 | 233,013,475 |
| Mean dep. variable | 0.241 | 0.008 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |
| *Panel (b): Instrument based on U.S. Imports* | | | |
| | %Δ w.r.t. mean import probability | | |
| | Neighbors | Co-workers | Friends |
| | (4) | (5) | (6) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 13.662 | 26.479 | 20.14 |
| | (1.832)*** | (4.162)*** | (4.274)*** |
| F-stat first stage | 54.56 | 53.19 | 22.21 |
| Observations | 53,575,094 | 172,695,347 | 545,063,293 |
| Mean dep. variable | 0.307 | 0.012 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |

*Notes*: This table displays the results of running our 2SLS for different networks. Panel (a) shows results based on the CEX, while Panel (b) relies on imports by U.S. customs districts. Robust standard errors, adjusted for clustering by network-product, are in parentheses. All regressions control for network×product, network×time, and product×time fixed-effects. Mean import probabilities are reported. Appendix D.5 presents details on the sample of products used in each regression. Data is quarterly and spans 2015-2019.

## D.5 Random Sample vs. Entire Sample of Products

For some regressions, the dimensionality of the sample (*network*×*product*×*quarter*), in addition to the battery of fixed-effects, would prevent estimations from running. Thus, for some specifications, we rely on a random sample of products. Table D.4 summarizes the samples used in each table and figure of the main paper.

First, for the case of the CEX, the total number of products imported by individuals in Costa Rica is relatively manageable (430; see Table B.1). Thus, *regressions for all networks are always run using the entire sample of products,* except for Table 1, as the distance-3 exercise involves a larger set of fixed effects which demands more computational power; this regression uses a 50% random sample of products.

Second, using the full sample of products for estimations in which we construct our instrument based on U.S. imports by customs districts is rarely possible. The reason is that there are 2,443 narrowly defined product codes in the U.S. imports data which are imported by individuals in CR; this would make most regressions have over one *billion* observations. Therefore, throughout the paper, results relying on U.S. imports are based on random samples. The size of these random samples is chosen to roughly match the total number of observations used when conducting estimations via the CEX. Thus, when relying on U.S. imports, we use the *entire* sample of product when defining networks as neighborhoods, but a random sample of 50% of products when considering networks of coworkers and friends.

Table D.4: Samples in Each Table and Figure of the Main Paper

| Table/Figure<br>(1) | Expenditures Source<br>(2) | Sample of Products<br>(3) | Unit of Observation<br>(4) | # Obs.<br>(5) |
|---|---|---|---|---|
| Figures 2,3,5<br>Tables 2, 3 | CEX | 100% sample | neighbors$\times p \times t$ | 12M |
| Figures 2,3<br>Table 2 | CEX | 100% sample | coworkers$\times p \times t$ | 78M |
| Figures 2,3<br>Tables 2 | CEX | 100% sample | friends$\times p \times t$ | 233M |
| Table 1 | CEX | 50% random sample | individual$\times p \times t$ | 349M |

*Notes*: Whenever the exercise does not include all products, the sub-sample is chosen at random. Table 1 has half of the sample of products and it includes a larger battery of fixed effects.

Results and robustness checks in the appendix follow a similar pattern as described above; for network level regressions via the CEX, we always use the entire sample of products, and for estimates based on U.S. imports, we use the entire sample for neighbors and a 50% random sample for coworkers and friends networks. Table E.1 is run at the individual-level with the full sample of products for estimates based on the CEX and a 25% random sample for estimates based on U.S. imports.

## D.6 Timing of the Specifications: Local Projections

We use Jordà (2005) local projections to better understand the timing of the propagation after a network is exposed to a product.[62] In particular, we consider the following set of panel local projections:

$$\text{Import}^{US\,exposure}_{bp,t+h} = \alpha^h + \beta^h \ln \widetilde{E}_{bpt} + \lambda^h x_{bpt} + \gamma^h_{bp} + \gamma^h_{bt} + \gamma^h_{pt} + \varepsilon_{bp,t+h}, \tag{19}$$

where $h = 0, 1, 2, 3$ and $x_{bpt}$ is a vector of controls with lags of the outcome variable and the shock. We present results using three lags of the outcome variable and two lags of the shock; pre-trends are controlled for by our lag specification. Propagation results are similar with less stringent specifications on the number of lags.

We are interested in the cumulative impulse response from an exogenous increase in exposure to a product. Panel (a) of Figure D.4 reports the results from estimating equation (19). It shows that an increase in the networks' exposure to a product from relatives abroad has a relatively small increase on impact. One quarter after the exogenous increase in exposure, import probability permanently stabilizes (vertical dashed line). Thus, in line with what an AIC would indicate, we include exposure with one lag in our main specification (equation (3)), which is the first period that would capture the full effect.

We then implement a similar local projection for the IV in equation (4), which considers people in a network who are unrelated to migrants abroad, as follows:[63]

$$\text{Import}_{bp,t+h} = \alpha^h + \beta^h \widehat{\text{Import}}^{US\,exposure}_{bpt} + \lambda^h x_{bpt} + \gamma^h_{bp} + \gamma^h_{bt} + \gamma^h_{pt} + \varepsilon_{bp,t+h}, \tag{20}$$

Panel (b) of Figure D.4 reports the results. The figure shows that the effects stabilize one quarter after the shock and persist afterward. Thus, equation (4)'s main independent variable is lagged one period. To better understand this result, recall our dependent variable equals one if someone in the network imports the product, thus, this result implies that after a person with a relative abroad introduces the product to the network, at least one person in the network (without relatives in the U.S.) imports the product thereafter.

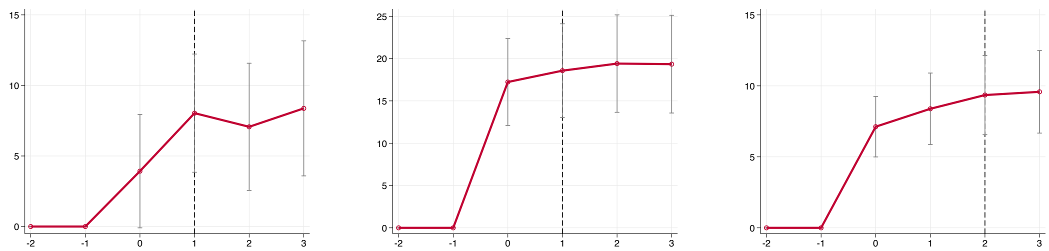
Lastly, we study the response of retailers to an increase exposure to a product using again an IV specification, as described in equation (7), using a local projection. Panel (c) shows that retailers respond to the observed local demand for foreign goods by importing those products on impact, but the effects grow and stabilize only two

[62] Local projections are based on sequential regressions of the endogenous variable shifted several steps ahead (Jordà, 2005). They are able to accommodate IV estimations (Jordà et al., 2020), and they can robustify inference and simplify the computation of standard errors (Montiel Olea and Plagborg-Møller, 2021).

[63] For examples of IV applications using local projections, see Jordà et al. (2020).

quarters after the shock, which aligns with the timing in equation (7). Importantly, note that effects do not differ significantly across quarters, so the choice of which lag to include in our main specification would not majorly alter the results.

Figure D.4: Local Projections of the Change in Import Probabilities




(a) First Stage (b) Second Stage (c) Retailers' Response

*Notes*: Panel (a) shows impulse responses of the probability of importing to an increase in network exposure from abroad, estimated using equation (19). Panel (b) shows impulse responses of the probability of importing for people without relatives abroad to an increase in network exposure from abroad, estimated using equation (20). Panel (c) shows impulse responses of the probability of importing for retailers to an increase in network exposure under an IV approach using a local projection.

## D.7 Intensive Margin Results

Our baseline regressions' dependent variables are quite conservative: if *anyone* in the network was importing the product, its value is one. Thus, changes in this variable should arise from unusual imports. It is possible, however, to conduct an intensive margin analysis, such that:

$$\text{Import Q}_{bpt} = \alpha + \beta_{int}\, \widehat{\text{Import Q}^{US\,exposure}_{bp,t-1}} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt}, \tag{21}$$

where the main variables consider the quantity of importing events; other variables are defined as in equation (4). Results are consistent with those of our main specification.

Table D.5: 2SLS: Intensive Margin Propagation Within Network

*Dep.var. first (second) stage: Imports p in b at t of those with (without) relatives in U.S.*

| | | %Δ w.r.t. mean import quantity | | | |
|---|---|---|---|---|---|
| | | Neighbors | Co-workers | Friends | Retailers |
| | | (1) | (2) | (3) | (4) |
| First Stage | $\ln \widetilde{E}_{bp,t-1}$ | 18.714 | 19.969 | 13.445 | 18.686 |
| | | (2.684)*** | (3.326)*** | (3.126)*** | (2.664)*** |
| | Mean dep. variable | 0.160 | 0.014 | 0.006 | 0.166 |
| Second Stage | $\widehat{\text{Import Q}}^{US\,exposure}_{bp,t-k}$ | 23.431 | 19.423 | 19.788 | 13.262 |
| | | (3.388)*** | (4.666)*** | (6.782)*** | (1.911)*** |
| | Mean dep. variable | 2.528 | 0.037 | 0.002 | 15.337 |
| | F-stat first stage | 48.62 | 36.06 | 18.50 | 49.19 |
| | Observations | 11,972,916 | 77,540,146 | 233,013,475 | 11,299,497 |

*Notes*: This table displays the intensive margin results per network. Robust standard errors, adjusted for clustering by network-product, are in parentheses. $k = 1$ for columns (1)-(3) and $k = 2$ for column (4). All regressions control for network×product, network×time, and product×time fixed-effects.

# E Robustness Exercises

## E.1 Instrument Using Distance-3 Nodes

**Specification** As explained in Section 3.3, we consider the following regression for individual $d$, which depends on product $p$ at time $t$:

$$\underbrace{Import_{dpt}}_{\text{individual } d} = \delta_0 + \underbrace{\theta_d \widehat{Import^d_{ip,t-1}}}_{\text{average exposure}} + \underbrace{\delta^d_{xpt}}_{\text{firm-product-time FE}} + \underbrace{\delta_d}_{\text{individual FE}} + \varepsilon_{dpt}$$

where the dependent variable $Import_{dpt}$ is the probability of consuming (importing) foreign good $p$ for individual $d$ at time $t$.[64] In terms of the right-hand side of the regression, $\widehat{Import^d_{ip,t-1}}$ is the average exposure of at individual $d$'s firm, defined as the mean residuals of equation (6) for employees of each firm. $\delta^d_{xpt}$ are firm$\times$product$\times$time fixed-effects; these fixed-effects are key, as they force the identifying variation to come from *differences* between the coworkers' spouses firms and the household's employers (see Figure 5 for reference). Finally, $\delta_d$ are individual fixed-effects.[65]

Thus, $\theta$ may identify the endogenous peer effects (our parameter of interest) in the absence of shared unobserved shocks or contextual effects if peers share traits. To isolate these effects, De Giorgi et al. (2019) exploit that social relationships are established along two lines: at the family level (e.g., husband and wife) and at the firm level (co-workers). The idea that the authors push forward is that firm-specific shocks at the firm of a coworker's spouse are a valid instrument for the household's consumption changes due to $\widehat{Import^d_{ip,t-1}}$. Therefore, we instrument for $\widehat{Import^d_{ip,t-1}}$ using, instead, the mean exposure at the firm of an individual's coworkers' spouses.

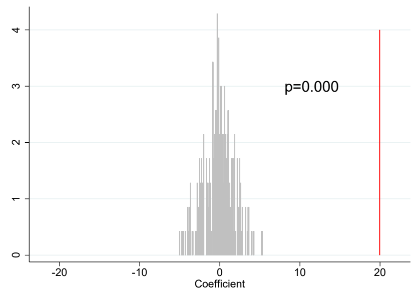
**Remarks on Data Construction** To construct the aforementioned instrument, we first identify couples in our sample where both spouses are employed. We then exclude couples who work at the same firm, and also coworkers whose spouses work at the same firm to avoid feedback effects. Information transmission, we assume, will occur across the remaining couples in the sample.

[64]Note that, unlike De Giorgi et al. (2019), who consider total consumption by households, this is a regression that will identify peer effects in the consumption of a particular product, which we think is beneficial in terms of identification of the effect. Moreover, we have data on individual consumption, thus we are able to run this regression at the individual level.

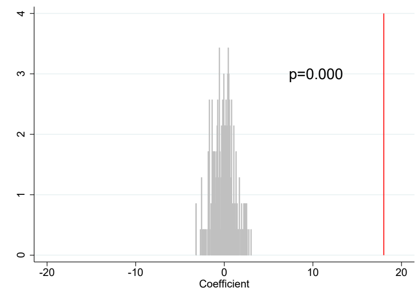
[65]While De Giorgi et al. (2019) run their regression in first differences and include changes in observables as dependent variables, we instead include much more demanding fixed-effects, which is similar (in spirit) to their approach. We depart from the first-differences approach because our dependent variable is an indicator, and because (unlike previous papers) we can include fixed-effects that can better discipline the identifying variation.

## E.2 Placebos: Additional Network Definitions

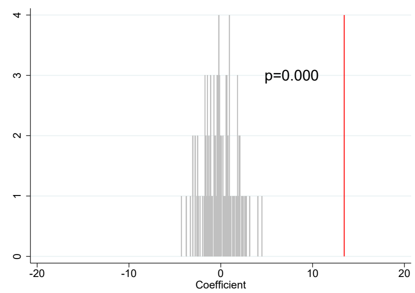
Figure E.1: Placebo vs. Actual Coefficients: Other Network Definitions



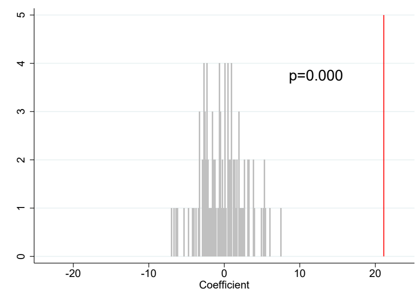

(a) First Stage: Coworkers




(b) Reduced Form: Coworkers




(c) First Stage: Friends




(d) Reduced Form: Friends

*Notes:* The figures plot the distribution of placebo coefficients based on placebo exposure measures defining networks as coworkers and friends. The red vertical lines plots the actual first stage and reduced form coefficients. While coworkers' results are based on 250 iterations, friends' results are based on 100 iterations, as its total observations is about three times larger and more demanding computationally. The p-values represent the share of 1,000 placebo coefficients that are larger in magnitude than the absolute coefficient for the actual first stage or reduced form.

## E.3 Results at the Individual-Level

Table E.1: Individual Imports and Relatives' Exposure to Products Abroad

*Dependent variable:* $Import^{US\,exposure}_{npt}$
*(Prob. importing product p for individual n at time t)*

| | Instrument based on | |
|---|---|---|
| | CEX | U.S. imports |
| | (1) | (2) |
| Relative's exposure abroad to product $p$ in $t-1$ | 17.976 | 13.403 |
| (%Δ w.r.t. mean import probability) | (2.755)$^{***}$ | (4.140)$^{***}$ |
| Observations | 512,567,440 | 592,314,592 |
| $np$, $nt$, $pt$ FE | Yes | Yes |
| F-statistic | 42.57 | 10.48 |

*Notes:* Column (1) shows results when constructing exposure measures based on the CEX, while column (2) relies on imports by U.S. customs districts. Robust standard errors, adjusted for clustering by individual-product, are in parentheses. Dependent variables are the probability that an *individual* imports a *specific* product code in a particular quarter *and* from the U.S., thus, the percentage mean import probability of a product is small; 0.0009 and 0.0004 for each column. Regressions control for individual-product, individual-time, and product-time fixed effects. Appendix D.5 details the sample used per regression.

## E.4 Results with Additional Fixed-Effects

Table E.2: Results with More Demanding Specifications for Neighborhoods

*Dependent variable: Prob. importing p in neighborhood b at time t for non-relatives*

| | %Δ w.r.t. mean import probability | |
|---|---|---|
| | (1) | (2) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 23.010 | 22.576 |
| | (3.383)*** | (3.195)*** |
| F-stat first stage | 45.80 | 49.50 |
| Mean dep. variable | 0.240 | 0.240 |
| District×$p$×$t$ FE | Yes | No |
| Network×HS-2×$t$ FE | No | Yes |
| Observations | 11,852,343 | 11,677,203 |

*Notes*: This table the results of running equation (4) defining networks as neighborhoods and adding more demanding controls. Column (1) includes District×$p$ × $t$ fixed effects, while column (2) includes network×HS-2×$t$ fixed effects. Robust standard errors, adjusted for clustering by network-product, are in parentheses. Mean import probabilities are reported. All regressions control for network×product, network×time, and product×time fixed-effects. Appendix D.5 presents details on the sample of products used in each regression. Data is quarterly and spans 2015-2019.

# F Determinants of Product Propagation

## F.1 Asymmetric Response of Positive and Negative Shocks

Table F.1: Effect of Positive and Negative Changes in Exposure

*Dependent variable: Prob. importing product p in network b at time t for importers with relatives in the U.S.*

| | %Δ w.r.t. mean import probability | | |
|---|---|---|---|
| | Neighbors | Co-workers | Friends |
| | (1) | (2) | (3) |
| $Positive_{cp,t-1} \times \ln \widetilde{E}_{bp,t-1}$ | 51.634 | 16.533 | 57.575 |
| | (9.397)*** | (3.956)*** | (12.767)*** |
| $\ln \widetilde{E}_{bp,t-1}$ | -2.574 | 0.750 | -10.876 |
| | (2.948) | (1.392) | (4.033)*** |
| F-stat first stage | 25.13 | 18.35 | 11.66 |
| Observations | 11,972,916 | 77,540,146 | 233,013,475 |
| Mean dependent variable | 0.241 | 0.008 | 0.001 |
| *bp*, *bt*, *pt* FE | Yes | Yes | Yes |

*Notes*: The table shows our first-stage regression interacting exposure with an indicator equal to one if the residual is positive. Robust standard errors, adjusted for clustering by network-product, are in parentheses. Mean import probabilities are reported. Regressions control for network×product, network×time, product×time fixed-effects. Appendix D.5 has details on the sample of products per regression.

## F.2 Dynamic vs. Established Products

Table F.2: Strength of Externalities According to Products' Dynamism

*Dependent variable: Prob. importing product p in network b at time t for non-relatives*

| | %Δ w.r.t. mean import probability | | | | | |
|---|---|---|---|---|---|---|
| | Jobs by New Establishments | | | Entry of Establishments | | |
| | Neighbors | Co-workers | Friends | Neighbors | Co-workers | Friends |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Dynamic_p \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 22.132 | 4.563 | 32.355 | 23.656 | 7.185 | 58.252 |
| | (4.466)*** | (9.044) | (38.490) | (5.613)*** | (10.066) | (99.804) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 1.354 | 13.168 | -9.849 | -1.469 | 10.639 | -36.602 |
| | (2.768) | (7.811)* | (37.702) | (4.504) | (9.216) | (100.069) |
| F-stat first stage | 22.00 | 11.41 | 4.26 | 24.65 | 11.65 | 5.21 |
| Observations | 11,928,214 | 77,379,405 | 232,565,929 | 11,928,214 | 77,379,405 | 232,565,929 |

*Notes*: The table shows the results of running equation (4), where the IV is interacted with an indicator equal to one if the good is classified as dynamic. Columns (1)-(3) classify a product as dynamic if the creation of jobs by new establishments is above the median of the sample, while columns (4)-(6) classify a product as dynamic if the entry of new establishments is above the median. Robust standard errors, adjusted for clustering by network-product, are in parentheses.

## F.3 Centrality of the Importer

Table F.3 displays results according to the centrality of importers with relatives in the U.S. Centrality is defined as degree centrality using our app-based definition of friendship. Given that we use friends' networks to define centrality, and as friend networks are defined at the individual level (see Appendix C), we do *not* run regressions defining networks as friends for this exercise.[66] While noisy, results in Table F.3 suggest that the more central the importers in the first stage, the stronger the propagation across the network in the second stage.[67]

Table F.3: Strength of Externalities and Importer's Centrality in Network of Friends

*Dependent variable: Prob. importing product p in network b at time t for non-relatives*

| | %Δ w.r.t. mean import probability | |
|---|---|---|
| | Neighbors | Co-workers |
| $Centrality_b \times \widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 3.032 | 1.536 |
| | (4.321) | (1.594) |
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-1}$ | 15.140 | 12.859 |
| | (8.305)* | (4.237)*** |
| F-stat first stage | 24.08 | 11.27 |
| Observations | 11,972,916 | 77,540,146 |

*Notes*: The table shows the results of running equation (4), but where the main independent variable was interacted with a measure of the average degree centrality of people in the network without relatives in the U.S. Robust standard errors, adjusted for clustering by network-product, are in parentheses. All regressions control for network×product, network×time, and product×time fixed-effects.

[66] Centrality would then necessarily measure how central a friend is *outside* of the network of friends considered, making results hard to interpret for networks of friends in particular.

[67] On column (1), one more connection in the mean degree centrality leads to a 3% higher probability of importing after an exogenous import, as compared with the mean network.

# G Retailers' Response: Additional Results

## G.1 Results Based on Imports from Any Country

Table G.1: Retailers' Imports from Any Country and Exposure

*Dependent variable: Prob. of retailers importing product p in neighborhood b at time t*

| | %Δ w.r.t. mean import probability |
|---|---|
| $\widehat{\text{Import}}^{US\,exposure}_{bp,t-2}$ | 10.271 |
| | (1.457)*** |
| F-stat first stage | 49.29 |
| Mean dependent variable | 0.480 |
| $bp$, $bt$, $pt$ FE | Yes |
| Observations | 11,299,497 |

*Notes*: Robust standard errors, adjusted for clustering by neighborhood-product, are in parentheses. Mean import probabilities are reported. The regression includes neighborhood×product, neighborhood×time, and product×time fixed-effects. Data is quarterly and spans 2015-2019.

## G.2 Mechanism Based on Employer-Employee Data

We can leverage our employer-employee data to better understand the mechanism behind retailers' response to individual-level imports. The idea behind this exercise is that employees can be exposed to foreign products in their neighborhoods and transmit information about the existence of these products to their employers. However, if employees live in areas which are relatively far away from the retailer, they should not be able to speak about the particular level of the local demand that their employer will face. This strategy exploits that there is an imperfect overlap between a retailer's location and the residence of its employees.

We construct measures of exogenous exposure to foreign products by employees depending on the exposure faced in the neighborhoods where they reside, and focusing on employees who reside in *districts* (denoted by D) other than the one where their employer is located; on average, each district spans between 3-4 neighborhoods. Namely, we will consider the following independent variable:

$$\widetilde{E}^{emp,far}_{bp,t} = \sum_{g \neq D} \frac{emp^F_g}{emp^F_b} \widetilde{E}_{gpt},$$

where $\frac{emp^F_g}{emp^F_b}$ is the share of employees of retailers in neighborhood $b$ in district $D$ who are living in neighborhoods $g$ outside of district $D$, so that we calculate the exposure of the firm as the average across the exposure faced in the neighborhoods that are outside the retailers' district but where its employees live. This variable would represent the exposure to product $p$ faced by employees of firms in $b$ in the "far away" neighborhoods where they reside. We then propose the following specification for imports of product $p$ by retail firms in neighborhood $b$ at time $t$:

$$\text{Import}^F_{bpt} = \delta + \kappa \widetilde{E}^{emp,far}_{bp,t-2} + \zeta \widetilde{E}^{US\,exposure}_{bp,t-2} + \gamma_{bp} + \gamma_{bt} + \gamma_{pt} + \varepsilon_{bpt},$$

where $\text{Import}^F_{bpt}$ is one if a firm in neighborhood $b$ imports product $p$ at time $t$. We then include as independent variable the firm's exposure from their employees' who reside "far away", $\widetilde{E}^{emp,far}_{bp,t-2}$, and control for the exposure faced by the firm's own neighborhood, $\widetilde{E}^{US\,exposure}_{bp,t-2}$; we again include a battery of fixed effects.

As shown in Table G.2, retailers do not show a meaningful response to the exposure of employees living far away; if anything, the coefficient is both small and negative. The latter aligns with firms learning about the level of the local demand for a product, as opposed to just a product-discovery story.

Table G.2: Retailers' Imports and Exposure of Employees Living Far Away

*Dependent variable: Prob. of retailers importing product p in neighborhood b at time t*

| | %Δ w.r.t. mean import probability |
|---|---|
| $\widetilde{E}^{emp,far}_{bp,t-2}$ | -2.095 |
| | (0.403)*** |
| $\widetilde{E}^{US\,exposure}_{bp,t-2}$ | 12.771 |
| | (0.720)*** |
| F-statistic | 287.94 |
| Mean dependent variable | 0.822 |
| *bp*, *bt*, *pt* FE | Yes |
| Observations | 2,316,575 |

*Notes*: Robust standard errors, adjusted for clustering by neighborhood-product, are in parentheses. Mean import probabilities are reported. The regression includes neighborhood×product, neighborhood×time, and product×time fixed-effects. Data is quarterly and spans 2015-2019.

# Supplementary References


Acosta, M. and Cox, L. (2019). The Regressive Nature of the U.S. Tariff Code: Origins and Implications. Technical report, Working Paper, Columbia University.

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Alvarez, F., Argente, D., Lippi, F., Méndez, E., and Van Patten, D. (2023). Strategic Complementarities in a Dynamic Model of Technology Adoption: P2P Digital Payments.

Caeyers, B. and Fafchamps, M. (2016). Exclusion Bias in the Estimation of Peer Effects. Technical report, National Bureau of Economic Research.

Caplin, A. and Leahy, J. (1998). Miracle on Sixth Avenue: Information Externalities and Search. *The Economic Journal*, 108(446):60–74.

De Giorgi, G., Frederiksen, A., and Pistaferri, L. (2019). Consumption Network Effects. *The Review of Economic Studies*, 87(1):130–163.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.