# Migrant Taste Bias in Gravity: Evidence from Inter-provincial Trade in China


Tiankuo Han (*Tsinghua University*)

Yushan Hu (*Boston College*)

Maozhen Wu (*Australian National University*)

Penglong Zhang* (*Tsinghua University*)


May 2021


**Abstract**

Do domestic migrants tend to buy products from their home regions? This paper embeds migrant taste bias into a structural gravity model. Bilateral market tastes are determined by combing bilateral migration and migrants' taste bias. Using inter-provincial manufacturing trade and migration data of China, this paper estimates the home bias of migrants in tastes. The results show that in each 1 dollar of the total expenditure, each consumer spends 46 cents more on products from her home province than other products, given all else equal. The estimated taste biases are larger when trading agricultural products and final products, while smaller when trading services and intermediate products. The migrants' taste bias is robust even with the assimilation effect. Taste bias is found to explain the trade bias more than trade cost. Counterfactuals suggest that anti-migration policy significantly impedes trade with migrants' province of origin.



**Keywords:** Gravity; Home bias; Taste bias; Migration; China.

**JEL Classification:** D10, F11, F14, F22.



*Corresponding author: Penglong Zhang, School of Public Policy and Management, Tsinghua University. Address: Qinghua West Road, Beijing 100084, China. E-mail address: zhangpenglong@tsinghua.edu.cn.


We would like to thank James Anderson, Pol Antrás, David Atkin, Susanto Basu, Pablo Fajgelbaum, Amit Khandelwal, Arthur Lewbel, Ben Li, James Rauch, Roberto Rigobon, Fabio Schiantarelli, Felix Tintelnot, Zhijie Xiao, Juanjuan Zhang and the seminar participants at Boston College, MIT Sloan School of Management, Tsinghua University, the meetings of the Midwest International Economics Group, as well as the European Trade Study Group. All errors are ours.

## 1 Introduction

Recent studies on international trade have documented the importance of consumer tastes in explaining both country-level and firm-level trade flows (Bronnenberg et al., 2012; Di Comite et al., 2014; Zhang, 2020). Consumers have home-biased preferences (home bias hereafter), which they retain when they migrate. Thus, migrants tend to buy products from their home countries. Ignoring this bias would lead us to overestimate the extent to which trade liberalization can foster market integration. Further, failure to account for preference heterogeneity causes problems in analyzing welfare gains across consumers. In addition, tastes usually conceptualize consumers' identity, which plays an important role in both the economic and political environments. Scholars have linked the effects of globalization, immigration, and rising inequality to a shift away from a now lower-status working-class identity toward a nationalist both Europe and the United States (Shayo, 2009).

Previous studies have linked taste differences to cross-country differences based on factors such as ethnicity, religion, language, and cultural history. In international trade, it is challenging to distinguish migrants' taste bias from language barriers because they both better know and better like the products of home brands than the others. Differently, in the domestic trade scenario, language barriers are removed, and brand taste plays a more essential role in consumption (Bronnenberg et al., 2012). In this study, we examine whether tastes differ within a country in which all other cross-country differences have been eliminated. Atkin (2013) provides the evidence based on agricultural consumption in India by attributing taste bias to habit formation, that is, adults favor the foods they consumed as a child, and the preferences developed in childhood persist into adulthood. In this study, we go beyond the agricultural sector by incorporating the manufacturing sector. We examine taste differences from domestic migration and inter-provincial trade in China, where consumers are much more homogeneous, and the market is more integrated than in India.

This paper replaces the assumption of the representative consumer with heterogeneous consumers in terms of taste bias. Specifically, any given consumer has a taste biased toward the goods produced by her province of origin, regardless of where she currently resides. Thus consumers are heterogeneous in terms of their taste biases. We extend the structural gravity model by building and estimating a structural component of home-biased preferences. Bilateral market tastes are determined by combing bilateral migration and migrants' taste bias. Using inter-provincial manufacturing trade and migration data of China, this paper estimates the home bias of migrants in tastes. The results show that migrants' tastes for products from their home provinces deviate from unbiased levels significantly. The estimated taste biases are larger when trading agricultural products and final products, while smaller when trading services and intermediate products. The migrants' taste bias is robust even with the assimilation effect.

The basic assumption in this paper is that a consumer's tastes only depend on her province of origin and are unrelated to where she currently lives. Thus, consumers from the same home province share the same tastes regardless of where they are located. One viable explanation of this is the habit formation in relation to tastes proposed by Atkin (2013), who argues that adult taste favors the foods they consumed as a child, and the preferences gained in childhood persist into adulthood. This assumption in relation to preferences yields an aggregate demand system that is feasible to estimate. With the taste structure and the preference assumption, we derive a home-biased gravity equation by aggregating expenditure shares across migrant groups within each province. In addition to trade costs, the gravity equation also captures the impact of the trade promoter that stimulates bilateral trade shares via migration. The intuition is that the biased migrants spend larger shares of their income on products from their home province than the native residents, thereby increasing the trade share from their home province to their host province.

In the empirical part of the paper, we estimate the home-biased gravity model. One of the related concerns is that the taste shifters may be overestimated because migration increases trade not only through the preference channel but also through the trade cost channel (Rauch, 2001). In order to control for the network channel, we follow Zhang (2020) to construct a network variable that represents the probability that, if we select an individual at random from each province, they will have a connection defined as common regional origin. Importantly, our measure extends the links in Rauch (2001) to a summation of all indirect and both direct links and thus captures the network effect as much as possible.

Another concern in our estimation is endogeneity. The migrant share may be correlated with trade shocks. It is possible that a positive shock to the value of bilateral trade between the two provinces leads to more migration between them. Migration may be correlated with unobserved factors that also affect trade, such as the trading partners' cultural similarity or bilateral economic policies. To address the possible endogeneity problem, we construct an instrumental variable by the exogenous migration push factor and conduct a two-stage least square estimation. Borrowing the idea in Llull (2018a), we instrument the migrant share by the interaction of natural disasters of origin and the distance to the destination. The underlying assumption is that the promoting effect of the push factor differs across distances.

Using inter-provincial manufacturing trade and migration data of China, this paper estimates the home bias of migrants in tastes. We refer to the migrants as all the people who currently live outside where they were born, i.g., both temporary and permanent migrants.[1] Different from most previous reduced-form studies on trade and migration, we use the trade share instead of the trade flow as the dependent variable, which is implied by the home-biased gravity equation derived in our structural model. The results show that in each 1 dollar of the total expenditure, each consumer spends 46 cents more on products from her home province than other products, given all else equal. To further investigate the taste biases, we also estimate the sectoral gravity. Compared with the manufacturing sector, consumers have larger taste biases in relation to agricultural products while less biased on services. Trade share is more sensitive to migrant share for final goods than for intermediate goods.

The assumption that tastes only depend on origin, not on the current province, is very strong. Migrants assimilate and change their tastes in order to adapt to the dominant society, the native people of the destination province. Thus we also extend the model by adding migrant assimilation. We redefine the taste structure. On the one hand, a consumer has a taste biased toward goods produced by her home province. On the other hand, her taste is also biased toward the goods produced locally. The gravity equation with assimilation also captures the local effect in the baseline model. It measures the size of the relative trade "promoter" that stimulates internal (local) trade due to the local bias in taste. The specification of the home-biased gravity equation with assimilation effect is identical to that without assimilation (our baseline results) for inter-provincial trade, but includes an internal promoter only for intra-provincial trade. To estimate the home-biased gravity equation with assimilation, we extend our sample by adding the observation of internal trade in each province. After controlling for local bias in the gravity equation, the

[1] In principle, there could be two kinds of domestic migrants in China. One kind includes individuals temporarily work in a host province and travel back to their home provinces yearly. The other kind includes individuals who permanently work in a province (with Hukou-registered) while their hometowns are in other provinces.

home taste shifter (migrant share coefficient) estimated is smaller, but still significantly positive. The difference is moderate, which implies that the baseline estimation results are robust.

Using the estimated parameters, we conduct two counterfactual experiments. First, we decompose the contribution of the taste effect and trade cost effect to the trade bias. About only one third is explained by the trade cost, and the other two thirds are explained by the taste bias. Second, we check how migration policy affects trade bias. The results show that countries decrease import shares significantly with a migration ban.

Our work contributes to the literature on the significance of taste bias or preferences in relation to trade. There is an increasing trend in the literature linking trade flows with the heterogeneous taste of trade bodies within countries and inter-provincially (Bronnenberg et al., 2012; Atkin, 2013; Di Comite et al., 2014; Zhang, 2020). And Caron et al. (2014) explore the relationship between characteristics of goods and services in production and characteristics of preferences. The preferences are assumed to be identical across countries but non-homothetic in their model. More recently, Morey (2016) demonstrates the existence of home bias in international trade using micro-evidence from the agricultural sector. Our study differs from those studies in two key ways. First, we provide a more general model and empirical evidence, focusing on the perspective of all migrants instead of a single ethnic group, and analyze macro-evidence from the manufacturing sector and service sector, as well as the agricultural sector, confirming the existence of taste bias across a broad range of traded goods. Second, and more importantly, instead of mainly focusing on home bias in international trade, we examine a much more specific range of economic activities at the domestic level. The intuition is that different results are obtained from inter-country and intra-country comparisons, which taste of domestic migrants is more homogeneous. Our findings suggest that taste bias exists in relation to domestic migration and can also promote bilateral inter-provincial trade.

This study also contributes to the large and growing body of literature quantifying the effects of migration (Bryan and Morten, 2019; Schmutz and Sidibe, 2019; Tombe and Zhu, 2019; Arkolakis et al., 2020), especially the effects of migration on bilateral trade flows (Iranzo and Peri, 2009; Felbermayr et al., 2015). In particular, immigrants promote communication among firms and reduce setup costs in the destination country, reducing the barriers to exporting, resulting in network effects (Rauch and Trindade, 2002; Combes et al., 2005; Bailey et al., 2021). In this study, we improve on the separate identification of the information and preference effects of migrants in China. Hao et al. (2020) find that the decline in China's internal migration costs reduced the costs of internal trade, and that changes in migration policy are central to China's structural changes and regional convergence. We provide support for these findings by providing an alternative explanation on the preference channel.

This paper helps to explain the "border puzzle" using home-biased preferences, which is supplementary to other explanations, e.g., trade costs (McCallum, 1995; Obstfeld and Rogoff, 2001), multilateral resistance (Anderson and van Wincoop, 2003), non-homothetic preferences (Caron et al., 2014), multi-stage production (Yi, 2010), and scale effects (Ramondo et al., 2016). Chaney (2014) and Allen (2014) discuss resistance as a result of information barriers. Head and Mayer (2013) summarize the sources of resistance as imperfect information, localized tastes, and distribution networks. It is also related to the work on border effect based on micro evidence (Coşar et al., 2018). Furthermore, there is a large body of literature on domestic fragmentation and trade friction within China (Young, 2000; Poncet, 2003; Ke, 2015; Tombe and Zhu, 2019; Jiang and Mei, 2020). We introduce consumers' home-biased preferences into the gravity approach, and our empirical results show that from the static perspective, taste bias as a result of domestic migration is an important part of the border puzzle in China, something that has been ignored by most previous studies.

The rest of this paper is organized as follows. Section 2 presents the theoretical model and our key predictions. Section 3 introduces the data, specification, and main empirical results. Section 4 presents estimates of home-bias gravity with assimilation. Section 5 presents the counterfactual experiments. Section 6 concludes.

## 2 Model

We present how to incorporate home-biased preferences into the trade framework in Section 2.1, and derive the home-biased gravity equation in Section 2.2.

## 2.1 Preferences

The country consists of $N$ regions, indexed by $i$ as exporter and $n$ as importer. Each region specialized production of a distinct variety. Let $p_{ni}$ be the price in region $n$ of the good imported from region $i$. The iceberg trade cost of exporting from $i$ to $n$ is $t_{ni}$. No-arbitrage condition implies that $p_{ni} = p_{ii}t_{ni}$. There are $N$ ethnic groups of consumers, indexed by $h$ as the ethnic origin. Each ethnic group is countrywide distributed and thus each region is populated by $N$ ethnic groups. Note that $n, i, h = 1, 2, ..., N$. Let $s_{nh}$ be the migrant population share in region $n$ from region $h$ if $h \neq n$, and $s_{nn}$ be the native population share in region $n$. Then $\sum_{h=1}^{N} s_{nh} = 1$ holds for all $n$.

Consumers are assumed to have the Translog preference introduced by Diewert (1976), which can be rationalized as a second-order approximation of an arbitrary expenditure system. Specifically, any consumer who is from region $h$ and resides at region $n$ has expenditure functions given in logarithmic form as

$$\ln e_n^h = \ln Q_n^h + \ln u_n^h, \tag{1}$$

where $e_n^h$ is the minimum expenditure at which the consumer can obtain utility $u_n^h$ given prices $p_{ni}$. The price index $\ln Q_n^h$ is given in logarithmic form as

$$\ln Q_n^h = \sum_{i=1}^{N} \alpha_i^h \ln p_{ni} + \frac{1}{2} \sum_{i=1}^{N} \sum_{i'=1}^{N} \gamma_{ii'} \ln p_{ni} \ln p_{ni'}. \tag{2}$$

Applying the Shephard's lemma and differentiating the expenditure function with respect to log price $p_{ni}$ generates the expenditure share in good (produced by region) $i$ for consumer (of ethnic group) $h$ at region $n$ equal to

$$w_{ni}^h = \alpha_i^h + \sum_{i'=1}^{N} \gamma_{ii'} \ln p_{ni'}, \tag{3}$$

for $i = 1, ..., N$. We assume that all shares are non-negative.

These expenditure shares have some nice features. First, $\alpha_i^h$ is a taste parameter of consumer $h$ for the good $i$ which shifts the expenditure share independently from the prices and income. Second, $\gamma_{ii'}$ is the cross-price elasticity for good $i$ when $i' \neq i$ and the own-price elasticity for good $i$ when $i' = i$.[2] To satisfy homogeneity of degree one, the parameters are constrained by

[2]Note that $\gamma_{ii'}$ are semi-elasticities since they relate expenditure shares to logs of prices, but we refer to them as elasticities to save notation. Actually the price elasticities are $-1 + \gamma_{ii'}/w_{ni}^h + w_{ni}^h$.

$\alpha_i^h \in [0,1]$, $\sum_{i=1}^N \alpha_i^h = 1$ for all $h$, and $\sum_{i=1}^N \gamma_{ii'} = 0$ for all $i'$. Symmetry is imposed to satisfy the Young's Theorem, $\gamma_{ii'} = \gamma_{i'i}$. Concavity is imposed by the requirement that $\{\gamma_{ii'}\}$ is negative semi-definite. When $\gamma_{ii'} = 0$ for all $i, i'$ and $\phi_i = 0$ for all $i$, the preference becomes the Cobb-Douglas preference.

In addition, we let all goods enter 'symmetrically' in the $\gamma_{ii'}$ coefficients. Following Feenstra (2003), we therefore impose the additional restrictions:

$$\gamma_{ii'} = \gamma(-\delta_{ii'} + \frac{1}{N}), \tag{4}$$

where the Kronecker delta $\delta_{ii'}$ is 1 when $i = i'$ and 0 otherwise. Specialization (4) satisfies all the general restrictions.[3] .

In equation (3), $\alpha_i^h$ is a taste parameter for good $i$ but heterogeneous across consumers. It is unvarying to where consumers consume but varies to where they are originally from. The explanation is the habit formation in tastes proposed by Atkin (2013), who argues that adults favor foods consumed as a child, and the preferences developed in childhood persist into adulthood.[4] Therefore, consumers from the same ethnic group share the same tastes regardless of where they live.[5] This leads to an important factor in consumer tastes, home bias. A large amount of empirical work using disaggregated level data provides evidence of its existence. For any consumer, she has a biased taste toward the goods produced by her home region (where she is originally from) whatever her host region (where she resides and consumes) is. In other words, consumers are biased toward the goods with the same origin of themselves all else equal regardless of where they are located.

First, we assume an unbiased consumer with tastes for good $1, 2, ..., N$ as $\{\alpha_1, \alpha_2, ..., \alpha_N\}$, and $\sum_{i=1}^N \alpha_i = 1$. Similarly to Zhang (2020), we set ethnic consumer $h$'s tastes as

$$\alpha_i^h = (1-\theta)\alpha_i + \theta\delta_i^h, \tag{5}$$

where the Kronecker delta $\delta_i^h$ is 1 when $i = h$ meaning home goods, and 0 otherwise, and $\theta \in [0,1]$

[3]This special case is followed by Novy (2013) and Fajgelbaum and Khandelwal (2016). Recently, Anderson and Zhang (2020) and Zhang (2020) were begun together to generate asymmetric price elasticities.

[4]He terms this process habit formation and provides ample evidence in the psychology and nutrition literature.

[5]Contrary to Atkin (2013), in this paper, tastes are assumed to be exogenous.

is the *taste shifter* which measures how biased ethnic consumer's taste is toward her home good. This specialization satisfies conditions $\sum_{i=1}^{N} \alpha_i^h = 1$ for all $h$. For any consumer, she raises her taste for home good by a shifter $\theta$, while reduces those for all the other goods by $1-\theta$ (as well as the home good to make sure taste aggregates equal to 1). When $\theta = 0$, ethnic consumers are unbiased. All consumers are homogeneous in preference, which is exactly the same to the case with a representative consumer. When $\theta = 1$, the consumers have fully biased tastes, i.e. 1 for home good while 0 for all the others. Consumers are heterogeneous in terms of tastes. Thus the taste shifter $\theta$ is the measurement of the taste biases. Now we rewrite the expenditure equations (3) as

$$w_{ni}^h = (1-\theta)\alpha_i + \theta\delta_i^h - \gamma \ln\left(\frac{p_{ni}}{\bar{p}_n}\right), \tag{6}$$

where average price in market $n$ is $\ln \bar{p}_n = \frac{1}{N}\sum_{i=1}^{N} \ln p_{ni}$. And $-\gamma$ is the semi price elasticity.

A different way to interpret the taste shifter $\theta$ is the fraction of fully biased consumers. In this case, we assume there are two types of consumers in each ethnic group. One is a cosmopolitan who is unbiased, and the other one is a parochial who spends all on home products while zero on the others given all else equal. If their population shares are $1-\theta$ and $\theta$ respectively, then the average taste of ethnic group $h$ for home good is $(1-\theta)\alpha_h + \theta * 1$, and tastes for the other goods is $(1-\theta)\alpha_i + \theta * 0$, which are exactly the same as equation (5). Thus $\theta$ could be interpreted as the parochial population share of each ethnic group, and they measure how ethnicities are biased. When $\theta = 0$, ethnic group are all cosmopolitans. When $\theta = 1$, ethnic group are all parochial.

### 2.2 Home-biased Gravity

In this part, we first aggregate individual expenditure shares across all ethnic groups within each market (region), and then derive the gravity equation for each region pair that depends on the demand parameters.

We assume that ethnic groups within a region share the same income. Aggregating individual expenditure shares across ethnic groups gives the market import share of region $n$ from region $i$ by $w_{ni} = \sum_{h=1}^{N} s_{nh} w_{ni}^h$, i.e.,

$$w_{ni} = (1-\theta)\alpha_i + \theta s_{ni} - \gamma \ln\left(\frac{p_{ni}}{\bar{p}_n}\right). \tag{7}$$

Equivalently, equation (7) could be taken as the expenditure shares of a "representative" consumer in market $n$ who is a "mixed-blood" with all ethnic origins. Since ethnic compositions differ, the representative consumers in different markets have different demand.

Next, we let $X_{ni}$ be the trade flow from region $i$ to region $n$, and let $E_n$ be the total expenditure of region $n$. Denote $w_{ni}$ as the share of aggregate expenditures in region $n$ devoted to goods from region $i$, i.e.,

$$w_{ni} = \frac{X_{ni}}{E_n}. \tag{8}$$

Then equation (7) could be rewritten as

$$\frac{X_{ni}}{E_n} = [(1-\theta)\alpha_i - \gamma \ln \left(\frac{p_{ii}}{\bar{p}}\right)] + \theta s_{ni} - \gamma \ln \left(\frac{t_{ni}}{\bar{P}_n}\right), \tag{9}$$

where $\ln \bar{p} = \frac{1}{N}\sum_{i=1}^{N} \ln p_{ii}$ and $\ln \bar{P}_n = \frac{1}{N}\sum_{i=1}^{N} \ln t_{ni}$. Income of region $i$ equals the sum of sales to every region,

$$Y_i = \sum_{n=1}^{N} X_{ni}, \tag{10}$$

where $Y_i$ measures region $i$'s total income. Substitute equation (9) into equation (10) and divide through by $Y$ which is the total country income. Subtract the resulting expression from equation (9) and simplify by canceling term $[(1-\theta)\alpha_i - \gamma \ln \left(\frac{p_{ii}}{\bar{p}}\right)]$. Then we term the resulting expression as the *home-biased gravity*

$$\frac{X_{ni}}{E_n} - \frac{Y_i}{Y} = \theta(s_{ni} - \Pi_i^s) - \gamma \ln \left(\frac{t_{ni}}{\Pi_i P_n}\right), \tag{11}$$

where

$$\ln \Pi_i = \sum_{n=1}^{N} \left(\frac{E_n}{Y}\right) \ln t_{ni}, \ \ln P_n = \frac{1}{N}\sum_{i=1}^{N} \ln \left(\frac{t_{ni}}{\Pi_i}\right), \tag{12}$$

and

$$\Pi_i^s = \sum_{n=1}^{N} \left(\frac{E_n}{Y}\right) s_{ni}. \tag{13}$$

On the left hand side, $\frac{X_{ni}}{E_n} - \frac{Y_i}{Y}$ is the deviation of bilateral trade per unit of $n$'s expenditure from its frictionless level $\frac{Y_i}{Y}$. There are two terms on the right hand side, which capture the *taste effect* and *price effect*, respectively. The second term, $-\gamma \ln \left(\frac{t_{ni}}{\Pi_i P_n}\right)$, is the effect of relative bilateral trade resistance from origin $i$ to destination $n$, where $\ln \Pi_i$ and $\ln P_n$ are the outward and inward multilateral resistances in log, respectively. This term is very similar to the CES structural gravity in Anderson and van Wincoop (2003).

The first term, $\theta_i(s_{ni} - \Pi_i^s)$, captures the effect of the relative trade "promoter" that stimulates bilateral trade via bilateral migration from origin $i$ to destination $n$. The intuition is that migrants raise the average market taste for their home region's products. This is because migrants with taste biases spend larger shares of their income on home-produced products, thereby increasing the market shares of the host region from their home regions. The magnitude of the promoter is increasing in how large the bilateral migrant share $s_{ni}$ is, as well as how large the taste shifter $\theta$ is. We refer to $\Pi_i^s$ as the outward "multilateral promoter" that summarizes the average trade promoters between a region and its trading partners. When consumers are unbiased, i.e., $\theta = 0$, the trade promoter is shut down, and this gravity equation corresponds to the translog gravity in Novy (2013).

### 2.3 Extension: Taste Bias with Assimilation

Migrants usually assimilate. In other words, migrant tastes depend on origins as well as destinations. Some studies find that consumers eventually converge to the consumption patterns of their new place of residence.[6]

In this section, we introduce "local bias" in addition to "home bias" into the model to allow migrants to assimilate. We redefine the taste parameter used in equation (3). On the one hand, a consumer has a biased taste toward goods produced by her home region (where she is originally from). On the other hand, her taste is also biased toward the locally-produced goods (where she is living). Specifically, we assume ethnic consumer (ethnic group) $h$'s tastes at location (region) $n$ to be

$$\alpha_{ni}^h = (1 - \theta - \phi)\alpha_i + \theta\delta_i^h + \phi\delta_{ni}, \tag{14}$$

where the Kronecker delta $\delta_{ni}$ is 1 when $i = n$, meaning local goods, and 0 otherwise, and $\phi \in [0, 1]$ is the *local shifter* which measures how the tastes of the consumers at each location is toward to local good. This specialization still satisfies conditions $\sum_{i=1}^N \alpha_{ni}^h = 1$ for all $h$ and all $n$. For any

[6] Bronnenberg et al. (2012) use micro-level data to infer that approximately 40% of the geographic variation in market shares is attributable to persistent brand preferences. They also show that their findings strongly reject the hypothesis that all that matters is where consumers lived in childhood: consumers who move after age 25 still eventually converge to the consumption patterns of their new place of residence.

consumer $h$ at location $n$, she raises her taste for home good $h$ by a shifter $\theta$, for local good $n$ by a shifter $\phi$, while reduces those for all the other goods by $1-\theta-\phi$. Thus the local shifter $\phi$ is the measurement of the local biases across locations. Now we rewrite the expenditure equations (3) as

$$w_{ni}^h = (1-\theta-\phi)\alpha_i + \theta\delta_i^h + \phi\delta_{ni} - \gamma \ln\left(\frac{p_{ni}}{\bar{p}_n}\right). \tag{15}$$

Aggregating the individual expenditure shares across ethnic groups gives the market import share of region $n$ from region $i$ by $w_{ni} = \sum_{h=1}^{N} s_{nh} w_{ni}^h$, i.e.,

$$w_{ni} = (1-\theta-\phi)\alpha_i + \theta s_{ni} + \phi\delta_{ni} - \gamma \ln\left(\frac{p_{ni}}{\bar{p}_n}\right). \tag{16}$$

Then equation (16) could be rewritten as

$$\frac{X_{ni}}{E_n} = [(1-\theta-\phi)\alpha_i - \gamma \ln\left(\frac{p_{ii}}{\bar{p}}\right)] + (\theta s_{ni} + \phi\delta_{ni}) - \gamma \ln\left(\frac{t_{ni}}{\bar{P}_n}\right). \tag{17}$$

Substitute equation (17) into equation (10) and divide through by $Y$ which is the total country income. Subtract the resulting expression from equation (17) and simplify by canceling term $[(1-\theta-\phi)\alpha_i - \gamma \ln\left(\frac{p_{ii}}{\bar{p}}\right)]$. Then the home-biased gravity with local bias becomes

$$\frac{X_{ni}}{E_n} - \frac{Y_i}{Y} = \theta(s_{ni} - \Pi_i^s) - \gamma \ln\left(\frac{t_{ni}}{\Pi_i P_n}\right) + \phi\left(\delta_{ni} - \frac{E_i}{Y}\right), \tag{18}$$

where $\ln \Pi_i$, $\ln P_n$ and $\Pi_i^s$ are the same as the baseline model.

On the left-hand side, $\frac{X_{ni}}{E_n} - \frac{Y_i}{Y}$ is the deviation of the bilateral trade per unit of $n$'s expenditure from its frictionless level $\frac{Y_i}{Y}$. Again, the first term, $\theta(s_{ni} - \Pi_i^s)$, captures the effect of the relative trade "promoter" that stimulates bilateral trade via bilateral migration from origin $i$ to destination $n$. When everybody is unbiased, i.e., $\theta = 0$, the trade promoters are shut down. This taste term is exactly the same as it is in the home-biased gravity without assimilation which is expressed by equation (11).

However, the home-biased gravity with assimilation has an additional term, $\phi\left(\delta_{ni} - \frac{E_i}{Y}\right)$. It captures the *local effect*, in addition to the taste effect and price effect in the baseline model. It measures the size of relative trade "promoter" that stimulates internal (local) trade due to the local bias in taste. Specially, when $i = n$, then $\delta_{ni} = 1$, equation (18) indicates the internal (domestic) trade for region $n$, in which the internal trade "promoter" is $\phi$. When $i \neq n$, then $\delta_{ni} = 0$, it indicates the inter-regional trade, in which the "promoter" becomes zero. This implies that internal trade is larger than inter-regional trade by a local bias, $\phi$, all else equal. The intuition is that the consumption pattern of the migrants converges to that of the native people. When consumers everywhere have no local bias, i.e., $\phi = 0$ for all $n$, the internal trade promoter is shut down, and this gravity equation corresponds to that in the baseline model, i.e., equation (11).

## 3 Baseline Estimation

In this section, we estimate the home-biased gravity derived in Section 2. Section 3.1 and Section 3.2 describe the data and the specification, respectively. Section 3.3 presents the main results by estimating the baseline model, i.e., equation (11). Section 3.4 discusses the disaggregate results.

### 3.1 Data

To estimate the home-biased gravity, we merged datasets on bilateral migration, trade, and gravity. First, the migration data are from the China Population Census (CPC), which presents estimates of total migrant stock by origin and destination provinces. For each individual, we define her Hukou-registered place (province) as the origin region, and her resident place (province) as the host region, which delivers all bilateral migrant population among provinces.[7] Hukou is an official document issued by the Chinese government, certifying that the holder is a legal resident of a particular area. This household registration record officially identifies a person as a permanent resident of an area. We pick the two most recent waves of CPC are in 2000 and 2010.[8]

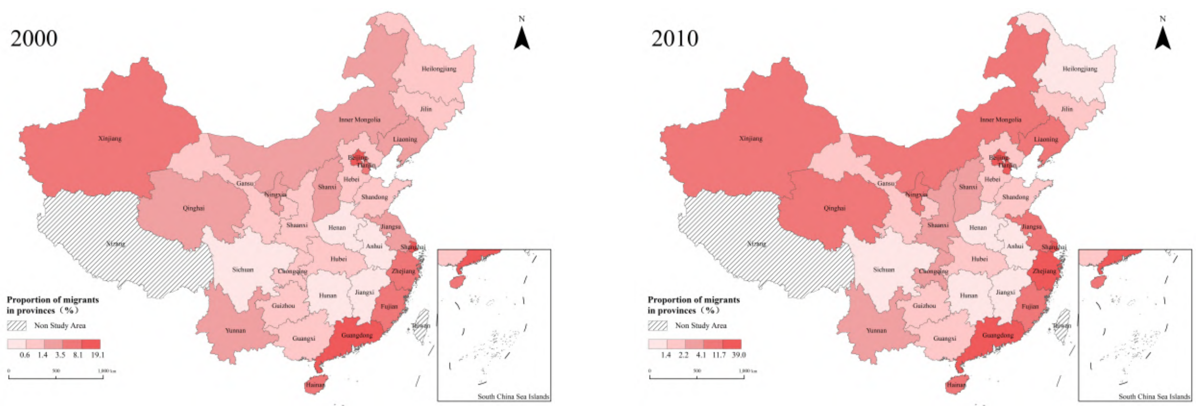
China is experiencing the largest internal migration in human history. Panel A and panel B in Figure 1 show the spatial distribution of migrants in the 30 provinces in 2000 and 2010, respectively.[9] In general, there are large regional differences in terms of migrant stock in China.

[7] Research on migration is often hampered by data limitations. Large-scale individual-level surveys do not ask about migrants' legal status, and government records on legal permanent residents are presented as aggregate tabulations with no individual-level information. (Caballero et al., 2018; Desmond and Kubrin, 2009; Lyons et al., 2013; Ravuri, 2014)

[8] The results of the 2020 CPC have been released recently, but the corresponding trade data are not available yet.

[9] The data for Tibet are not available.

For example, the five provinces with the largest inward migrant share in 2000 were Shanghai, Beijing, Guangdong, Zhejiang, and Xinjiang, suggesting that the main population inflows were concentrated in the eastern coastal areas. Conversely, the middle, western, and northeastern regions experienced the greatest population outflows. It can be seen from panel B in Figure 1 that these population migration trends were even more prominent in 2010. Figure 2 presents a more detailed view of the inter-provincial linkages established by migrants in 2000 and 2010. In general, the shorter the distance between the two provinces, the stronger the connection in terms of migration. For example, Guangdong and Hunan, Jiangsu and Anhui, and Beijing and Hebei are strongly connected.



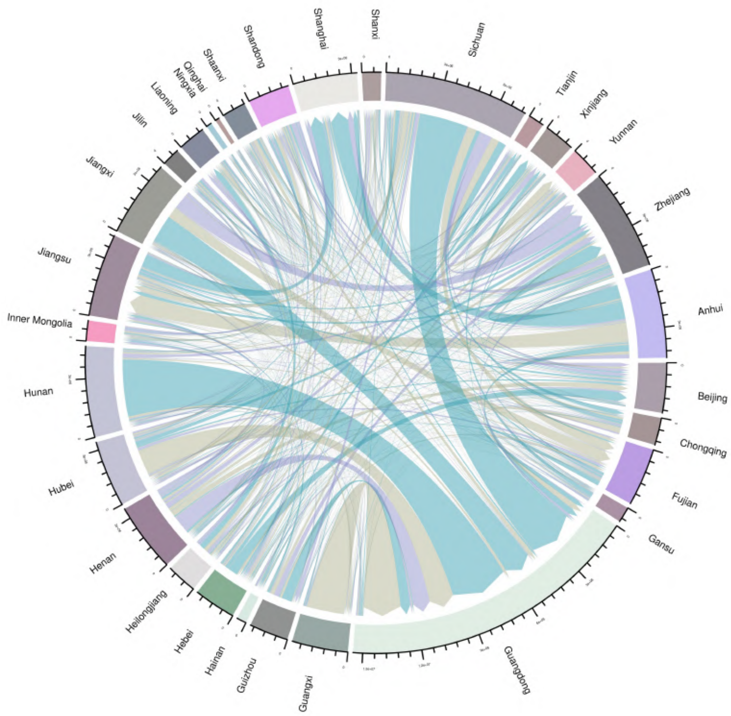

(a) Panel A: migrant share in 2000

(b) Panel B: migrant share in 2010

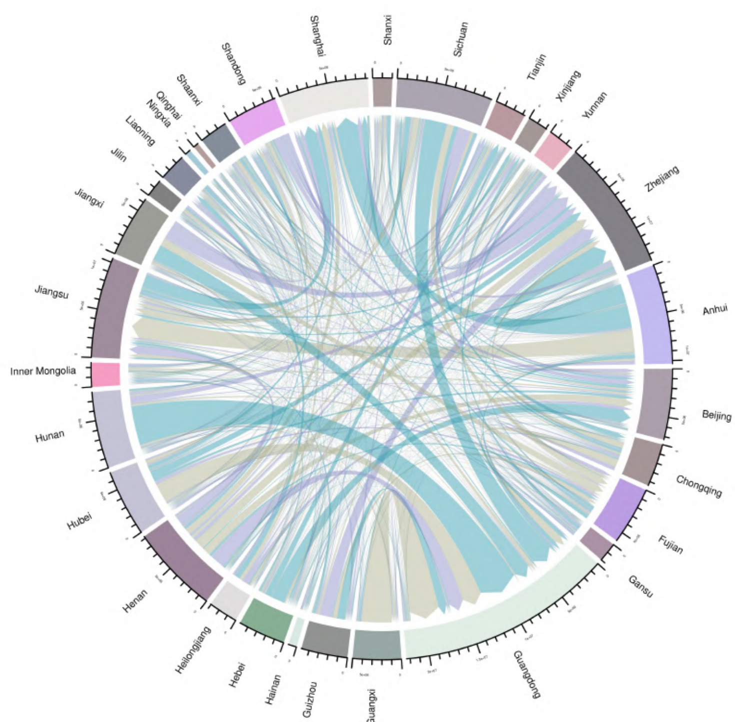
Figure 1: Spatial distribution of migration in China

Second, whereas the focus of this paper is on the relationship between migration and trade across provinces in China, the fact that there is no official record of inter-provincial trade flows poses a major challenge. Previous studies mainly approximate domestic trade across Chinese provinces by the volume of railway cargo, which is published annually in the *China Railway Yearbook* (Jiang and Mei, 2020). However, the volume of railway cargo does not fully account for the structure of trade flows.[10] Instead, we use the multi-regional input-output table (MRIO) constructed and used by Li (2010) and Mi et al. (2017). It records bilateral inter-provincial trade flows as well as production data for each province (excluding Tibet) and each sector in 2002, and 2012. We use the 2002 and 2012 trade flow to match the 2000 and 2010 migration flow, respectively.

[10]The volume of railway freight as a proportion of the total national freight volume is small, for example, 9.7% in 2016. Furthermore, the *China Railway Yearbook* only reports the total weight of the cargo, rather than the value of the cargo.




(a) Panel A: province linkages by migrants in 2000




(b) Panel B: province linkages by migrants in 2010

Figure 2: Inter-provincial migration in China

*Notes:* The arrows indicate the directions of migration.

Thus the migration is a two-year-lagged independent variable approximation in our paper. This approximation is appropriate because the migrant stock proportion is quite persistent in most regions over the years.

Third, data on distances between provinces and common border are obtained from the Database of Global Administrative Areas, which measures the bilateral distance among provinces in China based on the shortest railway route between their respective capital cities. In previous studies of inter-provincial trade in China, Wei (1996) uses one-quarter of the distance between the two capital cities as the inter-provincial distance, while Poncet (2003) treats the inter-provincial distance as a function of the province's area. We find that the measurement given by that both of these methods provide similar results in terms of estimating the home-biased gravity.

### 3.2 Specification

One of the related concerns is that the coefficient of migrant shares may capture the effects of factors other than taste bias. The taste shifters may be overestimated because migration increases trade not only through the preference channel but also through the trade cost channel. As Rauch (2001) argues, migration builds up business, and social networks across borders can help to alleviate problems related to contract enforcement and provide information about trading opportunities. Rauch and Trindade (2002) use international trade data and find that ethnic Chinese networks, as proxied by the product of the ethnic Chinese population shares, significantly increased bilateral trade, with a greater effect on differentiated products than on homogeneous products. In order to control for the network channel, we follow Zhang (2020) to construct a network variable

$$Network_{ni} = \sum_{j=1}^{N} s_{nj} s_{ij}, \tag{19}$$

where $s_{nj,t}$ denotes the migrant stock share from home province $j$ to host province $n$ in time (wave) $t$. This variable represents the probability that, if we select an individual at random from each province, they will have a connection defined as common regional origin. By construction, it is symmetric for province pairs. When $j = CHN$, it is an *indirect* link identical to that in Rauch and Trindade (2002), and when $j = n, i$, it is a summation of two *direct* links, similar to that in Combes et al. (2005). Importantly, our measure extends these links to a summation of all indirect and both direct links, and thus captures the network effect as much as possible.

Thus we can proxy bilateral trade costs with bilateral observables. Specifically, $\ln t_{ni,t} = \rho \ln d_{ni} + \rho^b border_{ni} + \rho^n network_{ni,t} + \varepsilon_{ni,t}$, where $d_{ni}$ is the bilateral distance between origin $i$ and destination $n$, $border_{ni}$ is a dummy variable that equals to 1 if the two provinces have common borders, and $\varepsilon_{ni,t}$ is idiosyncratic errors for trade costs. Then the specification of the home-biased gravity by equation (11) is

$$\frac{X_{ni,t}}{E_{n,t}} = \theta s_{ni,t} + b \ln d_{ni} + b^b border_{ni} + b^n network_{ni,t} + f_{i,t} + g_{n,t} + \epsilon_{ni,t}. \tag{20}$$

where $f_{i,t} = Y_{i,t}/Y_t - \theta \Pi^s_{i,t} + \gamma \ln \Pi_{i,t}$, and $g_{n,t} = \gamma \ln P_{n,t}$. Here, $X_{ni,t}$ represents trade flows from origin province $i$ to destination province $n$. $E_{n,t}$ denotes the total expenditure of destination province $n$.Note the total expenditure is a summation of purchases from the local, all the other provinces and abroad. Hence, $X_{ni,t}/E_{n,t}$ is bilateral inter-provincial trade share. Notably, we use trade volume in aggregate manufacturing goods in the baseline regression. The multilateral resistances and promoter are included in $f_{i,t}$ and $g_{n,t}$, that could be controlled by origin-time and destinations-time fixed effects, respectively.

Another concern in our estimation is endogeneity. The migrant share may be correlated with trade shocks. When this is the case, the OLS estimation leads to inconsistent results. Such endogeneity bias can arise from two sources. The first is reverse causality. It is possible that a positive shock to the value of bilateral trade between the two provinces leads to more migration between them. For example, more people might migrate across provinces to do business associated with more inter-provincial trade. The second is omitted variables. Migration may be correlated with unobserved factors that also affect trade, such as the trading partners' cultural similarity or bilateral economic policies.[11]

To address the possible endogeneity problem, we construct an instrumental variable by the exogenous migration push factor and conduct a two-stage least square estimation. We instrument the migrant share by the interaction of natural disasters of origin and the distance to the destination. Llull (2018a) proposes a novel approach by using the exogenous variation obtained from the interaction of push factors, distance, and skill cells to study the effect of migration on wages.

[11] There is a large literature that discusses the factors that affect migration. See more detailed discussion in De (2010).

Four push factors are considered in his paper: wars, political regime changes, natural disasters, and economic variables.[12] We use natural disasters in our paper because the other three variables might either affect or be affected by trade. The natural disaster of the migrant-sending province is origin-specific, and is thus captured by the origin fixed effect in the regression, implying that the natural disaster alone does not have enough variation. Thus we interact the natural disaster with distance. The underlying assumption is that the promoting effect of the push factor differs across distances. It is reasonable because the natural disaster pushes more migrants to neighboring provinces than to provinces that are further away. Natural disasters, calculated from the China Civil Affairs' Statistical Yearbook, are measured as the number of people requiring immediate assistance, having been displaced or evacuated following natural disasters such as floods, hailstorms, typhoons, earthquakes, freezing weather, or snowstorm. We use the multiplicative inverse of the log distance to make the instrument positively correlated with the migrant share.

The estimates of taste differences might be biased due to the good composition. For example, the migrants tend to be less biased if their origin province is more specialized in goods that are less tradable, or that serve as intermediate inputs and are not directly sold to consumers. To fix this composition effect, we also estimate sectoral (and industrial) gravity equations for 12 manufacturing industries, and the final and intermediate good sectors, as well as food and service sectors using more disaggregated data. Disaggregated estimation enables our results to be compared with the micro-level studies on home preferences in the literature. Specifically, we estimate

$$\frac{X_{ni,t}^k}{E_{n,t}^k} = \theta^k s_{ni,t} + b^k \ln d_{ni} + b^{kb} border_{ni} + b^{kn} network_{ni,t} + f_{i,t}^k + g_{n,t}^k + \epsilon_{ni,t}^k, \tag{21}$$

where all variables with a superscript $k$ are defined in the same way to those without any superscript but in good class $k$.[13] We run the regressions separately using corresponding data and obtain the taste bias estimates across in all good classes.

[12] Llull (2018b) builds a structural model in which immigration and wage are both endogenously determined.

[13] Good classes include final and intermediate goods, food, manufacturing, and services, and industrial goods.

### 3.3 Main Results

Table 1 reports the descriptive statistics for each wave, as well we the full sample. The regression results of our baseline analysis are shown in Table 2, in which OLS estimates are presented in columns (1) to (3). Specifically, column (1) shows the coefficient of migrant share without any gravity variables or fixed effects. The coefficient of migrant share is estimated to be 0.416, which is significantly different from zero, suggesting the existence of consumer taste bias. Then, gravity variables are introduced into our model, and time-variant multilateral resistance by origin-year and destination-year fixed effects are controlled for in the results presented in column (2). The network effect is further included in column (3). These results show that the distance elasticity is significantly negative, suggesting that increasing distance reduces trade flows between provinces. The coefficient of common borders is also significant, confirming the findings of previous studies. The migrant share coefficient is not significant in either column (2) or column (3), partially due to the endogeneity problem.

Table 1: Descriptive statistics

| Variable | Wave 1 | | | | Wave 2 | | | | Total | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Sd | Min | Max | Mean | Sd | Min | Max | Mean | Sd | Min | Max | N |
| Migrant share | 0.001 | 0.004 | 0.000 | 0.063 | 0.003 | 0.008 | 0.000 | 0.113 | 0.002 | 0.006 | 0.000 | 0.113 | 1740 |
| Agricultural trade share | 0.008 | 0.026 | 0.000 | 0.357 | 0.006 | 0.012 | 0.000 | 0.115 | 0.007 | 0.020 | 0.000 | 0.357 | 1740 |
| Manufacturing trade share | 0.007 | 0.015 | 0.000 | 0.168 | 0.012 | 0.015 | 0.000 | 0.127 | 0.009 | 0.016 | 0.000 | 0.168 | 1740 |
| Service trade share | 0.007 | 0.010 | 0.000 | 0.077 | 0.005 | 0.010 | 0.000 | 0.121 | 0.006 | 0.010 | 0.000 | 0.121 | 1740 |
| Total trade share | 0.007 | 0.010 | 0.000 | 0.099 | 0.008 | 0.010 | 0.000 | 0.095 | 0.008 | 0.010 | 0.000 | 0.099 | 1740 |
| Intermediate good share | 0.007 | 0.015 | 0.000 | 0.155 | 0.010 | 0.014 | 0.000 | 0.134 | 0.009 | 0.014 | 0.000 | 0.155 | 1740 |
| Final good share | 0.008 | 0.022 | 0.000 | 0.329 | 0.016 | 0.024 | 0.000 | 0.219 | 0.012 | 0.023 | 0.000 | 0.329 | 1740 |
| log Distance | 7.019 | 0.607 | 4.575 | 8.150 | 7.019 | 0.607 | 4.575 | 8.150 | 7.019 | 0.606 | 4.575 | 8.150 | 1740 |
| Common border | 0.151 | 0.358 | 0.000 | 1.000 | 0.151 | 0.358 | 0.000 | 1.000 | 0.151 | 0.358 | 0.000 | 1.000 | 1740 |
| Network | 0.003 | 0.006 | 0.000 | 0.063 | 0.005 | 0.010 | 0.000 | 0.112 | 0.004 | 0.008 | 0.000 | 0.112 | 1740 |
| Nature disaster $\times$ Distance (IV) | 2.318 | 0.287 | 1.422 | 3.505 | 2.239 | 0.401 | 0.853 | 3.383 | 2.278 | 0.351 | 0.853 | 3.505 | 1740 |

*Notes:* Wave 1 represents the sample of migration data in 2000 and trade data in 2002, and Wave 2 represents the sample of migration data in 2010 and trade data in 2012.

To deal with the endogeneity problem, Table 2 also reports the 2SLS estimates in columns (4) to (6). Using the interaction between natural disasters in the home province and distance from

Table 2: Home-biased gravity estimates: baseline

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent variable: trade share | | | | | | |
| Migrant share | 0.416*** | 0.081 | 0.023 | 1.462*** | 0.712* | 0.460** |
| | (0.091) | (0.051) | (0.071) | (0.282) | (0.379) | (0.202) |
| Distance | | -0.011*** | -0.010*** | | -0.008*** | -0.010*** |
| | | (0.001) | (0.001) | | (0.001) | (0.001) |
| Common border | | 0.003*** | 0.003*** | | 0.003** | 0.003*** |
| | | (0.001) | (0.001) | | (0.001) | (0.001) |
| Network | | | 0.061 | | | -0.164 |
| | | | (0.052) | | | (0.110) |
| First stage | | | | | | |
| Nature disaster × distance (IV) | | | | 0.005*** | 0.021*** | 0.039*** |
| | | | | (0.001) | (0.004) | (0.003) |
| *F*-stats | | | | 65.454 | 8.243 | 33.562 |
| Origin-time fixed effects | | ✓ | ✓ | | ✓ | ✓ |
| Destination-time fixed effects | | ✓ | ✓ | | ✓ | ✓ |
| Method | OLS | OLS | OLS | 2SLS | 2SLS | 2SLS |
| Observations | 1740 | 1740 | 1740 | 1740 | 1740 | 1740 |

*Notes:* Table reports the estimates of the home-biased gravity equation (20) using bilateral manufacturing trade data. Robust standard errors in parentheses. Significance * .10, ** .05, *** .01.

the origin province to province as the instrumental variable, the coefficient of the migrant share becomes significant in each column. After introducing the gravity variables, network effect, and fixed effects, the coefficient of immigrant share is 0.46, significantly from zero, as shown in column (6), which is much larger than the OLS result shown in column (3), which confirms the existence of consumer taste bias. Specifically, a 1% increase of migrant share increases the expenditure of the home products by 0.46 percentage point. Equivalently, according to the definition of taste shifter in the model, this result implies that in each 1 dollar of the total expenditure, each consumer spends 46 cents more on products from her home province than other products, given all else equal. Dropping the network effect leads to an overestimation of the taste bias, as shown in column (5). Table 2 also shows the first-stage results corresponding to the second-stage coefficients. They use the same control-variable settings. The coefficients of all instrumental variables are significantly positive.[14] The exclusion F-statistics are reported and the instrument is not weak.

The results presented in Table 2 show that the coefficients of network is insignificant for both

[14]Note the inverse of the log distance is used, and thus the instrument is positively correlated to the migrant share.

the OLS method in column (3), and the 2SLS method in (6). This is different from previous studies on network promotion effects on trade, e.g., Rauch and Trindade (2002) and Combes et al. (2005). Because we use translog gravity in which the dependent variable is trade share, rather than trade volume used in classical CES gravity specification.[15] All of this evidence indicates that the network channel has a negligible effect on the trade share of aggregate manufacturing products among province pairs. Moreover, different from trade flow which is more likely to be affected by the network, trade share is much less since the network effect in trade flow is partially offset by the total flows that is also affected by the network effect.

### 3.4 Sectoral Results

Table 3: Home-biased gravity estimates: by sector

| Dependent variable: trade share | (1) Manufacturing | (2) Agricultural | (3) Service | (4) Manufacturing | (5) Agricultural | (6) Service |
|---|---|---|---|---|---|---|
| Migrant share | 0.712$^{*}$ | 0.976 | 0.272 | 0.460$^{**}$ | 0.782$^{**}$ | 0.289$^{**}$ |
| | (0.379) | (0.609) | (0.234) | (0.202) | (0.334) | (0.129) |
| Distance | -0.008$^{***}$ | -0.009$^{***}$ | -0.006$^{***}$ | -0.010$^{***}$ | -0.010$^{***}$ | -0.006$^{***}$ |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) | (0.000) |
| Common border | 0.003$^{**}$ | -0.003$^{*}$ | 0.002$^{**}$ | 0.003$^{***}$ | -0.003$^{*}$ | 0.001$^{**}$ |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.002) | (0.001) |
| Network | | | | -0.164 | -0.126 | 0.011 |
| | | | | (0.110) | (0.183) | (0.071) |
| First stage | | | | | | |
| Nature disaster × distance (IV) | 0.021$^{***}$ | 0.021$^{***}$ | 0.021$^{***}$ | 0.039$^{***}$ | 0.039$^{***}$ | 0.039$^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.003) | (0.003) | (0.003) |
| *F*-stats | 8.243 | 8.243 | 8.243 | 33.562 | 33.562 | 33.562 |
| Origin-time fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Destination-time fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Method | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS |
| Observations | 1740 | 1740 | 1740 | 1740 | 1740 | 1740 |
| $R^2$ | 0.530 | 0.290 | 0.559 | 0.561 | 0.295 | 0.557 |

*Notes:* Table reports the estimates of the home-biased gravity equation (21) using sectoral trade flow data. Robust standard errors in parentheses. Significance * .10, ** .05, *** .01.

To further investigate the taste biases, we also estimate the sectoral gravity using equation (21) across the manufacturing, agricultural, and service sectors, as well as the final and intermediate

[15] We also check the CES gravity specification with our data and find that the network variable significantly increases the trade volume, suggesting our measure of the network effect is valid.

good sectors, using more disaggregated data. First, we estimate the taste biases in the three sectors in Table 3. The results for the manufacturing trade are reported in columns (1) and (4), which are corresponding to columns (5) and (6) in our baseline regression Table 2, respectively. Compared with the manufacturing sector, the agricultural sector tends to be more favored by home consumers, while the service sector is much less biased. In other words, consumers have larger taste biases (0.782) in relation to agricultural products, while less unbiased (0.289) on services. This makes sense based on the taste habit formation assumption that food tastes are even more persistent. Next, we estimate the results for three types of goods in Table 4. The results for the aggregate trade are reported in columns (1) and (4), which are corresponding to columns (5) and (6) in our baseline regression Table 2, respectively. Trade share is more sensitive to migrant share for final goods (0.751) than for intermediate goods (0.403).

Table 4: Home-biased gravity estimates: by good

| Dependent variable: trade share | (1) Aggregate | (2) Intermediate | (3) Final | (4) Aggregate | (5) Intermediate | (6) Final |
|---|---|---|---|---|---|---|
| Migrant share | 0.712* | 0.613* | 1.120* | 0.460** | 0.403** | 0.751** |
| | (0.379) | (0.349) | (0.596) | (0.202) | (0.187) | (0.319) |
| Distance | -0.008*** | -0.010*** | -0.009*** | -0.010*** | -0.011*** | -0.012*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Common border | 0.003** | 0.003*** | 0.003* | 0.003*** | 0.003*** | 0.003* |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.002) |
| Network | | | | -0.164 | -0.137 | -0.240 |
| | | | | (0.110) | (0.102) | (0.175) |
| First stage | | | | | | |
| Nature disaster × distance (IV) | 0.021*** | 0.021*** | 0.021*** | 0.039*** | 0.039*** | 0.039*** |
| | (0.004) | (0.004) | (0.004) | (0.003) | (0.003) | (0.003) |
| *F*-stats | 8.243 | 8.243 | 8.243 | 33.562 | 33.562 | 33.562 |
| Origin-time fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Destination-time fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Method | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS |
| Observations | 1740 | 1740 | 1740 | 1740 | 1740 | 1740 |
| $R^2$ | 0.530 | 0.537 | 0.478 | 0.561 | 0.561 | 0.506 |

*Notes:* Table reports the estimates of the home-biased gravity equation (21) using bilateral sectoral trade flow data. Robust standard errors in parentheses. Significance * .10, ** .05, *** .01.

Table 5 shows consumers' taste bias for various industrial goods. They are estimated by the industry-by-industry results. The results for the total manufacturing trade are reported in column

(1), which are corresponding to (6) in our baseline regression Table 2. Consumers' tastes are more biased for textiles, petroleum, chemicals, metallurgy, machinery, and food & tobacco, while they are unbiased in metals, transport, printing, electrical equipment, wood, and instrument & meter. More disaggregated estimation reduces the composition effect to some extent, but not all since disaggregation can never be as fine as reality.

Table 5: Home-biased gravity estimates: by industrial goods

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent variable: trade share | Manufacturing | Textiles | Petroleum | Chemicals | Metallurgy |
| Migrant share | 0.460** | 1.628*** | 1.304*** | 0.848*** | 0.701* |
| | (0.202) | (0.401) | (0.502) | (0.274) | (0.385) |
| Observations | 1740 | 1740 | 1740 | 1740 | 1740 |
| $R^2$ | 0.530 | 0.516 | 0.129 | 0.383 | 0.362 |
| | (6) | (7) | (8) | (9) | (10) |
| Dependent variable: trade share | Machinery | Food & tobaccos | Instrument & meter | Wood | Electrical equipment |
| Migrant share | 0.677** | 0.469** | 0.745 | 0.577 | 0.053 |
| | (0.317) | (0.225) | (0.558) | (0.353) | (0.396) |
| Observations | 1740 | 1740 | 1740 | 1740 | 1740 |
| $R^2$ | 0.443 | 0.571 | 0.674 | 0.473 | 0.612 |
| | (11) | (12) | (13) | | |
| Dependent variable: trade share | Printing | Transport equipment | Metals | | |
| Migrant share | -0.186 | -0.201 | -0.569 | | |
| | (0.306) | (0.341) | (0.357) | | |
| Observations | 1740 | 1740 | 1740 | | |
| $R^2$ | 0.508 | 0.530 | 0.325 | | |

*Notes:* Table reports the estimates of the home-biased gravity equation (21) using bilateral industrial trade flow data. All the results in the table are estimated by 2SLS method. Gravity variables, network, original-time fixed effects and destination-time fixed effects are included. Robust standard errors in parentheses. Significance * .10, ** .05, *** .01.

## 4 Estimation with Assimilation

Migrants can become assimilated into the host province. On the one hand, migrants may shift their tastes in accordance with the dominant society and consume products manufactured in the host province. On the other hand, influenced by the border effects related to internal trade, host provinces may alter their commodity production to cater to migrants rather than importing goods from their provinces of origin. Ignoring assimilation might lead to overestimation of taste bias. Thus this section estimates the results given by the home-biased gravity model with assimilation derived in Section 2.3.

To estimate the gravity equation with assimilation in (18), follow the methodology in the baseline estimation. Note that $\phi E_i/Y$ is captured by the exporter-specific fixed effects. $\delta_{ni}$ is a dummy for internal trade. In this case, the assimilation effect is fully captured by the internal trade dummy variable. Note that the internal trade dummy also captures all the other unobserved trade cost across region borders, e.g., local protectionism (Barwick et al., 2017). Thus the specification is the same as the baseline estimation. In other words, the specification of the home-biased gravity equation with assimilation effect is identical to that without assimilation (our baseline results) for inter-provincial trade, but includes an internal promoter only for intra-provincial trade. Specifically, the specification of the gravity equation (18) is

$$\frac{X_{ni,t}}{E_{n,t}} = \theta s_{ni,t} + b \ln d_{ni} + b^b border_{ni} + b^n network_{ni,t} + \phi\ internal_{ni} + f_{i,t} + g_{n,t} + \epsilon_{ni,t}. \tag{22}$$

In contrast to equation (20), the dummy variable *internal* here controls for the assimilation effect.

Thus we need to extend our sample by adding the observation of internal trade in each province, resulting in a sample size of $30 \times 30 \times 2 = 1800$. All the dependent variables need to be constructed for the additional observations of intra-provincial trade. Basically, all the control variables and the instrumental variables are constructed in the same way as our baseline regression. But one challenge is how to construct the distance measure for intra-provincial trade. Existing studies mainly use the distance between the two capital cities to measure the distance between the two corresponding provinces. Following this idea, we proxy the internal distance of each province by the distance between its largest two cities.[16] But this method does not work for the four directly controlled municipalities (DCMs) in China (Beijing, Shanghai, Tianjin, Chongqing). Instead, we approximately take each DCM as a disk (i.e., circle with interior) and measure its internal distance by its radius calculated from its area.[17]

The results with assimilation effect are reported in column (2) by the OLS method and column (4) by the 2SLS method in Table 6. For comparison, the results without assimilation effect are reported in column (1) by OLS method and column (3) by 2SLS method, which are corresponding

[16] The capital city of each province is usually one of its largest two cities.

[17] We also use the DCM-radius measure for all the other provinces as a robustness check and find that the estimation results are very similar.

Table 6: Home-biased gravity estimates: assimilation

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent variable: trade share | | | | |
| Migrant share | 0.023 | 0.304 | 0.460** | 0.357*** |
| | (0.071) | (0.189) | (0.202) | (0.094) |
| Distance | -0.010*** | -0.002 | -0.010*** | -0.002 |
| | (0.001) | (0.002) | (0.001) | (0.002) |
| Common border | 0.003*** | 0.008*** | 0.003*** | 0.008*** |
| | (0.001) | (0.003) | (0.001) | (0.000) |
| Network | 0.061 | 0.210* | -0.164 | 0.179*** |
| | (0.052) | (0.115) | (0.110) | (0.002) |
| Internal | | 0.178** | | 0.157*** |
| | | (0.082) | | (0.052) |
| First stage | | | | |
| Nature disaster × distance (IV) | | | 0.039*** | 0.028*** |
| | | | (0.003) | (0.002) |
| *F*-stats | | | 33.562 | 32330.61 |
| Origin-time fixed effects | ✓ | ✓ | ✓ | ✓ |
| Destination-time fixed effects | ✓ | ✓ | ✓ | ✓ |
| Method | OLS | OLS | 2SLS | 2SLS |
| Observations | 1740 | 1800 | 1740 | 1800 |
| $R^2$ | 0.571 | 0.945 | 0.561 | 0.945 |

*Notes:* Table reports the estimates of the home-biased gravity equation (22) after controlling for local bias. Robust standard errors in parentheses. Significance * .10, ** .05, *** .01.

to columns (3) and (6) in our baseline regression Table 2, respectively. The coefficients of the internal trade dummy are significantly positive, which suggests that consumers exhibit local bias (assimilating effect) in their tastes. Comparing columns (3) and (4), we find that after controlling for local bias in the gravity equation, the home taste shifter (migrant share coefficient) estimated is smaller, but still significantly positive. The difference (0.460-0.357 = 0.103) is smaller than one standard deviation of the coefficient (0.202).

Overall, compared with the baseline results, the taste bias estimated with assimilation is smaller. Without assimilation, the taste bias is overestimated. But the difference is moderate, which implies that the baseline estimation results are robust.

# 5 Counterfactuals

## 5.1 Variation Decomposition

The trade literature suggests that the home expenditure shares are much larger than the levels under hypothetical frictionless trades, which are equal to the region' production shares.[18] This implies a large home bias in trade (defined as *trade bias* hereafter). To study the trade bias in each province of China, we can write down the home share deviation based on the home-biased gravity equation (11) as

$$HB_n \equiv \frac{X_{nn}}{E_n} - \frac{Y_n}{Y} = \theta(s_{nn} - \Pi_n^s) - \gamma \ln\left(\frac{t_{nn}}{\Pi_n P_n}\right), \tag{23}$$

where $HB_n$ is defined as country $n$'s home bias in trade, or trade bias, that measures the deviation of home expenditure share from its frictionless level.[19] Equation (23) shows that trade bias can be explained by two separate effects: the taste effect and the price (trade cost) effect. Non-immigration regions have a larger trade bias than immigration regions.

Next, we examine the extent to which the trade bias can be explained by the taste effect and trade cost effect, respectively. We rewrite the equation (23) as

$$HB_n = \underbrace{\theta(s_{nn} - \Pi_n^s)}_{G_n^s} \underbrace{-\gamma \ln\left(\frac{t_{nn}}{\Pi_n P_n}\right)}_{G_n^t}.$$

Specifically, we term the taste effect and trade cost effect as $G_n^s$ and $G_n^t$, respectively. Then we calculate the contribution of each effect to the trade bias of each province by $Contribution_n^l = G_n^l / HB_n$ where $l = s, t$.

Column (1) in Table 7 lists the trade biases for 30 provinces, which are the deviations of the home shares from their frictionless levels in the manufacturing sector in 2012. The average bias is 56.7%. Columns (2)-(3) report the percentage contributions of the two effects on trade bias. On average, 38.8% of the market trade bias is explained by the trade cost, while 61.2% by the home-biased taste. In other words, the home-biased taste are the main causes of the trade bias, while

[18]In a standard benchmark in international trade, if foreign products were just as accessible and desirable as domestic ones, then under complete specialization each country would consume its expenditure share of every other country's production.

[19]In Anderson and Yotov (2010), they propose and estimate a similar index constructed home bias (CHB) by CES structural gravity: $CHB_n = \frac{\hat{X}_{ni}/Y_n}{Y_i/Y} = \left(\frac{\hat{t}_{nn}}{\hat{\Pi}_n \hat{P}_n}\right)^{1-\sigma}$.

Table 7: Counterfactuals: Contribution Decomposition

| Province | Share trade bias | Contribution of (%) trade cost | taste |
|---|---|---|---|
| Hubei | 86.2 | 60.8 | 39.2 |
| Shandong | 80.2 | 61.7 | 38.3 |
| Sichuan | 78.4 | 57.1 | 42.9 |
| Jiangxi | 77.5 | 55.9 | 44.1 |
| Fujian | 76.4 | 60.0 | 40.0 |
| Hebei | 72.4 | 53.8 | 46.2 |
| Jilin | 70.2 | 50.9 | 49.1 |
| Liaoning | 67.3 | 51.6 | 48.4 |
| Hunan | 67.0 | 49.2 | 50.8 |
| Zhejiang | 66.2 | 60.9 | 39.1 |
| Henan | 63.7 | 48.0 | 52.0 |
| Qinghai | 63.5 | 47.1 | 52.9 |
| Jiangsu | 61.1 | 53.8 | 46.2 |
| Tianjin | 60.8 | 56.0 | 44.0 |
| Chongqing | 59.5 | 43.1 | 56.9 |
| Guangxi | 56.7 | 39.7 | 60.3 |
| Ningxia | 56.0 | 40.1 | 59.9 |
| Anhui | 55.0 | 38.5 | 61.5 |
| Guangdong | 52.8 | 52.5 | 47.5 |
| Shanxi | 51.3 | 33.5 | 66.5 |
| Gansu | 47.1 | 26.3 | 73.7 |
| Xinjiang | 45.8 | 29.2 | 70.8 |
| Inner Mongolia | 42.4 | 22.4 | 77.6 |
| Heilongjiang | 40.9 | 15.4 | 84.6 |
| Shanghai | 40.1 | 47.9 | 52.1 |
| Shaanxi | 40.1 | 15.0 | 85.0 |
| Yunnan | 36.3 | 5.9 | 94.1 |
| Guizhou | 35.2 | 1.8 | 98.2 |
| Hainan | 27.7 | -19.4 | 119.4 |
| Beijing | 24.0 | 6.8 | 93.2 |
| Mean | 56.7 | 38.8 | 61.2 |
| St.d. | 16.2 | 20.9 | 20.9 |

*Notes:* Table reports the contribution decomposition of the home bias in trade.

the trade cost is less important. Head and Mayer (2013) show that between 50% and 85% of the distance effect on international trade flows are the result of the indirect trade costs (that they term as "dark trade costs"). Coşar et al. (2018) estimate the home bias in the global automobile market and conclude that about 2/3 of the reduced form home bias is due to the taste effect. Compared with those findings, our results of the taste effect (61.2%) is reasonable and consistent.

### 5.2 Policy Implications

In this subsection, we discuss how domestic migration policy affects the domestic trade bias using the parameter values estimated in the baseline regression. If the bilateral immigrant share $s_{nn}$ changes, it has a direct effect on the corresponding trade bias $HB_n$. However, the change in $s_{nn}$ also has an indirect effect on trade bias through a change in the multilateral promoter. Specifically,

$$\Delta HB_n = \underbrace{\theta_n \Delta s_{nn}}_{\text{direct effect}} \underbrace{-\theta_n \Delta \Pi_n^s}_{\text{indirect effect}} \tag{24}$$

The taste bias $\theta$ only captures the direct effect of a change in $s_{nn}$ on $HB_n$. The indirect effect is captured by the multilateral promoter $-\theta_n \Delta \Pi_n^s$.

Consider an extreme migration policy, such as a ban on migration (e.g., a completely strict Hukou system), for province $n$, in an effort to raise the native population share to 1. How does the province's trade bias respond to this shock? Using the equation above, we calculate the percentage change in trade bias $\Delta HB_n / HB_n$ for each province. The results are reported in column (2) in Table 8. The increase in trade bias is small, only 6.2% on average. However, the increase is much greater for immigrant regions, with Beijing's trade bias increasing by 52.3% and that of Shanghai increasing by 33.4%. A ban on migration in Guangdong province would increase the intra-national trade by 12.4%.

Conversely, if we removed all migration barriers (e.g., removal of the Hukou system) for province $n$ and allow consumers to freely migrate across its borders, province $n$ would eventually have exactly the same ethnic composition as the country population. The results are shown in column (3) in Table 8. The trade bias decreases by 61.2% on average, which is much greater than the size in the case of the migration ban.

Table 8: Counterfactuals: Migration Policy Implications

| Province | Shares Trade Bias | % Δ shares with zero migration | % Δ shares with full migration |
|---|---|---|---|
| Hubei | 86.2 | 1.0 | -39.2 |
| Shandong | 80.2 | 1.0 | -38.3 |
| Sichuan | 78.4 | 1.0 | -42.9 |
| Jiangxi | 77.5 | 0.9 | -44.1 |
| Fujian | 76.4 | 5.4 | -40.0 |
| Hebei | 72.4 | 1.1 | -46.2 |
| Jilin | 70.2 | 0.9 | -49.1 |
| Liaoning | 67.3 | 2.1 | -48.4 |
| Hunan | 67.0 | 1.0 | -50.8 |
| Zhejiang | 66.2 | 11.0 | -39.1 |
| Henan | 63.7 | 0.9 | -52.0 |
| Qinghai | 63.5 | 3.1 | -52.9 |
| Jiangsu | 61.1 | 5.0 | -46.2 |
| Tianjin | 60.8 | 13.3 | -44.0 |
| Chongqing | 59.5 | 2.1 | -56.9 |
| Guangxi | 56.7 | 1.4 | -60.3 |
| Ningxia | 56.0 | 3.7 | -59.9 |
| Anhui | 55.0 | 1.6 | -61.5 |
| Guangdong | 52.8 | 12.4 | -47.5 |
| Shanxi | 51.3 | 1.9 | -66.5 |
| Gansu | 47.1 | 1.4 | -73.7 |
| Xinjiang | 45.8 | 6.4 | -70.8 |
| Inner Mongolia | 42.4 | 4.9 | -77.6 |
| Heilongjiang | 40.9 | 1.4 | -84.6 |
| Shanghai | 40.1 | 33.4 | -52.1 |
| Shaanxi | 40.1 | 2.4 | -85.0 |
| Yunnan | 36.3 | 2.8 | -94.1 |
| Guizhou | 35.2 | 2.7 | -98.2 |
| Hainan | 27.7 | 8.7 | -119.4 |
| Beijing | 24.0 | 52.3 | -93.2 |
| Mean | 56.7 | 6.2 | -61.2 |
| St.d. | 16.2 | 10.8 | 20.9 |

*Notes:* Table reports the policy implications of the home-biased gravity.

## 6 Conclusion

Trade models usually assume identical preferences across consumers. Differences in demand across countries, such as home-biased tastes, are often included in unobservable "trade costs," which are indistinguishable. Existing studies on domestic trade even tend to ignore the consumers' taste differences. This paper provides a structural component of home-biased preferences that is feasible to estimate. Bilateral market tastes are determined by combing bilateral migration and migrants' taste bias. Using domestic migration data and trade data, we estimate the effects of taste bias on inter-provincial trade in China. Our results provide robust evidence that consumers' tastes are biased toward the products from their home province. Consumers also have larger taste biases in relation to agricultural products than in relation to the service sector, and the share of imports is more sensitive to migrant share in relation to final goods than in relation to intermediate and aggregated goods in China.

More importantly, our findings provide significant policy implications, given that reductions in trade costs and homogenization of tastes have different implications in relation to welfare and trade policies. In contrast to most previous studies, we separated and quantified home-bias tastes and unobservable trade costs. Domestic migration and trade policymakers should pay adequate attention to the heterogeneous preferences of various migrant groups.

## References

Allen, T. (2014). Information frictions in trade. *Econometrica*, 82(6):2041–2083.

Anderson, J. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93:170–192.

Anderson, J. and Yotov, Y. (2010). The changing incidence of geography. *American Economic Review*, 100(5):2157–86.

Anderson, J. E. and Zhang, P. (2020). Latent exports: Almost ideal gravity and zeros. Technical report, National Bureau of Economic Research.

Arkolakis, C., Lee, S. K., and Peters, M. (2020). European immigrants and the united states' rise to the technological frontier. Technical report, mimeo.

Atkin, D. (2013). Trade, tastes and nutrition in india. *American Economic Review*, 103:1629–1663.

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

Barwick, P. J., Cao, S., and Li, S. (2017). Local protectionism, market structure, and social welfare: China's automobile market. Technical report, National Bureau of Economic Research.

Bronnenberg, B., Dubé, J.-P., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508.

Bryan, G. and Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, 127(5):2229–2268.

Caballero, M. E., Cadena, B. C., and Kovak, B. K. (2018). Measuring geographic migration patterns using matrículas consulares. *Demography*, 55(3):1119–1145.

Caron, J., Fally, T., and Markusen, J. (2014). International trade puzzles: A solution linking production and preferences. *Quarterly Journal of Economics*, 129:1501–1552.

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, 104(11):3600–3634.

Combes, P.-P., Lafourcade, M., and Mayer, T. (2005). The trade-creating effects of business and social networks: evidence from france. *Journal of International Economics*, 66:1–29.

Coşar, A. K., Grieco, P. L., Li, S., and Tintelnot, F. (2018). What drives home market advantage? *Journal of International Economics*, 110:135–150.

De, H. H. (2010). Migration and development: A theoretical perspective. *The International Migration Review*, 44(1):227–264.

Desmond, S. A. and Kubrin, C. E. (2009). The power of place: Immigrant communities and adolescent violence. *The Sociological Quarterly*, 50(4):581–607.

Di Comite, F., Thisse, J.-F., and Vandenbussche, H. (2014). Verti-zontal differentiation in export markets. *Journal of International Economics*, 93(1):50–66.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of econometrics*, 4(2):115–145.

Fajgelbaum, P. and Khandelwal, A. (2016). Measuring the unequal gains from trade. *Quarterly Journal of Economics*, 131:1113–1180.

Feenstra, R. C. (2003). A homothetic utility function for monopolistic competition models. *Economics Letters*, 78:79–86.

Felbermayr, G., Grossmann, V., and Kohler, W. (2015). Migration, international trade, and capital formation: Cause or effect? *Handbook of the Economics of International Migration*, 1B:913–1025.

Hao, T., Sun, R., Tombe, T., and Zhu, X. (2020). The effect of migration policy on growth, structural change, and regional inequality in china. *Journal of Monetary Economics*, 113:112–134.

Head, K. and Mayer, T. (2013). What separates us? sources of resistance to globalization. *Canadian Journal of Economics*, 46:1196–1231.

Iranzo, S. and Peri, G. (2009). Migration and trade: Theory with an application to the eastern–western european integration. *Journal of International Economics*, 79(1):1–19.

Jiang, J. and Mei, Y. (2020). Mandarins make markets: Leadership rotations and inter-provincial trade in china. *Journal of Development Economics*, 147.

Ke, S. (2015). Domestic market integration and regional economic growth—china's recent experience from 1995–2011. *World Development*, 66:588–597.

Li, S. (2010). *Nian Zhongguo Diqu Kuozhan Touru Chanchubiao (2002 China Extended Regional Input-Output Tables)*.

Llull, J. (2018a). The effect of immigration on wages: exploiting exogenous variation at the national level. *Journal of Human Resources*, pages 0315–7032R2.

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

Lyons, C. J., Vélez, M. B., and Santoro, W. A. (2013). Neighborhood immigration, violence, and city-level immigrant political opportunities. *American Sociological Review*, 78(4):604–632.

McCallum, J. (1995). National borders matter: Canada-u.s. regional trade patterns. *American Economic Review*, 85:615–623.

Mi, Z., Meng, J., Guan, D., Shan, Y., Song, M., Wei, Y. M., Liu, Z., and Hubacek, K. (2017). Chinese co2 emission flows have reversed since the global financial crisis. *Nature Communication*, 8(1):1712.

Morey, M. (2016). Preferences and the home bias in trade. *Journal of Development Economics*, 121:24–37.

Novy, D. (2013). International trade without ces: Estimating translog gravity. *Journal of International Economics*, 89:271–282.

Obstfeld, M. and Rogoff, K. (2001). The six major puzzles in international macroeconomics: Is there a common cause? *NBER Macroeconomics Annual*, 15:339–390.

Poncet, S. (2003). Measuring chinese domestic and international integration. *China Economic Review*, 14(1):1–21.

Ramondo, N., Rodríguez-Clare, A., and Saborío-Rodríguez, M. (2016). Trade, domestic frictions, and scale effects. *The American Economic Review*, 106(10):3159–3184.

Rauch, J. (2001). Business and social networks in international trade. *Journal of Economic Literature*, XXXIX:1177–1203.

Rauch, J. and Trindade, V. (2002). Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84:116–130.

Ravuri, E. D. (2014). Return migration predictors for undocumented mexican immigrants living in dallas. *The Social Science Journal*, 51(1):35–43.

Schmutz, B. and Sidibe, M. (2019). Frictional labour mobility. *Review of Economic Studies*, 86(4):1779–1826.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political science review*, pages 147–174.

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–1872.

Wei, S. (1996). Intra-national versus international trade: how stubborn are nations in global integration. *NBER Working Paper*, No. 5531.

Yi, K.-M. (2010). Can multistage production explain the home bias in trade? *American Economic Review*, 100:364–393.

Young, A. (2000). The razor's edge: Distortions and incremental reform in the people's republic of china. *Quarterly Journal of Economics*, 115(4):1091–1135.

Zhang, P. (2020). Home-biased gravity: The role of migrant tastes in international trade. *World Development*, 129:104863.