Geographic proximity and the pro-trade effect of migration: State-level evidence from Mexican migrants in the United States

Michael Good[1]

Job Market Paper

1 October 2013

Abstract

I estimate the effect that migrants have on international trade between states of current residence and states of origin. The pro-trade effect of migration has been thoroughly examined since the mid-1990s, connecting both destination countries with origin countries and destination sub-national divisions with origin countries, respectively. However, a recent emphasis on the importance of geographic proximity to the migration-trade link leads me to pose the questions of how localized the trade-enhancing effect of migrants actually may be and how proximity matters for this relationship. My analysis provides the first results as to the migration-trade nexus at the state level for both places of destination and origin, relying on a unique data set allowing the mapping of Mexican-born migrants' US states of residence to Mexican states of origin; this ensures a more precise measurement of both migrant networks and other potential determinants of international trade, including the distance and mass variables fundamental to the standard gravity model. In addition to an augmented gravity model, I employ generalized propensity scores in examining the potential of nonlinearities in the migration-trade relationship. Furthermore, I unmask the distinct levels of geographic proximity that a single migration estimate disguises, estimating statistically significant elasticities of exports to both in-state and neighboring-state migration. These figures are not only qualitatively but also quantitatively important, corresponding to partial contributions of $1984 (in-state) and $538 (neighboring-state) to annual exports between respective US and Mexican states associated with each average additional migrant.

[1] *I thank Cem Karayalcin for useful discussion...* *Correspondence: Department of Economics, Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA. E-mail: mgood010@fiu.edu*

1. Introduction

Recent studies emphasize the link between migration and trade, both through theoretical models and empirical evidence. From the seminal works of Gould (1994) and Head and Ries (1998) to more recent articles such as Aleksynska and Peri (2011), Kugler and Rapoport (2011) and Hatzigeorgiou and Lodefalk (2011), all point to the same general conclusion: migrants do indeed promote international trade between the destination and origin countries. This robust positive relationship between human and goods mobility across studies is especially noteworthy given the variety of approaches employed and the number of countries studied. While most studies use a standard gravity model augmented with a migration variable as well as various controls for bilateral trade costs, specific methods vary, including pooled cross section or panel data OLS with fixed effects, 2SLS and generalized propensity scores. The US and Canada garner the majority of attention in terms of country-specific studies, focusing on migrants to the destination country and the subsequent trade from the destination country to all other countries; however, studies have also focused on the UK, Spain, Denmark and Bolivia, among others.[2] Furthermore, the geographic unit under examination varies, many measuring links at the country-country level while others narrow the focus to state-country connections.[3]

As to the channels through which migrants enhance trade, the consensus points to a preference channel normally associated with increased imports for the migrant-destination country, and an information channel associated with increased imports and exports for the migrant-destination country. Migrants may bring preferences for specific products with them to the destination region, leading to increased imports from the corresponding origin regions; on the other hand, migrants familiar with language, tastes, customs, or the workings of business and law in both the place of origin and destination may pass this information on to firms, thereby lowering the cost associated with entering or increasing presence in a particular foreign market, potentially increasing both imports and exports.

Two natural questions arise from this consistent body of evidence on the pro-trade effect of migration: (1) How localized are the preferences and information that migrants embody and potentially transmit to firms? and (2) How does geographic proximity matter for migration's pro-trade effect? While several previous studies focus on the state level for the given destination country, to my knowledge no studies examine the migration-trade link at the relatively localized state-state level, neither for one migration variable nor considering classification by geographic proximity of migrants.

[2]For example, see Girma and Yu (2002), Peri and Requena (2010), White (2007), and Erlich and Canavire Bacarreza (2006), respectively.

[3]In order to maintain consistency, I always refer to the geographic unit in terms of destination-origin throughout the paper. For example, I classify a study examining the connection between migrants living in the United States and their connection with exports from the US (entire country) to countries of migrants' origins as "country-country."

By employing the state-state level in the following analysis, this finer geographic disaggregation allows for (1) more precise measurement of migrant networks and other potential determinants of trade, including the distance and mass variables fundamental to any gravity model, and (2) differentiation between different states of residence *and* origin for migrants and the associated trade. The former more accurately depicts the variation in the covariates theorized to drive the migration-trade nexus, in turn permitting more precise estimates.[4] The latter crucially allows for the possibility that networks of migrants of similar state origin matter for the pro-trade effect of migration, above and beyond the traditionally-examined networks of compatriots. In other words, not only does a Mexican migrant now provide different information to the potential US export market than a Canadian migrant, a *Veracruzano* (from the Gulf coast state of Veracruz, Mexico) also provides different knowledge than a *Jalicense* (from the Pacific coast state of Jalisco, Mexico).

As far as geographic proximity, Herander and Saavedra (2005) and Artal-Tur et al. (2012) (hereafter HS and APR, respectively) both point to the importance of distance between the networks of migrants that ultimately supply the pro-trade effect of migration as a determinant of the extra associated trade. Further understanding the importance of geographic proximity clearly sheds light on the influence of migrants, specifically as to whether an increase in exports can be expected given an increase in migrants of a particular state in states adjacent to the state under examination. While both previous studies show that geographic proximity certainly matters (the pro-trade effect is stronger as distance between migrant networks decreases), they lack the finer geographic disaggregation on the side of migrant origin (export destination).

My approach examines empirically the migration-trade nexus at the state-state level, first intentionally ignoring geographic proximity and then classifying migrants into three categories based on distance between migrant networks. Using data linking migrants' states of origin with current states of residence, I determine how localized the information that migrants transmit to exporting firms actually is and how geographic proximity matters in this process of transmission. Specifically, this method maps Mexican migrants' Mexican state of origin to current state of residence in the United States, using data on *matrícula consular* (consulate registration) holders available from the Mexican government.[5] Use of a standard gravity model augmented to include migration as an additional explanatory variable for exports from US to Mexican states, as well as an extension relying on the generalized propensity scores (GPS) method, allow for estimation of the pro-trade effect that migrants have at the state-state level for the first time, both overall and classified by

[4]For example, Yilmazkuday (2013) finds that the estimated coefficient for distance in the gravity-type models of determinants of international trade suffers from greater bias if not considering production location within countries (at the state or local level).

[5]The *matrícula consular* is an identification card made available by the Mexican government to citizens residing abroad starting in 1871. The card must be renewed every five years, giving the holder access to opening a bank account, obtaining a driver's license, and other services, depending on the specific country and state of residence. I define state of origin as the last state of permanent residence before migration occurred from Mexico to the US.

geographic proximity.

I find that just as in previous studies at the country-country and state-country levels, migrants indeed have a statistically significant pro-trade effect at the state-state level, promoting exports from US states of residence to Mexican states of origin. The initial specification reveals an elasticity of exports to migration of 0.11, while the GPS extension signals a diminishing yet positive marginal contribution over nearly the entire range of actual measured migrant stocks. Classified by geographic proximity, the preferred specification reveals the expected partial pro-trade effects otherwise disguised by the single estimate of migration's pro-trade effect, corresponding to contributions of \$1984 (in-state) and \$538 (neighboring-state) to annual exports between respective US and Mexican states associated with each average additional migrant. Given these results, geographic proximity appears to matter just as in HS and APR, with the in-state effect surpassing the neighboring-state effect in magnitude; however, in contrast with previous results, I find that the neighboring-state contribution of migrants is indeed of statistical significance. All results vary minimally in magnitude and significance across several checks for robustness, providing clear evidence as to the additional benefit of migration that manifests itself through its nexus with international trade, highlighted at the relatively localized state-state level for the first time and classified by geographic proximity of migrant networks.

The rest of the paper continues with a discussion of several aspects unique to the US-Mexico relationship with relevance to the migration-trade nexus in Section 2. Section 3 highlights my initial empirical strategy and data sources, while Section 4 contains a discussion of the main results. Section 5 provides various checks for robustness; Section 6 details the GPS extension, prior to the Section 7 conclusion.

2. US-Mexico relationship

The US-Mexico relationship provides an especially interesting and appropriate setting conducive to examining the pro-trade effect of migration for several reasons. First and foremost is the fact that data is actually available, permitting the analysis at the state-state level for the first time. Detailed exports data from US to Mexican states are available for all years since 1994, coinciding with the implementation of the North American Free Trade Agreement. Perhaps most noteworthy is the availability of the *matrícula consular* data, uniquely allowing for the connection of Mexican state of origin to US state of residence for each migrant registered during the period examined.

Additionally, as both the US and Mexico are relatively large, heterogeneous countries, separation by state origin and destination is theoretically worthwhile; there is clearly wide differentiation within countries as to preferences and the knowledge and information that residents hold about markets, customs, and tastes, all important factors for the theorized channels through which the pro-trade effect operates. For example, an emigrant leaving the southeastern state of Chiapas to reside in a given US state undoubtedly has much different information than an emigrant leaving the northern state of Sonora, arguably similar to the level of differentiation existing across migrants of varying nationalities signaled by the previous literature.

Although this differentiation, depending on Mexican state of origin and US state of residence, points unequivocably to the theorized pro-trade effect of migration at the state-state level, several aspects of the US-Mexico relationship signal that this effect could potentially be minimized relative to the entire range of possible pro-trade effects across all countries (and the respective sub-national divisions). First, both trade and migration between the US and Mexico are relatively well-established, neither phenomenon being particularly new in its existence. HS, among others, find that the existence of a previous large migrant stock reduces any pro-trade effect of new migrants. However, the "newness" of migration from and to particular Mexican and US states, respectively, could potentially offset the fact that the Mexico-US migration is not novel at the country-country level.[6] Second, Mexican migration levels to the US are relatively high, especially relevant if beyond a certain level of migration, further migrants may not marginally stimulate trade between places of residence and origin. The mean state-state count of *matrículas consulares* for my sample is 3038, with the maximum of 270,201 corresponding to those *Michoacanos* registered in California. Previous evidence is divided on the existence of an "exhaustion point" of the pro-trade effect of migration, one of the possibilities explored in the GPS extension of Section 6; Serrano and Requena (2013, hereafter SR) finds that every migrant makes a positive marginal contribution to exports at the province-country level, while Egger et al. (2012, hereafter EVN) finds evidence of an exhaustion point for imports at the country-country level of approximately 4,000 migrants, above which additional migrants provide zero stimulus for imports. Finally, a majority of the Mexican migrant population in the US is relatively low-skilled and may not participate in any form of business network. These general characteristics are of potential importance given recent findings that being high-skilled and having access to business networks makes migrants particularly effective in their promotion of trade.[7]

Given the outlined aspects of the US-Mexico relationship, any pro-trade effect found at the state-state level between states of these neighbor countries can be hypothesized to fall at the lower end of the spectrum of potential worldwide effects across all countries. In turn, while any claims of external validity should clearly be met with due skepticism, the presence of a US-Mexico link between migration and trade would appear to suggest an even stronger potential of the general existence of a pro-trade effect of migration at the state-state level.

[6] Card and Lewis (2007) examines the choice of US states of destination for Mexican migrants, analyzing changing trends during the 1990s. Hatzigeorgiou and Lodefalk (2011) also highlights the importance of new migrants in updating information.

[7] See Felbermayr and Jung (2009) and Aleksynska and Peri (2011) for these respective emphases.

3. Empirical Strategy and Data

In estimating the effect of migration on international trade between states in the United States and in Mexico, I first employ a standard gravity model, the most common empirical strategy for studies examining not only factors affecting trade, but also the potential pro-trade stimulus provided by migration. Putting aside geographic proximity for a moment, I start with the standard model augmented with migration between respective state pairs as an additional explanatory variable; the resulting specification serves as a useful benchmark for the analysis to follow.

$$lnT_j^i = \alpha + u_i + m_j + \phi lnMig_j^i + \beta_1 lnY_iY_j + \beta_2 lnDist_j^i + \beta_3 Adj_j^i + e_j^i \tag{1}$$

Incorporating the potential importance of geographic proximity, I then further augment the gravity model to include two additional measures of Mexican migration to the United States.

$$\begin{aligned} lnT_j^i = \alpha + \upsilon_i + \mu_j + \gamma_1 lnMig_j^i + \gamma_2 lnMig_j^{adj} + \gamma_3 lnMig_j^{rest} + \delta_1 lnY_iY_j + \delta_2 lnDist_j^i \\ + \delta_3 Adj_j^i + \varepsilon_j^i \end{aligned} \tag{2}$$

T_{ij} measures exports from US state i to Mexican state j in terms of yearly total value, dependent on migration, size of market (income), distance, and adjacency. Mig_{ij} captures the stock of *matrícula consular* holders in each US state i from each Mexican state of origin j , Mig_j^{adj} measures the stock of *matrícula consular* holders in the states adjacent to each US state i from each Mexican state of origin j , while Mig_j^{rest} reflects the stock of *matrícula consular* holders in the rest of US states (exclusive of state i and adjacent states) from each Mexican state of origin j ; Y_i and Y_j are the gross state products of US state i and Mexican state j, respectively, $Dist_{ij}$ represents the distance by land from US state i capital to Mexican state j capital, while Adj_{ij} is a dummy variable taking the value of 1 for adjacent states and that of 0 for states not sharing a border. Given the log transformations, the coefficients of interest, γ_1, γ_2, and γ_3, thus pinpoint the percentage increase (decrease) in yearly exports flowing from a US state to a Mexican state associated with a 1% increase in the stock of migrants originating from the corresponding Mexican state and registered in the corresponding US state (i.e., in-state), in the neighbors of the corresponding US state (i.e., neighboring-state), and in the remainder of all other US states (i.e., other-state), respectively. Additionally, u_i , m_j , υ_i , and μ_j are US and Mexican state fixed effects, controlling for the multilateral resistance terms as recommended by Anderson and Van Wincoop (2003). Other variables commonly employed as controls for bilateral trade costs in previous migration-augmented gravity models, such as trade agreements, language, colonial ties, legal system, currency, and cultural distance, are not relevant in the current setting since these variables are generally not differentiated within a single country, this being true in the case of the United States and Mexico.[8] If migration

[8]There is a limited amount of heterogeneity for these potential variables, for example with the presence of a

indeed is conducive to trade, an expectation of $\phi > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$, and $\gamma_3 > 0$ holds, the latter two inequalities depending on how far-reaching migrant networks' influence is; if geographic proximity matters for the migration-trade nexus, the clear expectation is $\gamma_1 > \gamma_2 > \gamma_3$.

Values of state-state exports are obtained from the US Bureau of Transportation statistics; these statistics cover all exports from the US to Mexico at the state-state level, except for those transported by air or water, providing 90% coverage of total exports between the two nations.[9] Given this coverage, the non-contiguous US states of Alaska and Hawaii are excluded from the analysis. For the preferred specification, trade is measured by the state-state values from the year 2010.[10] As original export data are listed with current dollars as the unit, I use the US CPI-U series to convert all values to 2011 US dollars. Statistics on the number of *matrículas consulares* issued are calculated given the information provided by the *Insitituto de los Mexicanos en el Exterior* (IME). Since holders of the card must specify last state of Mexican residence as well as current state of US residence during the application process, these statistics uniquely allow for the construction of the necessary state-state migration data. As the identification cards have a renewal period of five years, I sum the available data from 2006 to 2010 in constructing the stock of Mexican migrants for each state-state combination.[11] I consult the US Bureau of Economic Analysis (BEA) and the Mexican *Instituto Nacional de Estadística y Geografía* (INEGI) statistics for the respective gross state products corresponding to 2010, while distance between capital cities is calculated using the shortest route by land expressed in number of miles.[12] The original data for Mexican gross state products are listed with the unit of 2003 pesos, therefore I initially convert the values to 2003 US dollars using the average of monthly historical peso-dollar exchange rates from 2003. Finally, just as with the US gross state products originally reported with the unit of 2005 US dollars, I again use the CPI-U series to convert all values to 2011 dollars in order to maintain uniformity with the export values.

As the *matrícular consular* data does not completely cover the population of Mexican origin in the US and could possibly present problems of selection, I closely examine the distribution of Mexican migrants across the US states of residence in attempting to determine whether this data sufficiently represents the actual distribution of residents of Mexican origin across the US

number of languages in Mexico, however in the sample at hand this differentiation is so minimal that it does not justify inclusion in the regression equation as an additional control variable.

[9] While imports data would clearly provide for useful comparison, unfortunately state-state data is not presently available.

[10] Section 5 details modifications to the trade value calculation used to check for sensitivity of the results to these changes.

[11] As of final revisions, data from 2010 is the most recent available.

[12] This differs from the standard measure used by similar studies, that of great circle distance, due to the fact that the trade data (and a majority of Mexico-US migration) is by land. However, if great circle distance is indeed employed as the measure of distance, results change only minimally, with a slight increase in the magnitude of the distance coefficient.

states. As there is no justifiable reason to expect that Mexican state of origin affects selection into obtaining a *matrícula consular*,[13] if the data's distribution is sufficiently close to the actual distribution of Mexican migrants (irrespective of Mexican state origin) across US states, the use of the *matrícular consular* data can be said to provide a certain level of representativeness for the state-state distribution, thereby minimizing any bias arising from selection problems. This thus allows the use of the 2010 US Census as a benchmark for comparison; I contrast the *matrícular consular* data with that of the Census, in which the number of residents in each US state claiming Mexican origin is detailed.

Figure 1: Percentage distribution of *matrículas consulares* versus US Census

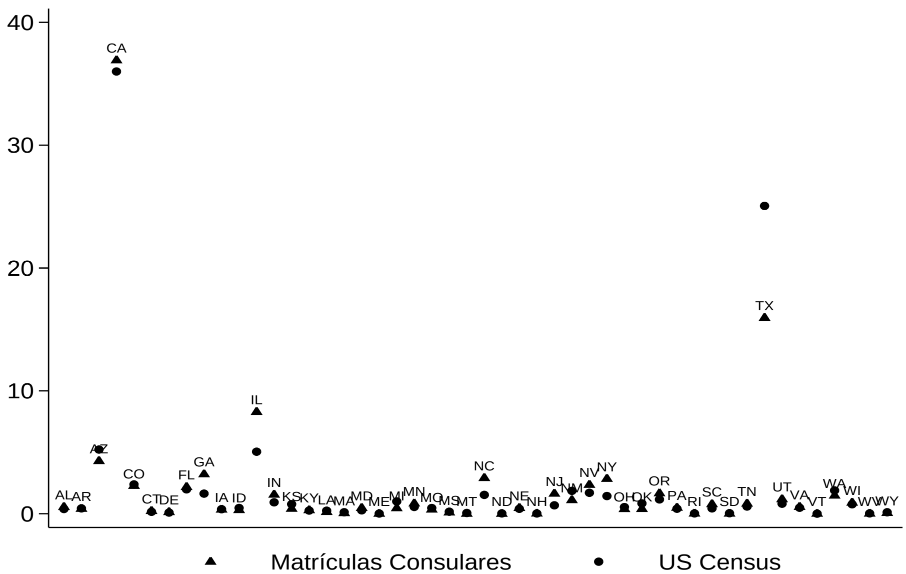


Figure 1 details the distribution of Mexican migrants in the US for both the *matrícula consular* data and the Census data; the data are expressed as the number of Mexican migrants in each state divided by the total stock of migrants from each respective source.[14] In fact, the *matrícula consular* data performs well in representing the actual distribution of Mexican migrants across US states, with most states' difference coordinates close to zero. Only two states, Texas and Illinois, suffer

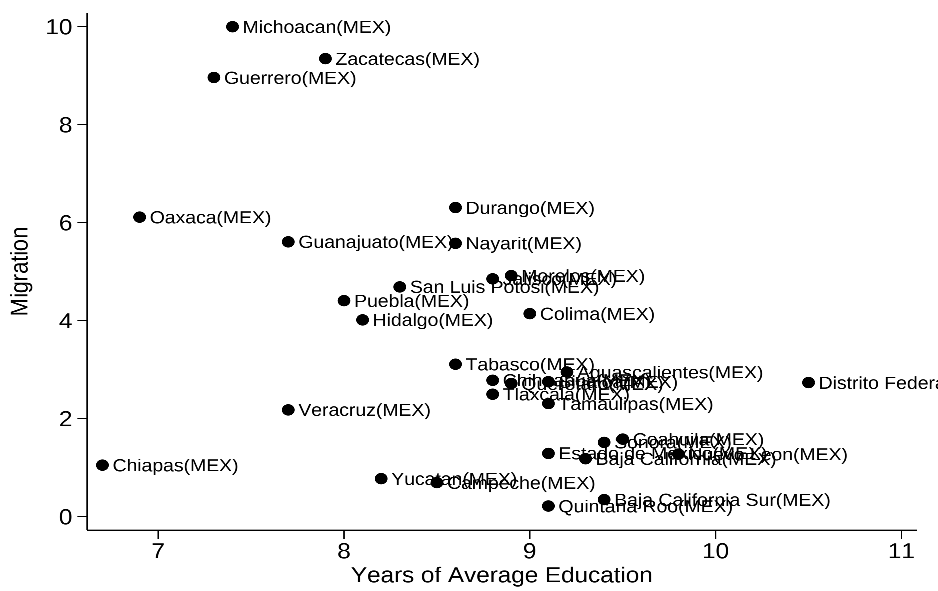
[13]A natural assumption may be that education level is associated with legal migration status, thereby making it more likely for individuals to obtain a *matrícula consular* if the state of origin corresponds to a low-education Mexican state on average. However, this assumption does not appear to be correct; see below for a related brief discussion.

[14]This fraction with an upper limit of 1 is then multiplied by 100, resulting in the numbers expressed on the y-axis of Figure 1.

from differences greater than 3%, while 43 of 48 states' differences in percentages are less than 1%.

Additionally, one may expect that the number of highly-educated migrants is underrepresented in the *matrícula consular* data, due to the fact that there is no clear incentive for a documented US resident to hold the identification card. This consideration is especially important given the previously mentioned studies emphasizing the extra relevance of highly-educated migrants in promoting trade above and beyond the average migrant contribution. Taking the average education level in Mexican states from INEGI statistics, dispersed over a range of 6.7 to 10.5 years of schooling with a mean of 8.6, a first check of the data indeed shows a negative correlation between Mexican state average education level and the percentage of origin state population registered with the *matrícula consular*.[15]

Figure 2: Mexican states average education and migration



Migration is measured as the ratio of *matrículas consulares* per Mexican state of origin to the total corresponding state population, then multiplied by 100 to express values as percentages.

However, this correlation gives no information as to the key question of how education level actually relates to legal migration status, and in turn to the *matrícula consular*. It is not clear that the expectation of underrepresentation is reasonable, given that the correlation between legal migration status and education level is anything but definitive for Mexican migrants in the US. Passel and

[15]See Figure 2 for a scatter plot of this correlated data.

Cohn (2009) determines that 47% of unauthorized migrants ages 25 to 64 in the US have completed high school or less, while Caponi and Plesca (2012) argues that documented Mexican migrants in the US are actually more likely to have a lower education level than undocumented migrants. Comparing the *matrícula consular* data with other representative data as to education level presents two problems. The IME only reports state-state statistics including education level for 2006 and 2007, thereby providing a smaller sample in representing the overall stock of migrants; in addition, the best data for comparison, that of the US Current Population Survey, is known to undercount undocumented migrants. Due to these difficulties and lack of available data, I do not empirically address the issue of state-state distribution by education level. Additional data availability would clearly allow for future exploration of this further rich level of detail.

Table A.1 details the number of *matrículas consulares* registered from 2006 to 2010, classified by both US state of residence and Mexican state of origin, while Table A.2 focuses on the state-state makeup of Mexican migration to the three top US destination states, California, Texas, and Illinois, and the corresponding exports to Mexico. Table A.2 and Figure A.1 provide an initial idea of the simple correlation between state-state migration and exports. Without any controls for bilateral trade costs or state fixed effects, the best-fit line displayed in Figure A.1 exhibits a slope of 0.44, providing preliminary evidence of a potential positive relationship between migration and exports at the state-state level. Table 1 shows the mean, standard deviation, maximum and minimum for variables in both the base and alternative samples.

4. Results and Discussion

Column 1 of Table 2 displays the results of the OLS regression employing the benchmark gravity equation listed in Equation (1). Including Mexico City, the base sample consists of 1536 observations, a result of all trading pairs of 48 US and 32 Mexican states. The coefficient estimate of migrants' effect on state-state exports is indeed significantly positive; holding all other factors constant, an increase of 1% in the number of state-state migrants is associated with a 0.11% increase in state-state exports, with $p < 0.01$. Distance, as expected, is significantly negative, reflecting a 1.60% decrease in state-state exports associated with a 1% increase in distance between the respective capitals of US and Mexican states. Holding other factors constant, states that are adjacent enjoy more than double the trade of nonadjacent states, while a 1% increase in combined economy size is associated with a 0.94% increase in state-state exports. All coefficient estimates have the expected positive (negative) relationship with state-state exports, and are highly significant.

Table 1: Descriptive statistics - base and alternative samples

Variable	Base sample $n = 1536$	Base sample minus Mexico City $n = 1488$	Base sample minus Texas and Illinois $n = 1472$
Exports (in USD)	81,993,615.91	73,115,404.78	50,679,742.75
	(674,482,678)	(658,528,331)	(385,569,081)
	0/17,900,000,000	*0/17,900,000,000*	*0/11,800,000,000*
Migration_j^i	3038	2938	2377
	(13,771)	(13,634)	(12,786)
	0/270,201	*0/270,201*	*0/270,201*
Migration_j^{adj}	12,409	12,070	12,770
	(27,591)	(27,410)	(28,108)
	1/308,918	*1/308,918*	*1/308,918*
$\text{Migration}_j^{rest}$	130,355	126,016	130,655
	(124,937)	(124,375)	(125,141)
	858/525,394	*858/525,394*	*858/525,394*
Distance (miles)	2077.56	2077.11	2105.75
	(597.34)	(599.59)	(585.22)
	239/3681	*239/3681*	*239/3681*
Adjacency	0.006	0.006	0.004
	(0.076)	(0.077)	(0.059)
	0/1	*0/1*	*0/1*
US GSP (bn. USD)	34,392.30	34,392.30	31,232.81
	(39,909.53)	(39,909.53)	(37,118.12)
	2,950.91/218,967.32	*2,950.91/218,967.32*	*2,950.91/218,967.32*
Mex. GSP (bn. USD)	2,969.87	2,514.13	2,969.87
	(3,197.62)	(1,975.85)	(3,197.62)
	529.45/17,097.79	*529.45/9,235.81*	*529.45/17,097.79*

For each variable, means are listed first, standard deviations are reported in parentheses, while minimum/maximum pairs are reported in italics.

While these initial results are the first to confirm the existence of the pro-trade effect of migration at the more localized state-state level, given the crucial findings of HS and APR and the main hypothesis of this paper, it is not surprising that Columns 2 to 5 unmask key complexities disguised by the estimates in Column 1. Focusing on the preferred specification in Column (5), the geographic proximity of migrants clearly appears to matter for the promotion of international trade. The coefficients of both in-state and neighboring-state migration are statistically significant, remaining so even after adding all relevant control variables. As expected, the elasticity with respect to in-state migration at 0.07 is lower than that of Column 1, due to the addition of neighboring-state migration. An increase of 1% in neighboring-state migration is associated with a 0.08% increase in state-state exports, while increased migration in the rest of US states is associated with a small, yet statistically insignificant negative effect on state-state exports. The remaining independent variables' coefficients and levels of significance are similar to those of Column 1, with attenuation in magnitude only for the distance estimate. Interpreting the coefficients of interest, Column 1 appears to capture the overall pro-trade effect of migration, masking the actual importance of geographic proximity. Column 5 sheds light on this importance; in-state migration indeed promotes trade between US states of residence and Mexican states of origin, however neighboring-state migration also has an essential role in this expansion of trade.

Table 2: Coefficient estimates using gravity equation (OLS, state-state fixed effects)

Dependent variable: US-Mexico state-state exports					
Independent variable	(1)	(2)	(3)	(4)	(5)
Migration_j^i	0.1134***	0.2403***	0.1635***	0.1372***	0.0735*
	(0.0382)	(0.0348)	(0.0420)	(0.0426)	(0.0432)
Migration_j^{adj}			0.1354***	0.0742	0.0814*
			(0.0418)	(0.0457)	(0.0457)
$\text{Migration}_j^{rest}$				-0.5031	-0.0215
				(0.1539)***	(0.1685)
Distance	-1.5952***				-1.4335***
	(0.2913)				(0.3177)
Adjacency	1.2308***				1.3146***
	(0.4449)				(0.4531)
Economy size	0.9365***				1.1589***
	(0.1391)				(0.1607)
R^2	0.8192	0.8116	0.8130	0.8143	0.8197
n	1536	1536	1536	1536	1536

Each estimate is from a separate OLS regression with the logarithm of US-Mexico state-state exports in US dollars plus one as the dependent variable, employing the base sample of the 48 contiguous US states and 32 Mexican states. All regressions include state fixed effects, controlling for any existing systematic differences across states that may affect all states' outcomes. Column (1) displays estimates for coefficients corresponding to Equation (1), while Columns (2) to (5) highlights those corresponding to Equation (2). Column (2) reports estimates using only migration_j^i as an explanatory variable, column (3) adds migration_j^{adj} as an explanatory variable, column (4) adds $\text{migration}_j^{rest}$ and column (5) displays the preferred specification, adding the remainder of the relevant controls. Heteroskedasticity-consistent robust standard errors are reported in parentheses. ***, ** and ** denote statistical significance at the 1%, 5% and 10% levels, respectively.

The OLS estimates in turn permit a simple calculation of the magnitude of the pro-trade effect of migration, highlighting the quantitative importance of this effect, as well as allowing for a comparison of the relative sizes of the benefit from in-state, neighboring-state, and other-state migration. This exercise carries extra importance given the fact that at first glance the estimates in Column 5 appear to point to a counterintuitive result, that in-state migration is associated with a *smaller* pro-trade effect than that of neighboring-state migration. Starting with results from Column 1, given a 10% increase in average immigration from a particular Mexican state to a particular US state, the average migrant stock increases from 3038 to 3342. Employing the estimated coefficient of approximately 0.11, this 10% increase in migration results in an increase in average state-state exports, settling on the new value of exports equal to $82,923,424. This translates into $3061 extra state-state exports per year associated with the average extra migrant.[16]

When different geographic proximities are separated in Column 5, I can now break down the distinct components of the pro-trade effect of migration. Once again assuming a 10% increase in average in-state migration, the average in-state migrant stock increases 3038 to 3342. Using the estimated coefficient of approximately 0.07 results in an increase in average state-state exports of $602,653, translating into $1984 extra state-state exports per year associated with the average extra in-state migrant. At the same time, these extra 304 migrants are neighboring-state migrants for an average of 4.39 states (the average number of adjacent states for all US states). Given the average neighboring-state migrant stock of 12,409, this increase is equivalent to a 2.45% increase. In turn, relying on the percentage increase and the estimate from Column 5 of approximately 0.08, average state-state exports increase by $538 per neighboring-state migrant. Finally, the extra 304 migrants are equivalent to a 0.23% increase in migration in the rest of the average of 43.61 states. Given the Column 5 estimate of approximately -0.02, this increase results in a decrease of $13 per average migrant.[17] Finally, collecting all calculations of separate components in order to compile an overall effect results in $3779 of extra state-state exports associated with the average extra immigrant from Mexico residing in the US; it is crucial to emphasize that this contribution to exports by the single average extra migrant is actually spread across US states, the three terms of Footnote 18 corresponding to state *i*, neighbors of state *i*, and the rest of US states, respectively.[18] This compiled result is similar in magnitude to that of Column 1, however the separation of distinct geographic proximities allows for the differentiation between the relatively larger in-state contribution of the average extra immigrant, the smaller neighboring-state contribution, and the minuscule other-state reduction of state-state exports.

[16] This figure can be alternatively calculated by multiplying the elasticity of exports to migration by the ratio of average state-state exports to average migrant stock.

[17] Trade diversion could be a simple explanation for this negative effect, however I intentionally do not explore this further given the small magnitude and lack of significance of the Mig_j^{rest} estimate.

[18] The calculation is as follows: $\$1984 + \$538(4.39) - \$13(43.61)$.

Table 3: Comparison of estimates for the elasticity of exports to migration

Authors	Elasticity of exports to migration	Extra annual exports generated per extra migrant	Sample	Specification-Method
My estimates	0.07 (in-state) 0.08 (neighboring-state)	$1985 (in-state) $3756 (across all states)	48 US states, 32 Mexican states, 2008-2010	Pooled cross section, OLS with state-state trading partner fixed effects
Aleksynska and Peri (2011)	0.25	$24,895	CEPII "square" gravity data set, 5230 observations	Pooled cross section, OLS with country-country fixed effects
Bandyopadhyay, Coughlin and Wall (2008)	0.14	—	50 US states and District of Columbia, 29 countries, 1990, 2000	Panel, OLS with country-time and trading partner pairs fixed effects
APR (2012)	0.02, 0.08, 0.07 (in-state) 0.00, 0.02, -0.04 (neighboring-state)	—	Italy, Portugal, and Spain, "about 100 countries," 2002-2010	Panel, OLS with country-time and trading partner pairs fixed effects
Briant et al. (2009)	0.10	$6590	94 French departments, 100 countries, 1998-2000	Pooled cross section, OLS, 2SLS with country-department fixed effects
Dunlevy (2006)	0.24-0.47	—	50 US states and District of Columbia, 87 countries, 1990-1992	Pooled cross section, OLS with country-state fixed effects
Felbermayr and Jung (2009)	0.11	$2717	21 "North" countries and 114 "South" countries, 1988-2000	Pooled cross section, OLS, differenced with country-country fixed effects
HS (2005)	0.16 (in-state) 0.07 (other-state)	—	50 US states, 36 countries, 1993-1996	Pooled cross section, Tobit and LAD
Peri and Requena (2010)	0.05-0.11	—	50 Spanish provinces, 77 countries, 1993-2008	Panel, OLS, 2SLS with with country-time and trading partner pairs fixed effects
Tadesse and White (2010)	0.04-0.05	$1034-$1267	50 US states and District of Columbia, 75 countries, 2000	OLS with state-country fixed effects
White (2007)	0.11	$2608	US, 73 countries, 1980-2001	Pooled cross section, OLS with country-country fixed effects

Estimates for elasticity are reported according to the preferred model specified by the authors in the corresponding articles, or if not specified, the most appropriate estimates for comparison to those of this paper. My estimates are those corresponding to the preferred base sample. Other articles' estimates are the following: the OLS fixed effects result for Aleksynska and Peri (2011), the benchmark OLS result for both APR and Briant et al. (2009), the fixed effects result for Bandyopadhyay et al. (2008), the differenced result for Felbermayr and Jung (2009), the Tobit result for HS (2005), the aggregate exports result for Tadesse and White (2009), and the full sample result for White (2007). Figures for column 3 are generated according to the reported elasticities, multiplying the respective elasticity by the ratio of average state-state exports to average state-state stock of migrants; — denotes that I found neither the corresponding summary statistics nor the estimate of the annual value of extra exports generated per migrant.

In comparing the estimates and magnitude of the pro-trade effect of migration to those of the literature, I rely on HS and APR, as well as previous state-country and country-country studies, as this paper is the first to examine the state-state level. While doing so provides a framework within which viewing my results is feasible, given the novel state-state level and the emphasis on distinct geographic proximities, any comparison must be accompanied by a disclaimer highlighting these differences in sampling and geographic disaggregation of data. Table 3 provides an update of Table 1 from Peri and Requena (2010) in order to include estimates from more recent studies, and those of this paper, as well as a comparison of extra annual exports generated per extra migrant. The elasticity of exports to in-state migration estimated as 0.07 echoes the positive, significant estimates from both HS and APR, however my significantly positive estimate for neighboring-state migration contrasts with that of HS (positive, but insignificant) and APR (not significantly different from zero).

Additionally, the finding of $3779 extra yearly exports generated by each extra migrant is similar to those of $2608 and $2717, detailed in White (2007) and Felbermayr and Jung (2009), respectively. While these estimates differ most dramatically from that of $24,895 found by Aleksynska and Peri (2011), it is worthwhile to signal that these numbers are not necessarily incompatible. As Aleksynska and Peri (2011) points out, factors such as average number of migrants in the sample and the specific measure of migrant stock contribute to these differentiated estimates. My measure based on the *matrícula consular* includes some migrants who may not be economically active, and does not classify migrants based on education level, which most likely further attenuates estimates as mentioned in Kugler and Rapoport (2011).

5. Robustness and Sensitivity Checks

As a first check for robustness of the obtained results, I outline a new set of estimates in Table 4, now excluding Mexico City from the sample under concerns of potential bias. Exports listed under the destination of Mexico City could create a bias, due to the fact that a large percentage of these exports actually has an alternative final destination within Mexico; the revised sample contains 1488 state-state observations. The magnitude and significance of the pro-trade effect of migration change minimally, the exclusion of Mexico City slightly decreasing the magnitude of the in-state coefficient to just under 0.07 and that of neighboring states to nearly 0.08. An additional concern arises from the comparison of the *matrícula consular* data and the US Census data highlighted in Section 3. Although a high level of representativeness is present, Texas and Illinois clearly are outliers in this respect, reflecting a difference of 9.11% and 3.24% between the data sets, respectively. Especially given the fact that both Texas and Illinois are two of the main destination states in the US for Mexican migrants, it is important to consider migration's pro-trade effect excluding the two outliers from the sample as an additional test of robustness. Table 4 also highlights the coefficient estimates generated excluding Texas and Illinois, using a sample of 1472 observations resulting from the combination of 46 US and 32 Mexican states. Compared to the results presented in Table 2, in-state migration's effect on state-state exports is slightly greater, while neighboring-state migration's estimate is slightly lower, both minimally less significant.

Table 4: Coefficient estimates using gravity equation (OLS, state-state fixed effects)

Dependent variable: US-Mexico state-state exports, $n = 1488$ and $n = 1472$								
Independent variable	(1)		(2)		(3)		(4)	
Migration$_j^i$	0.2394***	0.2349***	0.1616***	0.1554***	0.1351***	0.1355***	0.0671	0.0764*
	(0.0354)	(0.0360)	(0.0428)	(0.0436)	(0.0434)	(0.0440)	(0.0440)	(0.0444)
Migration$_j^{adj}$			0.1382***	0.1371***	0.0748	0.0725	0.0782*	0.0767
			(0.0431)	(0.0429)	(0.0472)	(0.0475)	(0.0471)	(0.0476)
Migration$_j^{rest}$					-0.5101***	-0.5032***	-0.0058	-0.0653
					(0.1565)	(0.1608)	(0.1713)	(0.1730)
Distance							-1.5813***	-1.3860***
							(0.3289)	(0.3351)
Adjacency							1.2479***	1.7183***
							(0.4603)	(0.6060)
Economy size							0.9927***	1.1865***
							(0.1510)	(0.1656)
R^2	0.7989	0.7965	0.8004	0.7980	0.8019	0.7994	0.8081	0.8052
n	1488	1472	1488	1472	1488	1472	1488	1472

Each estimate is from a separate OLS regression with the logarithm of US-Mexico state-state exports in US dollars plus one as the dependent variable, employing the two alternative samples. All regressions include state fixed effects, controlling for any existing systematic differences across states that may affect all states' outcomes. Column (1) reports estimates using only migration$_j^i$ as an explanatory variable, column (2) adds migration$_j^{adj}$ as an explanatory variable, column (3) adds migration$_j^{rest}$ and column (4) displays the preferred specification, adding the remainder of the relevant controls. Heteroskedasticity-consistent robust standard errors are reported in parentheses. ***, ** and ** denote statistical significance at the 1%, 5% and 10% levels, respectively.

Selection of state-state exports from the year 2010 as the measure for the dependent variable could be driving the obtained results; if estimates of migration's pro-trade effect on trade differ greatly across the use of varied individual years of trade data as alternative dependent variables, this would clearly be cause for concern. However, the estimates in fact vary only minimally when using exports data from 2008 and 2009 in lieu of 2010, as reported in Table 5. Migration's pro-trade effect remains significant and similar in magnitude across all alternative regressions accounted for, with the in-state estimate bottoming out at 0.0731 and peaking at 0.0935. Using the same simple method of calculation as in Section 4, these figures correspond to an extra \$1819 and \$2053 of annual exports, respectively, associated with the average extra in-state migrant, not considering the neighboring- and other-state contributions.

Table 5: γ_1 and γ_2 with alternative (in)dependent variables (OLS, state-state fixed effects)

(In)dependent variable	Base sample $n = 1536$	Base sample minus Mexico City $n = 1488$	Base sample minus Texas and Illinois $n = 1472$
US-Mexico state-state exports 2008			
Migration_j^i 2006-08	0.0929** (0.0445)	0.0849* (0.0454)	0.0935** (0.0455)
Migration_j^{adj} 2006-08	0.0859* (0.0484)	0.0850* (0.0498)	0.0875 (0.0489)*
R^2	0.8028	0.7908	0.7991
US-Mexico state-state exports 2009			
Migration_j^i 2006-09	0.0803* (0.0461)	0.0731 (0.0470)	0.0824* (0.0470)
Migration_j^{adj} 2006-09	0.0547 (0.0492)	0.0505 (0.0508)	0.0548 (0.0499)
R^2	0.7940	0.7806	0.7898

Heteroskedasticity-consistent robust standard errors are reported in parentheses, while ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively.

An additional concern is that trade and migration could be determined jointly, leaving forwarding the measure of exports as a clear strategy to alleviate this potential problem. I regress exports for periods $t+1$, $t+2$, and $t+3$, respectively, with the preferred sample and specification, using all possible corresponding measures of migrant stock (*matrícula consular* stock) to eliminate any possibility of joint determination.[19] This strategy results in six further regressions; Table 6 reports estimation results along with the corresponding exports and migration measures employed in each additional regression. The estimation of migration's pro-trade effect is consistent across these varied measures, both in magnitude and significance, with the elasticity of state-state exports to in-state migration ranging from 0.0789 to 0.1013.

[19] As Aleksynska and Peri (2011) mentions, since the migration measure is a stock accumulated over years, it is probable that it is determined before trade flows; however I forward exports to assure that joint determination is not a factor.

Table 6: Coefficient estimates using forwarded exports (OLS, state-state fixed effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Exports measure	2008	2009	2010	2009	2010	2010
Migration measure	2006-07	2006-07	2006-07	2006-08	2006-08	2006-09
$Migration_j^i$	0.1006**	0.1013**	0.0980**	0.0922**	0.0926**	0.0789*
	(0.0444)	(0.0446)	(0.0446)	(0.0447)	(0.0447)	(0.0461)
$Migration_j^{adj}$	0.0824*	0.0473	0.0533	0.0514	0.0558	0.0599
	(0.0484)	(0.0486)	(0.0486)	(0.0486)	(0.0486)	(0.0492)
$Migration_j^{rest}$	-0.0938	-0.1593	-0.0024	-0.1614	-0.0035	-0.0083
	(0.1809)	(0.1818)	(0.1816)	(0.1818)	(0.1816)	(0.1817)
Distance	-1.4556***	-1.3079***	-1.6222***	-1.3146***	-1.6242***	-1.6367***
	(0.3390)	(0.3405)	(0.3402)	(0.3410)	(0.3407)	(0.3417)
Adjacency	1.3076***	1.2909***	0.9733***	1.2990***	0.9791***	0.9907***
	(0.4866)	(0.4889)	(0.4884)	(0.4889)	(0.4884)	(0.4886)
Economy size	1.3044***	1.2673***	1.1223***	1.2717***	1.1242***	1.1360***
	(0.1723)	(0.1731)	(0.1729)	(0.1733)	(0.1731)	(0.1731)
R^2	0.8029	0.7943	0.8076	0.7941	0.8075	0.8073
n	1536	1536	1536	1536	1536	1536

Each estimate is from a separate OLS regression with the logarithm of US-Mexico state-state exports in US dollars plus one as the dependent variable, employing the base sample of the 48 contiguous US states and 32 Mexican states. All regressions include state fixed effects, controlling for any existing systematic differences across states that may affect all states' outcomes. Heteroskedasticity-consistent robust standard errors are reported in parentheses. ***, ** and ** denote statistical significance at the 1%, 5% and 10% levels, respectively.

6. Extension: An Application of Generalized Propensity Scores

EVN and SR point to the existence of nonlinearities in the migration-trade relationship, EVN finding an exhaustion point beyond which further migration no longer makes a positive marginal contribution to international trade. In this extension of the initial OLS examination, I apply generalized propensity scores to a continuous treatment (migrant stock levels), flexibly permitting the existence of nonlinearities. GPS estimation provides the advantage of describing the pro-trade effect in detail over the entire spectrum of observed migrant stocks as the resulting estimated dose-response function reflects the expected outcome (exports) associated with each and every observed treatment (migrant stocks) under examination, not just the general elasticity of exports to migration. The use of this methodology is particularly attractive given the importance of addressing three central questions of interest in the migration-trade link: (1) is there a minimum level of migration required to generate positive returns (measured in terms of marginal exports), (2) is there a level of migration corresponding to a saturation point, beyond which positive marginal exports are completely exhausted, and (3) is there a certain migrant stock size that maximizes the pro-trade effect of migration.

6.1. Method

As the GPS is based on the comparison of those observations demonstrating a certain level of homogeneity across observable characteristics, the method permits correction of the potential bias caused by selection into varying levels of treatment intensity, also allowing for the estimation of the trade outcomes associated with each of these different levels of treatment intensity. Propensity score methods have been applied to binary treatments (Heckman et al., 1997), multiple treatments (Imbens, 2000), and most recently, continuous treatments (Hirano and Imbens, 2004). As I consider differing levels of state-state migrant stocks as widely varying doses across the spectrum of a continuous treatment in a quasi-experimental setting, I rely on the methodology outlined in the latter of the propensity score applications.

Observing treatment dosis T_i, the vector of observable covariates X_i, and the outcome variable $Y_i = Y_i(T_i)$ associated with the received treatment for all state-state pairs $i = 1,...,N$, the goal of GPS estimation is ultimately pinpointing the dose-response function,

$$\mu(\tau) = E[Y_i(\tau)], \tag{3}$$

interpreted as the average outcome associated with the specific treatment intensity τ. This clearly highlights one advantage of GPS estimation; it allows for the estimation of the average outcome associated with each and every observed *treatment intensity* of the independent variable of focus.

The central assumption from Hirano and Imbens (2004) is that of weak unconfoundedness for continuous treatments, defined as

$$Y(\tau) \perp T \mid X \text{ for all } \tau \in T \ , \tag{4}$$

i.e. for all possible realizations of treatment intensity, the outcome variable must reflect conditional independence. This assures that any difference in treatment intensities is independent of the corresponding outcomes, after controlling for the observable covariates X. Assuming $g(\tau,x)$ to be the conditional density of the treatment given the set of covariates, i.e.

$$g(\tau,x) = f_{T|X}(\tau \mid x) \ , \tag{5}$$

the GPS is in turn defined as

$$G = g(T,X). \tag{6}$$

Just as in other applications of the propensity score, the GPS is characterized by a balancing property in which the probability that $T = \tau$ within strata of the GPS is independent of the set of covariates X. In turn, removing potential bias requires two steps: first, the estimation of the conditional expectation of the outcome given the treatment and the GPS,

$$\beta(\tau, g) = E[Y \mid T = \tau, G = g]; \tag{7}$$

second, the estimation of the dose-response function as the average of the conditional expectation over the GPS at a particular treatment intensity,[20]

$$\mu(\tau) = E[\beta(\tau, g(\tau, X))]. \tag{8}$$

6.2. Estimation of the effect of migration on exports

After logarithmic transformation, the treatment variable of state-state migrant stocks is approximately normal, with skewness of -0.14 and kurtosis of 2.52, so I assume a normal distribution in estimating the conditional distribution of migration given the vector of chosen covariates:

$$lnT_i|X_i \sim N\left(\beta_0 + X_i\beta_1, \sigma^2\right). \tag{9}$$

β_1 is a column vector, while X_i is a row vector consisting of a variety of observable push and pull determinants of treatment intensity. Population and gross state product (GSP) enter X_i as logarithmic transformations, as squares of those logarithmic transformations, and as growth variables for both Mexican states of origin and US states of destination, intentionally allowing for a flexible, non-linear relationship between these measures of market size and migrant stocks.[21] Furthermore, I include the standard geographic variables of adjacency and distance as covariates; the former as a binary variable taking the value of 1 for adjacent state-state pairs and 0 for non-adjacent state-state pairs, and the latter as a logarithmic transformation of the distance in miles by land from the respective Mexican state capital to the respective US state capital.[22] In addition, I insert unemployment rates and Gini coefficients for both Mexican states of origin and US states of destination into the X_i vector, controlling for scarcity of employment availability and

[20] See Hirano and Imbens (2004) for the proof that these two steps actually remove bias.

[21] In other specifications, I also included cubic terms of both population and GSP for Mexican and US states, however, while explanatory power $\left(R^2\right)$ increased slightly, none of the additional coefficients exhibited high statistical significance. Estimated coefficients did not change in any significant way compared to those reported.

[22] This differs from the standard circle distance used by much of the gravity literature, however I choose this measure given that the trade data captures only trade by land and a majority of Mexico-US migration is by land, as well.

income inequality, respectively.[23] Both measures are frequently included as push and pull factors in the determinants of migration (Clark et al., 2007; Grogger and Hanson, 2011; Mayda, 2010). I estimate equation (9) using scaled ordinary least squares, highlighting the results in Table 7.[24]

Table 7: OLS estimation, dependent variable: ln(migrants)

Covariates and statistics	Estimated coefficient	Standard error
US pop.	20.31***	4.18
US pop.^{2}	-0.75***	0.14
Mex. pop.	-5.82**	2.28
Mex. pop.^{2}	0.26***	0.08
US GSP	-7.46**	3.14
US GSP^2	0.48***	0.14
Mex. GSP	-0.03	1.75
Mex. GSP^2	-0.05	0.09
US pop. growth	6.58***	0.44
Mex. pop. growth	-7.12***	0.75
US GSP growth	2.11***	0.68
Mex. GSP growth	2.05***	0.46
Adjacency	-0.17	0.46
Distance	-3.00***	0.14
US unempl.	0.32***	0.04
Mex. unempl.	0.10**	0.04
US Gini	-33.60***	2.38
Mex. Gini	19.86***	13.32
Constant	-47.89***	16.70
Observations	1488	
R^2	0.76	

In general, the selected covariates are highly statistically significant, all independent variables being significant at the 2% level or lower except for the first- and second-order "Mex. GSP" variables and "Adjacency;" furthermore, the majority of the variation in treatement intensity (size of migrant stocks) is indeed explained by the covariates making up the vector X_i, the R^2 signaling this portion as 76%.

With the OLS estimation in hand, I construct the GPS as

[23] I take US population data, US unemployment rates, and US Gini coefficients from the BEA, Census Bureau and Bureau of Labor Statistics, respectively; data for Mexican states come from the INEGI.

[24] Both exports and migration are expressed as logarithms of the respective variable plus one. Results are equivalent to those obtained by maximum likelihood estimation, the method suggested for this initial estimation in Bia and Mattie (2008).

$$\hat{G}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{2\hat{\sigma}^2}\left(T_i - \hat{\beta}_0 - X_i\hat{\beta}_1\right)^2\right). \tag{10}$$

After generating the propensity scores, the balancing property must be tested in order to verify that the GPS indeed improves the balance of covariates, thereby providing confirmation of the first step necessary for bias removal. I follow the group and block method suggested by Hirano and Imbens (2004) in carrying out the balance check.

First, dividing the observations into four treatment intensity groups allows for the comparison of covariates across these quartiles of the migrant stock distribution before balancing on the GPS. The four groups of treatment intensity contain 373, 371, 372 and 372 state-state observations, respectively.[25] As no GPS adjustment has yet occurred, the left section of Table 10 clearly reflects the great disparities across groups in the covariates of vector X_i , showing an average t-statistic of 5.46 and 66% of observable covariate comparsions across treatment intensity groups being statistically different at the 5% level. If left alone, these disparities lead to obvious concerns of biased inference due to selection into treatment intensity groups determined by observable characteristics.

Second, dividing the observations into blocks or strata according to the GPS allows for comparison of covariates across the treatment intensity quartiles, but now balanced on the GPS. I evaluate the GPS for all observations $i = 1,...,N$ using the OLS estimates at the median level of treatment intensity T_{mj} for each of the four quartiles $j \in \{1,2,3,4\}$, then dividing the propensity scores into ten blocks based on the resulting GPS estimate deciles for each of the four corresponding T_{mj}. Just as before adjustment, I carry out two-tailed t-tests in order to measure the balance of covariates present comparing across groups, weighting the t-statistics by the respective number of observations in each block. However, having blocked on the GPS estimates allows me to now compare observations that have similar observable characteristics and hence the same *predicted* treatment intensity (same GPS block), but also have differing levels of *actual* treatment intensity (different migrant stock quartiles).[26] The middle section of Table 10 highlights the vast improvement in balancing the covariates achieved by employing the GPS; 79% of all covariate comparisons across treatment intensity groups exhibit no statistical differences at the 5% level, with average t-statistics displayed of 1.22.

[25] Appendix B details robustness checks using the GPS method with two alternative samples.

[26] The group and block method is highlighted in Tables 8 and 9, for the entire sample and for the modified common support, respectively.

Table 8: Groups and blocks, full sample $n = 1488$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	37	958	37	651	37	565	37	738
2	37	64	37	108	37	122	37	123
3	38	30	37	65	37	76	37	100
4	37	17	37	81	38	72	37	47
5	38	18	37	54	37	71	37	25
6	37	14	38	22	37	73	38	29
7	37	5	37	30	38	44	37	15
8	38	2	37	52	37	29	37	13
9	37	4	37	31	37	41	37	13
10	37	3	37	23	37	23	37	14

Table 9: Groups and blocks, modified common support sample $n = 1429$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	33	944	37	592	37	506	35	742
2	34	62	37	108	37	122	35	132
3	33	29	37	65	37	76	35	78
4	34	26	37	81	38	72	35	28
5	34	12	37	54	37	71	35	27
6	33	10	38	22	37	73	35	19
7	34	3	37	30	38	44	35	14
8	33	1	37	52	37	29	35	13
9	34	5	37	31	37	41	35	12
10	33	2	37	23	37	23	35	14

Some studies employing propensity score estimation additionally rely on the common support condition in order to improve comparability of observations. The use of the common support simply results in the exclusion of any observations in the given sample that do not demonstrate a certain level of similarity in the observable covariates. Given the group and block method, this translates into comparing the GPS calculations for $\hat{G}_k\left(T_{mj}, X_k\right)$ with those of $\hat{G}_l\left(T_{mj}, X_l\right)$, where $k \in j$ and $l \notin j$. In turn, the only observations used in the remainder of the estimation process are those l observations where

$$\min\left\{\hat{G}_k\left(T_{mj}, X_k\right)\right\} \leq \hat{G}_l\left(T_{mj}, X_l\right) \leq \max\left\{\hat{G}_k\left(T_{mj}, X_k\right)\right\} \text{ for all } j \in \{1,2,3,4\}. \tag{9}$$

A potential dilemma arises as only 390 of the 1488 state-state observations in fact meet the common support condition stated in (9). Faced with this large loss of information, a decision must be made among three standard solutions: (1) estimate the outcome variable only within the common support, thereby maximizing the reduction in bias, but also reducing the range over which

Table 10: Three-stage balancing comparison of covariates

T-stats	Before balancing on GPS $n = 1488$				After balancing on GPS (weighted t-stats) $n = 1488$				Common Support after balancing on GPS (weighted t-stats) $n = 1429$			
Covariate	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
US pop.	21.72	0.12	-4.04	-16.19	2.52	-0.41	0.41	-3.49	0.83	-0.30	0.42	-1.28
US pop.^{2}	21.19	0.29	-3.69	-16.37	2.39	-0.40	0.46	-3.62	0.79	-0.30	0.47	-1.26
Mex. pop.	9.45	4.74	-2.67	-11.69	2.25	1.41	-0.25	-2.65	1.21	1.58	-0.18	-1.82
Mex. pop.^{2}	9.30	4.80	-2.55	-11.74	2.17	1.42	-0.23	-2.69	1.17	1.57	-0.16	-1.86
US GSP	21.39	0.18	-3.63	-16.50	2.30	-0.42	0.41	-3.65	0.61	-0.30	0.44	-1.35
US GSP^2	20.56	0.41	-3.15	-16.66	2.11	-0.41	0.50	-3.79	0.56	-0.31	0.51	-1.33
Mex. GSP	5.40	0.08	-0.39	-5.07	1.35	0.32	0.39	-1.10	0.96	0.36	0.41	-0.36
Mex. GSP^2	5.26	0.08	-0.31	-5.02	1.31	0.30	0.40	-1.13	0.94	0.34	0.42	-0.38
US pop. growth	6.52	2.16	-2.22	-6.46	1.51	0.81	-0.23	-2.11	1.27	0.72	-0.28	-1.43
Mex. pop. growth	-2.87	2.67	0.27	-0.07	0.36	0.57	0.21	-0.28	0.51	0.55	0.19	-0.50
US GSP growth	-6.76	-0.12	3.64	3.18	0.10	0.77	0.92	-0.08	0.04	0.66	0.85	0.24
Mex. GSP growth	-0.88	1.08	1.61	-1.78	-0.29	0.44	0.81	0.31	-0.32	0.45	0.82	0.15
Adjacency	1.74	1.73	0.96	-4.48	-0.59	0.73	0.41	-6.25	0.26	0.34	-0.32	0.13
Distance	-14.03	-0.47	7.18	6.72	-4.75	-0.47	1.62	4.14	-4.18	-0.45	1.95	0.20
US unempl.	10.04	-0.10	-2.43	-7.26	1.03	0.32	0.55	-0.17	0.42	0.42	0.60	-0.34
Mex. unempl.	1.31	-1.67	-3.10	3.41	1.62	0.01	-0.70	0.55	1.51	0.00	-0.66	0.91
US Gini	9.40	-1.23	-1.75	-6.19	1.37	-0.31	0.47	-1.91	0.32	-0.26	0.46	-0.59
Mex. Gini	6.18	4.31	-3.50	-7.00	1.46	0.40	-0.87	-0.68	1.16	0.53	-0.83	-0.52
Avg. absolute t-stat	5.46, 24/72<| 1.96 |				1.22, 57/72<| 1.96 |				0.71, 71/72<| 1.96 |			

exports can be predicted given the observed migration levels; (2) estimate the outcome variable inside and outside the common support, thereby maximizing the reduction in bias, maximizing the range over which exports can be predicted given the observed migration levels, but also reducing the preciseness of the estimated outcomes; or (3) estimate the outcome variable with all available observations, thereby accepting a non-maximized reduction in potential bias, maximizing the range over which exports can be predicted given the observed migration levels, and maximizing the preciseness of the estimated outcomes. As exhibited in Table 10, much of the potential bias is indeed reduced simply by balancing on the GPS, without any consideration for the common support; 79% of all covariate comparisons across treatment intensity groups exhibit no statistical differences at the 5% level. On the other hand, using only the common support observations would greatly reduce the range of observed migrant stocks from 0 - 227,032 (all observations) to

2 - 5,878 (common support observations), corresponding to a dramatic reduction in means and medians of migrant stocks from 2406 to 347 and 210.5 to 134, respectively. Further evidence as to why the common support condition may be too stringent for our purposes is provided by the simple comparison of covariate means between those state-state combinations inside and outside the common support region. Table 11 highlights these comparisons, and perhaps surprisingly, the two groups do not appear to differ dramatically in the means of the covariates.

Table 11: Covariate means: included vs. excluded observations

Covariates	Mean of common support observations	Mean of excluded observations
US pop.	15.14	15.18
Mex. pop.	14.50	14.70
US GSP	12.04	12.09
Mex. GSP	9.77	9.86
US pop. growth	0.03	0.03
Mex. pop. growth	0.09	0.09
US GSP growth	0.07	0.07
Mex. GSP growth	0.08	0.08
Adjacency	0.00	0.01
Distance	7.58	7.60
US unempl.	5.66	5.67
Mex. unempl.	3.82	3.77
US Gini	0.45	0.45
Mex. Gini	0.47	0.48

This fact points to the observation that exclusion from the common support region is mostly attributed to a lack of similarity of generalized propensity scores corresponding to just one of the four median treatment GPS group calculations, not a general lack of comparability of observables for the excluded observations across all four median treatment GPS groups. As equation (9) highlights, the condition for inclusion in the common support region indeed depends on overlap of the GPS in each and every group j , therefore if any state-state observation has even only one exception to this rule, it is automatically excluded by the common support condition. Given the combination of the evidence mentioned, there is no clear best option of the three standard solutions, although accepting the trade-off of a non-maximized reduction in bias in exchange for a maximized amount of information, range over which exports can be predicted, and preciseness in estimation may be the most attractive.

However, following Lechner (2008) in exploring alternative solutions to the common support problem, I continue the estimation process by pursuing a fourth option, one that I argue permits researchers confronted with similar common support dilemmas a certain amount of flexibility that is extremely useful in modifying the common support condition according to the particular needs of the research undertaken. By slightly relaxing the stringent condition requiring presence of observations in the common support region for all four groups j , we can still assure a maximized reduction of bias, while not trading off as much coverage and preciseness in terms of estimated outcomes. If indeed the standard common support condition means the researcher must sacrifice a large portion of information as in this paper, this condition relaxation provides a second-best option that can be applied to the data. My modified common support rule simply relaxes equation (9), now proposing that the only observations used in the remainder of the estimation process are those l observations where

$$\min\left\{\hat{G}_k\left(T_{mj},X_k\right)\right\} \leq \hat{G}_l\left(T_{mj},X_l\right) \leq \max\left\{\hat{G}_k\left(T_{mj},X_k\right)\right\} \text{ for at least two } j \in \{1,2,3,4\}. \tag{10}$$

That is, state-state observations are included as long as the corresponding GPS median-treatment scores for at least two of the four treatment quartiles fall within the common support region of the particular quartile. Guaranteeing a high level of comparability, while relaxing the condition from equation (9) results in a new sample that does not suffer from the great loss of information previously seen; only 59 state-state observations are excluded by the modified common support, resulting in the preferred sample of 1429 observations, and the range of migrant stocks over which exports can be estimated is not reduced at all.[27] Additionally, the balancing of covariates is greatly improved through the use of the modified common support condition. The right section of Table 5 highlights this dramatic improvement; 71 of 72 covariate comparsion groups show no statistical differences at the 5% level, with an average t-statistic of 0.71. In turn, the high level of comparability of observations vindicates the use of the modified common support for this data set, as the guarantee of comparability of observations and the resulting reduction in bias is the very reason for adhering to a common support condition in the first place.

Moving on with only those observations meeting the modified common support condition, as a first step in estimating the dose-response function, I estimate the conditional expectation of exports given treatment intensity and the corresponding GPS,

$$E\left[Y_i|T_i,G_i\right] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 + \alpha_4 \hat{G}_i + \alpha_5 \hat{G}_i T_i. \tag{11}$$

[27] Appendix B lists the 59 state-state observations excluded from the $n = 1488$ sample, as well as those observations excluded from the $n = 1536$ and $n = 1426$ samples. It is important to point out that the selection of at least two js and four treatment quartiles is the result of trying several variations of the modified common support condition; however, rather than being a drawback of the approach, this actually provides the researcher with the advantage of flexibility while assuring comparability of observations. In other words, it is easy to consider modified conditions that differ along the two mentioned dimensions, the number of js and the number of original treatment groups, for example conditions requiring three of four quartiles, four of five quintiles, etc. Much as the Akaike Information Criterion (AIC) provides a means for model selection through trade-off between goodness of fit and model complexity, selection of the appropriate modified common support condition is dictated by the desired trade-off of the amount of balancing and bias reduction sacrificed compared to the amount of estimation power gained under each modification.

T_i values are the actual observed migrant stocks, while $\hat{G}_i$ values are those estimates calculated from equation (8). The resulting OLS estimates in Table 12 have no direct economic interpretation (Hirano and Imbens, 2004), however the individual and joint statistical significance of the GPS coefficients is noteworthy.[28] This significance signals that selection is important, confirming that the inclusion of the GPS terms and the GPS estimation process in general is indeed worthwhile in achieving some level of selection bias removal.

Table 12: OLS estimation, dependent variable: ln(exports)

Covariates and statistics	Estimated coefficient	Standard error
Migrant stock	1.06***	0.30
Migrant stock^2	-0.18***	0.07
Migrant stock^3	0.01**	0.00
GPS	-5.65**	2.54
Migrant stock*GPS	1.54***	0.44
Constant	11.75***	0.60
Observations	1429	
R^2	0.11	

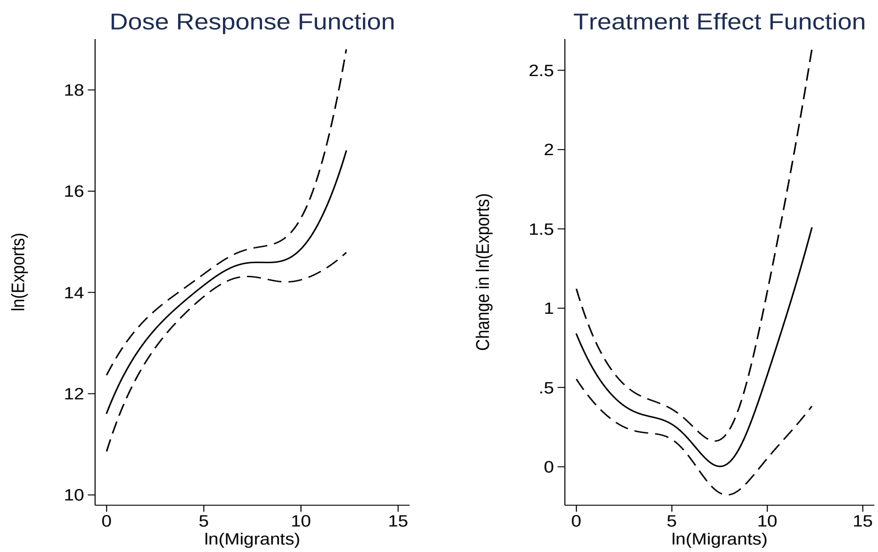
Finally, I estimate the dose-response function, capturing the average potential outcome at each and every treatment intensity τ :

$$\widehat{E\left[Y_\tau\right]} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\alpha}_0 + \hat{\alpha}_1\tau + \hat{\alpha}_2\tau^2 + \hat{\alpha}_3\tau^3 + \hat{\alpha}_4\hat{G}\left(\tau, X_i\right) + \hat{\alpha}_5\hat{G}\left(\tau, X_i\right)\tau\right]. \tag{12}$$

I report both the dose-response function and its derivative, the treatment effect function, in Figure 3.

[28] I choose to include the GPS only linearly as higher-order GPS terms (squared and cubed) do not add extra information and are not statistically significant.

Figure 3



90 percent confidence intervals, represented by dashed lines, are constructed by bootstrapping.

6.3. Marginal Contributions

By calculating the marginal exports associated with each treatment level,

$$MarExp_p = \left(E_{p+1} - E_p\right) \big/ \left(M_{p+1} - M_p\right) \ , \tag{13}$$

where E and M represent estimated exports and actual migrant stocks (backed out from the respective logarithmic transformations) and p denotes the ordinal value of treatment intensities employed in the estimation of the dose-response function, the nonlinearities present in the exports-migration relationship are now clearly on visual display in Figures 4 and 5.

Considering the benchmark OLS state-state estimate of \$3061 extra exports per extra migrant, this average contribution of migration to exports is clearly weighted by the first migrants from one respective Mexican state to the state of US residence. The first migrant makes a marginal contribution of \$76,297 to exports, the hundredth contributes \$3613, and the thousandth contributes \$175. Marginal contributions of \$8371, \$1879, and \$166, respectively, correspond to the 25th, 50th, and 75th percentiles ($M = 33,\ 216,$ and 1026) of the actual state-state migrant stock distribution for the $n = 1429$ sample. Marginal exports of \$3061 matches to a migrant stock of 123, representing the 42nd percentile of the actual state-state migrant stock distribution. Interestingly, at the level of 2276 migrants, the marginal contribution is temporarily "exhausted," and dips below zero until 3761 migrants; this negative marginal contribution to exports bottoms out at -$4.53. However, only 66 of the 1429 (less than 5%) state-state migrant stocks fall in this exhaustion zone range. Total exports quickly recovers from the shortlived negative effect of migration, reaching the pre-exhaustion zone level of exports as early as 4900 migrants, with subsequent migrants all making increasingly positive marginal contributions to state-state exports.

The estimated dose-response and treatment-effect functions provide clear answers to the three central questions of interest posed at the beginning of this paper. First, there is no minimum level of immigration (other than one) necessary to generate positive returns in terms of exports; in fact, an individual migrant has the largest pro-trade effect when there are few migrants of the corresponding state-state classification. Second, while there does exist a small range of migrant stocks over which the marginal contribution of migrants turns slightly negative, this negative contribution is extremely temporary, as further migration returns the marginal contribution increasingly positive over the remaining range of state-state migrant stocks. Finally, because the marginal contribution remains positive beyond the exhaustion zone, the pro-trade effect of migration can only be maximized by the maximum level of observed state-state migration.

Figure 4: Marginal contributions for migrants > 500

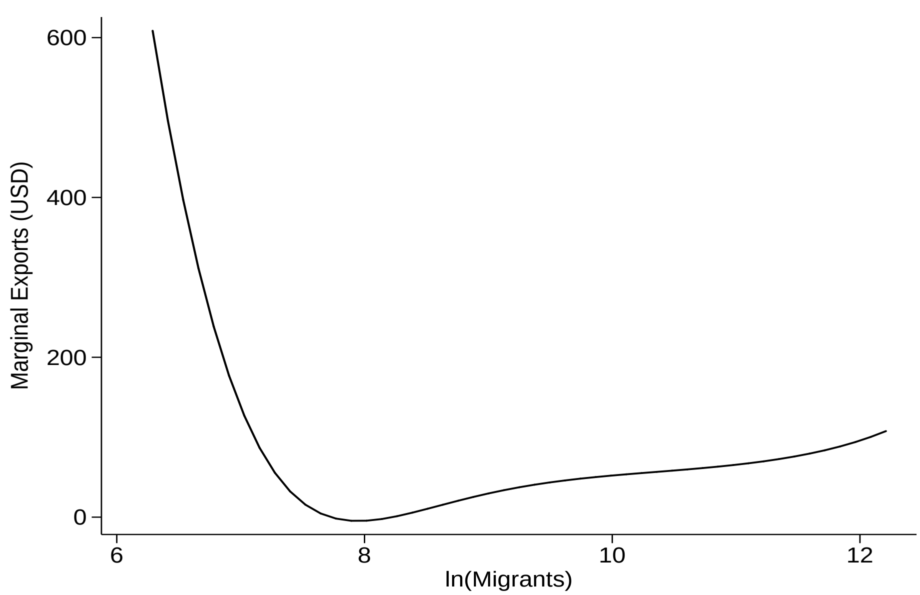
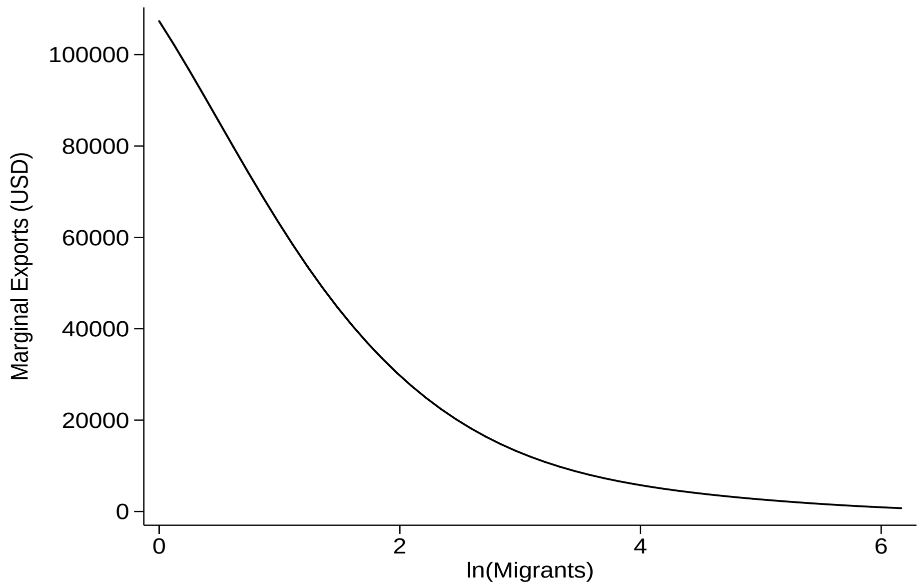


Figure 5: Marginal contributions for migrants < 500



7. Conclusion

Migrants indeed create a significant force in promoting extra trade from US states of residence to Mexican states of origin. This finding is empirically consistent not only in statistical significance, but also in magnitude across the varied methods, specifications, and samples employed in this paper. Without consideration of potential nonlinearities and differing geographic proximities in an augmented gravity model, the elasticity of state-state exports to immigration is 0.11, translating into \$3061 extra annual exports per average extra migrant for a particular US-Mexico state-state combination, holding other factors constant. On the other hand, the application of generalized propensity scores permits the potential of nonlinearities in the migration-trade relationship, results pointing to a diminishing yet positive marginal contribution of migration to exports as migrant stock size increases over most of the range of measured migrant stocks. These results contribute the first evidence of the pro-trade effect of migration at the state-state level, a relatively localized level capable of measuring more accurately the potential determinants of trade and differentiating between migrant networks of varied state origin.

Additionally, this paper unmasks the importance of distinct geographic proximities that the use of one migration variable disguises. Through the examination of not only in-state migraton, but also neighboring-state and other-state migration, geographic proximity is revealed to indeed be a relevant factor in determining the pro-trade effect of migration, with networks suffering lower amounts of spatial separation making larger contributions to trade. Both in-state and neighboring-state migration make significantly positive contributions to state-state exports, with estimated elasticities of 0.07 and 0.08, respectively, resulting in partial contributions to average state-state exports of $1984 and $538. Combining the three contributions from migration of separate geographic proximities gives an overall addition to state-state exports of $3779 by the average extra migrant.

Empirical studies employing data sets from countries other than the US and Mexico, as well as data detailing characteristics such as migrants' education level and participation in business networks, provide clear avenues for further research at the state-state level, just as they already have at the country-country level. In addition, the very existence of the exhaustion zone provides another direction for related research specifically focusing on the GPS method: first, in verifying that a similar shape of the dose-response and treatment-effect functions obtains using other data, and second, in hypothesizing why it is that the exhaustion zone may exist yet does not extend to the larger migrant stocks providing positive contributions to state-state exports. Finally, the results not only shed light on how localized migration's nexus with trade may be and how geographic proximity matters, they inevitably connect to the ongoing debate in a host of countries as to the economic costs and benefits of migration. Without a doubt, the pro-trade effect of both in-state and neighboring-state migration cannot be ignored in any careful analysis of the costs and benefits of migration.

References

Aleksynska, M., Peri, G., 2011. Isolating the network effect of immigrants on trade, Working Paper 2011/26, CEPII.

Anderson, J., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle, *American Economic Review*, 93 (1), 170-192.

Artal-Tur, A., Pallardo-López, V., Requena-Silvente, F., 2012. The trade-enhancing effect of immigration networks: New evidence on the role of geographic proximity, *Economics Letters*, 116 (3), 554-557.

Bandyopadhyay, S., Coughlin, C., Wall, H., 2008. Ethnic networks and US exports, *Review of International Economics*, 16 (1), 199-213.

Briant, A., Combes, P., Lafourcade, M., 2009. Product complexity, quality of institutions and the pro-trade effect of immigrants, Working Paper 2009/6, Paris School of Economics.

Caponi, V., Plesca, M., 2012. Empirical characteristics of legal and illegal immigrants in the US, *mimeo*, Ryerson University.

Card, D., Lewis, E., 2007. The diffusion of Mexican immigrants during the 1990s: explanations and impacts, in Mexican Immigration to the United States, Ed. G. Borjas, University of Chicago Press, Chicago, pp. 193-227.

Dunlevy, J., 2006. The influence of corruption and language on the pro-trade effect of immigrants: evidence from the American states, *The Review of Economics and Statistics*, 88 (1), 182-186.

Egger, P., von Ehrlich, M., Nelson D., 2012. Migration and trade, *World Economy*, 35 (2), 216-241.

Canavire Bacarreza, G., Ehrlich, L., 2006. The impact of migration on foreign trade: a developing country approach, MPRA Paper 1090, University Library of Munich.

Felbermayr, G., Jung, B., 2009. The pro-trade effect of the brain drain: sorting out confounding factors, *Economics Letters*, 104, 72-75.

Girma, S., Yu, Z., 2002. The link between immigration and trade: evidence from the United Kingdom, *Weltwirtschaftliches Archiv*, 138 (1), 115-130.

Gould, D., 1994. Immigrant links to the home country: empirical implications for US bilateral trade flows, *The Review of Economics and Statistics*, 76 (2), 302-316.

Hatzigeorgiou, A., Lodefalk, M., 2011. Trade and migration: firm-level evidence, Working Paper 2011/6, Swedish Business School.

Head, K., Ries, J., 1998. Immigration and trade creation: econometric evidence from Canada, *Canadian Journal of Economics*, 31 (1), 47-62.

Herander, M., Saavedra, L., 2005. Exports and the structure of immigrant-based networks: the role of geographic proximity, *The Review of Economics and Statistics*, 87 (2), 323-335.

Instituto de los Mexicanos en el Exterior, 2013. http://www.ime.gob.mx/index.php?option=com_content&view=article&id=19&Itemid=507&lang=es URL (accessed 23 January 2013).

Instituto Nacional de Estadística y Geografía, 2013. http://www.inegi.org.mx/sistemas/mexicocifras/default.aspx?e=1 URL (accessed 23 January 2013).

Kugler, M., Rapoport, H., 2011. Migration, FDI and the margins of trade, CID Working Paper 2011/222, Harvard University.

Passel, J., Cohn, D., 2009. A portrait of unauthorized immigrants in the United States, Research Report, Pew Hispanic Center.

Peri, G., Requena, F., 2010. The trade creation effect of immigrants: evidence from the remarkable case of Spain, *Canadian Journal of Economics*, 43 (4), 1433-1459.

Serrano-Domingo, G., Requena-Silvente, F., 2013. Re-examining the migration-trade link using province data: An application of the generalized propensity score, *Economic Modelling*, 32, 247-261.

Tadesse, B., White, R., 2010. Cultural distance as a determinant of bilateral trade flows: do immigrants counter the effect of cultural differences?, *Applied Economics Letters*, 17 (2), 147-152.

US Bureau of Economic Analysis, 2013. http://www.bea.gov/iTable/iTable.cfm?ReqID=70&step=1&isuri=1&acrdn=1 URL (accessed 23 January 2013).

US Bureau of Transportation, 2013. http://www.bts.gov/programs/international/transborder/TBDR_QA.html URL (accessed 23 January 2013).

White, R., 2007. An examination of the Danish immigrant-trade link, *International Migration*, 45 (5), 61-86.

Yilmazkuday, H., 2013. Understanding the distance puzzle through the location of production, *mimeo*, Florida International University.

Appendix A: Migration and Trade

Table A.1: *Matrículas consulares* registered 2006 to 2010, US states of residence and Mexican states of origin: *total of 4,659,656*

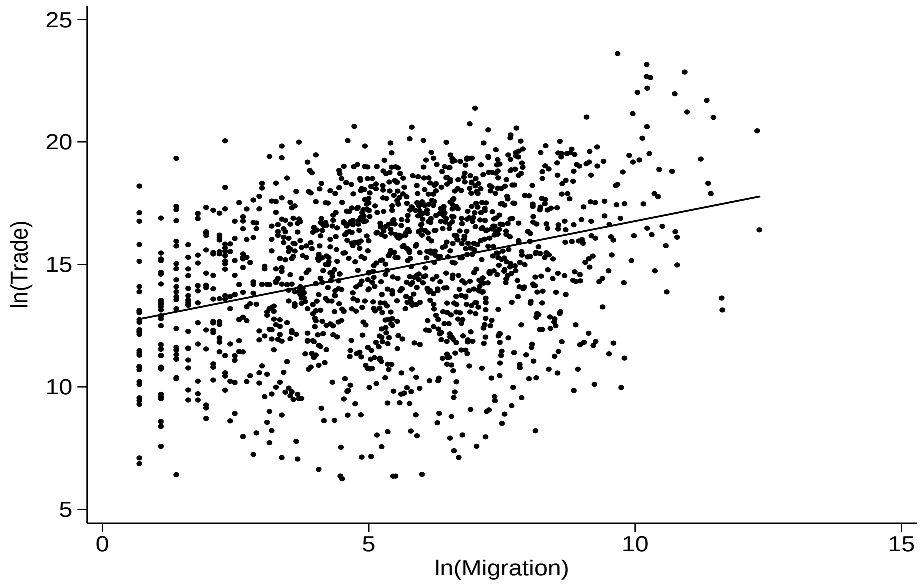
US		Mexico
Alabama 27,442	Oklahoma 19,867	Aguascalientes 42,799
Arizona 187,032	Oregon 74,103	Baja California 45,183
Arkansas 19,711	Pennsylvania 23,555	Baja California Sur 3,178
California 1,682,667	Rhode Island 931	Campeche 7,017
Colorado 105,125	South Carolina 38,551	Chiapas 62,697
Connecticut 10,645	South Dakota 723	Chihuahua 120,933
Delaware 6,637	Tennessee 38,736	Coahuila 56,687
Florida 107,392	Texas 779,636	Colima 32,326
Georgia 150,704	Utah 55,330	Durango 126,923
Idaho 16,340	Vermont 345	Guanajuato 377,674
Illinois 387,377	Virginia 24,492	Guerrero 371,279
Indiana 69,247	Washington 64,436	Hidalgo 131,280
Iowa 15,953	West Virginia 663	Jalisco 425,607
Kansas 21,981	Wisconsin 43,532	Mexico 238,343
Kentucky 14,428	Wyoming 3,563	Mexico City 293,920
Louisiana 8,074		Michoacan 525,514
Maine 232		Morelos 105,732
Maryland 20,729		Nayarit 72,227
Massachusetts 2,629		Nuevo Leon 77,824
Michigan 22,417		Oaxaca 283,295
Minnesota 38,019		Puebla 307,606
Missouri 17,103		Queretaro 58,608
Mississippi 5,600		Quintana Roo 3,470
Montana 164		San Luis Potosi 155,069
Nebraska 22,291		Sinaloa 91,019
New Hampshire 854		Sonora 49,074
New Jersey 73,881		Tabasco 72,502
New Mexico 53,212		Tamaulipas 98,290
New York 133,625		Tlaxcala 35,293
Nevada 108,310		Veracruz 205,799
North Carolina 142,813		Yucatan 17,837
North Dakota 54		Zacatecas 170,686
Ohio 18,505		

Table A.2 Migration and trade in top US states of Mexican migrant residence

	California		Texas		Illinois	
Ranking	Migration	Exports	Migration	Exports	Migration	Exports
1	Michoacan	Baja California	Guanajuato	Chihuahua	Michoacan	Mexico City
2	Jalisco	Mexico	San Luis Potosi	Tamaulipas	Guerrero	Nuevo Leon
3	Guerrero	Chihuahua	Tamaulipas	Mexico City	Guanajuato	Mexico
4	Oaxaca	Mexico City	Nuevo Leon	Mexico	Jalisco	Jalisco
5	Mexico City	Jalisco	Michoacan	Coahuila	Mexico City	Coahuila
6	Guanajuato	Sonora	Guerrero	Nuevo Leon	Mexico	San Luis Potosi
7	Puebla	Nuevo Leon	Zacatecas	Guanajuato	Veracruz	Chihuahua
8	Mexico	Sinaloa	Mexico	Jalisco	Durango	Sonora
9	Zacatecas	Tamaulipas	Coahuila	Queretaro	Puebla	Queretaro
10	Sinaloa	Puebla	Mexico City	Aguascalientes	Zacatecas	Baja California
11	Nayarit	Baja California Sur	Veracruz	San Luis Potosi	Morelos	Durango
12	Veracruz	Queretaro	Jalisco	Veracruz	Oaxaca	Tamaulipas
13	Morelos	Guanajuato	Durango	Hidalgo	San Luis Potosi	Guanajuato
14	Hidalgo	Coahuila	Chihuahua	Sonora	Hidalgo	Puebla
15	Baja California	Aguascalientes	Hidalgo	Durango	Tabasco	Aguascalientes
16	Durango	Durango	Puebla	Tabasco	Queretaro	Hidalgo
17	Tabasco	San Luis Potosi	Queretaro	Baja California	Aguascalientes	Veracruz
18	Colima	Quintana Roo	Oaxaca	Puebla	Chihuahua	Quintana Roo
19	Queretaro	Tlaxcala	Morelos	Michoacan	Nuevo Leon	Tlaxcala
20	Chiapas	Veracruz	Aguascalientes	Sinaloa	Tlaxcala	Sinaloa
21	Sonora	Hidalgo	Tabasco	Morelos	Tamaulipas	Morelos
22	Yucatan	Michoacan	Chiapas	Quintana Roo	Coahuila	Michoacan
23	Aguascalientes	Morelos	Tlaxcala	Campeche	Chiapas	Zacatecas
24	Tlaxcala	Yucatan	Sinaloa	Colima	Nayarit	Tabasco
25	Chihuahua	Campeche	Colima	Zacatecas	Sinaloa	Yucatan
26	San Luis Potosi	Nayarit	Nayarit	Tlaxcala	Baja California	Baja California Sur
27	Coahuila	Tabasco	Campeche	Yucatan	Colima	Oaxaca
28	Tamaulipas	Zacatecas	Baja California	Oaxaca	Sonora	Chiapas
29	Nuevo Leon	Chiapas	Yucatan	Chiapas	Campeche	Colima
30	Campeche	Colima	Sonora	Baja California Sur	Yucatan	Guerrero
31	Quintana Roo	Oaxaca	Quintana Roo	Guerrero	Quintana Roo	Campeche
32	Baja California Sur	Guerrero	Baja California Sur	Nayarit	Baja California Sur	Nayarit

States of origin are listed in order of number of *matrículas consulares* in the period of 2006 to 2010 and value of average state-state exports in the period of 2008 to 2010.

Figure A.1: Migration and trade, simple correlation



Migration is measured as the logarithm plus one of *matrículas consulares* from 2006-2010, while trade is measured as the logarithm plus one of trade in 2010; only values greater than zero are included.

Appendix B: GPS robustness checks

As a first check on the robustness of the results, I carry out the GPS estimation adding the 48 observations corresponding to Mexico City into the sample, resulting in a starting sample size of 1536 state-state relationships, later reduced to 1478 after exclusion according to the modified common support condition. If results are indeed robust to the inclusion of the extra Mexico City observations, the potential worries created by this sampling decision are minimized. Tables B.1 to B.5 and Figure B.1 display the key results for the larger sample, exhibiting only minor changes from the tables and figures highlighted in Section 6. The GPS estimation allows for an improvement in balance from only 36% before the GPS to 76% after the GPS adjustment of covariate comparisons reflecting a lack of statistically significant differences at the 5% level; all but 3 covariate comparisons show lack of statistical significance after both GPS adjustment and exclusion based on the modified common support condition. The dose-response function in Figure B.1 is shifted slightly downward from that displayed in Figure 3, due to a slight decrease across the board in marginal exports associated with the marginal migrant. The inclusion of the Mexico City observations results in a dose-response function that corresponds to marginal contributions of \$8118, \$1678, and \$158, respectively, for the 25th, 50th, and 75th percentiles ($M = 34,\ 228,$ and 1094) of the actual state-state migrant stock distribution for the $n = 1478$ sample. Furthermore, the downward shift results in an increased range for the exhaustion zone, now made up of migrant stocks between 2033 and 4465, corresponding to 7.4% of state-state migrant stocks.

Additionally, as detailed above, concerns may exist as to how well the *matrícula consular* data represents the actual state-state migrant distribution. In turn, I conduct a second check for robustness further excluding all observations associated with Texas and Illinois, the two outlier states from the original sample, resulting in a revised sample of 1426 state-state observations, then reduced to 1380 by exclusion following the modified common support condition.[29] The corresponding dose-response and treatment effect functions are highlighted in Figure B.2, once again reflecting slight changes in outcomes from the original generalized propensity scores estimation. The GPS adjustment and use of the modified common support condition again provide a balancing of covariates, improving the percentage lacking statistically significant differences at the 5% level from 33% to 87% to 94% of all possible covariate comparisons. Without the observations corresponding to Texas and Illinois, the dose-response and treatment effect functions again shift slightly downward, resulting in an exhaustion zone bordered by migrant stocks of 964 and 8845, corresponding to 21.0% of state-state migrant stocks. Marginal contributions of \$9238, \$1641, and \$18, respectively, correspond to the 25th, 50th, and 75th percentiles ($M = 31,\ 201,$ and 914) of the actual state-state migrant stock distribution for the $n = 1380$ sample.

[29]Tables B.6 to B.10 provide main results from the GPS estimation for the $n = 1380$ sample.

Table B.1: OLS estimation, dependent variable: ln(migrants)

Covariates and statistics	Estimated coefficient	Standard error
US pop.	19.33***	4.09
US pop.^{2}	-0.72***	0.14
Mex. pop.	-2.73	1.89
Mex. pop.^{2}	0.15**	0.06
US GSP	-6.61**	3.08
US GSP^2	0.44***	0.13
Mex. GSP	-3.39***	1.05
Mex. GSP^2	0.12**	0.05
US pop. growth	6.62***	0.43
Mex. pop. growth	-6.84***	0.74
US GSP growth	2.20***	0.66
Mex. GSP growth	1.94***	0.45
Adjacency	-0.08	0.45
Distance	-2.92***	0.14
US unempl.	0.32***	0.04
Mex. unempl.	0.10**	0.04
US Gini	-33.02***	2.33
Mex. Gini	20.95***	1.41
Constant	-53.49***	16.32
Observations	1536	
R^2	0.77	

Table B.2: Groups and blocks, full sample $n = 1536$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	38	994	38	643	38	592	38	777
2	39	57	38	130	38	126	38	114
3	39	38	39	68	39	65	39	107
4	38	17	38	88	38	83	39	44
5	39	16	38	46	38	66	38	29
6	39	15	39	43	39	62	38	26
7	38	5	38	26	38	45	39	19
8	39	2	39	52	39	50	39	11
9	39	3	38	29	38	38	38	9
10	38	3	38	28	38	26	38	16

Table B.3: Groups and blocks, modified common support sample $n = 1478$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	34	977	38	593	38	534	36	772
2	35	58	38	124	38	126	37	144
3	35	35	38	67	39	65	36	57
4	36	26	39	89	38	83	37	40
5	35	10	38	45	38	66	36	27
6	35	10	38	43	39	62	36	27
7	34	6	39	26	38	45	37	12
8	36	1	38	52	39	50	37	13
9	34	3	38	29	38	38	36	6
10	35	3	38	28	38	26	36	16

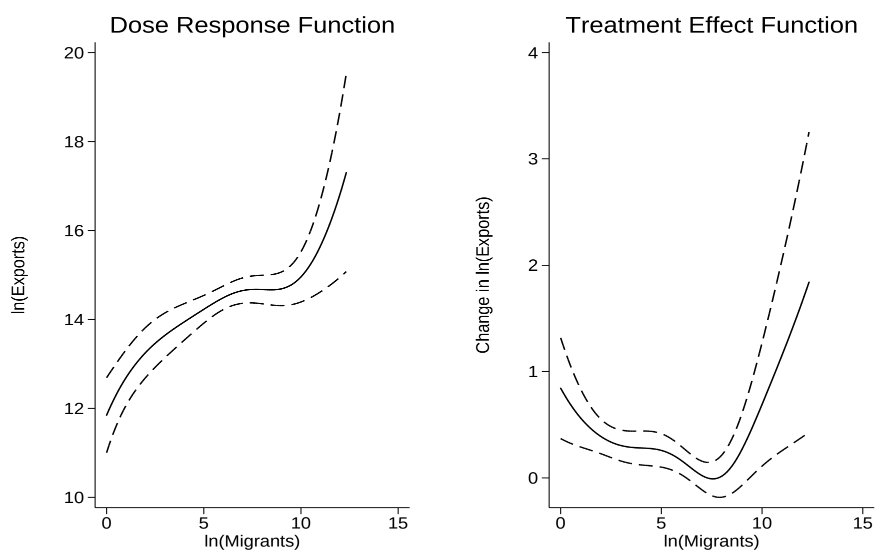
Table B.4: Three-stage balancing comparison of covariates

T-stats	Before balancing on GPS $n = 1536$				After balancing on GPS (weighted t-stats) $n = 1536$				Common Support after balancing on GPS (weighted t-stats) $n = 1478$			
Covariate	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
US pop.	22.15	0.02	-3.60	-17.21	2.66	-0.27	0.59	-4.05	1.07	-0.35	0.61	-1.93
US pop.^{2}	21.63	0.19	-3.26	-17.41	2.52	-0.25	0.64	-4.19	1.03	-0.34	0.65	-1.93
Mex. pop.	9.84	5.53	-3.38	-12.23	2.22	1.48	-0.57	-2.45	1.24	1.62	-0.48	-1.61
Mex. pop.^{2}	9.70	5.61	-3.27	-12.29	2.15	1.50	-0.57	-2.48	1.20	1.64	-0.48	-1.63
US GSP	21.90	0.05	-3.18	-17.56	2.39	-0.26	0.60	-4.19	0.80	-0.35	0.63	-2.01
US GSP^2	21.05	0.29	-2.70	-17.75	2.20	-0.23	0.68	-4.35	0.74	-0.33	0.70	-2.01
Mex. GSP	5.73	1.71	-1.03	-6.44	1.08	0.70	0.09	-1.13	0.69	0.75	0.13	-0.37
Mex. GSP^2	5.59	1.80	-0.97	-6.46	0.97	0.70	0.09	-1.15	0.66	0.76	0.12	-0.40
US pop. growth	6.26	2.13	-2.08	-6.32	1.40	0.83	0.03	-2.02	1.19	0.75	-0.05	-0.88
Mex. pop. growth	-3.03	1.47	0.75	0.81	0.35	-0.20	0.40	-0.27	0.53	-0.21	0.37	-0.31
US GSP growth	-6.91	0.47	2.87	3.50	0.05	-0.93	0.86	-0.04	0.09	0.78	0.78	0.37
Mex. GSP growth	-0.93	0.62	1.71	-1.39	-0.41	0.55	0.85	0.24	-0.45	0.81	0.85	0.12
Adjacency	1.74	1.73	0.96	-4.47	0.59	0.72	0.37	-6.34	0.32	0.42	-0.04	0.00
Distance	-14.13	0.04	7.17	6.31	-4.54	0.09	1.37	4.16	-3.92	0.25	1.64	0.02
US unempl.	10.37	-0.25	-2.23	-7.71	1.13	0.31	0.36	-0.50	0.58	0.36	0.43	-0.47
Mex. unempl.	1.81	-0.63	-2.72	1.54	1.28	0.10	-0.55	0.21	1.15	0.12	-0.51	0.52
US Gini	9.74	-1.24	-1.56	-6.77	1.34	0.08	0.84	-2.51	0.45	-0.13	0.82	-0.99
Mex. Gini	6.48	4.38	-4.11	-6.76	1.83	0.48	-1.00	-0.42	1.50	0.68	-0.93	-0.65
Avg. absolute t-stat	5.66, 26/72<| 1.96 |				1.28, 55/72<| 1.96 |				0.76, 69/72<| 1.96 |			

Table B.5: OLS estimation, dependent variable: ln(exports)

Covariates and statistics	Estimated coefficient	Standard error
Migrant stock	1.11***	0.30
Migrant stock^2	-0.21***	0.07
Migrant stock^3	0.01***	0.00
GPS	-7.58***	2.50
Migrant stock*GPS	1.91***	0.43
Constant	12.03***	0.59
Observations	1478	
R^2	0.13	

Figure B.1



90 percent confidence intervals, represented by dashed lines, are constructed by bootstrapping.

Table B.6: OLS estimation, dependent variable: ln(migrants)

Covariates and statistics	Estimated coefficient	Standard error
US pop.	20.11***	4.16
US pop.^{2}	-0.75***	0.14
Mex. pop.	-6.46***	2.31
Mex. pop.^{2}	0.28***	0.08
US GSP	-8.01***	3.15
US GSP^2	0.50***	0.14
Mex. GSP	0.11	1.77
Mex. GSP^2	-0.06	0.09
US pop. growth	6.71***	0.44
Mex. pop. growth	-7.11***	0.76
US GSP growth	2.14***	0.67
Mex. GSP growth	2.04***	0.46
Adjacency	-0.13	0.60
Distance	-3.15***	0.15
US unempl.	0.30***	0.04
Mex. unempl.	0.08**	0.04
US Gini	-33.48***	2.36
Mex. Gini	20.35***	1.51
Constant	-38.22**	16.78
Observations	1426	
R^2	0.75	

Table B.7: Groups and blocks, full sample $n = 1426$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	36	910	35	614	35	521	35	752
2	36	51	35	90	36	125	36	97
3	37	33	35	65	36	72	36	90
4	36	17	35	63	36	55	35	42
5	37	22	35	64	35	82	36	29
6	36	15	35	34	36	61	36	25
7	36	6	35	39	36	41	35	22
8	37	2	35	35	36	47	36	6
9	36	5	35	41	36	24	36	7
10	36	2	35	31	35	41	35	0

Table B.8: Groups and blocks, modified common support sample $n = 1380$

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	32	904	34	576	35	475	34	709
2	33	59	35	88	36	125	35	102
3	33	28	36	61	36	72	35	88
4	33	22	34	65	36	55	35	42
5	33	16	36	61	35	82	35	31
6	32	12	35	34	36	61	34	26
7	33	4	35	40	36	41	35	18
8	33	1	34	34	36	47	35	8
9	33	5	36	41	36	24	35	8
10	32	2	34	31	35	41	34	1

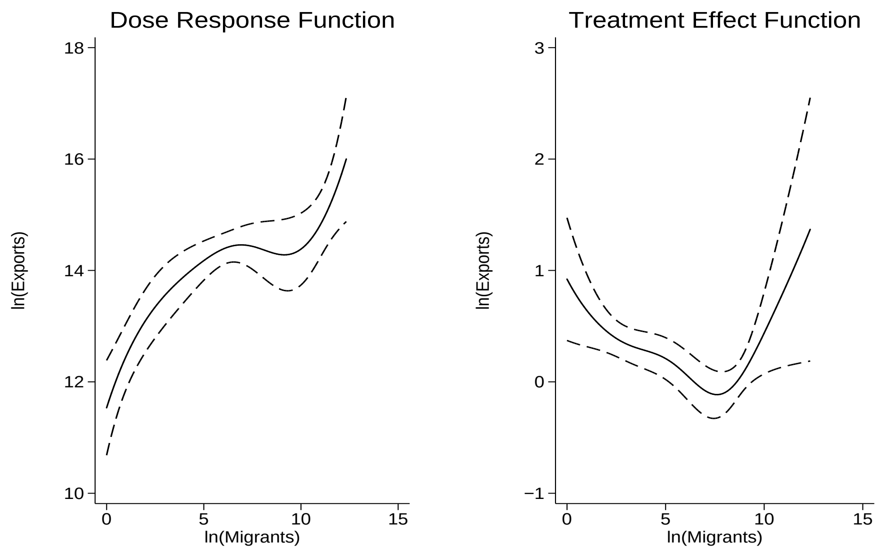
Table B.9: Three-stage balancing comparison of covariates

T-stats	Before balancing on GPS $n = 1426$				After balancing on GPS (weighted t-stats) $n = 1426$				Common Support after balancing on GPS (weighted t-stats) $n = 1380$			
Covariate	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
US pop.	21.14	-1.16	-3.86	-14.54	2.61	-0.87	0.29	-1.03	1.13	-0.85	0.40	-0.79
US pop.^{2}	20.62	-1.06	-3.53	-14.63	2.47	-0.89	0.35	-1.18	1.08	-0.88	0.45	-0.75
Mex. pop.	9.27	5.13	-2.88	-11.78	2.32	1.27	-0.40	-2.64	1.28	1.45	-0.30	-2.57
Mex. pop.^{2}	9.12	5.20	-2.74	-11.85	2.24	1.29	-0.37	-2.71	1.23	1.46	-0.27	-2.63
US GSP	20.79	-1.21	-3.43	-14.71	2.39	-0.90	0.30	-1.08	0.90	-0.88	0.42	-0.85
US GSP^2	19.97	-1.06	-2.97	-14.76	2.21	-0.92	0.38	-1.02	0.84	-0.91	0.49	-0.79
Mex. GSP	5.36	0.88	-1.11	-5.14	1.48	0.42	0.22	-1.36	1.16	0.51	0.25	-1.38
Mex. GSP^2	5.21	0.89	-1.03	-5.10	1.44	0.43	0.24	-1.40	1.14	0.51	0.26	-1.42
US pop. growth	6.00	2.59	-2.42	-6.20	1.16	1.09	-0.11	-1.44	1.12	1.04	-0.12	-1.32
Mex. pop. growth	-2.92	2.34	0.80	-0.19	0.45	0.73	0.38	-0.50	0.65	0.79	0.36	-0.53
US GSP growth	-6.90	0.02	3.11	3.73	-0.49	1.06	1.03	0.21	-0.25	0.97	0.98	0.21
Mex. GSP growth	-1.10	1.21	1.72	-1.82	-0.36	0.58	0.93	0.60	-0.37	0.81	0.91	0.52
Adjacency	1.31	1.28	0.26	-2.85	0.45	0.48	0.26	0.13	0.27	0.31	-0.28	0.12
Distance	-13.11	-0.71	9.23	4.23	-4.46	-0.75	1.98	0.69	-3.92	-0.79	2.08	0.34
US unempl.	10.08	-0.28	-1.65	-8.05	1.47	0.04	0.67	-0.38	0.88	0.05	0.69	-0.28
Mex. unempl.	1.19	-1.28	-3.53	3.62	1.68	0.02	-0.53	0.11	1.67	0.07	-0.52	0.13
US Gini	8.55	-1.58	-2.07	-4.79	1.31	-0.41	0.45	-0.52	0.59	-0.40	0.53	-0.36
Mex. Gini	6.08	3.95	-3.18	-6.89	1.31	-0.06	-1.24	0.34	0.76	0.14	-1.14	0.27
Avg. absolute t-stat	5.35, 24/72$<\|1.96\|$				0.97, 63/72$<\|1.96\|$				0.80, 68/72$<\|1.96\|$			

Table B.10: OLS estimation, dependent variable: ln(exports)

Covariates and statistics	Estimated coefficient	Standard error
Migrant stock	1.17***	0.31
Migrant stock^2	-0.20***	0.07
Migrant stock^3	0.01**	0.00
GPS	-6.26***	2.61
Migrant stock*GPS	1.60***	0.46
Constant	11.70***	0.62
Observations	1380	
R^2	0.10	

Figure B.2



90 percent confidence intervals, represented by dashed lines, are constructed by bootstrapping.

Table B.11: 59 State-state observations excluded by modified common support condition, equation (10)- starting sample of $n = 1488$

Arizona	Baja California	South Dakota	Baja California Sur	Texas	Zacatecas
Arizona	Sonora	South Dakota	Colima	Vermont	Aguascalientes
California	Baja California	South Dakota	Campeche	Vermont	Baja California
California	Veracruz	South Dakota	Quintana Roo	Vermont	Baja California Sur
Maine	Baja California Sur	Texas	Chihuahua	Vermont	Colima
Maine	Colima	Texas	Chiapas	Vermont	Campeche
Maine	Quintana Roo	Texas	Coahuila	Vermont	Coahuila
Montana	Colima	Texas	Durango	Vermont	Morelos
Montana	Campeche	Texas	Guerrero	Vermont	Nayarit
Montana	Quintana Roo	Texas	Guanajuato	Vermont	Nuevo Leon
North Dakota	Baja California Sur	Texas	Hidalgo	Vermont	Quintana Roo
North Dakota	Colima	Texas	Jalisco	Vermont	Queretaro
North Dakota	Campeche	Texas	Michoacan	Vermont	Sinaloa
North Dakota	Morelos	Texas	Mexico	Vermont	Sonora
North Dakota	Quintana Roo	Texas	Nuevo Leon	Vermont	Tlaxcala
North Dakota	Yucatan	Texas	Oaxaca	Vermont	Tamaulipas
New Hampshire	Baja California Sur	Texas	Puebla	Vermont	Yucatan
Rhode Island	Baja California Sur	Texas	San Luis Potosi	West Virginia	Baja California Sur
Rhode Island	Colima	Texas	Tamaulipas	West Virginia	Quintana Roo
Rhode Island	Quintana Roo	Texas	Veracruz		

Table B.12: 58 State-state observations excluded by modified common support condition, equation (10)- starting sample of $n = 1536$

Arizona	Baja California	South Dakota	Baja California Sur	Texas	Zacatecas
California	Baja California	South Dakota	Colima	Vermont	Aguascalientes
Maine	Baja California Sur	South Dakota	Campeche	Vermont	Baja California
Maine	Colima	South Dakota	Quintana Roo	Vermont	Baja California Sur
Maine	Quintana Roo	Texas	Chihuahua	Vermont	Colima
Montana	Colima	Texas	Chiapas	Vermont	Campeche
Montana	Campeche	Texas	Durango	Vermont	Coahuila
Montana	Quintana Roo	Texas	Guerrero	Vermont	Morelos
North Dakota	Baja California Sur	Texas	Guanajuato	Vermont	Quintana Roo
North Dakota	Colima	Texas	Hidalgo	Vermont	Queretaro
North Dakota	Campeche	Texas	Jalisco	Vermont	Sinaloa
North Dakota	Morelos	Texas	Michoacan	Vermont	Sonora
North Dakota	Quintana Roo	Texas	Mexico	Vermont	Tabasco
North Dakota	Yucatan	Texas	Mexico City	Vermont	Tlaxcala
New Hampshire	Baja California Sur	Texas	Nuevo Leon	Vermont	Tamaulipas
New Hampshire	Baja California Sur	Texas	Oaxaca	Vermont	Yucatan
New Hampshire	Quintana Roo	Texas	San Luis Potosi	Vermont	Tamaulipas
Rhode Island	Baja California Sur	Texas	Tamaulipas	West Virginia	Baja California Sur
Rhode Island	Colima	Texas	Veracruz	West Virginia	Quintana Roo
Rhode Island	Quintana Roo				

Table B.13: 46 State-state observations excluded by modified common support condition, equation (10)- starting sample of $n = 1426$

Arizona	Baja California	North Dakota	Colima	Vermont	Colima
Arizona	Chihuahua	North Dakota	Campeche	Vermont	Campeche
Arizona	Sonora	North Dakota	Quintana Roo	Vermont	Coahuila
California	Baja California	New Hampshire	Baja California Sur	Vermont	Morelos
California	Chihuahua	New Hampshire	Quintana Roo	Vermont	Nuevo Leon
California	Mexico	Rhode Island	Baja California Sur	Vermont	Quintana Roo
California	Oaxaca	Rhode Island	Colima	Vermont	Queretaro
California	Puebla	Rhode Island	Quintana Roo	Vermont	Sinaloa
California	Veracruz	South Dakota	Baja California Sur	Vermont	Sonora
Maine	Baja California Sur	South Dakota	Campeche	Vermont	Tabasco
Maine	Colima	South Dakota	Colima	Vermont	Tlaxcala
Maine	Quintana Roo	South Dakota	Quintana Roo	Vermont	Tamaulipas
Montana	Colima	Vermont	Aguascalientes	Vermont	Yucatan
Montana	Campeche	Vermont	Baja California	West Virginia	Baja California Sur
Montana	Quintana Roo	Vermont	Baja California Sur	West Virginia	Quintana Roo
North Dakota	Baja California Sur				