Migration and the location of MNE activities: evidence from Italian provinces

Article

Supplemental Material

Migration and the location of MNE activities. Evidence from Italian provinces

Online Appendix

A Data Appendix

The data we use originate from the linkage of different data sources. In our final sample, the observations for which we have complete information about our variables amount to 111,692, covering 1,113 investment projects, 52 countries of origin, and 107 provinces.

Table A.1 summarizes the data sources. Table A.2 reviews the main steps in the data construction that lead to the current sample size. We provide more details on the construction of the dataset below.

Table A.1: Data Sources

Variable	**Source**
Location choice; Parent co-location; FDI characteristics	*fDi Markets*
Log immigrants; log multilateral immigrants	ISTAT–Demo – Foreign residents `http://demo.istat.it`
Log emigrants; log multilateral emigrants	Anagrafe Italiana Residenti all'Estero (AIRE)
Log prov. GDP	ISTAT (pre-2008); Istituto Tagliacarne
Log prov. population	ISTAT – Demo – Population
Institutional quality	Nifo and Vecchione (2014)
Infrastructure endowment	Istituto Tagliacarne
Residents with tertiary education	ISTAT – 2011 Census `http://dati-censimentopopolazione.istat.it`
Log patent count	Eurostat
Unemployment rate	ISTAT
Log average wage (region)	WHIP (Work Histories Italian Panel)
Firm density; Sectoral diversity; Manuf. concentration	AIDA – Bureau van Dijk
Agglomeration (sector)	AIDA – Bureau van Dijk
Pre-2002 FDI stock	REPRINT - ICE `http://actea.ice.it/ide.aspx`
Log imports; Log exports	ISTAT – Coeweb `https://www.coeweb.it`
Log distance; Common border	`https://www.matematicamente.it/staticfiles/approfondimenti/astronomia/CoordGeogProvince.pdf`, `www.wikipedia.org`, and `http://thematicmapping.org/downloads/world_borders.php`.
Immigration instrument 1995	ISTAT – Demo – Residence permits
Emigration instrument 1995–1999	ISTAT – Demo – Residential cancellations

Main sources of data in our sample.

FDI flow data are drawn from the *fDi Markets* database, a comprehensive and regularly

Table A.2: Sample

Sample operation	Reduction	Obs.
1,147 investments × 110 Italian provinces		126,170
Drop provinces founded in 2009 from pre-2010 choice set	769 × 3	123,863
Drop provinces founded in 2005 from pre-2006 choice set	243 × 4	122,891
Drop investments from Hong Kong (no immigration data)	946	121,945
Missing population data for new provinces	733+107	121,105
Missing inst. quality/infrastr. endowment new prov	846+428	119,831
Missing match with sectors in AIDA	5,171+859	113,801
Zero trade flows	2,104	111,697
Missing immigration data from Bermuda	5	111,692
Final sample: 1,113 investment projects, 52 countries, 107 provinces		**111,692**

Main operations leading to sample size reductions in our sample.

updated online database of cross-border greenfield investments constructed by the Financial Times Intelligence Unit. It covers all countries and sectors worldwide. We extracted from this repository the data relating to inward FDI to Italian provinces for which the destination city was available. These correspond to $1,147$ individual foreign direct investments into 110 Italian provinces (NUTS3 level) occurring over the 2003–2015 period[1], i.e. a choice set of $126,170$ investment–province pairs. Over the same period, the total number of provinces in Italy varied between 103 and 110: four, all located in Sardinia, were founded in 2005, and three, namely Barletta-Andria-Trani, Fermo, and Monza-Brianza, were founded in 2009 and are located in Apulia, Marche, and Lombardy, respectively. Twenty-five provinces were never chosen as an investment location. Of these, six are the newly-founded provinces located in the Centre-South[2]. The new province of Monza-Brianza was chosen as the destination for 5 investment ventures occurring after 2009, the year of its establishment. Employing the choice set as such would bear the paradoxical implication that Monza-Brianza, Barletta-Andria-Trani, and Fermo were among the location options available to investors even in the years before they were constituted. Excluding these three provinces from the alternatives available to the 769 investments occurring before 2010, the choice set reduces to a maximum of $123,863$ feasible alternatives. Similarly, we exclude the four provinces established in 2005 from the feasible set of locations for the 243 investments occurring before 2006, leading the choice set to further reduce to $122,891$.

Data on the main variables of interest, i.e. immigrant and emigrant stocks, are respectively drawn from the demography unit of ISTAT (the Italian Statistical Institute), which publishes yearly data on the foreign residents in each province by nationality since 2002, and from the electoral register of Italians residing abroad, the AIRE (*Anagrafe Italiana dei Residenti all'Estero*, as in Murat and Pistoresi, 2009), available on a yearly basis and disaggregated by province of origin and foreign country of residence. Immigrant data are available for 13 years, from 2002 to 2015. The ISTAT data lack information about immigrants originating from Hong Kong, as many of these hold British or—to a lesser extent—Chinese passports. Hong Kong, however, is a significant partner of Italian provinces, being home to 9 investments occurring over the considered period (5 before 2005, 3 between 2006 and 2010, and 1 after 2010), and the missing data problem applies to immigrants but not to emigrants.

[1]The data extraction was done during the second quarter of 2015, so the coverage for 2015 is up to the first quarter of the year.

[2]The remaining ones are Aosta, Asti, Belluno, Benevento, Catanzaro, Cosenza, Crotone, Enna, Grosseto, Imperia, Isernia, Oristano, Pistoia, Ravenna, Rieti, Rimini, Sondrio, Teramo, and Vibo Valentia. Ten of these are located in the South, four in the Centre, and five in the North.

Subtracting the 946 alternatives (i.e. $103 \times 5 + 107 \times 3 + 110$) relating to Hong Kong's choice of Italian provinces, the number of observations available for immigrants reduces to $121,945$. Emigrant data cover the entire set of origin countries of FDI but are currently available for eight years only (i.e. from 2006 to 2013). To preserve sample size, the data have been imputed for the missing period.[3] A limitation of both variables is that they refer to officially registered residents. Hence, they probably underestimate the actual stocks of both immigrants and emigrants. Note that, as is standard in the relevant literature (e.g. Rauch and Trinidade, 2002), we measure immigration and emigration as stocks in order to more closely proxy for the probability of interaction, and hence for the information effect to materialise.

As both immigrants and emigrants are included in the model as log stocks, we add one unit to both variables in order to tackle the indeterminacy of the log of zero. To impute the pre-2006 data, the available emigration data are first regressed on time with province fixed effects to get an estimate of the trend effect and of the average emigrants for a specific province–country pair, then the out-of-sample prediction for the pre-2006 period is added to the estimated fixed effects.

We also include a set of control variables:

1. *Province-level controls: population, aggregate value added, institutional quality, infrastructure endowment, count of patent applications, average wage, unemployment rate, share of residents with a tertiary degree.* As for the aggregate value added of the provinces, the pre-2008 data are drawn from the Italian National Statistical Institute, ISTAT; the post-2008 data are computed by the Istituto Tagliacarne and are publicly available.[4] The data on the resident population over the 2002–2015 period are drawn from the demography unit of ISTAT.[5] Annual data on GDP and population are available until 2014 and 2015, respectively. Province population data are missing for the 'new' provinces established with the administrative reforms of 2005 and 2009 in the years after their establishment, i.e. for 2006 and 2010, respectively, leading to a loss of 733 observations. Missing data for Monza-Brianza imply, in one case, that a project that actually located there in 2010 will have no observation for which *Choice*=1 and will be dropped from the sample, leading to a further reduction of 107 observations and to an estimation sample of $121,105$.

   The measure of institutional quality that we employ is drawn from Nifo and Vecchione (2014). The measure refers to the year 2007. Hence, it is considered as time-invariant for the purposes of the present work, and it is missing for the new provinces established in 2009.

   Due to its limited time variation, the infrastructural endowment index is only calculated for a limited number of years. It is publicly available for the years 2007, 2009, 2010, 2011, and 2012 and has been interpolated and extrapolated for the remaining years to cover the entire period. To impute the missing data, the available infrastructure endowment data were first regressed on time with province fixed effects to get estimates of the trend effects and province-specific estimates of the average mean infrastructure endowment, then the out-of-sample predictions for the missing years were added to the estimated fixed effects and used for imputation. No imputation could

[3] The results of the specifications that include the original non-imputed emigration data support the findings of the paper and are available upon request.

[4] `http://dati.italiaitalie.it`

[5] `http://demo.istat.it`

be performed for the province of Monza-Brianza, however, as the information about infrastructure endowment is not available for the new provinces, even in the post-2009 years. Per se, this reduction involves 846 observations. In addition, missing data for the new provinces imply that the remaining 4 projects that targeted Monza-Brianza will have no observation for which *Choice*=1, hence they are dropped from the sample, leading to a further reduction of 428 observations and to an estimation sample of $119,831$.

In regards to the province-level shares of residents with a tertiary degree, meant to proxy the human capital available in the province, the data are drawn from the 2011 census and are publicly available from ISTAT at `http://dati-censimentopopolazione.istat.it`. Tertiary education data are standardised in the empirical analysis.

To add a measure of the R&D intensity of the province, publicly available Eurostat data on the number of patent applications to the European Patent Office by province have been included for the years 2002–2012 (currently, they are not available at the province level for later years) and extrapolated for the later years. The extrapolation is performed as in the other cases: the available patent data are first regressed on time with province fixed effects, then the out-of-sample prediction for the 2013–2015 period is added to the estimated fixed effects.

Finally, we include the province-level unemployment rate. As this variable is only available for the 2004–2013 period (due to changes in the computation rules at ISTAT) but is available at the regional level from Eurostat data, we employed the region-level variation to impute the missing data for 2002–2003 and 2014, leading to a complete coverage of the provinces. In addition, annual wage data in euros originating from the social security data of the Work Histories Italian Panel (WHIP) (Bena et al., 2012) are averaged by NUTS2 region to get a proxy for the regional labour costs (unfortunately, this could not be done at the province level due to limited information on firm locations).

2. *Urbanisation economies: firm density, sectoral diversity, manufacturing concentration.* In order to assess the relevance of Jacobian externalities to the location choices for FDI and to control for potential confounding factors in the interpretation of our variable of interest at the same time, we employ three variables, all based on the AIDA database, which includes the firms registered in Italy above a given turnover threshold and includes details about the NACE rev.2 sector of the firms.[6] The data cover the 2002–2014 period. Firm density is computed as the ratio between all firms registered in AIDA in the province and the area of the province in km^2. It is intended to capture the role of urban externalities that arise from the agglomeration of many firms. Parallel to this, we exploit the sectoral information in the AIDA data to construct a province-level measure of 2-digit sectoral diversity computed as $1 - H$, where H is a standard Hirschman–Herfindahl concentration index. The index was standardised in the empirical analysis and is available for all provinces. Finally, we recognise that immigrants tend to concentrate in provinces where there is a greater concentration of manufacturing activities and that provinces with a greater concentration of firms in these sectors may have a different capacity to attract FDI. Neglecting to control for this factor may confound the results. To capture the effect of the sectoral composition of the province, for each year we identify the firms that are operating in the

[6]The version of AIDA we use is the largest available, the so-called 'full' one that covers firms above a fairly low turnover threshold (one million euros).

manufacturing sector based on their NACE codes and compute a location quotient of manufacturing activities calculated as follows: $\frac{f_{sit}/f_{it}}{f_{st}/f_t}$, where f_{sit} is the number of firms in province i and active in the manufacturing sector s in year t, f_{it} is the total number of firms in province i in any sector in year t, f_{st} is the total number of firms in manufacturing sectors in year t country-wide, and f_t is the total number of firms country-wide operating in year t.

3. *Sectoral agglomeration.* Considering that agglomeration economies of the Marshallian kind are likely to play an attractive role for FDI, we matched the sector of the investment with the corresponding agglomeration in each province. Province-level agglomeration is drawn from the AIDA database and is computed as a location quotient of the number of firms in the same sector of operation as the investment. The sectoral classification used in AIDA is the NACE rev.2. To match this with the sectoral classification used in the *fDi Markets* database, which partly resembles the NAICS classification, a conversion table was prepared. However, as the correspondence is not exact, the available correspondence table for the NAICS and NACE classification[7] could not be applied as such and the match was done manually. It is worth noting that the classification provided by the *fDi Markets* database allows distinguishing the function (classified under the category *industry activity*, e.g. headquarters, business services, manufacturing) from the sector of operation (classified under the category *industry sector*, e.g. aerospace, automotive components, biotechnology, which is further detailed by the variable *sub-sector*). The match was operated using the combination of these three categories. The NACE codes corresponding to such combinations do not uniquely correspond to a single level of partitioning (e.g. 2, 3, 4 digits). While in many cases it was possible to associate investments with the corresponding sectoral agglomeration at the 3-digit level, it was only possible to obtain a complete correspondence with the 2-digit level. The match with AIDA agglomeration data was not possible for specific combinations of provinces and sectors (corresponding to NACE sectors 06, 09, 12, 14, 19, 21, 24, 29, 30, 35, 50, 53, 59, 61, 65, 74, and 78), leading to a further loss of 5,171 observations. Due to the removal of cases where *Choice*=1, this further eliminates 19 projects from the sample, and 859 observations.

   The combined effect of the missing information on population, institutional quality, infrastructure endowment, and sectoral agglomeration is to exclude the 'new' provinces from the feasible set of locations, with the resulting sample size shrinking to 113,801, with 1,114 projects and 107 provinces covered.

4. *Bilateral (province–country) controls: Bilateral FDI stocks, bilateral trade, distance, common border.* Using the REPRINT - ICE database developed by the Polytechnic of Milan (`http://actea.ice.it/ide.aspx`), we constructed a measure of the bilateral stock of manufacturing FDI from one country into the same province between 1985 and 1997. The resulting bilateral stock should capture most time-invariant unobservable factors driving preferential relationships between particular country-province dyads. As for trade-flows, data are drawn from Italian international trade data publicly available at the province–country pair level (`https://www.coeweb.it`). Because downloading the data is an extremely time-consuming manual process, we opted to exclude minor remote islands from the analysis, a choice that did not affect the quality of the merge with the FDI data. The data cover both import and export flows over

[7] e.g. `http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_NACE_2_US_NAICS_2007`

the 2002–2015 period; trade between specific country–province pairs is zero in $1,439$ cases for the import data and in $1,165$ cases for the export data, mainly in southern provinces and provinces established with the administrative reforms of 2005 and 2009. When these variables are log-transformed and included in the model, the sample size shrinks by $2,104$ units to $111,697$. This does not affect the number of projects involved.

The distances are calculated as great circle distances as in Bratti et al. (2014), based on latitude and longitude (in decimal degrees) of provinces and partner countries.[8] A dummy variable for a common border is equal to 1 if the destination province is located in a region that is bordering the country of origin of the FDI, and 0 otherwise.

5. *Co-location.* Recent studies (e.g. Defever, 2006; Castellani and Lavoratori, 2020) highlight the positive effect on location choice of previous investments of the same parent company in a given province. This variable is based on the *fDi Markets* database. For each investment n of investing company f in year t, we compute the cumulative number of investments that the parent company has made in province i until year $t-1$. Due to the limited number of observations in our data, we are unable to disentangle the function of the previous investment or to employ a continuous variable capturing the number of previous investments. We opted to construct a binary variable for co-location that is equal to 1 in the case that the parent of the investing company has invested at least once in the same province in the year before the investment, and 0 otherwise. A limitation of this variable is that the number of investments can only be computed for the period covered by *fDi Markets*—only from 2003 on. Therefore, it neglects any investments occurring before this date.

6. *Total immigration and emigration.* To distinguish the effect of bilateral immigrants from those of immigrants from any country, we aggregate the country-specific stocks of immigrants and emigrants described above and take the log.

These sample operations have the effect of removing all but 5 alternatives for the only investment from Bermuda in our sample. We therefore drop the remaining 5 observations, leading us to our final estimation sample of $111,692$ observations.

A.1 Imputation of immigration stocks by skills level

Many studies have highlighted the role of skills in driving immigrant effects on FDI. Unfortunately, detailed yearly data on immigrants and emigrants by province, country of origin, and level of education are not available. This information is available from the 2011 census at the NUTS2 level (rather than NUTS3 level), however, for immigrants only. Based on this, we can approximate the shares of bilateral immigrants for each level of educational attainment. We do so by computing the shares of immigrants by level of educational attainment in the NUTS2 regions and then multiply by the stock of immigrants by province. The log of this stock is a measure of the stock of these bilateral immigrants by province and level of education (similarly to Colombelli et al., 2020, 'high-skilled' captures the share of bilateral immigrants with high-school and tertiary education at the NUTS2 level, and 'low-skilled' captures the corresponding share of bilateral immigrants with primary and lower-secondary education). While

[8] Source websites for the geographic coordinates include `https://www.matematicamente.it/staticfiles/approfondimenti/astronomia/CoordGeogProvince.pdf`, `www.wikipedia.org`, and `http://thematicmapping.org/downloads/world_borders.php`.

the measure is imperfect, as more qualified immigrants are likely to concentrate in the administrative capitals of the NUTS2 regions, it is the most accurate given current data availability.

A.2 Immigration and emigration instruments

To address the potential threat to our identification strategy that is represented by endogeneity, we impute immigration from each country into each province, as well as emigration from each province into each country, in the spirit of an Altonji–Card type of instrument. Given their strength and wide applicability, these instruments have become very popular in the immigration literature in general (see, e.g. Jaeger et al., 2018, for a critical review) and in studies of the link between migration and trade in particular (e.g. Bratti et al., 2014; Javorcik et al., 2011; Briant et al., 2014). In the original application of the instrument, the variable of interest was total migration so imputed country-specific stocks were aggregated to predict total stocks. In studies on trade and FDI like ours, the variable of interest is bilateral, hence the country-specific stocks are not aggregated.

Data availability imposes a slight difference between the immigration and emigration sides in constructing the instrument. As for immigration, we rely on data on the 1995 stocks of residence permits, which are detailed by province and country, as in Bratti et al. (2014)[9]. Following their approach, and similarly to Ottaviano and Peri (2006), we use these data to compute shares of immigrants from each nationality in each province that are long pre-determined with respect to the occurrence of the FDI. Specifically, we define w_{io95}^{imm} the province-level shares of immigrants from country o located in province i in 1995, computed as $w_{io95}^{imm} = \frac{imm_{io95}}{imm_{o95}}$, where imm_{io95} is the total stock of immigrants from country o to province i in 1995, and imm_{o95} the total stock of immigrants from country o in 1995 in Italy.

These shares are then used as weights to impute the province-level distribution of the overall nationwide stocks of immigrants from each country in the 2003–2015 period: $\widehat{imm}_{iot} = w_{io95}^{imm} \times imm_{ot}$. We add one unit to this imputed immigration variable and take its log when using it to predict observed log immigration. This allows reflecting both a 'push' factor from the side of the origin country and a 'recursive' factor (see also Burchardi et al., 2018) in imputing the distribution of immigrants.

The 'recursive' factor, as originally noted by Altonji and Card (1991) and Card (2001), refers to the fact that immigrants tend to ceteris paribus locate where there is a larger community of co-ethnics, to benefit from social and family ties. Hence, the lagged shares w_{io95}^{imm} capture a driver of current immigrant settlements that operates regardless of the current economic performance of the destination regions. To the extent that these are uncorrelated with unobservable drivers of FDI location, they will satisfy the exclusion restrictions. A concern in this regard may be that, although the first investment in our sample occurs at least 8 years later, unobserved time-invariant factors predict both the lagged shares of immigrants and the current FDI flows. The impact of this possibility on our estimates should be minimized by the inclusion of start-of-period FDI stocks among the covariates, that should effectively control for most bilateral heterogeneity and for the most pressing concerns that reverse causality

[9] 1995 is the first year in which the administrative boundaries of the provinces correspond to the current ones (with the exception of the new provinces created after 2005).

runs from FDI to the lagged shares. Moreover, the robustness of our results to the inclusion of bilateral dummies (Table C.7) is reassuring in this regard.

The overall stocks of immigrants moving to Italy at time t imm_{ot} proxy for 'push' factors from the country of origin, that cause migration stocks to change but are arguably uncorrelated with the structure of opportunities in a specific destination province. The approximation will be valid to the extent that Italy represents a relatively small share of total outward migration from country o and that individual provinces are unable to drive major shifts in overall stocks of immigrants from country o that target Italy. As argued by Bratti et al. (2014), an advantage of studying Italian provinces is that they are very small geographic units, which are unlikely to exert a major impact on country-of-origin outflows.[10]

Turning to the instrument for emigration, we adopt a similar approach and impute 2003–2015 yearly emigration from province i to country o by multiplying lagged bilateral shares by the overall nationwide stocks of emigrants from any provinces to a specific country: $\widehat{em_{iot}} = w^{em}_{io95-99} \times em_{ot}$. Again, we add one unit to imputed emigration and take its log when using it to predict observed log emigration. In this case, the shares will capture the presence of bilateral ties, for instance driven by historical emigration, that will make it more likely that expatriates from a particular province will move to a given country of destination. Data availability drives us to construct the shares based on lagged flows, rather than stocks, of residential cancellations, by province and country of destination. A potential problem arising from the use of flow data is that specific dyads take a zero weight due to yearly fluctuations. To address this issue, we aggregate the 1995–1999 emigrant outflows by province of origin and country of destination and take these aggregate flows as the basis for computing weights, i.e. the ratio of emigration flows between each country–province pair over total emigration flows in 1995–1999: $w^{em}_{io95-99} = \frac{\sum_{t=1995}^{1999} em_{iot}}{\sum_{t=1995}^{1999} em_{ot}}$. As above, the validity of the shares to capture the 'recursive' factor is subject to the assumption that there are no omitted time-invariant factors at play, conditional on other regressors. Given the greater volatility of flow compared with stock data, we may expect more time-invariant factors to remain unobserved in the imputation of emigration compared with immigration. While this possibility cannot be ruled out, as discussed in Section 5, the flow-based instrument and the other covariates appear to fairly accurately predict observed emigration and, once again, bilateral pre-period stocks of FDI should capture most bilateral heterogeneity.

Total time-varying stocks of emigration from any provinces in Italy to country o em_{ot} will capture the 'pull' factors that attract emigrants to a particular foreign country, under the assumption that each province represents a relatively small share of total emigration from Italy and is unable to affect overall emigration. This is the case for the southern provinces that score the highest numbers of expatriates. Yet, it may seem a strong assumption for internationally open provinces such as Milan and Rome. From this perspective, the robustness of our results to excluding Milan and Rome is reassuring. Again, the high correlation of the stocks targeting the same country over time indicates that our instrument will jointly capture the long-and short-run effects of migration.

[10]Notice that, as argued by Jaeger et al. (2018), as the country-specific stocks of immigrants are highly correlated over time (all pairwise correlations > 0.7), our migration instrument will conflate the effects of long- and shorter-run immigration.

B Simulated maximum likelihood and estimation of individual coefficients

B.1 Simulated maximum likelihood

As discussed in section 3.1.1, the parameters $\bar{\delta}$ and σ_δ that maximise the likelihood stemming from equation 3 can be estimated via simulation based on assumptions about the distribution $g(\delta)$ across firms.

The simulation process is as follows. At each draw r, values of δ are drawn from $g(\delta_f; \bar{\delta}; \sigma_\delta)$, starting from the estimated coefficients of the corresponding conditional logit and from standard deviations equal to 1. The logit probability that corresponds to each draw $L_{fni}(\beta, \delta^r)$ is calculated. These steps are repeated R times, and the average $\breve{P}_{fni} = \frac{1}{R}\sum_{r=1}^{R} L_{fni}(\beta, \delta^r)$ is the simulated probability that province i is chosen by investor f at choice n when the parameters are $\bar{\delta}$ and σ_δ. The log of the simulated probabilities, weighted by whether the province was actually chosen, are inserted into the log-likelihood function to give a simulated log likelihood. The values of $\bar{\delta}$ and σ_δ that maximise the simulated log-likelihood will be our estimated distributional parameters. All mixed logit models were run in Stata using the user-written command mixlogit (Hole, 2007), implemented in each case using 500 Halton draws and taking into consideration the occurrence of repeated location choices of different investments by the same firm.

B.2 Computation of individual coefficients

The investor-specific effect of immigrants (emigrants) δ_f can be retrieved from the conditional distribution of the effects $h(\delta|\text{Choice}_{fin}, x_{fni}, z_{fni}, \hat{\beta}, \hat{\bar{\delta}}, \hat{\sigma}_\delta)$ for investors that would choose province i when facing a choice situation described by values x_{fni} and z_{fni} of the location determinants, and given the estimated parameters $\hat{\beta}$, $\hat{\bar{\delta}}$, and $\hat{\sigma}_\delta$. Being conditional not only on the covariates but also on the estimated distributional parameters and on the observed choice, this distribution will generally be different than the one estimated for the entire sample. It can be obtained by Bayes' rule and will be proportional to the product of the probability that province i is chosen given covariates and fixed coefficients when the random coefficients take value δ_f, times the estimated density of $g(\hat{\bar{\delta}}, \hat{\sigma}_\delta)$ in the entire sample. In other words, it will be proportional to the product of the probability of choosing i given δ_f, times the probability of δ_f (see Train, 2009, for a more formal and detailed explanation). The expected value of this distribution conditional on the investor-specific covariates, observed choice, and estimated parameters is an estimate of the individual-specific immigrant effects δ_f. Again, its value cannot be estimated directly but can be obtained via simulation. Likewise, we resort to parametric bootstrapping to estimate the standard error σ_{δ_f} of the estimated individual coefficients $\hat{\delta}_f$.[11]

[11] Computation of individual δ_f is implemented in Stata using the `mixlbeta` command (Hole, 2007) with 500 Halton draws. As for the estimation of σ_{δ_f}, we take $S = 1,000$ random draws from a multivariate normal distribution with means and covariances equal to the estimated parameters of the first stage, and for each draw we compute the individual-level coefficients $\hat{\delta}_f^s$ for the variables of interest (log immigrants and log emigrants). We then compute as standard error of the individual parameters the standard deviation $\hat{\sigma}_{\delta_f}$ of the simulated $\hat{\delta}_f^s$ over the $1,000$ simulations. We thank Arne Risa Hole for helpful guidance in this process.

C Additional results

C.1 Additional descriptive statistics

Table C.1 reports the first 20 countries of origin of Italian FDI, which account for about 87% of our sample of investments. High-income OECD countries represent the vast majority of the origin countries of Italian FDI, with more than half of overall investments originating from only four countries: the US, UK, Germany, and France. In this very concentrated distribution of origin countries of FDI, some relevant origin countries for immigrants appear to have a role, and in particular China, which ranks relatively high, but also the Philippines, India, and Russia, which rank among the first 20 countries, even though their contribution in absolute terms is limited (see Table C.3 below for the list of origin and destination countries of migrants). The right-hand panel of the table reports the composition of FDI in terms of function, displaying high heterogeneity across countries. Indeed, while market-access and business services FDIs represent a relevant share of the investments by most origin countries, manufacturing and R&D investments do not present a clearly discernible pattern according to country-level determinants such as GDP, distance, and institutional similarity, for instance. This suggests that function-specific considerations may matter more to the investment decisions than origin-country characteristics.

Table C.2 distinguishes Italian inward FDI by function and reports frequencies as well as the average capital investment in each function. The vast majority of inward FDIs in our sample (which excludes franchising FDI) is represented by what we call 'market-access' FDIs (i.e. those classified in *fDi Markets* as 'Sales, Marketing & Support' and 'Customer Contact Centres'). Several FDIs also classify as 'Business Services' and 'Manufacturing'. Instead, our definition of R&D FDI (corresponding to FDI in the functions of 'Research and Development' and 'Design, Development & testing' in *fDi Markets*) corresponds to a smaller number of ventures. *fDi Markets* data also provide the capital investment and the number of jobs created, yielding a measure of investment size.[12] Among market-access FDIs, those categorised as 'Sales, Marketing and Support' are characterised by a relatively small capital investment and a relatively low number of jobs created. Similar considerations apply to 'Business Services' investments, where the average capital investment is 7 million US$ higher than the previous category but the average number of jobs created is slightly lower. Detailed inspection of the microdata (not shown) reveals that the vast majority of 'Sales, Marketing and Support' investments consist of investments in sales representative offices intending to promote the sales of the parent company products, mainly ICT-related, in Italy. To the largest extent, they come from the US, the UK, and France. The majority of 'Business Services' comprise investment in advertising and financial services. Compared with the previous category, business services FDIs display a more diversified range of origin countries and many of them establish retail banking branches that are likely to mainly serve the immigrant population (e.g. Banque Centrale Populaire from Morocco, Bank of the Philippines, Bank of Communications Shanghai).

As for investments in R&D (investments in 'Research & Development' and 'Design, Development & Testing'), these are relatively labour-intensive but plausibly more reliant on localised knowledge.

In contrast, manufacturing FDIs, as well as FDIs in 'Electricity', 'Logistics, Distribution

[12] In some cases, these variables are estimated in *fDi Markets* on the basis of historical data from similar projects in similar sectors and activities.

Table C.1: **Origin countries of FDI in Italy**

Country	FDI count *(% of all inward FDI)*	R&D	Manuf.	Mkt Acc.	Bus. Serv.	Logistics	Constr.	Other
		Share of functions by country:						
United States	283 *(25.43 %)*	9.89	11.66	33.57	23.32	5.65	6.01	9.89
United Kingdom	130 *(11.68 %)*	3.85	2.31	29.23	43.08	3.85	10.77	6.92
Germany	116 *(10.42 %)*	4.31	11.21	50.86	18.1	7.76	0	7.76
France	101 *(9.07 %)*	3.96	17.82	43.56	16.83	6.93	1.98	8.91
Spain	91 *(8.18 %)*	1.1	4.4	37.36	8.79	4.4	34.07	9.89
Switzerland	58 *(5.21 %)*	10.34	15.52	31.03	15.52	10.34	3.45	13.79
Japan	35 *(3.14 %)*	5.71	40	31.43	8.57	8.57	0	5.71
China	31 *(2.79 %)*	29.03	0	58.06	12.9	0	0	0
Netherlands	29 *(2.61 %)*	0	27.59	20.69	10.34	17.24	20.69	3.45
Belgium	26 *(2.34 %)*	0	38.46	23.08	23.08	3.85	7.69	3.85
Austria	25 *(2.25 %)*	0	4	32	52	4	0	8
Ireland	24 *(2.16 %)*	12.5	0	12.5	20.83	50	0	4.17
Canada	20 *(1.8 %)*	5	5	65	15	0	5	5
Sweden	18 *(1.62 %)*	16.67	5.56	66.67	11.11	0	0	0
Finland	14 *(1.26 %)*	0	57.14	35.71	0	0	0	7.14
Philippines	11 *(0.99 %)*	0	0	0	100	0	0	0
United Arab Emirates	8 *(0.72 %)*	0	12.5	50	0	0	37.5	0
India	7 *(0.63 %)*	14.29	28.57	42.86	0	14.29	0	0
Morocco	7 *(0.63 %)*	0	0	0	100	0	0	0
Korea	5 *(0.45 %)*	20	0	40	0	0	0	40
Other countries	*79* (7.1%)	*3.8*	*18.99*	*35.44*	*17.72*	*8.86*	*5.06*	*10.13*
TOTAL	**1,113** ***(100%)***	**6.38**	**12.67**	**36.39**	**22.28**	**6.92**	**7.37**	**8**

Distribution of inward FDI in Italy by origin country and function, 2003–2015. Source: own elaborations on *fDi Markets*.

Table C.2: Functions of FDI

Function		FDI count	Share (%)	Capital investment[a]	Jobs created[b]
Market access:		*405*	*36.39*	*10.15*	*16.81*
	Sales, Marketing & Support	396	35.58	10.23	11.65
	Customer Contact Centre	9	0.81	6.86	243.89
Business Services		*248*	*22.28*	*17.08*	*10.14*
Manufacturing		*141*	*12.66*	*103.51*	*156.24*
Construction		*82*	*7.37*	*61.23*	*244.1*
Logistics, Distribution & Transportation		*77*	*6.92*	*105.72*	*121.34*
R&D:		*71*	*6.38*	*31.33*	*79.23*
	Design, Development & Testing	45	4.04	23.60	77.33
	Research & Development	26	2.34	44.71	82.50
Other functions:					
Electricity		24	2.16	163.68	58.88
Headquarters		22	1.98	31.71	192.09
Education & Training		16	1.44	11.66	31.94
ICT & Internet Infrastructure		14	1.26	105.98	49.79
Maintenance & Servicing		9	0.81	8.98	59.89
Recycling		2	0.18	26.00	40.50
Technical Support Centres		2	0.18	9.70	39.00
TOTAL		**1,113**	**100.0**	**40.23**	**66.38**

Characteristics of FDI by function. [a]Average value of capital invested per investment in million US$. [b]Average jobs created per investment. *Source: fDi Markets.*

& Transportation', 'ICT & internet infrastructure', and 'Extraction', display the highest average capital investments. The number of jobs created is also relatively high. Construction investments entail relatively lower investments but create a larger number of jobs, on average.

The geographic distribution of the different functions is reported in Figure C.1.

Table C.3 reports the breakdown of our migration variables by country. The left (resp. right) panel reports the top 20 origin countries of immigrants (resp. top 20 destination countries of emigrants) in 2011. The set of immigrant origin countries is very diversified and covers all continents, with only limited overlap with the set of origin countries for FDI, mainly due to China, the Philippines, and India. Germany and France, while not featured in the top 20 origin countries, display an above-average number of residents in Italian provinces. The set of destination countries for Italian expatriates is instead mainly represented by OECD countries, and it displays greater overlap with FDI origin countries—with the exception of China and Japan.

Table C.4 reports the top 20 provinces by multilateral immigration and emigration rates, i.e., by the stocks of immigrants and emigrants from any countries of origin (left) and destination (right) as a percentage of population. The very high immigration rates in Prato, which hosts a very large immigrant community from China, stand out from this table. Otherwise, immigration is concentrated in the economically most dynamic provinces of the North and Center, mainly located in Lombardy, Emilia-Romagna, and Veneto. Instead, multilateral emigration is disproportionately high in Southern provinces belonging to the regions of Sicily, Calabria, Molise, Basilicata, and Campania, as well as in provinces of the North-East with large historical diasporas (belonging to the regions of Veneto and Friuli-Venezia Giulia).

Table C.3: Origin and destination countries of immigrants and emigrants

Origin countries		Destination countries	
Country	Immigrants	Country	Emigrants
Romania	7,656.75	Argentina	6,089.07
Albania	4,117.84	Germany	5,900.72
Morocco	3,726.02	Switzerland	5,072.92
China	1,804.92	France	3,369.14
Ukraine	1,641.79	Brazil	2,774.87
Moldova	1,219.03	Belgium	2,339.62
Philippines	1,198.33	United States	2,009.59
India	1,094.36	United Kingdom	1,864.19
Peru	845.41	Canada	1,257.11
Poland	779.17	Australia	1,215.28
Tunisia	763.63	Spain	1,089.28
Bangladesh	745.42	Venezuela	1,051.23
Ecuador	712.91	Uruguay	822.13
FYR Macedonia	684.85	Chile	462.34
Senegal	675.05	Netherlands	315.53
Sri Lanka	652.85	South Africa	287.76
Pakistan	629.74	Peru	280.86
Egypt	605.58	Luxembourg	215.31
Nigeria	446.64	Austria	184.72
Ghana	411.35	Ecuador	132.17
Entire sample[(a)]	*206.11*	*Entire sample*[(b)]	*218.66*

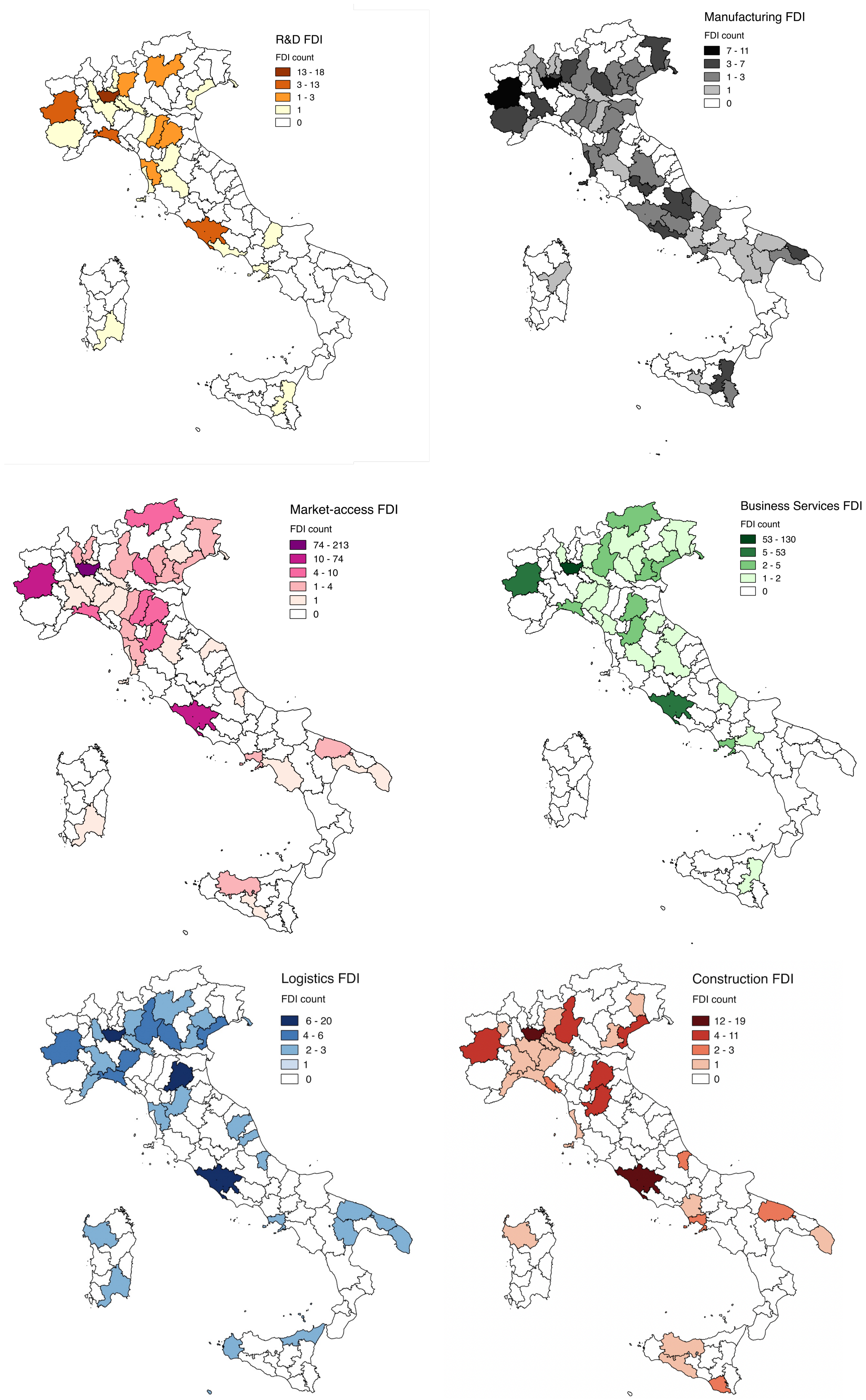
Ranking of the top 20 origin and destination countries of immigrants and emigrants in Italy in Italian provinces, 2011. Observations: 107. Left panel: Average province-level stocks of foreign residents by country of origin (Source: ISTAT). Right panel: Average province-level stocks of residents registered abroad by destination country (Source: AIRE). [(a)]Average bilateral (country–province) stocks of foreign residents in the 107 provinces in our sample. [(b)]Average bilateral (country–province) stocks of Italian citizens residing abroad in the 107 provinces in our sample.

Table C.4: Top 20 provinces by multilateral immigration and emigration rate

Province	Immigration rate	Province	Emigration rate
Prato	13.40	Enna	40.85
Piacenza	12.56	Agrigento	32.00
Brescia	12.45	Vibo Valentia	28.35
Mantova	12.08	Isernia	27.54
Reggio Emilia	11.96	Caltanissetta	25.68
Modena	11.91	Potenza	24.59
Parma	11.01	Campobasso	24.26
Treviso	10.68	Avellino	23.51
Milano	10.66	Cosenza	20.67
Pordenone	10.57	Belluno	20.43
Asti	10.43	Benevento	17.43
Perugia	10.37	Chieti	17.37
Bergamo	10.30	Catanzaro	16.60
Verona	10.25	Pordenone	15.07
Cremona	10.17	Reggio Calabria	14.83
Ravenna	10.17	Crotone	14.59
Lodi	10.09	Udine	13.25
Macerata	10.05	L'Aquila	12.36
Firenze	9.92	Messina	11.78
Forlì-Cesena	9.89	Lecce	11.42
Ave. Italy$^{(a)}$	*6.33*	*Ave. Italy*$^{(b)}$	*8.39*

Ranking of the top 20 provinces by multilateral immigration and emigration rates, 2011. Observations: 110. Left panel: Percentage ratio of total immigration stocks from any countries of origin to province population (Source: ISTAT). Right panel: Percentage ratio of total emigration stocks towards any countries of destination to province population (Sources: AIRE/ISTAT). $^{(a)}$Average percentage immigration rate across the 110 Italian provinces in our sample; $^{(b)}$Average percentage emigration rate across the 110 Italian provinces in our sample.

Figure C.1: **Distribution of FDI by province for selected functions (2003-2015)**



Total number of inward FDI projects by province and function over the 2003–2015 period. Source: Own elaborations on fDi Markets data.

C.2 Robustness checks

A concern about the robustness of the estimates may derive from our descriptive analysis. Indeed, the high concentration of investment ventures in the provinces of Milan and Rome leads to a disproportionately higher probability of choosing these destinations for a venture seeking to locate in Italy. The global cities located in these provinces, which host many immigrants, may drive our results. The peculiarities of the decision process leading foreign firms to locate in one of these two global cities may be not fully captured by the inclusion of the 'Rome/Milan' dummy. Therefore, we re-estimate Model 1 in Table 5 on the subsample that excludes all investments with Milan or Rome as a destination and report the results in panel (a) of Table C.5. This reduces our sample size by about half but the main results are confirmed, thereby refuting that the prominent role of Rome and Milan is the main driver of the immigrant effects.

Our second robustness check refers to the literature that has highlighted an effect for migrant skills in promoting FDI (e.g. Docquier and Lodigiani, 2010; El Yaman et al., 2007; Foad, 2012; Gheasi et al., 2013; Javorcik et al., 2011; Kugler and Rapoport, 2007). Unfortunately, yearly data on immigrants and emigrants by province, country of origin, and level of education are not available but can be imputed—for immigrants only—based on census data (see Appendix A for details). In panel (b) of Table C.5, we substitute the log of the immigrants' stock with the imputed stocks of high- and low-skilled immigrants. In line with previous literature, the effect of bilateral immigrants is positive and highly significant in regions where the share of highly skilled immigrants from the same country is greater. The effects are significantly heterogeneous, confirming that the heterogeneity in the effects is not exhausted by the immigrants' skills. Both the mean and standard deviations are instead insignificant in regions with greater shares of low-skilled immigrants.

In a third set of robustness checks, we address the potential concerns arising from geographic spillover effects. Indeed, omitted variables may be operating on a wider geographic scale than the provinces and may correlate with the included regressors, inducing correlation across the errors related to different provinces (Bratti et al., 2014). Immigrants of the same nationality located in neighbouring provinces may affect the location of FDI in a specific province, for instance, if they are mobile across provinces. To allow for this possibility, we augment Model 1 in Table 5 with the stocks of immigrants from the same country of origin as the investment that are located within a specified radius from the province centroid: less than 50 km, between 50 and 100 km, between 100 and 200 km, and over 200 km (see Bratti et al., 2014, on the construction of the variable); we also construct a corresponding variable for emigrants targeting the country of the investment from neighbouring provinces and include spatial spillover effects for GDP, as a proxy for market access.

The results reported in Table C.6 confirm the picture painted so far. The importance of market access in the location choice for FDI is confirmed by the positive and significant effect of the GDP in neighbouring provinces (i.e. within a 50-km radius; see Model 1a). As we might expect, this result is be driven by the two most economically dynamic provinces in Italy, i.e. Rome and Milan, and disappears when these provinces are excluded from the sample (see Model 1b). As for immigrants, their positive, significant, and significantly heterogeneous effect in the destination province is confirmed and turns out to be highly localised. A less precise negative effect of immigrants within a 100–200 km radius emerges, suggesting the presence of patterns of competition among provinces, which is again driven by the presence of Rome and Milan in the sample. As for emigrants, we are still unable to detect a positive role for them, on average. Nonetheless, many emigrants in more distant provinces are found to discourage location. This may suggest that emigrant effects operate at a different scale, possibly less localised, than immigrant effects (similarly to what is observed by D'Ambrosio and Montresor, 2022, for trade), and calls for further research in this regard.[13] These results support the interpretation that the location choice for FDI operates at a very fine-grained geographical scale and that the choice of the NUTS3 level as a unit of analysis is appropriate, at least for the immigrant effects.

Finally, we recognise that while the results are robust to the inclusion of province dummies and to addressing endogeneity, omitted variables might be operating at the dyadic province–country level. In our application, we face the limitation that for a vast majority of province–country pairs (75.5%), the number of investments per dyad is zero. Within the subset of dyads with at least one investment, about 89% do not exceed 5 investments, with 55% being tied by a single investment and another 18% by only two investments. Overall, this does not allow us to meaningfully estimate the effect of migration conditional on pair dummies. Yet, bilateral omitted variables can be addressed for dyads that are tied by several investments. Therefore, we adopt an intermediate solution and include within our control function approach 15 province–country dummies capturing the dyads that share more than 10 investments. This approach allows us to estimate the variables of interest while controlling for the most pressing potential sources of time-invariant bilateral omitted variable bias. Results are reported in Table C.7, and suggest that our previous findings are not substantially affected by omitted variables operating at the dyadic level.

Overall, the robustness checks confirm a significant, positive, and significantly heterogeneous effect of immigration on the location choice for FDI, and are reassuring about the validity of our main estimates.

[13] Emigration is highest in Southern provinces but also in the North-East and in large metropolitan areas country-wide. Therefore, this result is unlikely to come from the North-South divide.

Table C.5: Robustness checks - Sample and skills

	Mean	**SD**
(a) Sample without Rome and Milan		
Log immigrants$_{ijt-1}$	$0{\cdot}340^{***}$	$0{\cdot}411^{***}$
	(0·106)	(0·093)
Log emigrants$_{ijt-1}$	0·067	0·020
	(0·085)	(0·234)
(b) Immigrants by qualification level at the NUTS2 level		
Log (immigrants$_{ijt-1} \times s_{rj}^{\mathrm{hq}}$)	$0{\cdot}522^{**}$	$0{\cdot}366^{***}$
	(0·224)	(0·086)
Log (immigrants$_{ijt-1} \times s_{rj}^{\mathrm{lq}}$)	$-0{\cdot}132$	$-0{\cdot}150$
	(0·218)	(0·238)
Log emigrants$_{ijt-1}$	0·107	0·007
	(0·082)	(0·113)

Location choice models for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. For each model, the table reports the estimated average effect of the variable (Mean) and standard deviation (SD). Model (a) replicates the specification in Model 1, Table 5, but excludes all investments having targeted Rome or Milan and Rome and Milan from the choice set, and therefore the Rome/Milan dummy. Covariates with fixed parameters: parent co-location, unemployment rate, infrastructure endowment. Covariates with random parameters: residents with tertiary education, log prov. population, institutional quality, common border, log prov. GDP, sectoral diversity, log patent count, log average wage (region), log distance, agglomeration (sector), log imports, log exports, pre-2002 FDI stock, log multilateral immigrants, log multilateral emigrants, firm density, manufacturing concentration.
Model (b) distinguishes the effect of bilateral immigrants between highly qualified (*hq*) and less qualified (*lq*), based on an imputation procedure which draws on regional data and is described in Appendix A. The specification contains the same set of regressors as Model (a) but a slightly different set of random parameters: log immigrants$\times s^{\mathrm{hq}}$, log immigrants$\times s^{\mathrm{lq}}$, log prov. GDP, log distance, pre-2002 FDI stock, log multilateral immigrants, firm density, manufacturing concentration, and the Rome/Milan dummy. Standard errors (in parentheses) take into account the correlation among the investments by the same company. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Robustness checks - Geographic spillovers

Dep. var: Choice	**Model 1a**		**Model 1b**	
	Full sample		**No Rome Milan**	
	Mean	**SD**	**Mean**	**SD**
Log prov. GDP	0·685	−0·038	0·993	−0·000
	(0·504)	(0·198)	(0·766)	(0·320)
Log prov. GDP <50km	0·109***	−0·001	0·050	−0·001
	(0·042)	(0·035)	(0·053)	(0·035)
Log prov. GDP 50–100km	0·281	0·047	0·270	−0·144
	(0·192)	(0·424)	(0·217)	(0·213)
Log prov. GDP 100–200km	0·287	0·039	0·226	−0·037
	(0·265)	(0·148)	(0·306)	(0·148)
Log prov. GDP >200km	−0·165	−0·740	−0·936	0·292
	(0·965)	(0·810)	(1·110)	(1·213)
Log immigrants	0·463***	0·383***	0·426***	0·431***
	(0·109)	(0·076)	(0·113)	(0·090)
Log imm. <50km	−0·094	0·007	0·028	0·001
	(0·072)	(0·058)	(0·088)	(0·050)
Log imm. 50–100km	−0·067	0·028	−0·065	0·008
	(0·144)	(0·239)	(0·163)	(0·214)
Log imm. 100–200km	−0·342*	−0·026	−0·302	−0·001
	(0·180)	(0·098)	(0·218)	(0·126)
Log imm. >200km	−0·168	0·184	0·445	0·000
	(0·657)	(0·659)	(0·798)	(0·728)
Log emigrants	0·030	−0·009	0·016	0·004
	(0·084)	(0·122)	(0·099)	(0·165)
Log em. <50km	−0·003	0·001	−0·006	−0·000
	(0·052)	(0·062)	(0·065)	(0·042)
Log em. 50–100km	−0·056	0·024	−0·068	−0·008
	(0·092)	(0·182)	(0·118)	(0·140)
Log em. 100–200km	−0·012	0·002	0·004	−0·002
	(0·106)	(0·087)	(0·133)	(0·101)
Log em. >200km	−1·532***	0·001	−1·629**	−0·005
	(0·562)	(0·587)	(0·689)	(0·594)
Observations	111,692		50,267	
AIC	4,121.794		2,514.822	
BIC	4,651.086		2,982.552	

Location choice models for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. For each model, the table reports the estimated average effect of the variable (Mean) and standard deviation (SD). In Model 1a, we include the values of log GDP, log of immigrants, and log of emigrants for the focal province along with the value of the neighbouring provinces within a set of radiuses from the province centroid (<50 km,50–100 km,100–200 km,>200 km). Control variables correspond to that of Model 1 in Table 5. Covariates with fixed parameters: parent co-location, unemployment rate, infrastructure endowment, residents with tertiary education, log prov. population, institutional quality, common border, log prov. GDP, sectoral diversity, log patent count, log average wage (region), log distance, agglomeration (sector), log imports, log exports, pre-2002 FDI stock, log multilateral immigrants, log multilateral emigrants, firm density, manufacturing concentration. Covariates with random parameters: log distance, pre-2002 FDI stock, firm density, manufacturing concentration, and the Rome/Milan dummy. Model 1b employs the same specification but excludes all investments having targeted Rome or Milan and Rome and Milan from the choice set, and therefore the Rome/Milan dummy. Standard errors (in parentheses) take into account the correlation among the investments by the same company. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: Control function results including province–country dummies for pairs with more than 10 investments

Dep. var.: Choice	**Model 4a**	
	Mean	**SD**
Log immigrants	0·396***	0·328***
	(0·115)	(0·081)
Log emigrants	−0·199	0·005
	(0·162)	(0·120)
Residuals (immigrants)	−0·101	0·004
	(0·118)	(0·166)
Residuals (emigrants)	0·313**	−0·168
	(0·152)	(0·179)
Observations	111,692	
AIC	4,154.256	
BIC	4,731.666	
LR test of joint significance of the SD	100.255	
Degrees of freedom	20	
Test p-value	0.000	

Location choice model for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. The table reports the estimated average effect of the variable (Mean) and standard deviation (SD). The reported specification adds 15 province–country dummies capturing the dyads tied by more than 10 investments, included with fixed coefficients, to the variables in Model 4 of Table 5. Covariates with fixed parameters: parent co-location, unemployment rate, infrastructure endowment, residents with tertiary education, sectoral diversity. Covariates with random parameters: log prov. population, institutional quality, common border, log prov. GDP, log patent count, log average wage (region), log distance, agglomeration (sector), log imports, log exports, pre-2002 FDI stock, log multilateral immigrants, log multilateral emigrants, firm density, manufacturing concentration, Rome/Milan dummy. The second-stage regression associated with this model is reported in Online Appendix Table C.11. Standard errors (in parentheses) take into account the correlation among the investments by the same company, but are not bootstrapped (see note 22 in the main text). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Additional econometric evidence

Table C.8: Control function - all migration variables

Dep. var.: Choice	Mean	SD
Log Immigrants	0·413***	0·355***
	(0·112)	(0·083)
Log Emigrants	−0·094	−0·020
	(0·169)	(0·132)
Parent colocation	5·353***	
	(0·159)	
Unemployment rate	−0·022	
	(0·037)	
Infrastructure endowment	0·001	
	(0·001)	
Residents with tertiary education	0·087	
	(0·121)	
Log Prov. Population	0·148	
	(0·608)	
Institutional Quality	−0·235	−0·109
	(0·766)	(0·758)
Common border	0·209	0·250
	(0·207)	(0·446)
Log prov. GDP	0·376	
	(0·595)	
Sectoral diversity	0·050	
	(0·107)	
Log Patent Count	0·070	−0·091
	(0·145)	(0·112)
Log average wage (region)	−1·046	−0·949
	(2·147)	(1·785)
[0.3em] Log Distance	−0·193	0·987**
	(0·302)	(0·451)
Agglomeration (Sector)	0·197***	−0·003
	(0·032)	(0·070)
Log Imports	0·192***	0·145*
	(0·073)	(0·082)
Log Exports	0·095	0·009
	(0·078)	(0·101)
Pre-2002 FDI stock	0·009***	0·007
	(0·003)	(0·010)
Log multilateral immigrants	0·040	0·110
	(0·503)	(0·157)
Log multilateral emigrants	−0·115	0·062
	(0·309)	(0·175)
Firm density	0·001	0·015***
	(0·006)	(0·003)
Manufacturing concentration	−0·521	−0·101
	(0·356)	(0·301)
Rome/Milan	−0·182	−0·844**
	(0·539)	(0·355)
Residuals (immigrants)	−0·047	0·036
	(0·118)	(0·183)
Residuals (emigrants)	0·180	−0·154
	(0·155)	(0·193)
Residuals (multilateral immigrants)	−0·450	0·008
	(0·577)	(0·516)
Residuals (multilateral emigrants)	−0·111	−0·346
	(0·322)	(0·375)
Observations	110,269	
AIC	4,135.291	
BIC	4,586.993	
LR test of joint significance of the SD	116.41	
Degrees of freedom	20	
Test p-value	0.000	

Location choice models for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. The table reports the estimated average effect of the variable (Mean) and standard deviation (SD). The set of regressors corresponds to the one in Model 4 of Table 5, plus two additional control functions: (i) Residual (multilateral immigrants) are the residuals from a regression of multilateral immigration on its multilateral instrument, the multilateral emigration instrument, the two bilateral migration instruments and the remaining exogenous covariates. (ii) Residual (multilateral emigrants) are the residuals from a regression of multilateral emigration on its multilateral instrument, the multilateral immigration instrument, the two bilateral migration instruments and the remaining exogenous covariates. Standard errors (in parentheses) take into account the correlation among the investments by the same company, but are not bootstrapped (see note 22 in the main text). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: Mixed logit results without co-location

Dep. var.: Choice	**Model B1**		**Model B2**	
	Mean	**SD**	**Mean**	**SD**
Log immigrants	0·332***	0·395***	0·436***	0·456***
	(0·095)	(0·107)	(0·109)	(0·100)
Log emigrants	0·139*	0·209**	−0·174	0·131
	(0·077)	(0·106)	(0·143)	(0·131)
Unemployment rate	−0·017		−0·014	
	(0·033)		(0·033)	
Infrastructure endowment	0·001		0·001	
	(0·001)		(0·001)	
Residents with tertiary education	0·165*	0·363***	0·142*	
	(0·090)	(0·110)	(0·085)	
Log prov. population	−0·404	−0·357**	−0·353	−0·265
	(0·466)	(0·181)	(0·467)	(0·193)
Institutional quality	0·648	−0·596	0·297	−0·449
	(0·604)	(0·926)	(0·615)	(1·186)
Common border	0·169	0·970***	0·108	0·891**
	(0·177)	(0·299)	(0·179)	(0·369)
Log prov. GDP	1·136***	0·076	1·302***	0·374 *
	(0·436)	(0·211)	(0·434)	(0·197)
Sectoral diversity	0·053	−0·123	0·029	
	(0·079)	(0·169)	(0·074)	
Log patent count	0·167	0·238*	0·205	0·340**
	(0·135)	(0·124)	(0·152)	(0·136)
Log average wage (region)	−0·693	−3·679*	−0·123	2·693
	(1·572)	(1·950)	(1·478)	(1·916)
Log distance	−0·580*	1·744***	−0·609*	2·283***
	(0·309)	(0·443)	(0·327)	(0·454)
Agglomeration (sector)	0·266***	0·057	0·271***	−0·073
	(0·031)	(0·061)	(0·031)	(0·059)
Log imports	0·231***	0·106	0·232***	0·110
	(0·058)	(0·091)	(0·062)	(0·096)
Log exports	0·024	0·010	0·084	0·054
	(0·066)	(0·094)	(0·071)	(0·081)
Pre-2002 FDI stock	0·011***	0·013	0·008***	0·010
	(0·003)	(0·011)	(0·003)	(0·008)
Log multilateral immigrants	0·088	0·488***	−0·084	0·368**
	(0·238)	(0·131)	(0·236)	(0·182)
Log multilateral emigrants	−0·077	−0·004	0·072	0·053
	(0·109)	(0·096)	(0·129)	(0·123)
Firm density	0·003	−0·002	0·004	0·011**
	(0·004)	(0·010)	(0·004)	(0·005)
Manufacturing concentration	−0·583**	0·110	−0·658***	0·480
	(0·250)	(0·857)	(0·249)	(0·401)
Rome/Milan	−0·858*	1·460***	−1·092**	0·957**
	(0·448)	(0·511)	(0·465)	(0·483)
Residuals (immigrants)			−0·180*	−0·146
			(0·103)	(0·216)
Residuals (emigrants)			0·301**	0·356**
			(0·131)	(0·136)
Observations	111,692		111,692	
AIC	6,037.297		6,040.483	
BIC	6,441.484		6,463.917	
LR test of joint significance of the SD	129.469		127.871	
Degrees of freedom	20		20	
Test p-value	0.000		0.000	

Location choice models for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. For each model, the table reports the estimated average effect of the variable (Mean) and standard deviation (SD). The set of covariates in Model B1 corresponds to the one in Model 1 of Table 5, except co-location. The set of covariates in Model B2 corresponds to the one in Model 4 of Table 5, except co-location. Standard errors (in parentheses) take into account the correlation among the investments by the same company, but are not bootstrapped (see note 22 in the main text). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: Mixed logit estimates: 2008–2015 subperiod

Dep. var.: Choice	Model B3	
	Mean	**SD**
Log immigrants	0·348**	0·444***
	(0·148)	(0·110)
Log emigrants	0·090	−0·030
	(0·120)	(0·162)
Parent co-location	5·607***	
	(0·229)	
Prov. unemployment rate	0·003	
	(0·049)	
Infrastructure endowment	0·002	
	(0·001)	
Residents with tertiary education	0·155	−0·002
	(0·126)	(0·182)
Log prov. population	0·154	0·006
	(0·757)	(0·192)
Institutional quality	−0·127	0·025
	(0·952)	(0·856)
Common border	−0·049	0·600
	(0·259)	(0·510)
Log prov. GDP	0·799	0·069
	(0·718)	(0·245)
Sectoral diversity	0·067	0·038
	(0·117)	(0·222)
Log patent count	0·020	−0·041
	(0·201)	(0·221)
Log average wage (region)	3·739*	−1·393
	(2·262)	(3·131)
Log distance	−0·319	1·818***
	(0·459)	(0·596)
Agglomeration (sector)	0·200***	0·006
	(0·044)	(0·104)
Log imports	0·155	−0·127
	(0·097)	(0·132)
Log exports	−0·000	0·003
	(0·097)	(0·102)
Pre-2002 FDI stock	0·013***	0·014**
	(0·004)	(0·006)
Log multilateral immigrants	−0·261	0·071
	(0·334)	(0·172)
Log multilateral emigrants	−0·087	−0·035
	(0·172)	(0·158)
Firm density	0·001	−0·015***
	(0·005)	(0·005)
Manufacturing concentration	−0·437	0·140
	(0·381)	(0·439)
Rome/Milan	−0·681	−0·487
	(0·605)	(0·731)
Observations	65,171	
AIC	2253.311	
BIC	2643.956	
LR test of joint significance of the SD	83.735	
Degrees of freedom	20	
Test p-value	0.000	

Location choice models for inward FDI targeting Italian provinces. Mixed logit estimates based on 500 Halton draws each. For each model, the table reports the estimated average effect of the variable (Mean) and standard deviation (SD). The set of covariates in Model B3 corresponds to the one in Model 1 of Table 5, but is limited to the 2008–2015 subperiod to mitigate concerns about possible measurement error in the co-location variable. Standard errors (in parentheses) take into account the correlation among the investments by the same company. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: Model including province–country dummies for pairs with more than 10 investments - Sources of heterogeneity in the immigrant effects $\hat{\delta}_f^{\text{Immi}}$

Dep. var: $\hat{\delta}_f^{\text{Immi}}$	**Model 4a**
Jobs/mln US$ invested	$-0{\cdot}001^{**}$
	(0·000)
Type of sector (ref: Services)	
Final goods	$-0{\cdot}030^{*}$
	(0·018)
Intermediate goods	$-0{\cdot}033^{***}$
	(0·010)
Other goods	$-0{\cdot}027^{**}$
	(0·013)
More than one investment in Italy	$-0{\cdot}036^{***}$
	(0·012)
Log total capital investment worldwide	$-0{\cdot}014^{***}$
	(0·003)
Italy share of capital investment worldwide	$-0{\cdot}053^{***}$
	(0·013)
Area of origin (ref: EU)	
South-East Asia	0·019
	(0·015)
Non-EU Europe	0·008
	(0·017)
North America	0·002
	(0·009)
Rest of the world	0·017
	(0·024)
Constant	$0{\cdot}489^{***}$
	(0·016)
Observations	895
Test of the joint significance of the regressors	90·682
Degrees of freedom	11
Test p-value	0·000

Variance-weighted least-squares regression of firm-specific immigrant effects on investing company characteristics. Dependent variable: Firm-specific coefficients for the bilateral immigrants' effects $\hat{\delta}_f^{\text{Immi}}$ simulated from the distributional parameters estimated in the mixed logit model reported in Model 4a, Table C.7. Simulation procedure described in Section 3.2.1. Variances of the individual parameters $\hat{\delta}_f^{\text{Immi}}$ estimated by parametric bootstrapping. Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: Heterogeneity in the immigrant effect $\hat{\delta}_f^{\mathrm{Immi}}$ - Specifications including sectoral dummies

Dep. var: $\hat{\delta}_f^{\mathrm{Immi}}$	**Model 4**
Jobs/mln US$ invested	$-0{\cdot}001^{**}$
	(0·000)
Dummy: more than one investment in Italy	$-0{\cdot}039^{***}$
	(0·012)
Log total capital investment worldwide	$-0{\cdot}013^{***}$
	(0·003)
Italy share of capital investment worldwide	$-0{\cdot}052^{***}$
	(0·014)
Area of origin (ref: EU)	
South-East Asia	0·016
	(0·015)
Non-EU Europe	0·004
	(0·017)
North America	−0·001
	(0·009)
Rest of the world	0·023
	(0·024)
Sector (ref: Industrial)	
Construction	−0·037
	(0·028)
Consumer goods	0·008
	(0·027)
Creative industries	$0{\cdot}048^{**}$
	(0·021)
Energy	−0·026
	(0·032)
Environmental technology	0·004
	(0·022)
Financial services	$0{\cdot}052^{***}$
	(0·020)
Food, beverages & tobacco	−0·016
	(0·026)
ICT & electronics	$0{\cdot}036^{*}$
	(0·019)
Life sciences	0·013
	(0·023)
Physical sciences	$-0{\cdot}047^{*}$
	(0·027)
Professional services	$0{\cdot}048^{**}$
	(0·023)
Retail trade	0·103
	(0·091)
Tourism	0·023
	(0·024)
Transport equipment	−0·005
	(0·022)
Transportation, warehousing & storage	0·017
	(0·023)
Wood, apparel & related products	0·052
	(0·034)
Constant	$0{\cdot}466^{***}$
	(0·022)
Observations	895
Test of the joint significance of the regressors	143·730
Degrees of freedom	24
Test p-value	0·000

Variance-weighted least-squares regression of firm-specific immigrant effects on investing company characteristics. Dependent variable: Firm-specific coefficients for the bilateral immigrants' effects $\hat{\delta}_f^{\mathrm{Immi}}$ simulated from the distributional parameters estimated in the mixed logit model reported in Model 4 of Table 5. Simulation procedure described in Section 3.2.1. Variances of the individual parameters $\hat{\delta}_f^{\mathrm{Immi}}$ estimated by parametric bootstrapping. Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Is there substitution between migration and manufacturing FDI?

While our results are clearly supporting an information channel, they less clear-cut for what concerns the country-specific labour supply channel. Indeed, although the net effects of bilateral immigration for manufacturing FDI are positive and very close to zero, the estimated magnitude is significantly smaller than that of other comparably low-cost-labour intensive functions such as Construction and Logistics. This is somewhat unexpected, given that our theoretical arguments suggest that for these investments the labour supply channel may actually add up to the information effects, so that the estimated effect could even be larger.

This suggests that, in addition to the hypothesized channels, there may be an unexplored mechanism that leads to a substitution between migration and FDI in manufacturing. A possible explanation may refer to the distinction between horizontal and vertical FDI (Markusen, 2002, 2005). Indeed, one could argue that, depending on whether the manufacturing investment is horizontal or vertical, immigration may complement or substitute FDI. For vertical FDI, we expect immigration to complement FDI due to the standard labour cost and information channels discussed in Section 2 (Markusen, 2002). On the other hand, there may be substitution between immigration and horizontal FDI if we consider that immigrants have been proven to promote trade, as well as FDI. As mentioned in Section 2, the trade literature has long identified the home country bias that leads immigrants to increase imports from their countries of origin due to their demand of home country goods and to network effects. Given the substantially higher costs entailed in FDI compared to trade, an MNE engaging in horizontal FDI may find it more profitable to locate in provinces which are not already served via trade flows, either to avoid competition with other firms from its country of origin or to avoid duplicating efforts in serving a particular region (Blonigen, 2005).

Unfortunately, our data do not allow us to distinguish between horizontal and vertical FDI, but the evidence that we report in Table D.1 is consistent with the above interpretation. If export-substituting manufacturing FDI tend to avoid provinces where their co-nationals have helped developing trade ties, we should observe a negative interaction between province bilateral imports from the FDI country of origin and bilateral immigration. To test this hypothesis, we augment the final specification in Model 7 of Table 6 with a further set of interactions of immigration, trade, and the dummy for manufacturing FDI.

The results show that imports and immigration are generally complements, except for manufacturing investments. For manufacturing, we find stronger effects of immigrants in provinces that import less from their countries of origin. Moreover, in this specification, the coefficient of *Log immigrants*$\times$*Manufacturing*, which captures the (hypothetical) effect of immigration in provinces where bilateral imports are zero, is positive, significant and very large, even in comparison with the estimated coefficients for other functions, consistent with the expectations about a comparatively large effect we get from the literature review.

Overall, the direction of our findings suggest that manufacturing investments in Italy are mainly horizontal investments, which could also explain why we find little evidence for multilateral immigration and the low-cost labour supply channel. Moreover, these considerations highlight that further research should investigate the differences in location determinants between the determinants horizontal and vertical FDI.(Blonigen, 2005).

Table D.1: Interaction between imports, immigration and functions

Dep. var.: Choice	**Model C1**	
	Mean	**SD**
Log immigrants	−0·43	
	0·343	
Log emigrants	−0·184	
	0·219	
Log imports	0·009	
	0·099	
Residuals (immigration)	−0·081	
	0·09	
Residuals (emigration)	0·319∗∗∗	
	0·121	
Log immigrants × RD	0·066	
	0·209	
Log immigrants × Manufacturing	1·234∗	
	0·646	
Log immigrants × Market-Access	0·507∗∗∗	
	0·163	
Log immigrants × Business Services	0·631∗∗∗	
	0·19	
Log immigrants × Logistics	0·006	
	0·191	
Log immigrants × Construction	−0·042	
	0·246	
Log immigrants × Already invested in Italy	−0·81∗∗∗	
	0·131	
Log imports × Manufacturing	0·262	
	0·164	
Log immigrants × Import	0·039∗∗	
	0·016	
Log immigrants × Log Imports × Manufacturing	−0·077∗∗	
	0·031	
Log imports × Manufacturing	0·009	
	0·099	
Observations	111,692	

Location choice models for inward FDI targeting Italian provinces. Covariates: parent co-location, unemployment rate, infrastructure endowment, residents with tertiary education, sectoral diversity, log prov. population, institutional quality, common border, log prov. GDP, log patent count, log average wage (region), log distance, agglomeration (sector), log imports, log exports, pre-2002 FDI stock, log multilateral immigrants, log multilateral emigrants, firm density, manufacturing concentration, Rome/Milan dummy, and interactions of log emigrants with the dummies for R&D, Manufacturing, Market Access, Business Services, Logistics, Construction, Already invested in Italy. The model is estimated by conditional logit with standard errors based on 2,000 bootstrap samples, hence assumes fixed parameters. Bootstrapped standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Altonji, J. G. and Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market*, pages 201–234. University of Chicago Press.

Bena, A., Leombruni, R., Giraudo, M., and Costa, G. (2012). A new Italian surveillance system for occupational injuries: characteristics and initial results. *American Journal of Industrial Medicine*, 55:584–92.

Blonigen, B. A. (2005). A Review of the Empirical Literature on FDI Determinants. *Atlantic Economic Journal*, 33(4):383–403.

Bratti, M., De Benedictis, L., and Santoni, G. (2014). On the Pro–Trade Effects of Immigrants. *Review of World Economics*, 150:557–594.

Briant, A., Combes, P.-P., and Lafourcade, M. (2014). Product Complexity, Quality of Institutions and the Protrade Effect of Immigrants. *The World Economy*, 37(1):63–85.

Burchardi, K., Chaney, T., and Hassan, T. (2018). Migrants, Ancestors, and Foreign investments. *The Review of Economic Studies*, 86(4):1448–1486.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Castellani, D. and Lavoratori, K. (2020). The lab and the plant: Offshore r&d and co-location with production activities. *Journal of International Business Studies*, 51(1):121–137.

Colombelli, A., D'Ambrosio, A., Meliciani, V., and Quatraro, F. (2020). Newborn Firms and Regional Diversification Patterns: The Role of Cultural Diversity. *Economic Geography*, 96(4):336–362.

D'Ambrosio, A. and Montresor, S. (2022). The pro-export effect of subnational migration networks: new evidence from Spanish provinces. *Review of World Economics*, 158:53–107.

Defever, F. (2006). Functional fragmentation and the location of multinational firms in the enlarged Europe. *Regional Science and Urban Economics*, 36(5):658–677.

Docquier, F. and Lodigiani, E. (2010). Skilled Migration and Business Networks. *Open Economies Review*, 21:565–588.

El Yaman, S., Kugler, M., and Rapoport, H. (2007). Migrations et Investissements Directs Étrangers dans l'espace Européen (UE-15). *Revue Économique*, 58(3):725–733.

Foad, H. (2012). FDI and immigration: a regional analysis. *The Annals of Regional Science*, 49(1):237–259.

Gheasi, M., Nijkamp, P., and Rietveld, P. (2013). Migration and Foreign Direct Investment: education matters. *The Annals of Regional Science*, 51(1):73–87.

Hole, A. (2007). Estimating mixed logit models using maximum simulated likelihood. *Stata Journal*, 7:388–401.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

Javorcik, B. S., Çağlar Özden, Spatareanu, M., and Neagu, C. (2011). Migrant networks and Foreign Direct Investment. *Journal of Development Economics*, 94(2):231 – 241.

Kugler, M. and Rapoport, H. (2007). International labor and capital flows: Complements or substitutes? *Economics Letters*, 94(2):155 – 162.

Markusen, J. (2002). *Multinational firms and the theory of international trade*. MIT Press, Cambridge, MA.

Markusen, J. (2005). Modeling the offshoring of white-collar services: from comparative advantage to the new theories of trade and FDI.

Murat, M. and Pistoresi, B. (2009). Migrant Networks: Empirical Implications for the Italian Bilateral Trade. *International Economic Journal*, 23(3):371–390.

Nifo, A. and Vecchione, G. (2014). Do Institutions Play a Role in Skilled Migration? The Case of Italy. *Regional Studies*, 48(10):1628–1649.

Ottaviano, G. and Peri, G. (2006). The Economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, pages 9–44.

Rauch, J. and Trinidade, V. (2002). Ethnic Chinese Networks in International Trade. *Review of Economics and Statistics*, 84(1):116–30.

Train, K. E. (2009). *Discrete choice methods with simulation (2nd Edition)*. Cambridge University Press.