

The impact of Immigrant Stocks on Bilateral Exports: A Global Gravity Analysis

Nathan Lea

Professional Economist BSc and Apprenticeship Level 6

School of Economics

University of Kent, 2026

Executive Summary

This dissertation asks whether immigrant stocks are positively associated with origin-to-destination exports, and whether that association is robust to country-pair fixed effects. Theoretically, immigrant stocks are expected to increase bilateral trade through the information and preference channels (Gould, 1994). Existing literature generally finds a positive migration-trade relationship, but Parsons (2012) shows this relationship weakens on a global panel once country-pair fixed effects are introduced. More recent work uses stronger methods but is OECD-focused. This dissertation uses a global country-pair panel observed at three-year intervals for 1996 to 2017. It combines CEPII Gravity data with the Standaert and Rayp (2025) migration dataset to provide broader country-pair coverage than recent OECD-focused work. It estimates PPML gravity models with exporter-year and importer-year fixed effects, then adds country-pair fixed effects as a stricter specification. The model without country-pair fixed effects finds a positive and statistically significant association, implying a 10% increase in immigrant stock is associated with 1% higher exports. However, the coefficient becomes close to zero, turns negative and is statistically insignificant once country-pair fixed effects are included. This indicates that the initial association may reflect fixed differences between country-pairs rather than a robust relationship between migration and exports. Therefore, the dissertation supports the specification concern raised by Parsons (2012) on a more recent global panel using a modern PPML estimator. The main conclusion is not that migration has no relationship with exports, but that this dissertation does not find robust evidence of a positive association in a global panel once persistent bilateral heterogeneity is accounted for.

AI Statement

I acknowledge the use of generative AI in code development in this paper. However, the work reported remains my own.

Acknowledgements

I would like to thank my family and friends for their support throughout my degree apprenticeship. I am extremely grateful to all my colleagues, past and present, for their guidance and support throughout my professional journey. I would particularly like to thank Alex Hutchison, Joel Kirby, Louis Ariss, Jack Herbert and Dan Wells for their support during my final year and this dissertation. I would also like to thank the academic staff at the University of Kent. Finally, I am incredibly grateful to my fellow apprentices for sharing this experience with me, and I feel fortunate to have made lifelong friends along the way.

1. Introduction

International trade in goods and international migration are two of the most consequential and contested domains of economic policy. The past decade has seen a marked shift in the international trade environment, with rising protectionism and renewed debates over deglobalisation. Over the same period, migration has remained a prominent and politically sensitive issue. Yet while both these issues attract sustained attention in their own right, less attention is given to the relationship between them. Examining the relationship between migration and trade is therefore the focus of this dissertation. Whether migration increases trade matters because it could suggest that immigrant communities strengthen economic links between origin and destination countries. If robust, such a relationship could have implications for export promotion, diaspora engagement and how policy makers assess the wider economic effects of migration.

Theoretically, immigrant stocks are expected to increase bilateral trade through the information and preference channels (Gould, 1994). A large body of empirical work using gravity models finds a broad positive association. Whether this association is robust to more demanding specifications and whether it reflects a causal effect rather than endogenous association, remains an open debate (Hatzigeorgiou and Lodefalk, 2021). For example, Parsons (2012) shows on a global panel that the immigrant stock coefficient changes from positive and strongly statistically significant to negative and only weakly statistically significant once country-pair fixed effects are included. More recent work has adopted more advanced methods, but has focused on OECD destination countries rather than a broader global migration panel (Elguezabal et al., 2025). The specification-sensitivity concern raised by Parsons has not, to the knowledge of the author, been revisited on a global panel with more recent bilateral migration data using a modern Poisson pseudo-maximum likelihood estimator (PPML).

This dissertation addresses that gap. It asks whether immigrant stocks are positively associated with bilateral exports¹, and whether that association is robust to a specification that controls for persistent bilateral heterogeneity. The dependent variable is origin-country

¹ Bilateral exports in this dissertation refers to exports from the immigrant's origin (country of birth) to the destination (where they live) and not the opposite trade flow.

exports to the destination². The main explanatory variable is the bilateral immigrant stock, defined as the number of origin-born individuals residing in the destination in a given year. The dissertation estimates a PPML gravity specification with exporter-year and importer-year fixed effects and then introduces country-pair fixed effects as a sensitivity check. Drawing on a global bilateral migration dataset with broader country-pair coverage than recent work, it tests whether the specification-sensitivity concern raised by Parsons is reproduced using a modern PPML gravity specification and more recent bilateral migration data.

The remainder of the dissertation is structured as follows: Section 2 reviews the literature; Section 3 describes the methodology; Section 4 discusses the data; Section 5 discusses the results and limitations; Section 6 considers policy implications; and Section 7 concludes.

2. Literature Review

2.1 Theoretical Mechanisms Linking Migration and Trade

The theoretical foundations of the migration-trade relationship are most commonly attributed to Gould (1994) who identified two main channels through which immigrant stocks may raise bilateral trade. He identifies an information channel, which operates through a reduction in the transaction costs of trade, and a preference channel, which operates through demand for origin-country goods in the host country. Rauch (2001) subsequently placed the information channel on a firmer network-theoretic footing. This section discusses each in turn and closes by relating the channels to the empirical setup in this dissertation.

The information channel rests on the observation that international trade is costly not only because of transport and tariffs but because cross-border transactions require buyers and sellers to search for counterparties. Immigrants living in a host country reduce these frictions in relation to their country-of-origin. They bring knowledge of origin-country producers, product specifications, regulatory environments and business practices as well as fluency in the relevant languages. They are also often embedded in social networks that support informal contract enforcement through reputation. Because the information channel reduces

² This is equivalent to destination-country imports from the origin.

friction costs for trade in both directions, it is theoretically expected to raise both origin-country exports to the destination and origin-country imports from the destination (Gould, 1994).

The preference channel rests on a different mechanism. Immigrants carry consumption preferences from their country of origin to their host country, which raises host country demand for origin-country goods. This demand may spill over to native consumers as unfamiliar products become more widely available. Unlike the information channel, the preference channel is directionally asymmetric. It raises origin-country exports to the destination because that is the flow that satisfies migrant demand but has no symmetric effect on origin-country imports from the destination (Gould, 1994).

Rauch (2001) provides the network foundation for the information channel. In his account, international trade in differentiated products is characterised by search frictions in a large characteristic space, where goods are not traded on centralised exchanges and buyers and sellers must find each other directly. Co-ethnic networks reduce these frictions by providing trusted channels of information and referral between the two countries.

Taken together, these mechanisms have a clear implication for the empirical setup used in this dissertation. The dependent variable is origin-country exports to the destination country. Both channels act on this flow. The information channel reduces trade friction between the two countries. The preference channel raises destination-country demand for origin-country goods. Therefore, the theoretical expectation is a positive effect of immigrant stocks on bilateral exports, supported by two reinforcing mechanisms.

2.2 Empirical evidence

A substantial empirical literature examines the relationship between migration and international trade using gravity models. The majority of empirical studies find evidence consistent with theoretical predictions, with immigrant stocks positively associated with bilateral trade flows. Hatzigeorgiou and Lodefalk (2021, Table 1) review this body of work and report a median elasticity of 0.15 for exports and 0.18 for imports³. The size of the estimated effect varies across studies, depending on the host countries examined, the time

³ Many studies report trade flows from the host-country perspective, and this section follows each paper's terminology when summarising its findings. In those studies, host-country imports from origin countries are equivalent to the flow analysed in this dissertation: origin-country exports to the destination.

period covered, the specification used and whether trade is measured in aggregate or by direction.

A substantial proportion of the empirical literature examines the migration-trade relationship using data from individual countries and their trading partners. Head and Ries (1998, p.47) estimate gravity models for Canadian imports and exports and find that a 10% increase in bilateral immigrant stock is associated with 1% higher exports and 3% higher imports. Bratti et al. (2014) and Hatzigeorgiou (2010) report similar patterns for Italy and Sweden respectively, with positive effects on both flows but larger effects on imports. This asymmetry is consistent with the theoretical channels in Section 2.1: the information channel reduces trade costs in both directions while the preference channel raises only host-country imports from the origin. Other single-country studies also find positive effects. Peri and Requena-Silvente (2010) report positive effects for Spanish exports, while Girma and Yu (2002) find that immigration from non-Commonwealth countries has a significant impact on UK exports and imports, whereas immigration from Commonwealth countries does not.

A related strand of the literature examines the migration-trade relationship using datasets that include multiple destination countries. Lewer and Van den Berg (2009) examine immigration and trade between 16 OECD destination countries and a large set of immigrant source countries between 1991 and 2000. Using a simultaneous equation model to estimate several channels of influence, they find that a 10% increase in immigration is associated with a 4.5% increase in international trade (Lewer and Van den Berg, 2009, p.212). When exports and imports are estimated separately, the effect is larger for imports. This provides broader multi-destination evidence that immigration is positively associated with trade, while also reinforcing the import-export asymmetry found in single-host-country studies.

Other work has qualified the broad consensus by showing the size and statistical significance of the estimated relationship is sensitive to modelling choices. Using a comprehensive panel dataset covering up to 178 countries over the period 1960 to 2000, Parsons (2012) finds that immigrant stocks are positively associated with trade in gravity specifications with country-year fixed effects. However, when Parsons includes country-pair fixed effects to control for unobserved heterogeneity, the estimated coefficient becomes negative and less statistically significant. The paper argues that part of the positive relationship identified in earlier studies

may reflect unobserved pair-specific factors or endogeneity, rather than a direct causal effect of migration on trade.

More recent work continues to apply gravity models to large cross-country panels. For example, Elguezabal et al. (2025) analyse trade between 30 OECD countries and their trading partners using data from 1995 to 2023. They estimate a modern PPML gravity specification and use a control function approach to address the potential endogeneity of immigrant stocks. In their baseline PPML specification, a 10% increase in migration is associated with a 0.28% increase in exports and a 0.46% increase in imports (Elguezabal et al., 2025, p.17). However, once endogeneity is addressed using their preferred control function approach, the export effect is no longer statistically significant in the short run, while the import effect remains positive and significant. The paper therefore reinforces a common finding in the literature: migration tends to be more strongly associated with imports than exports, although the authors also find that export effects become more apparent in longer-run specifications.

Taken as a whole, the literature establishes a broad positive association between immigrant stocks and bilateral trade, consistent with theoretical channels. However, the size and robustness of the relationship varies across host countries, time periods and specification choices.

2.3 Methodological development of the Gravity Model

While Section 2.2 reviewed the main empirical findings of the migration-trade literature, those findings depend heavily on how the gravity model is specified and estimated. This section traces the developments in the gravity framework that bear directly on the specification used in this dissertation: the theoretical foundation for modelling bilateral trade costs; the need to control for multilateral resistance; the case for country-pair fixed effects as a stricter specification; and the case for PPML as the appropriate estimator.

The origin of the gravity model of trade is normally attributed to Tinbergen (1962) who suggested an analogy between international trade flows and Newton's law of gravity. In its simplest form, the model predicts that bilateral trade between two countries increases with their economic size and decreases with the geographical distance between them. Larger economies are expected to trade more due to higher production and consumption capacity. Distance is a proxy for trade costs such as transport expenses, time and information frictions.

Tinbergen's model was empirically successful, but did not have a strong theoretical foundation.

Anderson (1979) provided that foundation, showing that gravity-type trade relationships can be derived from standard models of consumer behaviour and product differentiation, with bilateral trade depending systematically on economic size and bilateral trade costs.

Anderson's framework establishes the conceptual foundation for the gravity equation but leaves open how the bilateral trade costs that determine trade flows should be measured in practice.

Bilateral trade costs are not directly observable in aggregate and may include transport costs, tariffs and non-tariff barriers, communication frictions, institutional differences, and the costs of cross-border information and contract enforcement (Anderson and van Wincoop, 2004).

Empirical gravity models therefore proxy trade costs using observable bilateral characteristics. Yotov et al. (2016) identify distance, contiguity, common language, colonial ties and regional trade agreements (RTAs) as standard variables used to approximate bilateral trade costs in gravity specifications. Geographic distance captures the transport, time and information costs that rise with separation between two economies and is expected to enter the gravity equation with a negative sign. By contrast, contiguity, common official language, colonial ties and regional trade agreement membership are expected to facilitate trade. Contiguity captures the effect of sharing a border, common official language reduces communication costs, colonial ties proxy historical and institutional links, and regional trade agreements reduce formal policy barriers to trade. Within this framework, immigrant stocks are introduced as an additional bilateral linkage. They may reduce informal trade costs through the information channel, while also increasing demand for origin-country goods through the preference channel discussed in Section 2.1.

A major development in the gravity literature came with Anderson and van Wincoop (2003). They argued that bilateral trade depends not only on direct costs between two countries, but also on each country's trade barriers relative to all other trading partners. They formalised this through the concept of multilateral resistance and showed that omitting it leads to biased and unstable coefficient estimates. In panel settings, multilateral resistance is controlled for through exporter-year and importer-year fixed effects, which the specification in this

dissertation includes. Studies using simpler specifications without these fixed effects may incorrectly attribute country-specific trade resistance effects to immigrant networks.

A subsequent development in panel gravity work was the use of country-pair fixed effects alongside country-year fixed effects. Baier and Bergstrand (2007) argue that estimates of FTA effects may be biased because countries enter into trade agreements on the basis of observable and unobservable bilateral characteristics that also affect trade. Their preferred panel specification uses bilateral fixed effects to absorb time-invariant pair-specific heterogeneity. This approach has since been recommended in the structural gravity literature (Yotov et al., 2016), particularly in settings where the variable of interest is plausibly correlated with persistent bilateral factors. It is a stricter specification rather than an always-required correction. The second stage of the empirical test in this dissertation adopts this approach.

A further development concerned the econometric validity of log-linear estimation. For many years, gravity equations were estimated by taking logarithms of bilateral trade flows and applying ordinary least squares (OLS). Santos Silva and Tenreyro (2006) showed that this practice is problematic in trade settings for two reasons. First, country pairs with zero trade flows are automatically excluded when trade is logged, because the logarithm of zero is undefined, creating a source of sample-selection bias. Second, under the heteroskedasticity typical of trade data, log-linear OLS no longer recovers consistent estimates. They proposed PPML estimation as a more appropriate alternative. PPML is now standard in contemporary gravity work (Yotov et al., 2016) and is the estimator adopted in this dissertation.

2.4 Summary of the literature and contribution

The literature reviewed in this section establishes a broad positive association between immigrant stocks and bilateral trade, consistent with the information and preference channels. However, it also shows that the estimated size and significance of this relationship depend on the specification used. In particular, Parsons (2012) raises the concern that positive migration-trade estimates may weaken once country-pair fixed effects are used to control for persistent unobserved bilateral heterogeneity.

However, Parsons's evidence is based on an earlier generation of data and estimation practice. His analysis uses a 1960–2000 panel and estimates log-linear gravity models using a least squares dummy estimator approach rather than PPML. More recent work has adopted

more advanced gravity methods and has addressed endogeneity more directly. For example, Elguezabal et al. (2025) use a control-function approach on a 1995–2023 panel, but their analysis is focused on trade between 30 OECD destination countries and their trading partners. This leaves space for an updated reassessment of Parsons’s specification-sensitivity concern on a global bilateral panel.

This dissertation therefore tests whether the specification-sensitivity concern raised by Parsons is reproduced using a modern PPML estimator on a global bilateral panel. It uses the Standaert and Rayp (2025) bilateral migration dataset, which provides annual country-pair coverage from 1960 to 2022 and allows the analysis to focus on the period 1996 to 2017.

3. Methodology

3.1 Empirical focus and overall strategy

This dissertation examines whether immigrant stocks are associated with bilateral export flows. The hypothesis is framed as an association rather than a causal relationship because the empirical strategy cannot fully rule out the endogeneity concerns discussed in Section 3.4.

The theoretical expectation, derived from the preference and information channels discussed in the literature review, is that larger immigrant stocks should be associated with higher exports from the origin to the destination. The empirical analysis tests this theoretical prediction in two stages. The first stage estimates a baseline PPML gravity model with standard gravity controls and exporter-year and importer-year fixed effects and asks whether bilateral immigrant stock is positively and statistically significantly associated with bilateral exports. The second stage adds country-pair fixed effects to the baseline specification and asks whether any such association is robust after controlling for persistent bilateral heterogeneity.

3.2 Econometric specifications

The first gravity specification (Model 1) estimated can be written as:

$$X_{ijt} = \exp(\beta_1 \ln MigStock_{ijt} + \beta_2 RTA_{ijt} + \beta_3 \ln Dist_{ij} + \beta_4 Comlang_{ij} + \beta_5 Contig_{ij} + \beta_6 Colony_{ij} + \delta_{it} + \gamma_{jt}) + \varepsilon_{ijt}$$

where the dependent variable X_{ijt} is exports from country i to country j in year t , the main explanatory variable $MigStock_{ijt}$ is the stock of immigrants from i residing in j in year t . $Dist_{ij}$ is bilateral distance between the two countries, while $Comlang_{ij}$, $Contig_{ij}$, $Colony_{ij}$, are dummy variables that equal 1 when countries share a common official language, a border or have had a colonial relationship, and zero otherwise. RTA_{ijt} is a time-varying dummy equal to 1 when both countries are members of the same regional trade agreement in year t . These controls are used because they proxy for bilateral trade costs or bilateral links that are expected to affect exports independently of migration. Similar gravity control set-ups are widely used in the migration-trade literature, including Parsons (2012), although the exact set of controls varies.

Consistent with Anderson and van Wincoop (2003) and the subsequent structural gravity literature, exporter-year fixed effects δ_{it} and importer-year fixed effects γ_{jt} are included to control for multilateral resistance and other country-time specific factors affecting trade. ε_{ijt} is the error term.

Immigrant stock is entered in natural logarithms allowing β_1 to be interpreted as an elasticity. This approach is consistent with much of the empirical literature, which models immigrant stocks in logarithmic form (Elguezabal et al., 2025).

A second specification (Model 2) is then estimated where country-pair fixed effects are added:

$$X_{ijt} = \exp(\beta_1 \ln MigStock_{ijt} + \beta_2 RTA_{ijt} + \mu_{ij} + \delta_{it} + \gamma_{jt}) + \varepsilon_{ijt}$$

where μ_{ij} denotes country-pair fixed effects. The purpose of this second model is to absorb time-invariant bilateral heterogeneity that may be correlated with both migration and trade, such as persistent historical, cultural, political or geographic ties. Parsons (2012) argues that such unobserved bilateral characteristics are a major concern and substantially affect the estimated migration coefficient. Egger and Nigai (2015) and Agnosteva et al. (2014) show that pair fixed effects are a better measure of bilateral trade costs than the standard set of gravity variables. Once these pair fixed effects are included, time invariant bilateral controls such as distance, contiguity, common language and colonial ties are absorbed by the fixed effects and thus cannot be estimated. As a result, only variables that vary over time within

country pairs like immigrant stocks and RTA membership remain. This identifies the migration effect from within-pair changes over time rather than cross-sectional bilateral differences.

3.3 Estimation strategy

Although gravity equations have historically been estimated in log-linear form using OLS, the models in this dissertation are estimated using PPML for two main reasons. First, Santos Silva and Tenreyro (2006) show that log-linear OLS estimation can lead to misleading results when the error term is heteroskedastic, whereas PPML remains consistent under more general forms of heteroskedasticity. As shown in the summary statistics in Section 4.3, the dispersion of bilateral exports is substantial, which is the kind of pattern in trade data that typically generates heteroskedasticity in regression residuals.

Second, Santos Silva and Tenreyro (2006) show log-linear OLS cannot naturally accommodate zero trade flows. In the baseline specifications estimated here, zero trade flows are not present after sample construction. However, by utilising PPML over log-linear OLS it allows zero trade flows to be introduced as a robustness check.

Standard errors are clustered by country-pair to allow for correlation in the error term within bilateral pairs over time. This is consistent with the approaches adopted in the literature, such as Parsons (2012) and Peri and Requena-Silvente (2010).

3.4 Methodological limitations

This methodological approach has some limitations which bear directly on the interpretation of the results.

The first limitation concerns the scope of the research question. The analysis focuses on one directed trade flow rather than bilateral trade in both directions. Specifically, it examines origin-country exports to the destination country. This flow was chosen as both channels discussed in Section 2.1 can operate on it. However, the question does not test whether the same relationship holds in the other direction, namely origin-country imports from the destination country. The estimated coefficient should therefore not be interpreted as evidence of the full two-way migration-trade relationship between the two countries.

A second limitation is that the specification includes only the bilateral immigrant stock and not the corresponding bilateral emigrant stock, defined as the stock of destination-born individuals residing in the origin-country. Parsons (2012) argues that emigrant stocks also plausibly facilitate bilateral trade, and their omission is likely to bias the estimated immigrant-stock coefficient through standard omitted-variable arguments. This specific concern is directly addressed in Section 5.3. However, the baseline coefficient should still be read as a conditional association between immigrant stocks and directional exports rather than a complete estimate of the role of migration on bilateral trade.

Third, the analysis remains vulnerable to time-varying bilateral factors that affect both migration and trade. Exporter-year and importer-year fixed effects control for country-level shocks, while country-pair fixed effects control for persistent bilateral characteristics. However, they do not control for bilateral factors that change over time. Examples could include changes in visa policy, diplomatic relations, institutional cooperation, transport links or bilateral agreements that affect both migration and exports. If these factors are correlated with immigrant stocks, the estimated coefficient may still be biased.

Reverse causality is also possible. The dissertation examines whether immigrant stocks are associated with exports, but stronger trade relationships may themselves encourage future migration by creating economic opportunities, business networks or stronger links between countries.

A stronger identification strategy would be required to address these endogeneity concerns more fully. Elguezal et al. (2025), for example, use a control-function approach to address the endogeneity of immigrant stocks within a gravity framework. Applying a comparable approach to a global bilateral panel would require credible instruments or control-function variables that are available consistently across countries and over time, and that affect migration without directly affecting trade. Developing and validating such an identification strategy is beyond the scope of this dissertation because of the time available and technical complexity involved. The results should therefore be interpreted as conditional associations rather than causal estimates.

The baseline specification enters immigrant stock in natural logarithms. This is consistent with much of the migration-trade literature, but it creates a limitation because the logarithm of zero is undefined. Country pairs with zero immigrant stock are therefore dropped from the

baseline sample. The reported coefficient is identified only from bilateral pairs where a positive immigrant stock already exists. This introduces potential sample-selection bias, because the sample over-represents bilateral relationships with established migration links and under-represents those without. This concern is directly addressed in Section 5.3.

4 Data

4.1 Data Sources

CEPII Gravity

The CEPII Gravity dataset provides over 80 geographic, historical, cultural and policy variables covering 1948 to 2020. It is designed specifically for gravity model estimation, providing harmonised bilateral measures that are consistent across countries and over time, allowing researchers to select theoretically appropriate controls and trade values (Conte et al., 2022).

The Gravity dataset integrates several bilateral trade-flow series. In this dissertation, exports (X_{ijt}) are drawn from its BACI-sourced trade flow variable (*trade_flow_baci*), measured in thousands of current US dollars and constructed from the CEPII BACI International Trade Database (Gaulier and Zignago, 2010). BACI provides harmonised goods bilateral trade flows constructed from UN Comtrade data, covers over 200 reporting countries and is a recommended source for structural gravity models (Yotov et al., 2016). BACI is preferred to raw Comtrade data because CEPII applies a reconciliation procedure to reduce inconsistencies between exporter- and importer-reported flows and to improve comparability across country pairs.

The gravity controls are sourced from the Gravity dataset (Conte et al., 2022). Distance is measured using *distw_harmonic*, a population-weighted distance measured between the countries' most populated cities (harmonic mean). This better captures the economic geography of trade than a simple point-to-point measure. Common language is measured using *comlang_off* which records whether two countries share a common official or primary language. Colonial relationship is measured using *col_dep_ever*, which records whether two countries have ever been in a colonial relationship. RTA membership is measured using

fta_wto, a binary indicator of whether two countries are both members of an RTA in the relevant year. Contiguity is measured by *contig*, a binary indicator of whether the two countries share a border.

Immigrant stocks

Immigrant stocks ($MigStock_{ijt}$) are drawn from the Annual Bilateral Migration Data (1960–2022) dataset constructed by Standaert and Rayp (2025). The dataset provides a global, annual, bilateral measure of immigrant stocks defined by country of birth. It covers over 56,000 country pairs and approximately 2.9 million observations from 1960 to 2022. The dataset is constructed using a Bayesian state-space model that integrates information from multiple sources on immigrant stocks, net flows and gross flows. Rather than filling gaps through simple demographic accounting or differencing of stock data, the state-space model exploits the demographic relationship between stocks, flows, births and deaths to produce internally consistent annual estimates. This approach allows the dataset to be a ten-fold increase in observations relative to the second largest dataset.

4.2 Panel construction

The empirical analysis uses a directed bilateral country-pair panel observed at three-year intervals over the period 1996 to 2017 ($T=8$), specifically for the years 1996, 1999, 2002, 2005, 2008, 2011, 2014 and 2017. This period provides a more recent assessment of the migration-trade relationship on a global panel than Parsons. The start year reflects the availability of BACI trade data and the end year avoids the deterioration in migration country-pair coverage after 2019. Three-year intervals are used rather than consecutive annual observations because Olivero and Yotov (2012) show that gravity estimates based on three-year and five-year intervals are very similar, whereas consecutive annual data can generate less credible estimates.

Table 1 summarises how the panel is constructed. The CEPII Gravity dataset defines the initial sample, with 508,032 directed country-pair-year observations across the eight selected years. Because the dataset is squared, it includes all possible origin-destination pairs in each year, including self-pairs and entities that do not exist in a given year (for example, Soviet Union in 2000). The panel is then restricted to pairs where both countries existed in the relevant year and where origin and destination are not the same country, leaving 434,788

observations. Next, observations with missing values for the bilateral controls used in the baseline specification are removed, leaving 400,468 observations. Most of this loss is concentrated in the common language variable, where coverage is thinner for a subset of entities in the Gravity dataset. A smaller loss from the colony variable is driven by a handful of politically contested country-pair classifications (for example, Hong Kong and China).

The most involved stage of panel construction was the merge with immigrant stock data. Because country codes were not harmonised between the two datasets, the merge was performed on country name.

Most migration country names aligned with the gravity dataset directly (198 of 239). In the remaining 41 cases, 32 were simple manual fixes (for example, “Bahamas” to “The Bahamas”) while 5 Gravity countries did not exist in the migration data (for example, Christmas Island), although these were all small territories. In the remaining four cases a substantive judgement was required. In two cases, the two datasets classified similar territorial units in different ways: so Palestine was matched to West Bank and Gaza; and Western Sahara to Sahrawi Arab Democratic Republic. Two further cases required explicit treatment because of territorial change over time. The migration dataset records pre-secession Sudan and South Sudan as separate entities even in years prior to the 2011 secession, and likewise records Indonesia and Timor-Leste separately for years prior to Timor-Leste’s independence in 2002. To maintain consistency with the Gravity panel, immigrant stocks were combined for the periods where these were separate entities. After this harmonisation process, the migration data was joined to the Gravity data on origin, destination and year. Once observations with no recorded bilateral immigrant stock were removed, the panel contained 333,643 observations. This loss reflects genuine differences in dataset coverage rather than unresolved matching errors. The baseline models use the logarithm of immigrant stock, so observations with zero immigrant stocks are not included in the estimation, reducing the sample to 172,617. Observations with missing trade flows are then removed in the baseline sample. The final panel used to estimate both models contains 132,369 unique origin-destination-year observations.

The result of this process produces a much broader global panel than recent OECD-focused work. Elguezabal et al. (2025, Table 1) record 72,907 observations for their baseline import model across 29 years, equivalent to approximately 2,514 observations per year. By contrast,

the final estimation sample contains 132,369 observations across eight selected years, equivalent to approximately 16,546 observations per year.

Table 1: Panel Construction Process

Stage	Observations	Loss
Squared Gravity Universe (1)	508,032	-
Existing international pairs only (2)	434,788	73,244
After baseline gravity-control completeness (3)	400,468	34,320
After migration merge and removal of migration NAs (4)	333,643	66,825
After removal of zero immigrant stocks (5)	172,617	161,026
Baseline sample after removal of NA trade flows (6)	132,369	40,248

4.3 Summary statistics

Table 2 reports summary statistics for the variables used in the estimation. Bilateral exports display a substantial right skew: the maximum trade value is roughly 680 times the mean and the standard deviation is around 8.5 times the mean. Such dispersion typically generates heteroskedastic residuals. The residuals in trade data tend to scale with the level of bilateral trade. Country-pairs with very large trade values tend to generate larger residuals. PPML is better suited than log-linear OLS in this setting (Santos Silva and Tenreyro, 2006). The dummy variables show that contiguity, common official language, colonial relationships and RTAs apply to between 2% and 18% of the country-pair-year observations in the estimation sample.

Table 2: Summary Statistics

Variable	Mean	Standard Deviation	Min	Max
Bilateral exports (USD thousands)	668,830	5,657,755	0.001	455,536,421
Bilateral Immigrant Stock	11,105	120,505	1	12,200,000
Harmonic Distance (km)	6,855	4,480	23	19,839
Contiguity	0.031	0.172	0	1
Common official language	0.175	0.380	0	1
Colonial relationship	0.022	0.147	0	1
RTA	0.172	0.377	0	1

Notes: Based on the final estimation sample (row 6 of Table 1).

4.4 Data Limitations

The panel construction process and the underlying data sources introduce important limitations.

As outlined in Table 1, 375,663 observations were lost during the panel construction process. This means that the final panel does not fully reproduce total exports and migration recorded in the broader usable source data. This matters because systematic sample loss could affect how representative the final estimation sample is. However, the aggregate loss is relatively modest. The final estimation sample retains 92% of the total goods trade recorded in the relevant trade panel and 89% of the total immigrant stock recorded in the migration data prior to merging. These figures suggest that, although matching and sample construction does exclude some observations, most excluded country-pairs have little or no recorded trade or migration. The final panel therefore still captures the majority of observed trade and migration contained in the datasets used for this study.

A deeper limitation concerns the migration data itself. As Parsons (2012) notes, patchy coverage of bilateral migration data has long been a constraint on empirical work in this area. The Standaert and Rayp (2025) dataset used here addresses this constraint through a substantially broader country-pair coverage, with over 10 times more observations than the second largest dataset. It achieves this through a Bayesian state-space model that imputes missing observations. Therefore, the key explanatory variable in the dissertation is modelled rather than directly observed immigrant stock. This introduces an extra layer of uncertainty. Any measurement error or modelling assumption embedded in the migration dataset may carry through into the estimated coefficient. The subsequent estimates may therefore be biased if the imputation process does not fully reflect the true underlying pattern of bilateral migration. The trade-off is therefore between coverage and measurement certainty. This dissertation favours broader and more usable panel coverage based on modelled migration data over a smaller panel restricted to directly observed immigrant stocks.

A further limitation concerns the treatment of unobserved trade flows. Because BACI does not identify whether missing bilateral trade values are true zeros or simply unreported flows, either treatment risks bias. Excluding them may omit genuine zero-trade relationships and tilt the sample toward observed trading pairs, while treating them all as zeros may understate trade where reporting is incomplete. This uncertainty affects the composition of the

estimation sample and therefore the interpretation of the estimated bilateral immigrant stock coefficient. Parsons (2012) highlights a similar issue in aggregate trade data and treats such values as missing rather than zero. The main specifications follow this approach, while an alternative treatment is considered in Section 5.3.

5 Results

5.1 Main Results

Table 3 reports the main estimation results. Column (1) represents Model 1, the baseline PPML-gravity specification, which includes the standard gravity controls, exporter-year and importer-year fixed effects and the immigrant stock variable. Column (2) presents Model 2, which adds country-pair fixed effects to the baseline specification and therefore time-invariant bilateral variables are absorbed and no longer separately estimated.

Table 3: Baseline results

Variables	(1) Model 1: Baseline PPML	(2) Model 2: Pair fixed effects
Log (Distance)	-0.5966 *** (0.0268)	-
Contiguity	0.3962 *** (0.0585)	-
Common Official Language	0.0305 (0.0538)	-
Colony	-0.0499 (0.0941)	-
RTA membership	0.3447 *** (0.0500)	0.0570* (0.0246)
Log (Stock)	0.0987 *** (0.0105)	-0.0049 (0.0082)
Exporter-year FE	Yes	Yes
Importer-year FE	Yes	Yes
Country-pair FE	No	Yes
Observations	132,359	130,368
Pseudo R ²	0.9362	0.9924

Notes: PPML estimates with standard errors clustered by country-pair in parentheses. The dependent variable is bilateral exports. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Observations fall below the constructed panel because the PPML estimator drops fixed-effect groups with only zero outcomes or singleton observations. This is a feature of the estimator rather than a data error.

Column (1) reports the baseline PPML gravity specification, corresponding to the first stage of the empirical test set out in Section 3.1. The coefficients on the standard gravity controls are broadly consistent with theoretical expectations. Distance enters negatively and is statistically significant, consistent with greater bilateral distance proxying higher trade costs and lower exports. Contiguity is positive and statistically significant, indicating that origin countries that share a border with the destination tend to export more to them. Common official language is positively associated with exports, but it is not statistically significant. RTA membership is also positive and statistically significant, suggesting that origin countries that are in an RTA with the destination are more likely to export to them. The coefficient on colony is negative, which is theoretically unexpected and likely reflects overlap between colonial relationships and bilateral immigrant networks. Despite this, the pattern of coefficients and significance suggests that the baseline model behaves plausibly and is consistent with conventional gravity theory.

The main variable of interest, the logged immigrant stock, is positive and statistically significant at the 0.1% level. The estimated coefficient of 0.0987 implies that a 10% increase in immigrant stock is associated with approximately 0.99% higher exports. This result is consistent with the broader migration-trade literature, which generally finds a positive association between immigrant stocks and exports.

Column (2) presents the results from the second stage of the empirical test, which adds country-pair fixed effects to the baseline specification. The coefficient on RTA membership remains positive but falls sharply from 0.3447 to 0.0570 and is only statistically significant at the 5% level. More importantly for the central question, the coefficient on immigrant stock falls from 0.0987 to -0.0049 and becomes statistically insignificant. The negative sign mirrors the sign change in Parsons (2012), although unlike his result the coefficient in this dissertation is not statistically distinguishable from zero.

5.2 Interpretation of the main findings

The main finding of the dissertation is that the positive migration coefficient in the baseline model does not survive the introduction of country-pair fixed effects. The results mirror the broad pattern in Parsons (2012). Model 1 produces the conventional positive and statistically significant association between immigrant stock and origin-to-destination exports, while

Model 2 reduces the coefficient to a value close to zero and renders it statistically insignificant.

This difference reflects a change in identification. In Model 1, the migration coefficient is estimated partly from cross-sectional differences between country pairs, so it captures whether pairs with larger immigrant stocks tend to have larger export flows. Once country-pair fixed effects are introduced, all time-invariant bilateral characteristics are absorbed, and the migration coefficient is identified only from changes in immigrant stocks within the same bilateral relationship over time. Model 2 is therefore a more demanding test of whether immigrant stock retains independent explanatory power once persistent bilateral features are controlled for.

A plausible interpretation is that immigrant stocks are positively correlated with bilateral factors that do not vary over time. These could include deeper historical, political, cultural or institutional ties that also support trade, but are not fully captured by the standard gravity controls. The disappearance of the migration coefficient once pair fixed effects are added therefore implies that the baseline result may overstate the independent role of immigrant networks. This does not invalidate the theoretical expectation developed in Section 2.1, where immigrant stocks are expected to increase exports through the preference and information channels. Rather, it weakens the evidence that these mechanisms generate a robust independent association in this global panel once persistent bilateral heterogeneity is controlled for.

This interpretation also helps explain how the results relate to more recent work. Elguezabal et al. (2025) use country-pair fixed effects and a control-function approach, and still find positive migration-trade effects. However, their analysis is focused on OECD destination countries, whereas this dissertation tests the relationship on a broader global panel. One possible explanation is that the migration-trade relationship may be more robust in OECD-focused samples than in wider global panels, perhaps reflecting differences in migration composition or institutional context. This should be treated as a hypothesis rather than a firm conclusion of this dissertation. The evidence here therefore supports Parsons's specification-sensitivity concern on a newer global panel using PPML, while suggesting that the robustness of the migration-trade relationship may depend on the country composition of the panel being studied.

5.3 Robustness and remaining limitations

Sections 3.4 and 4.4 discuss important limitations that might impact the reliability of the results. This section does not resolve all of those concerns, but it does test whether the main findings are sensitive to several specification and sample-construction choices.

As discussed in the methodology section, excluding zero immigrant stocks because of the log transformation may create sample-selection bias. To assess the impact of this, the baseline specifications were re-estimated by adding the zero immigrant stocks back to the baseline sample and using $\log(1+\text{Stock})$ in place of the logged stock variable. The results remain similar for both Model 1 and Model 2. In Model 1, the migration coefficient remains positive but falls from 0.0987 to 0.0678 and remains statistically significant at the 1% level. In Model 2, the migration coefficient is negative, very close to zero and remains statistically insignificant. This suggests that the exclusion of zero immigrant stocks does not materially alter the main findings.

Similarly, excluding missing trade flows may affect sample composition. To assess this, the baseline specifications were re-estimated by turning these missing values to zero instead of removing them. The results are very similar for both Model 1 and Model 2. In Model 1, the migration coefficient remains positive and statistically significant and is almost identical to the baseline estimate, increasing only marginally from 0.0987 to 0.0989. In Model 2, the migration coefficient remains statistically insignificant, negative and close to zero. This indicates that the main findings are not materially sensitive to the treatment of missing trade observations in the baseline sample.

As argued by Parsons (2012), excluding emigrant stock from the baseline results can overstate the immigrant coefficient. To examine this, the baseline specifications were augmented by including the emigrant stock, defined as the stock of migrants born in the destination country and residing in the origin-country. In Model 1, both variables are positive and highly statistically significant. The coefficient of immigrant stock falls from 0.0987 to 0.0712 suggesting that part of the baseline coefficient may have captured the omission of migration in the opposite direction. In Model 2, both migration coefficients become statistically insignificant and close to zero. Therefore, the inclusion of the reverse migration corridor does not alter the central conclusions.

Four concerns discussed in Section 3.4 and 4.4 are not resolved by these tests, and they bear directly on how the results should be interpreted.

First, the analysis examines origin-country exports to the destination and does not estimate the opposite trade flow. Second, the empirical strategy does not fully resolve endogeneity. Time-varying unobservable factors and reverse causality remain plausible sources of bias. Third, the panel construction reduces the sample from its initial size, although the final sample retains the majority of trade and migration in the source datasets. Fourth, the key explanatory variable is a modelled rather than directly observed bilateral immigrant stock, so any imputation error in the underlying dataset may carry through to the estimated coefficient.

Taken together, these remaining limitations mean that the reported results should be interpreted cautiously.

6. Policy implications

Despite the question's potential policy relevance set out in the introduction, the findings of this dissertation do not have strong direct implications for policy. In the baseline specification, immigrant stocks are positively and significantly associated with bilateral goods exports, suggesting that larger origin-country communities in destination markets may be linked to higher export flows. However, this relationship does not remain robust once country-pair fixed effects are introduced to account for unobserved bilateral heterogeneity. The coefficient on immigrant stock becomes close to zero and statistically insignificant, indicating that the initial association may reflect fixed differences between country-pairs rather than a robust relationship between migration and exports. As a result, the findings do not support strong policy claims that immigrant communities can be treated as a direct lever for increasing bilateral goods exports.

However, the results still have an important implication for how policymakers should interpret empirical evidence. The main lesson is not that migration has no relationship with trade, but that estimated relationships can be highly sensitive to modelling choices. In this dissertation, the inclusion of country-pair fixed effects changes the migration coefficient from positive and statistically significant to negative and statistically insignificant. This highlights the risk of drawing policy conclusions from a single specification without testing whether the

result is robust to more demanding empirical approaches. The broader policy relevance of the dissertation is therefore cautionary. Empirical findings on migration and trade, and similar areas of economic policy, should therefore be interpreted carefully. Particular attention should be paid to data construction, model specification and the possibility that apparent relationships may be driven by unobserved bilateral factors.

7. Conclusion

This dissertation has asked whether immigrant stocks are positively associated with bilateral exports, and whether that association is robust to a specification that controls for persistent bilateral heterogeneity. The findings are mixed. A standard modern gravity specification with exporter-year and importer-year fixed effects produces a positive and statistically significant association between immigrant stocks and bilateral exports. This is consistent with the theoretical predictions of Section 2.1 and is broadly consistent with the majority of the existing migration-trade literature. However, once country-pair fixed effects are introduced, the estimated coefficient falls close to zero and is no longer statistically significant. This mirrors the broad pattern in Parsons (2012). This result suggests that the baseline association may partly reflect persistent bilateral characteristics such as historical, cultural or political ties, rather than an independent within-pair relationship between immigrant stock and exports. This pattern is robust across sensitivity tests that vary the treatment of zero immigrant stocks, missing trade flows and the omission of reverse bilateral migration. Therefore, the contribution of this dissertation is that the sensitivity concern raised by Parsons (2012) also appears in a more recent global bilateral panel estimated using PPML.

There are no direct policy implications from this dissertation. However, the results have important implications for how policymakers interpret empirical evidence. The main lesson is that the estimated migration-trade relationship can be highly sensitive to modelling choices. This highlights the risk of drawing policy conclusions from a single specification without testing whether the result is robust to more demanding empirical approaches.

This dissertation has two primary limitations that bear directly on how the findings should be read. First, the empirical strategy cannot fully resolve endogeneity. Country-pair fixed effects absorb time-invariant bilateral factors, but time-varying bilateral factors such as visa policy

may still bias the estimated coefficient. Reverse causality from trade to migration remains plausible. Second, the immigrant stock used here is itself a modelled value which adds imputation uncertainty into the main explanatory variable.

The work in this dissertation could be extended in two main ways. First, future research could re-estimate the analysis using the opposite trade flow, origin-country imports from the destination. This would give a fuller picture of the migration-trade relationship. Second, future work could combine the broader global bilateral coverage used here with a stronger identification strategy along the lines of the control function approach applied by Elguezabal et al. (2025). This would allow the specification sensitivity question to be tested with a more direct treatment of endogeneity.

Overall, this dissertation shows that the positive association between immigrant stocks and bilateral exports is sensitive to model specification. While the baseline model finds a positive and statistically significant relationship, that relationship does not remain statistically significant once country-pair fixed effects are introduced. The main conclusion is therefore not that immigrant stocks have no relationship with exports, but that this dissertation does not find robust evidence of a positive association in a global panel once persistent bilateral heterogeneity is accounted for.

Bibliography:

Agnosteva, D., Anderson, J. E. and Yotov, Y. V. (2014) *Intra-national trade costs: measures and aggregation*. NBER Working Paper No. 19872. Cambridge, MA: National Bureau of Economic Research. Available at: <https://www.nber.org/papers/w19872>.

Anderson, J. E. (1979) *A theoretical foundation for the gravity equation*. American Economic Review, 69(1), pp. 106–116. Available at: <https://www.jstor.org/stable/1802501>.

Anderson, J. E. and van Wincoop, E. (2003) *Gravity with gravitas: a solution to the border puzzle*. American Economic Review, 93(1), pp. 170–192. doi: <https://doi.org/10.1257/000282803321455214>.

Anderson, J. E. and van Wincoop, E. (2004) *Trade costs*. Journal of Economic Literature, 42(3), pp. 691–751. doi: <https://doi.org/10.1257/0022051042177649>.

Baier, S. L. and Bergstrand, J. H. (2007) *Do free trade agreements actually increase members' international trade?*. Journal of International Economics, 71(1), pp. 72–95. doi: <https://doi.org/10.1016/j.jinteco.2006.02.005>

Bratti, M., De Benedictis, L. and Santoni, G. (2014) *On the pro-trade effects of immigrants*. Review of World Economics, 150(3), pp. 557–594. doi: <https://doi.org/10.1007/s10290-014-0191-8>.

Conte, M., Cotterlaz, P. and Mayer, T. (2022) *The CEPII gravity database*. CEPII Working Paper No. 2022-05. Paris: CEPII. Available at: <http://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=13432>.

Egger, P. and Nigai, S. (2015) *Structural gravity with dummies only: constrained ANOVA-type estimation of gravity models*. Journal of International Economics, 97(1), pp. 86–99. doi: <https://doi.org/10.1016/j.jinteco.2015.05.004>.

Elguezabal, P., Martínez-Zarzoso, I. and Nowak-Lehmann, F. (2025) *Immigration as a trade facilitator: a review and new insights*. Cuadernos Económicos de ICE, 109. doi: <https://doi.org/10.32796/cice.2025.109.7919>.

Gaulier, G. and Zignago, S. (2010) *BACI: international trade database at the product-level. The 1994–2007 version*. CEPII Working Paper No. 2010-23. Paris: CEPII. Available at: <http://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=2726>.

Girma, S. and Yu, Z. (2002) *The link between immigration and trade: evidence from the United Kingdom*. Weltwirtschaftliches Archiv, 138(1), pp. 115–130. doi: <https://doi.org/10.1007/BF02707326>.

Gould, D. M. (1994) *Immigrant links to the home country: empirical implications for U.S. bilateral trade flows*. The Review of Economics and Statistics, 76(2), pp. 302–316. doi: <https://doi.org/10.2307/2109884>.

Hatzigeorgiou, A. (2010) *Does immigration stimulate foreign trade? Evidence from Sweden*. Journal of Economic Integration, 25(2), pp. 376–402. Available at: <https://www.jstor.org/stable/23000981>.

Hatzigeorgiou, A. and Lodefalk, M. (2021) *A literature review of the nexus between migration and internationalization*. The Journal of International Trade & Economic Development, 30(3), pp. 319–340. doi: <https://doi.org/10.1080/09638199.2021.1878257>.

Head, K. and Ries, J. (1998) *Immigration and trade creation: econometric evidence from Canada*. The Canadian Journal of Economics, 31(1), pp. 47–62. doi: <https://doi.org/10.2307/136376>.

Lewer, J. J. and Van den Berg, H. (2009) *Does immigration stimulate international trade? Measuring the channels of influence*. The International Trade Journal, 23(2), pp. 187–230. doi: <https://doi.org/10.1080/08853900902779422>.

Olivero, M. P. and Yotov, Y. V. (2012) *Dynamic gravity: endogenous country size and asset accumulation*. Canadian Journal of Economics, 45(1), pp. 64–92. doi: <https://doi.org/10.1111/j.1540-5982.2011.01687.x>.

Parsons, C. R. (2012) *Do migrants really foster trade? The trade-migration nexus, a panel approach 1960–2000*. Policy Research Working Paper No. 6034. Washington, DC: World Bank. Available at: <http://documents.worldbank.org/curated/en/251531468321559329>.

Peri, G. and Requena-Silvente, F. (2010) *The trade creation effect of immigrants: evidence from the remarkable case of Spain*. Canadian Journal of Economics, 43(4), pp. 1433–1459. doi: <https://doi.org/10.1111/j.1540-5982.2010.01620.x>.

Rauch, J. E. (2001) *Business and social networks in international trade*. Journal of Economic Literature, 39(4), pp. 1177–1203. doi: <https://doi.org/10.1257/jel.39.4.1177>.

Santos Silva, J. M. C. and Tenreyro, S. (2006) *The log of gravity*. The Review of Economics and Statistics, 88(4), pp. 641–658. doi: <https://doi.org/10.1162/rest.88.4.641>.

Standaert, S. and Rayp, G. (2025) *Annual bilateral migration data – 1960–2022*. Mendeley Data, V2. doi: <https://doi.org/10.17632/cpt3nh6jct.2>.

Tinbergen, J. (1962) *Shaping the world economy: suggestions for an international economic policy*. New York: Twentieth Century Fund.

Yotov, Y. V., Piermartini, R., Monteiro, J. A. and Larch, M. (2016) *An advanced guide to trade policy analysis: the structural gravity model*. Geneva: World Trade Organization. Available at: https://www.wto.org/english/res_e/booksp_e/advancedwtountad2016_e.pdf.