Migration and the Location of MNEs Activities. Evidence from Italian Provinces

The John H Dunning Centre for International Business

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Migration and the Location of MNEs Activities. Evidence from Italian Provinces

December 2019

Luigi Benfratello
Politecnico di Torino and CSEF

Davide Castellani
University of Reading, Henley Business School

Anna D'Ambrosio
Politecnico di Torino

Discussion Paper Number:
JHD-2019-04

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

dunning@henley.ac.uk

www.henley.ac.uk/dunning

Migration and the location of MNEs activities. Evidence from Italian provinces*

[1], Luigi Benfratello[†1], Davide Castellani[‡2], and Anna D'Ambrosio[§3]

[1]Politecnico di Torino and CSEF
[2]University of Reading, Henley Business School
[3]Politecnico di Torino

Abstract

This paper investigates the link between migration and inward FDI in narrow geographies. Our results, based on 1,147 greenfield investment projects made by 895 MNEs into Italian provinces (NUTS3) over the 2003-2015 period, confirm a positive effect of the stock of immigrants on FDI, but no robust effects of emigrants. However, beyond this average effect lies significant heterogeneity. By unraveling this heterogeneity, we shed light on the potential mechanisms underlying this relation. Our results are consistent with an important role of demand and information channels, but not with an effect through the labour market. On the one hand, immigrants are not a factor that attracts more labour-intensive investments. On the other hand, the effect of immigrants is stronger when information and, to a lesser extent, market demand are more important. Overall, our paper bears significant implications for local development policy that partially contrast with the current public discourse on immigration.

Keywords: Foreign Direct Investment; Migration; Location Choice; Information Effect; Demand Effect; Conditional Logit; Mixed Logit
JEL classification: F22; F21; R30

*We wish to thank Marco Mutinelli and Roberto Leombruni for providing part of the data we employ. We are indebted to Arne Risa Hole for his indications on the estimation of the standard errors of the individual random parameters. We also wish to thank Laura Abrardi, Alessandro Arrighetti, Marco Bellandi, Luca de Benedictis, Ivan Etzo, Fabio Landini, Andrea Lasagni, Katiuscia Lavoratori, Francesco Lissoni, Alessia Lo Turco, Giulio Pedrini, Valentina Meliciani, Gianluca Santoni and the participants in the 2018 AIB, ERSA, EARIE, AISRE, SIE and to the 2019 SIEPI and ITSG conferences and workshops for helpful comments and suggestions. All usual disclaimers apply. Anna D'Ambrosio and Luigi Benfratello gratefully acknowledge financial support by Compagnia di San Paolo under the Starting Grant initiative.

†luigi.benfratello@polito.it
‡davide.castellani@henley.ac.uk
§anna.dambrosio@polito.it

1 Introduction

Migration and the fragmentation of the multinational enterprises (MNEs) activities along the value chain are making the world increasingly interconnected. In this paper, we study whether such interconnections affect each other and, specifically, whether migration contributes to attract inward Foreign Direct Investments (FDIs) into narrow geographies.

This topic is relevant on multiple levels. Firstly, considering the prominence of migration and of FDI attraction policies in the public debate, our study bears important implications for policymaking, as it contributes to highlight migration as a potentially under-explored mechanism to increase the attractiveness of narrowly defined geographies for FDI. This also helps framing the debate about the economic costs and benefits of migration. Secondly, this topic is intriguing from a theoretical point of view. While standard neoclassic models view the relationship between migration and capital flows as one of substitution, a growing literature has documented a robust empirical complementarity between the two. Such theoretical mismatch points to the need to better understand the mechanisms underlying the effects of migration, which has stimulated prominent theoretical and empirical contributions (Jayet and Marchal, 2016; Burchardi et al., 2018; Felbermayr et al., 2015; Javorcik et al., 2011; Buch et al., 2006). Different theoretical approaches highlight different mechanism underlying the effects of migration on FDI. However, to the best of our knowledge, only Burchardi et al. (2018) specifically address the different possible explanations for the migration effect on FDI in an attempt to learn about the underlying mechanisms. In particular, migrants can contribute to explain inward FDI due to a combination of labour market, information, enforcement and demand effects. We submit that these effects may play a different role according to the type of FDI and the type of multinational enterprise (MNE).

To address this issue, we draw on a rich dataset combining detailed investment-level data —from the *fDi markets* database— with information at a highly disaggregated geographical scale, such as the Italian provinces, which correspond to EU NUTS3 regional level. Investment-level data allow us to distinguish the function/activity in which the investment is taking place (Defever, 2006). We focus in particular on manufacturing, R&D, market-access and business services FDI [1]. The fine geographical disaggregation of our analysis allows to accurately reflect the geography of factor market integration

[1] We use the terms "function" or "activity" interchangeably to "refer to a stage or an activity which is part of the firm's value chain, and where the production stage itself only accounts for one stage" (Defever, 2006, p. 658–659)

and the localised nature of the information effects (Buch et al., 2006; Rauch, 2001), as well as to address the existence of spatial decay effects in our variables of interest (see Bratti et al., 2014, for an application to trade). The geography of global factor movements, and of FDI in particular, is increasingly spiky and concentrated in specific urban and even suburban areas (Goerzen et al., 2013; Belderbos et al., 2020). Failing to study the phenomenon at the right scale leads to the modifiable areal unit problem (MAUP), which arises when the aggregation of data into arbitrary units leads to loss of variation (Openshaw, 1983). Italy provides a relevant case study for this analysis, considering its highly fragmented geography of (domestic) production and its relatively recent and growing immigration experience that co-exists with a consolidated emigration history.

Furthermore, we distinguish the two sides of migration, namely immigrants and emigrants. Considering that, if any, their effects on FDI are likely to be not symmetric, including both variables should yield insights on the underlying mechanisms.

Finally, we employ a novel econometric strategy, based on a two-step estimation approach, which allows us first to assess whether the effect of migration on the location of inward FDI is indeed heterogeneous and then to tease out the sources of this heterogeneity thus allowing us to dig deeper into the determinants of these effects (Castellani and Lavoratori, 2019; Alcácer et al., 2018; Hornstein and Greene, 2012; Saxonhouse, 1976). More specifically, we first estimate a random parameter (mixed) logit model, then we predict the individual-specific random parameters derived by the mixed logit model and examine the determinants of their distribution via regression models.

Our results confirm a positive and robust effect of immigration on FDI but no effect of emigrants. However, the effect of immigration on FDI is indeed heterogeneous along the value chain and across firms. First, we do not find any effect of immigration on FDI in production activities. Instead we find a robustly positive and significant effect of immigrants on R&D, market-access and business services FDI. Taken together, these results do not provide support for the effect of migration via the labour market, and point to the co-existence of an information and a demand channel. Furthermore, we find that the effect of immigrants is highly spatially confined and stronger for firms that are investing in Italy for the first time, for firms with fewer worldwide investments, that devote smaller shares of their investment portfolios to the Italian market, and for firms coming from more culturally distant countries and from East Asia in particular. These results are consistent with idea that firms rely on the immigrants' effects when they have less information of the Italian market. On the whole, our results support the interpretation that immigrants effectively contribute to reducing the investors' lack of knowledge of local contexts.

The remainder of the paper is organised as follows. Section 2 discusses the theoretical framework for our study. Section 3 introduces the empirical model, presents our data and variables, as well as summary and descriptive statistics. Section 4 presents our results alongside with a set of robustness checks. A discussion of their implications and some concluding remarks follow in Section 5. A detailed data appendix concludes the paper.

2 Theoretical framework

A vast literature has addressed the determinants of FDI location choice. Typically, their underlying models assume that a firm decides to locate its subsidiaries where the achievable profits outweigh those that can be gained in all other available locations (e.g. Nielsen et al., 2017; Head and Mayer, 2004; Head et al., 1999; Spies, 2010; Defever, 2006; Basile et al., 2008; Basile, 2004). In broad terms, reduced-form formulations of the profit functions faced by the firm assert that profits depend on production costs (Amiti and Javorcik, 2008; Tintelnot, 2017), market potential of the location (Head and Mayer, 2004), fixed costs of market entry (Helpman et al., 2004; Spies, 2010) and achievable information spillovers arising from agglomeration (Fujita and Thisse, 1996).

A growing literature addresses the global sourcing decisions of MNEs recognising the heterogeneity in the determinants that is introduced by functions (e.g. Castellani and Lavoratori, 2019; Defever and Toubal, 2013; Defever, 2012; Nefussi and Schwellnus, 2010; Markusen, 2006). Specifically, the location determinants have been argued to differ for upstream and downstream service activities—with upstream activities such as headquarters and R&D being mainly determined by localized production factors, especially skilled labour—, and downstream activities such as marketing being more sensitive to the proximity to markets. Defever (2012) formalised this idea in a simple model attributing function-specific weights to the different production factors.

The increasingly compelling evidence about the effect of migration on FDI (e.g. Burchardi et al., 2018; Javorcik et al., 2011; Jayet and Marchal, 2016; Buch et al., 2006; De Simone and Manchin, 2012; Kugler and Rapoport, 2007; Docquier and Lodigiani, 2010; Gao, 2003; Etzo and Takaoka, 2018) can be easily integrated in this framework, which ultimately implies focussing on the extensive margin of FDI. Different assumptions about the mechanisms underlying the migration effects have translated into different theoretical models and empirical methodologies (see the review in Jayet and Marchal, 2016); yet, in most cases, the migration effect can be traced back to an effect on either transnational information costs (Buch et al., 2006; Gao, 2003; Kugler and Rapoport, 2007; Docquier and Lodigiani, 2010; Javorcik et al., 2011) or on labour costs (Jayet and Marchal, 2016). If immigrants affect a specific component of production costs such as low-skilled wages, we should expect investments that are more sensitive to such costs to respond more to the presence of migrants. Similarly, if immigrants affect information costs, which usually enter the profit function among the fixed costs of FDI, investments that are more sensitive to information asymmetries should respond more to the presence of migrants.

Despite such wealth of perspectives, as mentioned, we are aware of only one study that confronts the possible explanations for the mechanism underlying the migration effect on FDI. Burchardi et al. (2018) identify four potential channels: information effects, enforcement effects, similarity in skills and similarity in preferences. These can be easily integrated in the profit function for the investor: the information and enforcement effects affect the fixed costs of establishing a plant abroad; the similarity in skills affects production costs; and the similarity in preferences affects the revenues generated by the MNEs' sales.

More specifically, the information and the enforcement effects were originally proposed as mechanisms underlying the migrants' effect on trade, rather than on FDI (Gould, 1994; Wagner et al., 2002; Rauch and Trinidade, 2002). Thanks to their knowledge of the home country institutions and language, migrants effectively decrease bilateral trade barriers by facilitating the flow of information between the origin and the destination country regarding procedures and business opportunities—the "information effect". Furthermore, migrants' embeddedness within co-ethnic networks generates reputational bounds that ensure the enforcement of transnational contracts and can be especially valuable in countries where the rule of law is weakly enforced—the "enforcement effect" (Rauch and Trinidade, 2002; Dunlevy, 2006). Those pioneering contributions about the mediating role of contacts and networks for the establishment of trade relationships have now been fully incorporated in trade models (Arkolakis, 2010; Chaney, 2014). As to FDI, migrants' brokering role is potentially even more salient, as FDI are subject to substantially higher capital investment, information asymmetries and cognitive barriers than foreign sales, and depend critically on the knowledge of foreign institutions, business opportunities and labour market pools of specific skills (Head et al., 1995; Jayet and Marchal, 2016; Javorcik et al., 2011; Daude and Fratzscher, 2008). FDI require comparatively greater fixed costs and lower variable costs than trade (Buckley and Casson, 1981; Helpman et al., 2004), and migrants' effect has been found to operate precisely at the level of fixed, rather than variable costs (Peri and Requena-Silvente, 2010).

As regards the similarity in skills, Burchardi et al. (2018) argue that "migrants may bring with them a specific skill-mix or other factors abundant in their origin country, so that firms can more easily outsource production, using the same skill-mix at home and abroad" (p.28). More generally, migrants could affect the location choice of FDI as providers of labour for the new establishments. This link has been proposed by a branch of the literature attempting to accommodate the observed complementarity between migration and FDI through extensions of the Heckscher-Ohlin model, often distinguishing between skilled and unskilled migrants (Jayet and Marchal, 2016; Markusen, 2006). This literature has mainly shown that the complementarity operates between FDI and skilled migration: FDI respond to the availability of (foreign) skilled labour and, in turn, attract skilled labour from the source country of FDI. Instead, there is evidence of substitution between unskilled migration and (outward) FDI. Most studies highlight different effects according to the direction of FDI that is considered (outward vs. inward) which ultimately implies recognizing heterogeneity in the kind of investment, in their motives and in the determinants of their location choice.

The fourth channel that has been highlighted is that migrants may play a role in the location of FDI through the similarity in preferences, so ultimately by directly adding up to a firm's demand and hence revenues. This channel has received comparatively less attention in the migration literature on FDI. Yet, when discussing location choice, Nefussi and Schwellnus (2010, p.184) have emphasized that the assessment of the market potential for an affiliate should take into consideration the demand originating from co-ethnic firms and consumers residing in a potential location. This is because, they argue, these firms and consumers may have a consumption bias in favour of goods and services offered by firms from their countries of origin. Most basically, this effect is analogous to what the trade literature has labelled the "preference effect" or the "transplanted-home bias effect" (Gould, 1994; White, 2007): migrants' preference for consuming home country goods contributes to increasing imports from their home country. In a similar way, foreign investors may recognise the potential market demand represented by co-ethnic expatriates and, provided that the proximity-concentration trade-off is in favour of foreign production (Horstmann and Markusen, 1992), target horizontal FDI to serve this market, as well as FDI intending to support the sales of the establishments, such as sales support, marketing, and customer contact services. Furthermore, migrants' home bias in consumption may apply to market-driven services such as banking, insurance and real estate services. According to Nefussi and Schwellnus (2010), the high level of customisation of such services and the high extent of complex face-to-face communication required in their delivery may be responsible for such consumption bias.

As for the actual role of the four potential channels, Burchardi et al. (2018) document a robust and causal information effect while they cannot find robust evidence of the other effects. However, while they detect heterogeneity in the effects of migration on FDI at the NAICS sector level, they do not study the differences in FDI functions.[2]

The different effects may clearly co-exist. The relative salience of each of them, in turn, obviously depends on the elasticities of labour costs, information costs and demand to migration. For instance, if labour costs in the formal sector do not react much to the presence of migrants, as could be the case in a rigid labour market such as Italy, the resulting migration effect associated with this channel will be limited. We can only observe the cumulative result of the four potential channels, but the weight of each will be larger depending on the importance of labour costs, information, enforcement costs, and of market demand in the specific activity that the MNE is undertaking.

Distinguishing the investments by function and studying the heterogeneity in the associated immigrants' effect can therefore yield insights as to the prevailing mechanisms underlying the migration effect.[3] We consider four functions: R&D, manufacturing, market-access and business services. R&D investments can be viewed as ventures seeking to exploit localised knowledge assets, the access to which requires complex interactions between local and expatriate labour. Migrants' effect on this type of FDI will arguably depend mostly on information effects mediating cultural differences and facilitating the transfer of knowledge and know-how. Manufacturing ventures, among the considered activities, are expected to rely the least on information and the most on the availability of specific inputs and on the labour effects of migrants. Finally, migrants' effects on demand should affect mostly market-seeking investments, such as investments in sales support (e.g. sales, marketing and support and customer contact centres) and business services (e.g. retail banking, advertising and real estate services). In this case, the migrants' effect in bridging information about the market potential of the location may add up to the demand effects exerted by coethnic firms and customers.

[2]Kugler and Rapoport (2007) similarly disaggregate the immigrants' effect by sector and analyse manufacturing and services FDI separately. However, this is not done in response to a specific hypothesis and their results are not exploited to yield specific insights on the underlying mechanism.

[3]As enforcement effects are likely not to be function-specific, we will concentrate in what follows on disentangling three of the four potential mechanisms: information, demand and labour market effects.

Similarly, the relative importance of the labour, information and preference effects may be firm specific. Less labour-intensive and more productive firms may rely less on labour cost savings deriving from migration. As argued by Burchardi et al. (2018), firms producing intermediate goods, or firms producing reference-priced products, may rely less on the home bias in consumption of their co-ethnics when planning foreign sales. In addition, firm-level heterogeneity may affect the extent to which migrants attenuate the "liability of foreignness" (Hymer, 1976; Zaheer, 1995; Nachum, 2003) central in the International Business literature, i.e. the costs faced by foreign firms when operating in an overseas market, which underlie several barriers and possibly even a differential treatment with respect to local firms.

In this respect, the availability of data on both immigration and emigration is convenient due to the arguably asymmetric effects of the two. The demand channel and the labour channel obviously only apply to immigrants. Instead, information effects may in principle operate both ways. Expatriates may communicate with foreign investors about the availability of business opportunities in their home countries, provide contacts and facilitate administrative procedures when launching a new venture. Immigrants may also facilitate the flow of information and of specific knowledge as well as the access to social networks to fellow countrymen seeking for advice in a foreign country. Consistently with this argument, Flisi and Murat (2011), with a country-level focus, find that inward Italian FDI are actually mainly driven by emigration.

3 Empirical Application

3.1 Model

Our paper positions in between the literature on the location choice of FDI and on the migrant's effects on trade and FDI. Both branches of the literature have a fairly established set of estimation strategies of reference. The literature on the migration-FDI nexus mainly applies gravity-like models (see for instance De Simone and Manchin, 2012; Javorcik et al., 2011; Buch et al., 2006; Felbermayr et al., 2015; Burchardi et al., 2018). Addressing the question at stake from a gravity perspective, however, necessarily implies aggregating investments at some geographic scale, loosing potentially insightful information about the decision-making process underlying the location choice. Hence, we follow the literature on the location choice of FDI (e.g. Head et al., 1995; Du et al., 2008; Spies, 2010; Defever, 2006), which usually appeals to discrete-choice models (conditional, nested and mixed logit models, see Marschak, 1974; McFadden, 1974; McFadden and Train, 2000; Train, 2009) and studies the location choice for each individual investment[4]. These models share an underlying Random Utility Model, i.e. a model assuming in a partial equilibrium setting that the location chosen by a multinational firm yields the highest utility compared to the other possible locations, subject to uncertainty deriving from unobservables (Train, 2009). As it is standard in discrete choice models, the dependent variable *"Choice"* is equal to one if a specific alternative is selected, and zero for all other alternatives in the choice set. In our case, the alternatives are constituted by the set of Italian provinces where the FDI could locate (which is in turn composed of any provinces selected at least once in our data). The decision-maker is in our case the investing company f facing investment decision n. Hence, the total number of choices under consideration for our analysis is equal to the number of potential locations in the choice set J times the number of investment projects N. The probability to choose a specific province depends only on the difference in utility that a specific province j yields to firm f in investing decision n compared with the other alternatives. The absolute value of utility does not matter. Hence, any attributes of the alternative that do not induce a difference in utility, are not affecting the choice and are not estimated. This implies that variables that are invariant by firm (e.g. its country of origin, the GDP of the origin country, its size, knowledge, capital investment, etc.) will be included in the specification only if interacted with alternative-varying variables (see Train, 2009). On the other hand, bilateral variables such as the migration from a given country to a given province induce a difference in utility and are therefore estimated.

The simplest and most widely used discrete choice model is the conditional logit model, which relies on the independence from irrelevant alternatives (IIA) assumption and on an assumption of homogeneous effects of the parameters across decision makers. Along with it, in the light of our discussion about the potential heterogeneity of the migrants' effect on FDI, we will implement a set of mixed logit models (Train, 2009) to analyse the determinants of the choice of a specific province as a destination for the investment. Mixed logit models, or random parameter logit, indeed, allow estimating the heterogeneity in the effects of the parameters across decision makers—in our case, the heterogeneity in the migrants' effects across investing firms. This implies being able to make a distinction between those variables that have a significant mean from those

[4]Due to data unavailability on specific investment projects, several works have resorted to aggregating data and estimated Poisson and Negative Binomial models (e.g. Kogut and Chang, 1991; Coughlin and Segev, 2000; Barry et al., 2003; Blonigen, 1997; Basile, 2004).

having a significant variance (Alcácer et al., 2018). Seen another way, mixed logit models allow relaxing the independence from irrelevant alternatives (IIA) assumption typical of conditional logit models (Train, 2009). Furthermore, according to Alcácer et al. (2018), random coefficient models are better suited than standard linear or non-linear regression models to explain why some factors may affect an outcome differently depending on the individual considered, thereby helping to address key questions in strategy research. Indeed, the distributional estimates can be used as the starting point for a second-stage analysis of the sources of heterogeneity in the effects, as we will do in Section 4. Although discrete choice models are fairly standard in the FDI literature, their application to the analysis of the migration effect on FDI, as well as the exploration of the sources of the heterogeneity in the migrants' effects with investment and firm-level determinants are, to the best of our knowledge, novel contributions.

Each firm f makes a series of location decisions. At each decision n the utility deriving from investing in province j is assumed to be linear in the parameters. In the case of conditional logit, utility is modelled as a function of alternative-specific regressors, varying by province or by investment and province. The coefficients of these regressors are taken as fixed in the conditional logit whereas all, or some of them, are interpreted as varying by decision-maker in the mixed logit model. With this background, we model utility as follows:

$$U_{fnj} = \alpha' x_{nj} + \beta' y_{fnj} + \gamma_f' w_{nj} + \delta_f' z_{fnj} + \epsilon_{fnj} \tag{1}$$

where α and β are vectors of fixed coefficients, while γ_f and δ_f are vectors of random (i.e. investor specific) coefficients; x_{nj} and w_{nj} are vectors of observed variables relating to the province, while y_{fnj} and z_{fnj} are vectors of observed variables varying by firm, investment decision and province; ϵ_{fnj} is iid extreme value. The random coefficients γ_f and δ_f are unobservable to the researcher and they are assumed to vary over investing companies with densities $f(\gamma)$ and $g(\delta)$, respectively. Conditional on these parameters, the investment project locates in a specific province i if the utility associated with it exceeds that of all other provinces $j \neq i$, which can be modelled as a logit choice probability. Because, however, the parameters are unknown to the researcher, the unconditional probability that a given project locates in a specific province is the integral of these logits over all possible values of $f(\gamma)$ and $g(\delta)$:

$$P_{fni} = P(Choice_{fni} = 1|x, y, z, w) = \int_{\delta} \left[\int_{\gamma} \left(\frac{e^{\alpha' x_{ni} + \beta' y_{fni} + \gamma'_f w_{ni} + \delta'_f z_{fni}}}{\sum_j e^{\alpha' x_{nj} + \beta' y_{fnj} + \gamma'_f w_{nj} + \delta'_f z_{fnj}}} \right) f(\gamma|\delta)\, d\gamma \right] g(\delta)\, d\delta \quad (2)$$

Drawing on simulation, we can estimate the distribution parameters of the random coefficients γ and δ; i.e. the means and standard deviations of their distributions $f(\gamma)$ and $g(\delta)$, assumed normal. The magnitude and significance of their standard deviations are measures of the heterogeneity of the effects of w_{nj} and z_{fnj} on the location choice of FDI. Model 2 reduces to a conditional logit model if $f(\gamma)$ and $g(\delta)$ are degenerate at fixed parameters c and d: $f(\gamma) = 1$ for $\gamma = c$ and 0 for $\gamma \neq c$; $g(\delta) = 1$ for $\delta = d$ and 0 for $\delta \neq d$ (see Train, 2009, for a more formal and detailed explanation).

3.2 Variables

This study analyzes the location choices of $1,147$ inward FDI into 85 Italian provinces occurred over the 2003-2015 period. As discussed, the dependent variable in our models is a binary variable *Choice* equal to 1 if firm f chose to locate investment project n in province j, an zero otherwise.

Our variables of interest are *Log Immigrants* and *Log Emigrants*. *Log Immigrants* is the log of the stock of immigrants who come originally from the same origin country o as the investment and reside in province j at time $t-1$, where t is the year of occurrence of the investment. For brevity, in what follows we will refer to them as "bilateral immigrants". *Log Emigrants* represents log of the stock of emigrants from province j and residing, at time $t-1$, in the same country o from where the investment originated (we will label them "bilateral emigrants")[5].

Based on the literature on location choice as well as on the migration effect on FDI, we run different specifications of model (2) and include province-level and province-investment variables. To proxy for localised market potential, we include province-level GDP and province population, both measured in logs. To proxy for the costs of labour, we include the log of the average wage at regional level and the province-level unemployment rates; as proxies for the human capital endowment of the location, we include the log of the count of the patent applications filed by the province to the EPO, as well as the share of residents holding a tertiary degree. We include an index of the infrastruc-

[5] A detailed description of all data sources and variables is provided in the Data Appendix.

ture endowment of province j, a dummy for investors located in countries on the border with the Italian regions where the investment has taken place, the log of the distance between the centroid of the province j and of the capital city of the country of origin of the investment o, an index for the institutional quality of province j drawn from Nifo and Vecchione (2014) as well as, following Fujita and Thisse (1996), two measures of agglomeration aiming to capture the effects of, respectively, Jacobian and Marshallian externalities. Jacobian externalities are measured through a standard sectoral diversity measure calculated as $1 - H$, where H is the sectoral Herfindahl-Hirschmann concentration index calculated at the NACE 2-digit level; Marshallian externalities are measured through a province-level specialisation index in the same 2-digit sector as the investment project. Considering their central role in the Italian administration and economy, respectively, we also include a dummy variable "Rome-Milan" equal to one if the investment is located in one of these two cities, and zero otherwise.

In some specifications we also include other bilateral time-varying variables that may correlate with bilateral migration: the log of bilateral imports and exports (e.g. Gould, 1994; Rauch and Trinidade, 2002; Bratti et al., 2014), the pre-2000 stock of FDI from the same country to the same province (Head et al., 1995) as well as a binary variable equal to 1 if the same parent company has already invested in the province and zero otherwise ("Parent Co-location"; see Defever, 2006; Nefussi and Schwellnus, 2010; Castellani and Lavoratori, 2019), the log of immigrants from any origin country residing in province n (net of bilateral immigrants) and the log of emigrants having moved from province n to any destinations (net of bilateral emigrants). All regressors are lagged one year when included in the model (unless otherwise specified).

Table 1 reports the summary statistics and Table 6 the correlation matrix of our variables. GDP, bilateral FDI stock and parent colocation, as well as the dummy for Rome and Milan, display the highest correlations with the dependent variable *Choice*. Furthermore, *Choice* is positively and highly correlated with the log of total immigrants, suggesting that the determinants of the location choice of FDI may be similar as those for the location choice of immigrants. Correlations patterns support the expectation that immigration reacts to the structure of opportunities prevailing locally, as it is highly correlated with province GDP, patent count, infrastructure endowment and wages, while it is negatively correlated with the province-specific unemployment rates. Emigration is positively correlated with immigration, which, along with the positive signs of the correlations with province GDP, patent count, imports and exports, suggest that the decision to expatriate relates with the openness of the local context, while it also constitutes a reaction to unemployment (with which the correlation is positive and quite high). Coherently with these arguments, the dummy for Rome and Milan results positively correlated with both immigration and emigration, but more strongly with immigration. Finally, geography matters for both immigrants and emigrants, as shown by the negative correlation of both variables with distance, but the correlation coefficient is larger for emigrants, suggesting that immigrants travel longer distances than emigrants.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Choice	96,726	0.012	0.108	0.000	1.000
Log Immigrants	95,969	4.346	1.604	0.000	11.725
Log Emigrants	96,726	6.014	2.388	0.000	11.017
Log GDP	96,726	9.248	0.759	7.789	11.868
Log wage (region)	96,726	9.745	0.129	9.410	10.003
Unempl. rate	96,726	7.847	4.644	1.855	26.103
Log Patent Count	96,726	3.090	1.382	0.000	6.402
Share tertiary educated	96,726	−0.003	1.003	−1.507	3.366
Infrastructure Endowment	96,348	102.638	66.620	23.087	522.210
Common Border	97,495	0.388	0.487	0.000	1.000
Log Distance	96,726	7.774	1.092	4.545	9.840
Institutional Quality	96,348	0.601	0.212	0.077	1.000
Agglomeration (sector)	93,783	1.032	0.887	0.025	33.628
Sectoral diversity	96,726	0.015	0.989	−4.939	1.943
Rome-Milan	97,520	0.024	0.152	0.000	1.000
Log Imports	96,034	17.285	2.198	1.946	23.680
Log Exports	96,467	17.933	1.991	4.159	22.326
Pre-2002 FDI stock	96,726	2.293	9.051	0.000	125.000
Parent Colocation	96,726	0.003	0.053	0.000	1.000
Log Total Immigrants	96,645	9.886	1.373	0.000	13.163
Log Total Emigrants	97,495	10.001	1.329	0.000	12.630

Table 3 reports the first 15 countries of origin of Italian FDI, which account for about 87% of our sample of investments. Unsurprisingly, 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: US, UK, Germany and France. In this very concentrated distribution of origin countries for FDI, yet, some relevant origin countries for immigrants appear to have a role, in particular China which ranks relatively high but also Philippines, India and Russia which rank among the first 20 countries, even if their contribution in absolute terms is limited (see Table 7 below for the list of origin and destination countries for 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 FDI 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

Table 2: Correlation matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	Choice	1.00																				
2	Log Distance	−0.01	1.00																			
3	Common Border	0.06	−0.11	1.00																		
4	Log GDP	0.23	−0.03	0.20	1.00																	
5	Log Patent Count	0.16	−0.10	0.45	0.76	1.00																
6	Infrastructure Endowment	0.07	−0.02	0.21	0.21	0.23	1.00															
7	Institutional Quality	0.09	−0.12	0.44	0.28	0.68	0.34	1.00														
8	Log wage (region)	0.07	−0.11	0.57	0.26	0.54	0.18	0.68	1.00													
9	Unempl. rate	−0.04	0.13	−0.50	−0.15	−0.59	−0.15	−0.78	−0.54	1.00												
10	Agglomeration (sector)	0.05	−0.02	0.09	0.03	0.07	0.06	0.08	0.06	−0.08	1.00											
11	Sectoral diversity	−0.03	0.04	−0.09	−0.12	−0.19	0.06	−0.19	−0.16	0.15	0.06	1.00										
12	Share tertiary educated	0.19	−0.01	−0.16	0.39	0.37	0.40	0.37	0.10	−0.22	0.05	0.20	1.00									
13	Log Immigrants	0.14	−0.12	0.09	0.51	0.43	0.19	0.26	0.23	−0.15	0.01	−0.04	0.32	1.00								
14	Log Emigrants	0.06	−0.21	−0.06	0.24	0.06	0.04	−0.15	−0.06	0.18	−0.01	0.06	0.09	0.39	1.00							
15	Log Total Immigrants	0.19	−0.07	0.32	0.84	0.82	0.20	0.54	0.62	−0.41	0.03	−0.21	0.38	0.49	0.13	1.00						
16	Log Total Emigrants	0.10	0.10	−0.13	0.49	0.06	0.04	−0.46	−0.31	0.47	−0.00	0.27	0.12	0.16	0.31	0.17	1.00					
17	Pre-2002 FDI stock	0.33	0.04	0.17	0.42	0.33	0.09	0.20	0.17	−0.11	0.05	−0.06	0.28	0.27	0.20	0.36	0.14	1.00				
18	Log Imports	0.13	−0.26	0.25	0.51	0.53	0.20	0.39	0.33	−0.30	0.06	−0.08	0.23	0.60	0.47	0.51	0.07	0.32	1.00			
19	Log Exports	0.10	−0.18	0.27	0.49	0.59	0.11	0.47	0.39	−0.38	0.05	−0.10	0.19	0.53	0.51	0.54	−0.01	0.32	0.78	1.00		
20	Parent Colocation	0.29	−0.00	0.02	0.10	0.07	0.04	0.04	0.02	−0.02	0.03	−0.01	0.09	0.07	0.03	0.08	0.04	0.12	0.06	0.04	1.00	
21	Rome-Milan	0.37	−0.01	0.04	0.52	0.30	0.17	0.16	0.13	−0.05	0.03	−0.05	0.48	0.31	0.14	0.41	0.29	0.56	0.27	0.19	0.15	1.00

Table 3: **Origin countries of FDI**

Country	FDI count (%*)	of which (%): R&D	Manuf.	Market Acc.	Business S.	Logistics	Other funct.
United States	290 *(25.28%)*	10.00	12.07	32.76	23.45	5.52	16.21
United Kingdom	134 *(11.68%)*	3.73	3.73	28.36	41.79	5.22	17.16
Germany	117 *(10.20%)*	4.27	11.11	51.28	17.95	7.69	7.69
France	101 *(8.81%)*	3.96	17.82	43.56	16.83	6.93	10.89
Spain	93 *(8.11%)*	1.08	6.45	36.56	8.60	4.30	43.01
Switzerland	59 *(5.14%)*	10.17	16.95	30.51	15.25	10.17	16.95
Japan	36 *(3.14%)*	5.56	41.67	30.56	8.33	8.33	5.56
China	31 *(2.70%)*	29.03	0.00	58.06	12.90	0.00	0.00
Netherlands	29 *(2.53%)*	0.00	27.59	20.69	10.34	17.24	24.14
Belgium	27 *(2.35%)*	0.00	40.74	22.22	22.22	3.70	11.11
Austria	25 *(2.18%)*	0.00	4.00	32.00	52.00	4.00	8.00
Ireland	24 *(2.09%)*	12.50	0.00	12.50	20.83	50.00	4.17
Canada	20 *(1.74%)*	5.00	5.00	65.00	15.00	0.00	10.00
Sweden	18 *(1.57%)*	16.67	5.56	66.67	11.11	0.00	0.00
Finland	14 *(1.22%)*	0.00	57.14	35.71	0.00	0.00	7.14
Other countries	*143* *(12.47%)*	*2.80*	*20.28*	*33.57*	*22.38*	*7.69*	*13.29*
TOTAL	**1,147** ***(100 %)***	**6.28**	**13.34**	**36.09**	**21.80**	**7.15**	**15.34**

*Of total inward FDI into Italy, 2003-2015. Source: FDI markets

such as for instance GDP, distance, institutional similarity. This suggests that function-specific considerations may matter more to the investment decisions than origin country characteristics.

Table 4 distinguishes Italian inward FDI by function and reports their frequencies as well as the average capital investment in each function. The vast majority of inward FDI in our sample (which excludes franchising FDI) is represented by what we call "Market-access" FDI (i.e. those classified in *fDI Markets* as "Sales, Marketing & Support" and "Customer Contact Centres"). Several FDI 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 an estimate of the capital investment and of the jobs created, yielding a measure of investment size[6]. Among Market-access FDI, those categorised as "Sales, Marketing and Support" are characterized 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. These "lighter" kinds of investments can be expected to be more reactive to changes in information costs. Detailed inspection of the microdata (not shown) reveals that "Sales, Marketing and Support" investments consist in their wide majority in 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 US, UK and France. "Business services" comprise in their largest majority investment in advertising and financial services. Compared with the previous category, business services FDI 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). Along with the access to information, it is reasonable to expect these kinds of investments to respond to demand considerations.

As to investments in R&D (Research and Development investments and Design, Development and Testing), these are relatively labour intensive but plausibly more reliant on localised knowledge, which may require the mediating role of migrants. Similarly to FDI establishing Headquarters, which are unfortunately too few to run a separate statistical analysis, we expect these to react to the availability of highly skilled labour in the location and on the presence of agglomeration externalities of the Marshallian or of the Jacobian kind (Defever, 2006).

By contrast, Manufacturing FDI, as well as FDI in "Electricity", "Logistics, Distribution and Transportation", "ICT and internet infrastructure" and "Extraction", display the highest average capital investments. The number of jobs created is also relatively high. Hence, while information asymmetries and demand considerations may play a role, the accessibility of the location in terms of infrastructures and the availability and costs of labour may be among the more binding drivers for the location of these investments.

In fig. 1 we report the distribution of Italian inward FDI by province. The sub-

[6]These variables are estimated in *fDI Markets* on the basis of on historical data of similar projects in similar sectors and activities.

Table 4: Functions of FDI

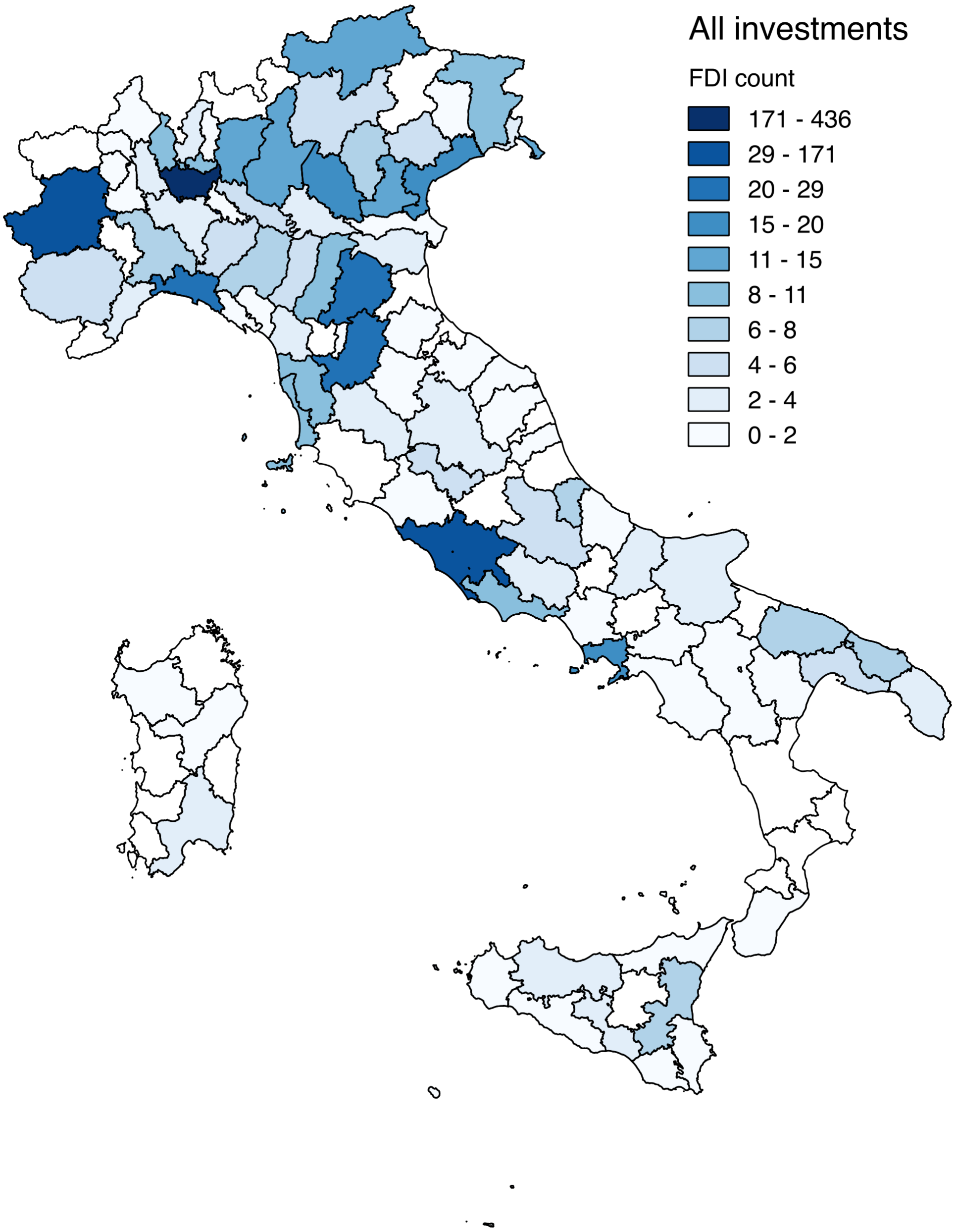
Function	FDI count	%	Capital investment*	Jobs created**
Market-access:	*414*	*36.1*	*11.43*	*17.22*
Customer Contact Centre	10	0.9	61.77	239.50
Sales, Marketing & Support	404	35.2	10.18	11.72
Business Services	*250*	*21.8*	*17.01*	*10.14*
Manufacturing	*153*	13.3	*112.20*	*163.52*
R&D:	*72*	*6.3*	*31.41*	*78.61*
Design, Development & Testing	46	4.0	23.89	76.41
Research & Development	26	2.3	44.71	82.50
Other functions:				
Construction	83	7.2	60.89	242.47
Logistics, Distribution & Transportation	82	7.2	111.76	127.57
Electricity	24	2.1	163.68	58.88
Headquarters	24	2.1	31.74	177.83
Education & Training	17	1.5	11.30	34.18
ICT & Internet Infrastructure	14	1.2	105.98	49.79
Maintenance & Servicing	9	0.8	8.98	59.89
Recycling	2	0.2	26.00	40.50
Technical Support Centres	2	0.2	9.70	39.00
Extraction	1	0.1	521.10	214.00
TOTAL	**1,147**	**100.0**	**43.31**	**68.70**

*Average, Millions US$. **Average, jobs. *Source: FDI Markets*

national heterogeneity in the distribution of FDI is striking, the vast majority of FDI being directed to the province of Milan. The provinces of Rome, Turin, Bologna, Genova, Florence, Verona and Naples also result as comparatively important attraction poles. Some geographical clustering of the investments in Northern provinces could be identified, while it seems almost absent in Southern provinces. Fig 2 reports the geographic distribution of FDI broken down by selected functions (R&D, Manufacturing, Market-access and Business Services). The leading roles of Milan, followed by Rome and Turin are confirmed in this figure, while a more dispersed geographical distribution, with some indication of a spatial clustering effect, emerges for Manufacturing FDI. R&D FDI display the highest concentration, while relatively similar geographical distributions emerge for Market-access and Business services FDI.

Finally, Table 5 reports some summary statistics drawn from *fDi Markets* that refer to the characteristics of the 895 investing companies in our sample. Average capital investments for these companies worldwide amount to 44.40 million US$ in the considered period and are marked by a highly right-skewed distribution. The total capital

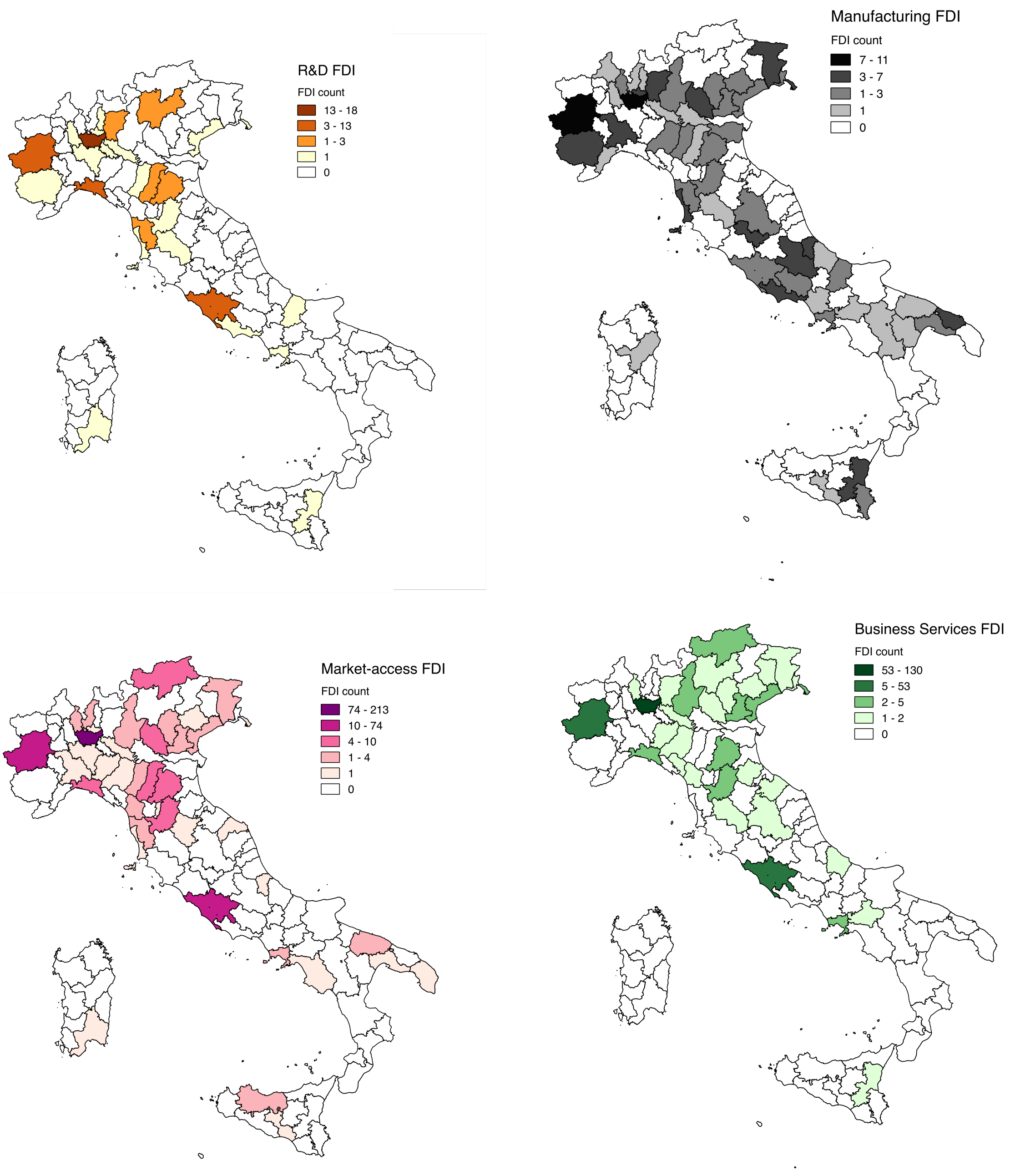
Figure 1: **Distribution of FDI by province (2003-2015)**



Source: Own elaborations on fDI markets

investment worldwide of these companies amounts on average to 913.81 Mln US$. In order to capture the relative importance of investments in Italy with respect to the overall investment portfolio of our investing companies, we compute the share of Italy over total investments. These represent on average a 39% of total investments worldwide. The number of countries where these firms invest, which we take as a proxy of their

Figure 2: **Distribution of FDI by province for selected FDI functions (2003-2015)**



Source: Own elaborations on fDI markets.

coordination capacity, ranges from 1 to 76, with an average of 6.9. We also compute a proxy for the international experience of these firms at the time of their first investment in Italy, which is based on the difference between the year of their first occurrence in *fDI Markets* and the year of their first investment in Italy. Due to data availability, we assume that the earliest year of investment is 2003. On average, the years of expe-

Table 5: Summary Statistics - Investing Companies (2003–2015)

Variable	Mean	Std. Dev.	Min	Max
Average investment size (mln US$)	44.40	116.25	0.20	2 561.00
Worldwide capital investment (mln US$)	913.81	3 049.80	0.20	33 426.85
Italy share over total investments	0.39	0.39	0.00	1.00
Number of investments in Italy	1.34	1.04	1.00	13.00
Dummy: More than one investment in Italy	0.17	0.38	0.00	1.00
Number of countries worldwide	6.90	8.74	1.00	76.00
Years of experience at first investment in Italy	1.62	2.33	0.00	11.84
Number of activities worldwide	2.16	1.98	1.00	15.00
Jobs/Mln US$ invested	4.52	8.06	0.04	166.67
Sectoral category dummies:				
Services	0.59		0.00	1.00
Final goods	0.05		0.00	1.00
Intermediate goods	0.24		0.00	1.00
Other	0.12		0.00	1.00
World region dummies:				
EU	0.55		0.00	1.00
South & East Asia	0.09		0.00	1.00
Non-EU Europe	0.06		0.00	1.00
North America	0.28		0.00	1.00
Rest of the world	0.03		0.00	1.00
Linguistic distance[a]	0.07	0.39	−0.74	0.53
Religious distance[a]	−0.79	0.65	−1.29	1.28

Observations: 895. [a]Observations: 891. Source: *fDI Markets*

rience are 1.62. As regards the labour intensity of the new ventures, on average, the investing companies create 4.52 jobs per million US$ investments, again, however, with substantial heterogeneity across firms.

A 17% of the firms in our sample invested in Italy more than once, with a maximum number of investments of 13. In terms of sectoral distribution, the majority of firms in our sample (59%) are operating in the services sectors. As for manufacturing firms, 5% of them produce final goods, a 24% intermediate goods, and a 12% of them produce goods that can enter both categories.

Table 6: Correlation matrix - Investing companies variables

		1	2	3	4	5	6	7	8	9	10	11
1	Average investment size (mln US$)	1.00										
2	Worldwide capital investment (mln US$)	0.34	1.00									
3	Italy share over total investments	−0.05	−0.25	1.00								
4	Number of investments in Italy	0.08	0.29	−0.06	1.00							
5	Dummy: More than one investment in Italy	0.11	0.29	−0.06	0.71	1.00						
6	Number of countries worldwide	0.10	0.64	−0.52	0.29	0.28	1.00					
7	Years of experience at first investment in Italy	0.04	0.16	−0.48	−0.01	−0.01	0.37	1.00				
8	Number of activities worldwide	0.16	0.60	−0.45	0.21	0.27	0.78	0.36	1.00			
9	Jobs/Mln US$ invested	−0.11	−0.07	0.05	−0.07	−0.08	−0.05	−0.03	−0.03	1.00		
10	Linguistic distance	0.01	0.02	−0.05	−0.00	−0.00	0.06	0.11	0.07	−0.09	1.00	
11	Religious distance	0.01	0.02	−0.07	−0.04	−0.03	0.06	0.10	0.07	0.01	0.43	1.00

As regards the world areas of origin, 55% of investing companies come from the EU, 28% from North America, 9% from South and Eastern Asia, 6% from non-EU Europe 3% from the rest of the world.

Table 7: Origin and destination countries of immigrants and emigrants

Origin countries		Destination countries	
Country	Immigrants*	Country	Emigrants**
Romania	8,939.35	Argentina	6,274.01
Albania	4,716.65	Germany	6,090.79
Morocco	4,380.58	Switzerland	5,245.26
China	2,163.16	France	3,588.38
Ukraine	1,928.51	Brazil	3,119.32
Philippines	1,483.14	Belgium	2,499.87
Moldova	1,454.08	United States	2,236.11
India	1,354.48	United Kingdom	2,085.95
Peru	1,071.14	Australia	1,280.80
Bangladesh	940.46	Canada	1,240.38
Ecuador	925.27	Spain	1,228.42
Tunisia	914.64	Venezuela	1,184.64
Poland	879.65	Uruguay	923.22
Sri Lanka	833.44	Chile	558.75
Pakistan	816.62	Peru	342.40
Senegal	798.16	Netherlands	339.20
Egypt	777.18	South Africa	338.67
FYR Macedonia	733.27	Luxembourg	249.56
Nigeria	534.95	Austria	220.62
Ghana	519.42	Ecuador	153.32
*Average****	*245.0*	*Average****	*234.5*

*Province-level averages, stocks of residents, 2011. *Source: ISTAT*;
** Province-level averages, stocks of registered citizens residing abroad, 2011. *Source: AIRE.*
*** Average bilateral (country-province) stocks of residents/registered citizens residing abroad in provinces where at least one FDI is observed

The correlation matrix of our investing company variables shows especially high correlations between the total worldwide capital investment of the firm and the number of countries it invests worldwide and the years of experience of the MNE at its first investment in Italy, as well as with the number of different MNE activities in which it is active globally. These findings suggest that firms with greater managerial capacity have a larger and more diversified portfolio of international investments. Italy's share of total investments—as well as the sheer number of investments in Italy—are also larger for firms with a more diversified portfolio in terms of countries and activities. The linguistic and religious distance between Italy and the country of origin of the investment are only marginally correlated with our investor-level variables.

The breakdown of our two migration variables by country is reported in Table 7. The left (resp. right) panel reports the first 20 origin countries of immigrants (resp. first 20 destination countries of emigrants) in 2011. The set of immigrants' origin countries is very diversified and covers all continents, with only limited overlap with the set of origin countries for FDI (yet, Germany and France, while not featuring in the top-20 origin countries, display an above average number of residents in Italian provinces), mainly due to China, the Philippines and India. The set of destination countries for Italian expatriates, instead, is mainly represented by OECD countries, and it displays greater overlap with FDI origin countries—with the exception of China and Japan. This heterogeneity may underlie different effects of immigrants' and emigrants' networks on FDI, which will be addressed in the next Section.

4 Results

4.1 Conditional Logit Results

In order to study the role of immigrants on the location choice of FDI, we first implement a conditional logit model. When the variables are log-transformed, the coefficients can be interpreted as approximate elasticities (Train, 2009)[7].

The results of the conditional logit model describing the location choice for FDI among Italian provinces are reported in Table 8. In column (1) we include standard province-level variables considered to promote the attractiveness of investments in the literature. Among the variables capturing market size, the log of the province GDP results positive and significant, while the log of population is not statistically significant. The two variables have been included separately in order not to impose parametric restrictions to their relationships and their estimated coefficients reveal that province GDP is the main driver for FDI. The sign of the coefficient of the average wage could be considered to be ambiguous ex ante, as it depends on the underlying motives for the investment decision —if the FDI is intended to save on labour costs, higher wages may have a negative effect on the location of FDI; if, instead, human capital and skills considerations prevail, we may expect to observe a positive coefficient. In the specification of column (1), the variable results negative and significant. The province-level unemploy-

[7] Indeed, denoting with b the coefficients estimated for the variable $log(x_i)$ via conditional logit, it can be easily shown that $\frac{\partial P_i}{\partial x_i} \frac{x_i}{P_i} = b(1 - P_i)$. As long as the probability to choose a specific province i is small, which is generally the case when there are many provinces in the choice set, the estimated coefficients will be approximately equal to the elasticities; in Table 9 we report the elasticities estimated for the provinces which are more likely to be chosen (e.g. Milan, Rome, Turin). The corresponding cross-elasticity, i.e., the change in the probability to choose province i associated with a change in x_j in another province j, is $-bP_j$, which will be negligible as long as the probability to choose j is small.

Table 8: Estimation results - Conditional Logit

Dep. var: Choice	(1)	(2)	(3)	(4)	(5)	(6)
Log Immigrants		$0{\cdot}321^{***}$		$0{\cdot}220^{***}$	$0{\cdot}218^{***}$	$0{\cdot}214^{***}$
		(0·058)		(0·065)	(0·068)	(0·068)
Log Emigrants			$0{\cdot}290^{***}$	$0{\cdot}202^{***}$	$0{\cdot}106^{*}$	$0{\cdot}135^{**}$
			(0·052)	(0·058)	(0·059)	(0·061)
Log Prov. GDP	$1{\cdot}226^{***}$	$1{\cdot}048^{***}$	$1{\cdot}114^{***}$	$1{\cdot}014^{***}$	$1{\cdot}092^{***}$	$1{\cdot}262^{***}$
	(0·306)	(0·307)	(0·305)	(0·307)	(0·324)	(0·333)
Log Prov. Population	−0·274	−0·464	−0·422	−0·497	$-0{\cdot}595^{*}$	−0·111
	(0·331)	(0·333)	(0·331)	(0·332)	(0·350)	(0·376)
Log average wage (region)	$-3{\cdot}089^{***}$	$-3{\cdot}408^{***}$	$-2{\cdot}698^{**}$	$-3{\cdot}037^{***}$	$-1{\cdot}914^{*}$	−0·981
	(1·083)	(1·085)	(1·096)	(1·098)	(1·130)	(1·206)
Prov. unemployment rate	0·028	0·035	0·027	0·032	0·018	−0·013
	(0·028)	(0·028)	(0·028)	(0·028)	(0·028)	(0·030)
Log Patent Count	0·091	0·138	0·097	0·129	−0·001	0·017
	(0·101)	(0·101)	(0·101)	(0·101)	(0·106)	(0·107)
Share tertiary educated	$0{\cdot}165^{**}$	$0{\cdot}147^{**}$	$0{\cdot}158^{**}$	$0{\cdot}147^{**}$	$0{\cdot}128^{*}$	0·101
	(0·070)	(0·072)	(0·071)	(0·071)	(0·073)	(0·072)
Infrastructure endowment	$0{\cdot}001^{*}$	0·001	0·001	0·001	$0{\cdot}001^{**}$	$0{\cdot}002^{**}$
	(0·001)	(0·001)	(0·001)	(0·001)	(0·001)	(0·001)
Common border	$0{\cdot}471^{***}$	$0{\cdot}610^{***}$	$0{\cdot}445^{***}$	$0{\cdot}548^{***}$	$0{\cdot}299^{**}$	$0{\cdot}232^{*}$
	(0·112)	(0·116)	(0·112)	(0·117)	(0·127)	(0·128)
Log Distance	$-1{\cdot}019^{***}$	$-0{\cdot}704^{***}$	$-0{\cdot}885^{***}$	$-0{\cdot}693^{***}$	$-0{\cdot}491^{**}$	$-0{\cdot}467^{**}$
	(0·193)	(0·197)	(0·188)	(0·195)	(0·206)	(0·207)
Institutional Quality	$1{\cdot}296^{***}$	$0{\cdot}985^{**}$	$1{\cdot}705^{***}$	$1{\cdot}383^{***}$	0·332	−0·141
	(0·485)	(0·478)	(0·484)	(0·489)	(0·540)	(0·583)
Agglomeration (Sector)	$0{\cdot}294^{***}$	$0{\cdot}280^{***}$	$0{\cdot}288^{***}$	$0{\cdot}279^{***}$	$0{\cdot}247^{***}$	$0{\cdot}243^{***}$
	(0·023)	(0·023)	(0·023)	(0·023)	(0·025)	(0·025)
Sectoral diversity	0·032	0·025	−0·000	0·005	0·039	$0{\cdot}110^{*}$
	(0·056)	(0·057)	(0·057)	(0·057)	(0·059)	(0·065)
Rome-Milan	$0{\cdot}834^{***}$	$0{\cdot}776^{***}$	$0{\cdot}738^{***}$	$0{\cdot}730^{***}$	0·339	$0{\cdot}558^{**}$
	(0·230)	(0·231)	(0·231)	(0·231)	(0·250)	(0·260)
Log Imports					$0{\cdot}142^{***}$	$0{\cdot}131^{***}$
					(0·049)	(0·050)
Log Exports					0·029	−0·006
					(0·060)	(0·061)
Pre-2002 FDI stock					$0{\cdot}008^{***}$	$0{\cdot}007^{***}$
					(0·001)	(0·002)
Parent Colocation					$3{\cdot}863^{***}$	$3{\cdot}830^{***}$
					(0·198)	(0·198)
Log Total Immigrants						$-0{\cdot}365^{**}$
						(0·185)
Log Total Emigrants						$-0{\cdot}340^{***}$
						(0·110)
Observations	91, 502	91, 502	91, 502	91, 502	90, 915	90, 910
AIC	6, 179·710	6, 152·149	6, 151·183	6, 141·948	5, 652·059	5, 639·221
BIC	6, 302·224	6, 284·086	6, 283·121	6, 283·309	5, 830·995	5, 836·991
Pseudo R^2	0·372	0·375	0·375	0·376	0·426	0·428

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Estimated elasticities, conditional logit

Province	Probability to choose province	Elasticities Immigration	Elasticities Emigration
Milan	0.44	0.12	0.08
Rome	0.15	0.18	0.12
Turin	0.04	0.20	0.13
Bologne	0.02	0.21	0.13
Genova	0.02	0.21	0.13
Florence	0.02	0.21	0.13
Verona	0.01	0.21	0.13
Naples	0.02	0.21	0.13
Venice	0.01	0.21	0.13
Brescia	0.01	0.21	0.13

Estimates based on the predicted probabilities to choose each province deriving from conditional logit estimates reported in column (6) of Table 8; first 10 provinces.

ment rate results, in this first specification but also throughout all the conditional logit estimates, insignificant. The human capital endowment of the province, as measured by the share of tertiary degrees among residents in the province, is positively and significantly related with location choice. Among the factors influencing the information and enforcement costs, institutional quality, as well as Marshallian externalities, proxied by the agglomeration of domestic firms operating in the same sector as the MNE, are found to positively and significantly affect location choice[8]. Instead, Jacobian externalities ("Sectoral diversity") are not found to significantly affect location choice conditional on the included regressors. Among the other control variables, infrastructure endowment has a positive and weakly significant effect on the location choice of FDI; common border is positive and significant while distance is negative and significant, as in standard gravity models. Instead, after controlling for all these province characteristics, the log of the count of patent applications filed by the province to the European Patent Office in the previous year, which can proxy for the level of innovativeness of the province, is positive but insignificant. Finally, the dummy for Rome and Milan is highly positive and significant, reflecting the *ceteris paribus* greater attractiveness of these provinces for investors.

In column (2) we augment the model with the first of our variables of interest: the log of the bilateral immigrants' stock from the same origin country as the investment in each province. A positive and statistically significant effect is observed, confirming our hypotheses on the positive effect of migrants on FDI. The effect of the inclusion of

[8]The results are robust when substituting the sectoral agglomeration variable with a functional agglomeration variable as suggested by Duranton and Puga (2005)

immigrants on the other variables is in line with the picture outlined in the discussion of the correlations: the coefficient of the variables that positively affect both migration and FDI location (e.g. GDP, wage, share of highly educated, institutional quality, agglomeration, Rome-Milan) diminishes, while those affecting the two in opposing ways (common border, distance) increase. The direction of the implied omitted variable bias deriving from the omission of immigrants in column (1) suggests once again that immigrants come from distant, non-contiguous countries to settle in wealthier provinces with higher salaries, greater institutional quality and more educated residents, particularly Rome and Milan. In other words, unsurprisingly, migrants are attracted by provinces arguably offering more opportunities.

In column (3), we add the log of bilateral emigration, i.e. emigration from each province to the country of origin of the investment, to the model in column (1). The coefficient is positive and significant, supporting the interpretation of an information effect operated by emigrants. Including this variable confirms the broad picture of location determinants outlined so far, indicating again that emigrants mainly originate from wealthier provinces with greater human capital, but that they also react to worse labour conditions as measured by wages and worse institutions. Indeed, including emigrants' stocks reduces the coefficients of GDP and tertiary education and makes the coefficient of wages less negative, while it substantially increases the coefficient of institutional quality. From these results, one could expect that skilled emigration coexists with emigration stimulated by less economic opportunities. The smaller increase in the distance effect and the decrease in the border effect suggest that geography matters more to emigrants than to immigrants, with emigrants moving to relatively close countries and in particular to neighbouring countries.

In column (4) we include both migration variables in the same specification. Both coefficients decrease, confirming the positive correlation between the two variables, with immigrants' effect resulting about as large as the emigrants' effect. The previous discussion about the effects on the other regressors is confirmed.

One may question that the detected positive effect of migrants be due to other bilateral province-country variables which are correlated with migration. To assess the robustness of our results to this potential critique, in column (5) we further augment the model with the log of imports and exports (Blonigen, 2001), with the pre-2002 stock of Manufacturing FDI originating from the same country of origin and directed to the same province (e.g. Head et al., 1999), as well as with a dummy for the co-location of the same parent company in the same province (Defever, 2006; Nefussi and Schwellnus, 2010). Imports, pre-existing FDI stocks and, most relevantly, co-location, do result to positively and significantly affect the location choice of FDI. Because, in particular, co-location acts as a lagged dependent variable, including these variables reduces the effect of wages, education and institutional quality; because a majority of the co-locations are located in Rome and Milan, the relevant dummy looses significance; the import, exports and other FDI also respond to geography, decreasing the coefficients for border and distance in absolute terms. More importantly for the purposes of this paper, the effect of immigration is almost unaffected, while the coefficient for emigration drops by about one half and remains only weakly significant.

Finally, we recognise that it may be important to distinguish between the effect of bilateral migration and the effect of overall immigration and emigration stocks. For this reason, in column (6), we add the log of the immigration stock from any countries in the province ("*Log Total Immigrants*") and the log of the overall stock of emigrants in that province to any countries ("*Log Total Emigrants*"). Both variables are net of the migrants from and to the country of the investor. The resulting coefficients are negative. Considering this result in combination with the ones for the bilateral variables, we may argue that the share of co-nationals over total migrants, rather than their absolute size, is what enters the profit function of the firm. The correlation of total migration with population, wages and education makes these variables loose significance, suggesting that the coefficients of total immigration and emigration also capture factors relating to the local structure of opportunities.

The magnitude of the coefficients suggest that, on average, a 10% increase in the immigrant population from the same origin country o of investment n in a given province i, keeping all other regressors constant, would increase the probability to choose province i by about 2.14%. The same increase in the bilateral emigrant stocks from i to o would be reflected in a 1.35% increase in the probability to choose i as a location. As we discussed, though, the estimated migration elasticities depend on the predicted probability to choose i, hence they may be lower for the provinces that are more likely to be chosen, primarily Milan and Rome. Table 9 reports the exactly estimated elasticities for the 10 provinces with the highest probability to be chosen. The table indicates that the lower bounds of the estimated elasticities are 0.12 for immigration and 0.08 for emigration. As to the cross-elasticities, these will be negligible as long as the probability to choose a specific alternative province j is small (see footnote 7). However, the corresponding decrease in the probability to choose i may be more relevant if the change occurs in a province j that is relatively more likely to be chosen. For instance, a 10% increase in the immigrant population from country o in the province of Milan would decrease the probability to choose any other province i by about 0.94%; the same increase in the province of Rome would lead to a 0.32% decrease in the probability to choose any other province i.

4.2 Heterogeneity in migration effects

In order to explore the heterogeneity in the immigrants' and emigrants' effects, we re-estimate our model via mixed logit[9]. In random coefficient models such as the mixed logit, the effect of some or all of the explanatory variables is allowed to vary by individual (i.e. at the investor level; Train, 2009). In the realm of discrete choice models, this is equivalent to relaxing the assumption of Independence from Irrelevant Alternatives (IIA) implied in the conditional logit. Indeed, modelling the heterogeneity in the effects of the regressors and interpreting their coefficients as random parameters is equivalent to allowing some locations to be closer substitutes than other (Train, 2009).

Model 1 in Table 10 reports the results of our full specification (with the same regressors that were included in column (6) of Table 8) estimated by mixed logit, where all the independent variables enter our model with random parameters[10]. For random parameters, it is possible to estimate both the means and the standard deviation. Means are reported in columns (1) and (3), while standard deviations are in columns (2) and (4). It is important to highlight that a standard deviation of a random parameter significantly different from zero is evidence of heterogeneity in the extent to which the variable associated with that parameter can explain the location of inward FDI in Italian provinces. Also, it is worth mentioning that the inclusion of the dummy for Rome and Milan with a random parameter in a mixed logit is equivalent to allowing for a nest in a nested logit model, i.e. a grouping of alternatives within which error terms are correlated (Train, 2009). By including this variable with a random parameter, we are recognising the high sub-national concentration of investments and we are allowing a foreign firm to face a two-step decision: first, whether to invest in either Rome or Milan or elsewhere in Italy; second, in which specific province to invest[11].

Allowing for imperfect substitutability between alternative provinces, the positive

[9]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.

[10]Only co-location is included as a fixed parameter because its limited variation hampers convergence of a highly computationally demanding model.

[11]The estimates are robust to different specifications of the random parameters, as well as to the omission of the Rome-Milan dummy.

effect of immigration is confirmed, while the effect of emigrants decreases and looses significance. Otherwise, with regards to the average effects (column (1)), the picture is broadly in line with the one highlighted by the conditional logit estimates. Co-location, province-level GDP, infrastructure endowment, agglomeration, imports, and pre-2002 FDI stocks confirm their role as significant attraction factors for FDI, the negative signs of distance and total emigrants are confirmed. The weak significance of the effects of contiguity and of sectoral diversity resulting from the conditional logit vanishes.

Coming to the heterogeneity in the effects, the significant likelihood ratio test on the joint significance of the standard deviations allows rejecting the null hypothesis of fixed coefficients and supports the expectation that there is heterogeneity in the parameters, consistent with the idea of heterogeneous weights attributed to different location determinants by different investing companies. Specifically, the results in column (2) reveal standard deviations significantly different from zero for the effects of immigration, wages, distance, pre-2002 FDI stocks and total immigration and emigration. Our estimates imply that, while heterogeneous, the effect of immigrants is positive for a majority of the companies (about 79%, i.e. $100 \times \Phi(0.329/0.415)$), an issue that we shall further explore in what follows. Wages result having on average an insignificant coefficient but a significant heterogeneity. The effect is positive for about 49% of the firms and negative for the rest. This is consistent with the interpretation that some firms may value labour cost savings (for instance if they are more intensive in manual labour) while others may be seeking for knowledge and highly skilled human capital, hence they would locate where wages are higher. As discussed, functions may be the main drivers of this heterogeneity.

Functional, sectoral and knowledge base heterogeneity may also explain the significant standard deviations in imports and FDI stocks, whose effects are estimated to be positive for 83% and 63% of the firms, respectively. Indeed, both variables may indicate a preferential bilateral tie between the country of origin of the investment and the destination province, hence their effects may be interpreted to be the result of heterogeneity in the effects of information costs on location choice. Similarly, the effect of distance is estimated to be negative, as expected, for about two thirds of the firms, but it is actually positive for the remaining third. This finding is consistent with the results by Castellani et al. (2013) who find that specific types of FDI, namely R&D investments, travel longer distances. Heterogeneity is also detected in the effects of bilateral imports and FDI stocks, i.e. in the total stock of immigrants (estimated positive in 50% of the cases) and emigrants (positive in 13% of the cases). Finally, the standard deviation estimated for the dummy for Rome and Milan is significant at the 10% level. The estimated coefficients indicate that, consistent with their distinctive role as the economic and administrative capitals of Italy, about one third of the investors still consider these cities as a preferential location for their investment, even conditionally on all other covariates.

The heterogeneity detected in the immigrants' effects may be due to different factors. A way to test our hypothesis that the immigrants' effect is heterogeneous by function is to interact the log of immigrants with a set of dummy variables representing the main functions that we observe in our data (R&D, Manufacturing, Market Access, Business Services) and to enter the interaction terms as fixed parameters in our specification (Chung and Alcácer, 2002). The results are shown in column (3) and (4) of Table 10. These interaction terms express how different functions move the immigrants' effect away from the baseline estimated mean of the immigrants' effects, which, in this specification, becomes insignificantly different from zero. The interaction term is positive for R&D, Market Access and Business Services FDI, indicating a larger effect of immigrants in this kind of investments. On the contrary, the interaction effect is negative for Manufacturing FDI. These results suggest FDI in information and knowledge-intensive activities as well as those in market access activities are more sensitive to the presence of immigrants from the same home country. Instead, FDI in manufacturing, which are arguably more intensive of unskilled labour, turn out to be less sensitive. Still, the standard deviation of the immigrants' effect remains relatively large and significant, suggesting that the sources of heterogeneity in the immigrants' effects are not exhausted by the inclusion of the function-specific dummies and that, even controlling for the heterogeneity deriving from FDI functions, the immigrants' effects are still positive for about 50% of the investors.

To explore the remaining sources of heterogeneity, we predict the firm-specific random parameters $\hat{\delta}_f^{\text{Immi}}$ for *Log Immigrants* derived by the mixed logit model and examine the determinants of their distribution via regression models (Castellani and Lavoratori, 2019; Alcácer et al., 2018; Hornstein and Greene, 2012; Saxonhouse, 1976). Specifically, we study the distribution of the $\hat{\delta}_n^{\text{Immi}}$ predicted from the interacted specification reported in column (3) of Table 10.

Our second stage regression takes the following form:

$$\hat{\delta}_f^{\text{Immi}} = \theta' X_f + u_f \tag{3}$$

where X_f is a vector of firm, industry and home-country level characteristics.

To take into account the heteroskedasticity in this second-stage model, the estimated

Table 10: Estimation results - Mixed Logit

	(1)	(2)	(3)	(4)
	Model 1		**Model 2**	
	Means	**SD**	**Means**	**SD**
Log Immigrants	0·329***	0·415***	0·007	−0·379***
	(0·088)	(0·085)	(0·122)	(0·088)
Log Emigrants	0·052	−0·052	0·151	0·004
	(0·074)	(0·143)	(0·117)	(0·132)
Log Immigrants × R&D			0·341	
			(0·220)	
Log Immigrants × Manufacturing			−0·437***	
			(0·153)	
Log Immigrants × Market Access			0·808***	
			(0·135)	
Log Immigrants × Business Services			0·907***	
			(0·168)	
Log Emigrants × R&D			−0·510**	
			(0·236)	
Log Emigrants × Manufacturing			−0·157	
			(0·152)	
Log Emigrants × Market Access			−0·057	
			(0·145)	
Log Emigrants × Business Services			−0·035	
			(0·173)	
Parent Colocation	5·041***		4·737***	
	(0·276)		(0·252)	
Log Prov. GDP	1·546***	−0·153	1·217***	−0·027
	(0·436)	(0·174)	(0·408)	(0·164)
Log Prov. Population	−0·356	0·136	−0·092	0·044
	(0·474)	(0·224)	(0·444)	(0·164)
Log Patent Count	0·026	0·020	0·066	−0·075
	(0·113)	0·174	(0·115)	(0·126)
Share tertiary educated	0·116	−0·026	0·138*	−0·031
	(0·078)	0·130	(0·078)	(0·120)
Log average wage (region)	−0·085	−3·099***	−0·848	0·611
	(1·417)	(1·173)	(1·335)	(1·575)
Prov. unemployment rate	−0·019	−0·009	−0·013	−0·012
	(0·033)	(0·034)	(0·032)	(0·036)
Infrastructure endowment	0·002**	−0·000	0·002***	−0·000
	(0·001)	(0·001)	(0·001)	(0·002)
Common border	0·155	0·141	0·245*	0·215
	(0·144)	(0·293)	(0·146)	(0·295)
Log Distance	−0·850***	1·979***	−0·567**	1·525***
	(0·312)	(0·411)	(0·286)	(0·363)
Institutional Quality	−0·213	0·055	−0·150	0·282
	(0·622)	(0·742)	(0·608)	(0·683)
Agglomeration (Sector)	0·242***	−0·030	0·209***	−0·014
	(0·031)	(0·072)	(0·030)	(0·093)
Sectoral diversity	0·080	−0·024	0·074	−0·048
	(0·069)	(0·107)	(0·066)	(0·123)
Rome-Milan	−0·366	−0·801*	−0·417	0·685
	(0·367)	(0·420)	(0·313)	(0·438)
Log Imports	0·189***	−0·195**	0·168***	0·155**
	(0·065)	(0·088)	(0·058)	(0·073)
Log Exports	−0·043	0·049	−0·006	0·099
	(0·067)	(0·107)	(0·071)	(0·079)
Pre-2002 FDI stock	0·010***	0·030***	0·010***	−0·017***
	(0·003)	(0·005)	(0·002)	(0·004)
Log Total Immigrants	−0·019	−0·538***	−0·164	0·250**
	(0·240)	(0·156)	(0·205)	(0·123)
Log Total Emigrants	−0·242**	0·214**	−0·269**	−0·176
	(0·121)	(0·089)	(0·119)	(0·137)
Observations	90,910		90,910	
LR test χ^2	184.81		106.40	
df	20		20	
p-value	0.000		0.000	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

individual parameters must be weighted by their estimated variance (Saxonhouse, 1976; Hornstein and Greene, 2012). Because our parameters were obtained via simulation, the variance of the individual parameters $\hat{\delta}_f^{\text{Immi}}$ was estimated via parametric bootstrapping[12]. Results from OLS regressions of equation (3) are reported in Table 11.

The average $\hat{\delta}_n^{\text{Immi}}$ corresponds, with minor deviations attributable to numerical issues in simulation, to the average coefficient of the log of immigrants estimated in the first stage. Hence, in the second-stage regressions, the constant of the model corresponds to the average effect of immigrants when the covariates are zero, net of the role of functions that is captured in the first-stage interaction effects. Consistently with the first-stage results, the constant-only model yields a coefficient for the constant that is insignificantly different from zero. The coefficients of the covariates reported in Table 11 can be interpreted as factors that drive these effects to be larger or smaller[13].

Based on the arguments in Section 2, we expect that firms' knowledge of the Italian market will be among these factors. Firms with more than one investment in Italy will presumably have better knowledge and a more established network of reference supporting the new venture setup. Hence, we expect, they will rely less on the immigrants' facilitating role. For this reason, we include a dummy for whether the firm has more than one investment in Italy in column (1). The coefficient, indeed, results negative and significant. The overall distribution of the $\hat{\delta}_f^{\text{Immi}}$ is illustrated in Figure 3, along with the distribution of the coefficient for the two cases that the firm has only one, or more than one investment in Italy. Clearly, the distribution of the $\hat{\delta}_f^{\text{Immi}}$ for the firms with more than one investment is more spread and has more mass around the smaller values of the coefficient than the one for the firms with only one investment. These results are robust to substituting the dummy with a continuos variable as well as with a categorical grouping of firms by the number of investments they have in Italy (1, 2, 3, 4 or more). The relationship between the average coefficient and the number of investments is represented in Figure 4.

In column (2), we add a measure of the size of the firm, proxied by the total (log) capital expenditure in FDI worldwide (column 2), as well as the firm-level share of worldwide capital investments targeting Italy. Larger firms that, presumably, have more

[12]We would like to thank Arne Risa Hole for offering valuable guidance in this process.

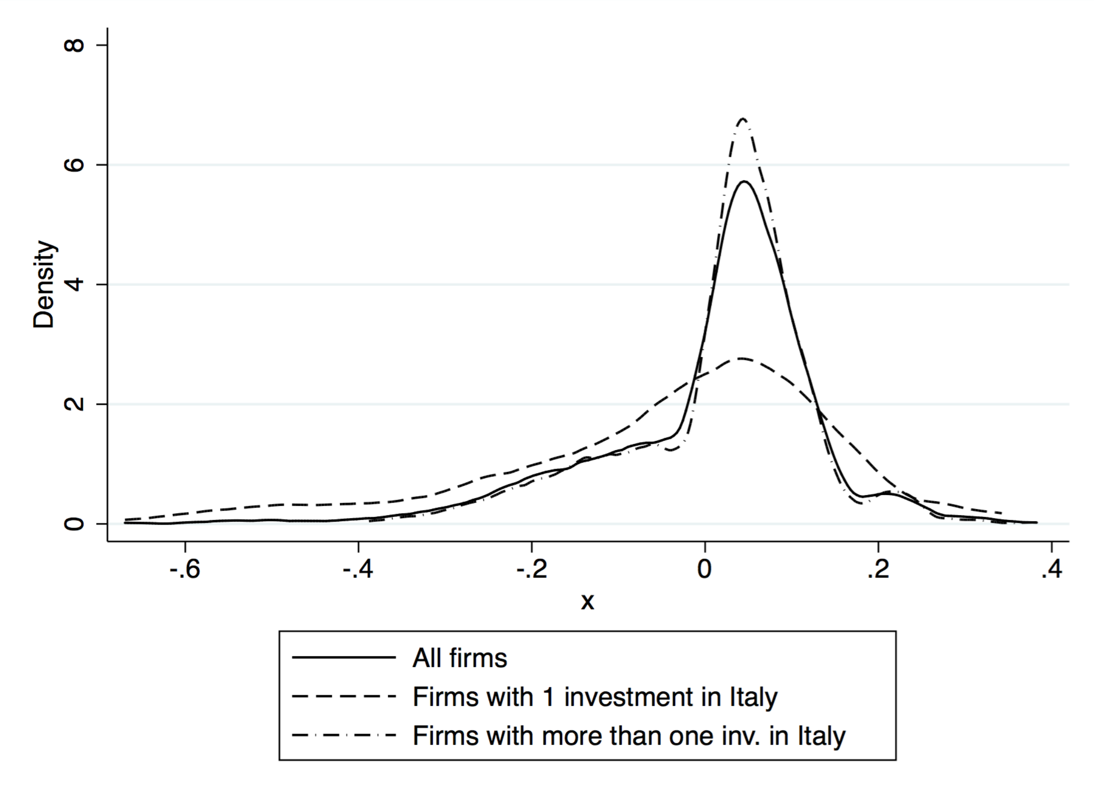
[13]The results of second stage regressions are robust to changing the baseline category for the interaction effects in the first stage. Changing the baseline category for the interaction effects in the first stage mainly impacts the constant of the second-stage model, while the second-stage covariate coefficients are changing very little (within an order of magnitude of 20%, attributable to numerical issues in the simulations) and the standard errors remain very stable. The second-stage regressions are also qualitatively the same when the dependent variable $\hat{\delta}_f^{\text{Immi}}$ is derived from a first stage with no interactions such as the one in column (1) of Table 10.

Table 11: Sources of heterogeneity in the immigrants' effects $\hat{\delta}_n^{\text{Immi}}$

Dep. var: $\hat{\delta}_n^{\text{Immi}}$	(1)	(2)	(3)	(4)	(5)	(6)
Dummy: more than one investment in Italy	$-0{\cdot}050^{***}$	$-0{\cdot}033^{**}$	$-0{\cdot}031^{**}$	$-0{\cdot}030^{**}$	$-0{\cdot}034^{***}$	$-0{\cdot}034^{**}$
	(0·012)	(0·013)	(0·013)	(0·013)	(0·013)	(0·013)
Log total capital investment worldwide		$-0{\cdot}008^{***}$	$-0{\cdot}010^{***}$	$-0{\cdot}011^{***}$	$-0{\cdot}009^{***}$	$-0{\cdot}009^{***}$
		(0·003)	(0·003)	(0·003)	(0·003)	(0·003)
Italy share of capital investment worldwide		$-0{\cdot}033^{**}$	$-0{\cdot}038^{***}$	$-0{\cdot}040^{***}$	$-0{\cdot}032^{**}$	$-0{\cdot}033^{**}$
		(0·014)	(0·015)	(0·015)	(0·015)	(0·015)
Jobs/Mln US$ invested			$-0{\cdot}002^{***}$	$-0{\cdot}002^{***}$	$-0{\cdot}002^{***}$	$-0{\cdot}002^{***}$
			(0·001)	(0·001)	(0·001)	(0·001)
World areas of origin (ref: EU)						
South & East Asia				$0{\cdot}032^{*}$	$0{\cdot}041^{**}$	
				(0·016)	(0·017)	
Non-EU Europe				−0·002	0·000	
				(0·019)	(0·019)	
North America				0·002	0·001	
				(0·010)	(0·010)	
Rest of the world				0·022	0·022	
				(0·026)	(0·026)	
Type of sector (ref: Services)						
Final goods					−0·012	−0·011
					(0·021)	(0·021)
Intermediate goods					$-0{\cdot}036^{***}$	$-0{\cdot}034^{***}$
					(0·011)	(0·011)
Other goods					0·002	0·002
					(0·014)	(0·014)
Linguistic distance						−0·010
						(0·013)
Religious distance						$0{\cdot}021^{***}$
						(0·008)
Constant	0·004	$0{\cdot}050^{***}$	$0{\cdot}068^{***}$	$0{\cdot}068^{***}$	$0{\cdot}067^{***}$	$0{\cdot}089^{***}$
	(0·005)	(0·016)	(0·017)	(0·017)	(0·017)	(0·019)
Observations	895	895	895	895	895	891
chi2	18·477	28·131	35·870	40·244	52·184	53·535
df	1	3	4	8	11	9
p-value	0·000	0·000	0·000	0·000	0·000	0·000

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Variance-weighted least-squares regression. Dependent variable: estimated coefficients $\hat{\delta}_n^{\text{Immi}}$ from the mixed logit model reported in Table 10. Variances of the individual parameters $\hat{\delta}_n^{\text{Immi}}$ estimated by parametric bootstrapping.

Figure 3: **Distribution of the estimated coefficients** $\hat{\delta}_n^{\mathbf{Immi}}$

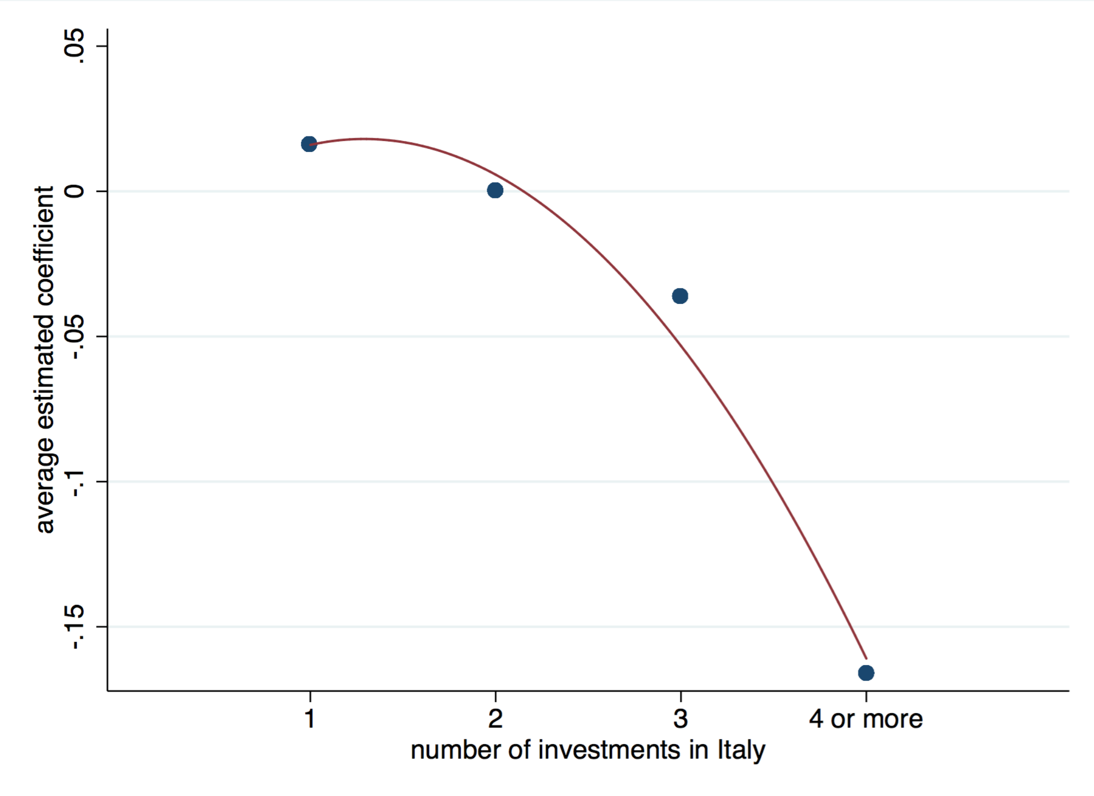


Source: Own elaborations on fDI markets

efficient management structures and better information processing capacity, benefit less from the facilitating effects of immigrants. The relative importance of Italy in the portfolio of a firm's worldwide investments also decreases the salience of the immigrants' effect. In column (3), we also include a measure of the labour intensity of the investment, proxied by the number of jobs created per million US$ spent in the firms' worldwide investments. This could be interpreted as another size effect, or as another indication against the conjecture of a labour effect of immigrants that confirms the results in the first stage about the smaller effect of immigrants for investments in manufacturing. Firms that make more labour intensive investments tend to be less sensitive to the number of migrants from their home country to the destination province of their Italian FDI.

In column (5), we further include a set of dummies representing the areas of origin of firms. Among the set of origin-areas, immigrants' effect is found to be stronger for firms from South and East Asia (relative to the reference EU countries) at a 10% significance level. The coefficient for South and East Asia becomes larger and more significant when

Figure 4: **Average estimated coefficient $\hat{\delta}_n^{\mathbf{Immi}}$ by number of investments in Italy**



we further augment the specification with a categorical variable distinguishing the broad type of sector in which the investment is operating: services, intermediate goods, final goods, and other (column (5)). The negative and significant coefficient detected for intermediate goods could be seen as an indication that sectors which are more likely to be reference-priced and less information-intensive rely more on the immigrants' effects (see Rauch and Trinidade, 2002). The alternative interpretation of this result as a demand effect would imply that the effect for intermediate goods is significantly smaller than the one for final goods, as suggested by Burchardi et al. (2018). This is supported in terms of the relative magnitudes of the coefficients but not to a statistically significant extent.

The significantly positive coefficient of the dummy for South and East Asian firms could be attributed to the larger cultural distance existing between Italy and these origin countries. This interpretation is supported in column (6), where we substitute the world area dummies with the measures of linguistic and religious distance computed by Douglas Dow[14] (see Dow and Karunaratna, 2006, and related papers using these measures; the regressions are unreported for brevity but are available upon request). Among

[14]http://dow.net.au/

these, religious distance, which could be taken as the closest proxy for cultural distance, turns out to be positively and highly significantly correlated with our estimated $\hat{\delta}_n^{\text{Immi}}$. Hence, the migrants' effects are stronger towards more culturally distant countries, mirroring an established finding in the trade literature (e.g. Girma and Yu, 2002; Tadesse and White, 2008). Alternatively, the result could be attributed to the strong role of co-ethnic business networks in supporting the international activities of the investors from countries belonging to these areas, particularly China and India (Weidenbaum and Hughes, 1996; Redding, 1995; Rauch and Casella, 2003; Özden et al., 2011; see also for instance Gao, 2003; Tong, 2005 on the role of Chinese networks on FDI, Rauch and Trinidade, 2002 on the role of Chinese networks on trade).

In a further set of unreported regressions, we regressed the $\hat{\delta}_n^{\text{Immi}}$ on similar specifications as in columns (5) and (6) where the type of sector dummies were replaced by a set of 16 dummies indicating the main cluster of economic activity of the investing firm (Construction, Consumer Goods, Creative Industries, Energy, Environmental Technology, Financial Services, Food, Beverages & Tobacco, ICT & Electronics, Industrial, Life Sciences, Physical Sciences, Professional Services, Tourism, Transport Equipment, Transportation Warehousing & Storage, Wood, Apparel & Related Products). Sectoral dummies erode the effects of the log capital investment and of the share of Italian investments, but otherwise confirm the main results. Relative to the reference Industrial sector, the results indicate that the immigrants' effects are largest and most highly significant for Creative Industries. They are also large and significant at the 10% level for ICT & Electronics, and Wood, Apparel & Related Products. This could be seen as an indication that firms that rely more on intangible and firm-specific knowledge rely more on the immigrants' effects to facilitate the knowledge transfer with the target location.

Overall, the results of our second-stage regressions strongly support the information effect of immigrants, in line with the interpretation that firms with less knowledge of the Italian context, originating from more culturally distant countries and whose activities have greater information content rely more on the immigrants' effects.

Clearly, a corresponding analysis of the emigrants' effects would be pointless due to the insignificant mean and standard deviation of its coefficient reported in Table 10.

4.3 Robustness checks

A primary concern about the robustness of the estimates may derive directly from our descriptive statistics and from the inspection of Table 9 above. Indeed, the high concentration of investment ventures in the province of Milan, and, to a lesser extent, of Rome, leads to a disproportionately higher probability to choose these destinations for an investment venture that seeks to locate into Italy. Furthermore, our model implicitly assumes that the investor has chosen Italy as a destination, and that, conditional on this decision, chooses one among the Italian provinces. In reality, however, depending on the activity that the investment is going to implement, the alternative may not be located in Italy; for instance, an investment in an automotive R&D centre may consider as alternative locations Milan, Munich, and Paris. The extent to which our estimates of the immigration and emigration elasticities are driven by the prominence of Milan and Rome as destination provinces needs to be addressed.

To investigate this issue, in the upper part of Table 12, we report the results of our estimates conducted on the subsample of our data where all investments with Milan and Rome as destination were excluded. This reduces the numerosity of our sample by about one half, but the main results are confirmed. The positive, significant, and significantly heterogeneous effect of immigration is robust to the exclusion of Milan and Rome.

In spite of the robustness of our results to different specifications of the mixed logit models and of the random parameters, a second potential source of concern is that the estimated migrants' effects may be due to a reverse causality running from FDI to migration. To address this issue, we follow previous studies (Bratti et al., 2014; Javorcik et al., 2011; Briant et al., 2014, e.g.) and 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. Data availability imposes a slight difference between the immigration and emigration sides in constructing the instrument. As to 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). Following their approach (for more details, see Bratti et al., 2014, p. 580–585), 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. These shares are then used as weights to impute the province-level distribution of the overall nation-wide stocks of immigrants from each country in the 2003-2015 period. This allows reflecting both a "push" factor from the side of the origin country and a "recursive" factor (due to the fact that there is strong path dependence in the location of immigrant communities in specific provinces) in imputing the distribution of immigrants; this measure can however be considered as unrelated with the current attractiveness of the provinces to FDI (cfr. also Burchardi et al., 2018). We construct the imputed emigration variable in a similar way, but in this case the available data on

Table 12: Robustness checks

	Means	**SD**
Sample without Rome and Milan		
Log Immigrants$_{ijt-1}$	0·349***	0·439***
	(0·088)	(0·083)
Log Emigrants$_{ijt-1}$	0·053	−0·017
	(0·073)	(0·114)
Migration variables imputed based on pre-determined distributions		
Log Immigrants$^{\text{imputed}}_{ijt-1}$	0·207***	0·181***
	(0·055)	(0·061)
Log Emigrants$^{\text{imputed}}_{ijt-1}$	−0·015	0·306**
	(0·106)	(0·145)
2-year lags in migration variables		
Log Immigrants$_{ijt-2}$	0·308***	0·437***
	(0·097)	(0·110)
Log Emigrants$_{ijt-2}$	0·014	0·007
	(0·075)	(0·143)
Immigrants by qualification levels at NUTS2 level		
Log Immigrants$_{ijt-1} \times s^{\text{hq}}_{rjt-1}$	0·540***	0·302***
	(0·205)	(0·117)
Log Immigrants$_{ijt-1} \times s^{\text{lq}}_{rjt-1}$	−0·198	0·318***
	(0·196)	(0·099)
Log Emigrants$_{ijt-1}$	0·099	−0·025
	(0·076)	(0·124)
Dummy for Centre-South included		
Log Immigrants$_{ijt-1}$	0·315***	0·412***
	(0·091)	(0·085)
Log Emigrants$_{ijt-1}$	0·054	−0·049
	(0·074)	(0·145)

Mixed logit estimates. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The list of regressors, not reported for brevity, includes: Log Prov. GDP, Log Prov. Population, Log average wage (region), Prov. unemployment rate, Log Patent Count, Share tertiary educated, Infrastructure endowment, Common border, Log Distance, Institutional Quality, Agglomeration (Sector), Sectoral diversity, Log Imports, Log Exports, Pre-2002 FDI stock, Parent Colocation, Log Total Immigrants, Log Total Emigrants. Based on the results of the model in Table 10, besides the Log of GDP, the Log of Immigrants and the Log of emigrants and their spatial lags, the list of random parameters also includes Log Distance, Rome-Milan, Pre-2002 FDI stock, Log Total Immigrants.

Table 13: Estimation results - Mixed Logit - Geographic spillovers

Dep. var: Choice	**Means**	**SD**
Log Prov. GDP	1·797***	−0·170
	(0·472)	(0·524)
Log pr. GDP <50km	0·101***	−0·004
	(0·039)	(0·039)
Log pr. GDP 50-100km	0·228	−0·114
	(0·173)	(0·170)
Log pr. GDP 100-200km	0·475*	−0·084
	(0·245)	(0·168)
Log pr. GDP >200km	0·251	0·210
	(0·861)	(0·904)
Log Immigrants	0·426***	0·465***
	(0·122)	(0·140)
Log Imm. <50km	−0·127**	0·013
	(0·059)	(0·059)
Log Imm. 50-100km	0·068	0·106
	(0·126)	(0·136)
Log Imm. 100-200km	−0·277*	0·019
	(0·165)	(0·124)
Log Imm. >200km	0·164	−0·694
	(0·597)	(0·721)
Log Emigrants	0·051	−0·107
	(0·081)	(0·195)
Log Em. <50km	−0·006	−0·015
	(0·046)	(0·035)
Log Em. 50-100km	−0·085	−0·288***
	(0·089)	(0·092)
Log Em. 100-200km	−0·012	0·085
	(0·115)	(0·144)
Log Em. >200km	−0·816	−0·287
	(0·512)	(0·852)

Mixed logit estimates. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The list of regressors, not reported for brevity, includes: Log Prov. GDP, Log Prov. Population, Log average wage (region), Prov. unemployment rate, Log Patent Count, Share tertiary educated, Infrastructure endowment, Common border, Log Distance, Institutional Quality, Agglomeration (Sector), Sectoral diversity, Log Imports, Log Exports, Pre-2002 FDI stock, Parent Colocation, Log Total Immigrants, Log Total Emigrants. Based on the results of the model in Table 10, besides the Log of GDP, the Log of Immigrants and the Log of emigrants and their spatial lags, the list of random parameters also includes Log Distance, Rome-Milan, Pre-2002 FDI stock, Log Total Immigrants.

past emigration by province and country of destination are yearly flows of residential cancellations rather than stocks. A potential problem arising from these data is that specific destination countries are given a zero weight due to yearly fluctuations in the emigration data. To address this, we aggregate the 1995-1999 distributions of emigrants outflows by provinces of origin and countries of destination and take these aggregate flows as the base for computing weights, i.e., the ratio of emigration flows between each country-province pair over total emigration flows in 1995-1999.[15] The 2003-2015 emigration data are then similarly imputed by multiplying this weight by the overall nation-wide stocks of emigrants from any provinces to a specific country. Again, this approach allows considering the "pull" factor of the foreign countries for the emigrants as well as the "recursive" factor suggesting that people from the same province may locate in the same country where other residents in the same province have moved, but is arguably unrelated to the current attractiveness of provinces to investments. As a first attempt to address endogeneity concerns, we include the imputed immigration and emigration stocks directly in our mixed logit, whose results are reported in Table 12. The results support the main findings and in particular the positive and significant, and significantly heterogeneous, effect of immigration on the location of MNE activities[16]. To address reverse causality further, we also report the results of specifications where the regressors are lagged 2 years. The results are in line with the previous findings.

Our fourth set of robustness checks refers to the literature that highlighted an effect for migrants' 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, detailed yearly data on immigrants and emigrants by province, country of origin and level of education are not available. This kind of information is however available from the 2011 Census at the NUTS2, rather than NUTS3 level, for immigrants only. Based on this information, we can approximate the shares of bilateral immigrants for each level of educational attainment as follows. We compute the shares of immigrants by level of educational attainment in the NUTS2 regions and then multiply by the immigrants' stock by province. The log of this stock is a measure of the stock of these bilateral immigrants by province and level of education ("high-skilled" corresponding to the share of bilateral immigrants with high-school and tertiary

[15] 1995 is the first year when the administrative boundaries of the provinces correspond to the current ones (with the exception of the new provinces created after 2005).

[16] We are aware that this approach does not deliver consistent estimates and we view it mainly as a first step towards fully addressing endogeneity. We have attempted to use our imputed migration variables as instruments in a Generalized Structural Equation Modelling - Correlated Random Effects (GSEM - CRE) approach without reaching convergence in the model.

education at the NUTS2 level, and "low-skilled" corresponding to the corresponding share of bilateral immigrants with primary and lower-secondary education). While 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. In Table 12, we include the results of a set of mixed logit estimates where we substitute the log of the immigrants' stock with these measures. In line with the findings of previous literature (e.g. Javorcik et al., 2011), the effect of bilateral immigrants results positive and highly significant in regions where the share of highly skilled immigrants from the same country is greater. The effect of immigrants in regions with higher shares of low-skilled immigrants is instead insignificant. Both effects result significantly heterogeneous, confirming that the heterogeneity in the immigrants' effects is not exhausted by the immigrants' skills.

As a further robustness check, the lower panel of Table 12 reports results of a specification where we include a dummy for regions located in the Centre-South of Italy. Also in this case the results are in line with the main findings.

Finally, the presence of geographic spillover effects of migration may also represent a source of bias to our estimates. As argued by Bratti et al. (2014) with regards to the migration-trade link, omitted variables operating at a wider geographic scale than the provinces which are correlated with the included regressors may induce correlation across the errors of observations related to different provinces. 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, in Table 13 we augment our specification 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 to the same country of destination but originating from different provinces and include spatial spillover effects for GDP as well, as a proxy for market access.

The results reported in Table 13 confirm the picture sketched so far. The importance of market access in the location choice of FDI is confirmed by the positive and significant effect of the GDP in neighbouring provinces. As to immigrants, their positive, significant and significantly heterogeneous effect in the destination province is confirmed. Furthermore, the effect turns out to be highly localized and some patterns of competition among provinces appear, especially for provinces located within a radius of 50 km. No role is still detectable for emigrants on average, but the heterogeneity in its effect is highly significant in the radius range 50-100 km, which is about the size of an average NUTS2 region. These results support the interpretation that the location choice of FDI operates at a very fine-grained scale and that the choice of the NUTS3 level as a unit of analysis is appropriate[17].

Overall, our results strongly support a significant, positive and significantly heterogeneous effect of immigration on the location choice of FDI but no robust effect for emigration. In spite of the robustness of the results, though, we can still not entirely rule out that our results are driven by omitted variables that affect both the location of the FDI and the stocks of immigrants and emigrants. Yet, a number of variables proxying for the dynamism of the local systems and their labour markets are already included (e.g. unemployment rates, wages, share of tertiary degrees, dummy for Milan and Rome) and they only limitedly affect the immigration coefficients, while they more significantly reduced the emigration coefficients. Furthermore, they often did not result very significant. Overall, these considerations are reassuring that the regressors that we included are controlling for most of the unobserved factors that may bias our results.

5 Discussion and conclusions

This paper investigates the link between migration and FDI. Our results, based on 1,147 greenfield investment projects made by 895 MNEs into Italian provinces over the 2003-2015 period, confirm a positive, significant and robust effect of immigration on FDI, in line with the previous literature, but no robust emigrants' effect. Beyond this average effect lies significant heterogeneity. We contribute to the literature by providing an in-depth account of the heterogeneity in the immigrants' effects on inward FDI, by looking at the activity in which FDI take place and the characteristics of the investing firms. This allows us to shed light on the mechanisms underlying this relation.

The extant literature has highlighted that immigrants can attract MNE investments from their home countries through three mechanisms: a labour market effect, a demand effect and an information effect.

Our results are consistent with an important role of demand and information channels, but not with an effect through the labour market. On the one hand, immigrants are

[17] Admittedly, the leading role of Rome and Milan as attractors of FDI may drive these results. Re-estimating this specification excluding Rome and Milan, the crucial result of a localised effect of immigrants is confirmed, whereas the spatial spillovers from immigrants in neighbouring regions loose statistical significance. The results are available upon request to the authors.

not a factor that attracts more FDI in manufacturing facilities in Italian provinces. Since these investments are typically more labour intensive than other types of investments, we conclude that MNE do not follow migrants from their home country, to employ them in their overseas manufacturing plants. Furthermore, the elasticity of inward FDI to the stock of immigrants is lower for MNE making relatively more labour-intensive investments. Altogether, these findings do not support a role for an effect of migration on inward FDI through the labour market.

On the other hand, the strongly positive and significant effect of immigrants detected for R&D, market-access and business services FDI points to the co-existence of an information channel, in line with the extant interpretations of the immigrants' effects, with a demand channel. The latter would support the interpretation that the immigrants effect operates in terms of creating a market for home country products, in line with what the trade literature calls a "preference effect" or "transplanted home-bias effect". While the demand and the information effects may clearly co-exist, our analysis of the geographic spillover effects of immigrants suggest that the information effects prevail. Indeed, we find that the effects are highly geographically localised within the immigrants' province of residence. If the immigrants' demand was the main driver of such effect, we should expect positive, even if weaker, effects of immigrants in neighbouring provinces. Instead, larger immigrants' stocks in neighbouring provinces appear to displace the location of FDI away from a focal province onto its neighbours, which seems more consistent with the prevalence of a localized information effect. Moreover, in accordance with the previous literature, we find that the effect is mainly to be attributed to skilled immigrants. This is also more likely to reflect an information rather than a demand effect.

The information mechanism is also supported by our findings on firm-level heterogeneity. In fact, we show that the immigrants' effect is stronger for firms that are investing in Italy for the first time, for less internationalised firms, that devote smaller shares of their investment portfolios to the Italian market, and for firms coming from more culturally distant countries and from East Asia in particular. These results strongly suggest that firms rely on the immigrants' effects when they have less knowledge of the Italian market. On the whole, our results support the interpretation that immigrants effectively contribute to reducing the investors' "liability of foreignness" in international business.

It seems interesting to note that the effects of immigration are robust and strongest for investments that could be considered to target intangible assets[18]. Studying the

[18] "Intangible assets consist of the stock of immaterial resources that enter the production process and are

sectoral sources of heterogeneity supports this interpretation, as the effects are larger for firms operating in the Creative Industries and ICT & Electronics. These investments are strongly tied to the firm-specific human capital, organisational resources and capabilities, and corporate culture of the firm (Arrighetti et al., 2014). In the global sourcing of services, the role of firm specificities has been argued to lead MNEs to locate different stages of the value chain close to each other in order to save on coordination costs and to benefit from complementarities (Markusen, 2006; Defever, 2006; Nefussi and Schwellnus, 2010; Castellani and Lavoratori, 2019), as also confirmed in our results by the strong role of firm co-location. The strong role of immigrants in driving the location choice of FDI provides indirect support to the interpretation that the culture of the origin country affects corporate culture of the MNEs and that divergence in cultural approaches between the source and the destination economy creates barriers to international investment (Head et al., 1995; Nefussi and Schwellnus, 2010). Indeed, several cultural barriers may intervene in hampering access and exploitation of the location-specific technological know-how, and of the market knowledge required to develop the MNE brand in the new market. According to our results, immigrants, especially skilled immigrants, are well suited to mediate between the culture-related organizational routines of the MNE and the knowledge assets available in the location. Our results essentially complement previous findings about the role of cultural factors in reducing information frictions that lead firms to locate where other firms from the same countries have already located (e.g. Head et al., 1995).

Overall, our paper bears significant implications for policymaking that partially contrast with the current public discourse on immigration. Indeed, beyond the usual interpretation of migrants as a burden to public welfare, our results suggest that immigrants can play a significant role in attracting FDI in narrowly defined geographies. To the extent that more FDI contribute, through spillover effects, to diffusion of knowledge and innovation locally and nationally, migrants can play a key role for local economic development. Comparatively low-cost measures to promote the migrants' information effects by simplifying the communication channels to their homeland could actually be seen as supporting the activities of regional investment promotion agencies.

necessary for the creation and sale of new or improved products and processes. They include both internally produced assets—e.g., designs, blueprints, brand equity, in-house software, and construction projects—and assets acquired externally—e.g., technology licenses, patents and copy-rights, and the economic competencies acquired through purchases of management and consulting services." (Arrighetti et al., 2014, p.202). Hence, foreign direct investments in R&D, Design Development and Testing, but also in activities oriented to improve the company reputation and branding such as those categorised as Market Access and Business services would fall in this category (see also Montresor et al., 2013).

While providing some new insights on the mechanism underlying the immigrants' effect on the location of FDI, our approach still has some limitations. First of all, our arguments about the different channels through which migrants may affect FDI rest on the assumption that migrants affect the channels themselves, i.e. labour costs, information costs and demand. The lack of evidence on the labour channel may be due to the fact that, in a rigid labour market such as the Italian one, immigrants are not significantly affecting labour costs in the formal sector. A second limitation of the study is its inability, due to data limitations, to distinguish between horizontal and vertical manufacturing FDI. In principle, the effect of immigrants may be very different in the two cases. Future research addressing this issue may yield insights and contribute to more neatly disentangling the demand effect from the information effect. Data limitations have also hampered our ability to study the emigrants' skills and to explore the localization of the migration effects at an even finer geographical level than the province. These issues may be addressed in future research if the relevant data become available. Finally, future studies may fully address some remaining endogeneity concerns.

References

Alcácer, J., Chung, W., Hawk, A., and Pacheco-de Almeida, G. (2018). Applying random coefficient models to strategy research: Identifying and exploring firm heterogeneous effects. *Strategy Science*, 3(3):533–553.

Amiti, M. and Javorcik, B. (2008). Trade costs and location of foreign firms in China. *Journal of Development Economics*, 85(1-2):129–149.

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118(6):1151–1199.

Arrighetti, A., Landini, F., and Lasagni, A. (2014). Intangible assets and firm heterogeneity: Evidence from Italy. *Research Policy*, 43(1):202–213.

Barry, F., Görg, H., and Strobl, E. (2003). Foreign direct investment, agglomerations, and demonstration effects: An empirical investigation. *Review of World Economics*, 139(4):583–600.

Basile, R. (2004). Acquisition versus greenfield investment: the location of foreign manufacturers in Italy. *Regional Science and Urban Economics*, 34(1):3 – 25.

Basile, R., Castellani, D., and Zanfei, A. (2008). Location choices of multinational firms in Europe: The role of EU cohesion policy. *Journal of International Economics*, 74(2):328–340.

Belderbos, R., Du, H. S., and Slangen, A. (2020). When do firms choose global cities as foreign investment locations within countries? The roles of contextual distance, knowledge intensity, and target-country experience. *Journal of World Business*, 55(1):101022.

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. (1997). Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment. *The American Economic Review*, 87(3):447–465.

Blonigen, B. A. (2001). In search of substitution between foreign production and exports. *Journal of International Economics*, 53(1):81–104.

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.

Buch, C., Kleinert, J., and Toubal, F. (2006). Where enterprises lead, people follow? Links between migration and FDI in Germany. *European Economic Review*, 50:2017–2036.

Buckley, P. and Casson, M. (1981). The optimal timing of a Foreign Direct Investment. *Economic Journal*, 91:75–87.

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

Castellani, D., Jimenez, A., and Zanfei, A. (2013). How remote are R&D labs? Distance factors and international innovative activities. *Journal of International Business Studies*, 44(7):649–675.

Castellani, D. and Lavoratori, K. (2019). The lab and the plant: offshore r&d and co-location with production activities. *Journal of International Business Studies*.

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

Chung, W. and Alcácer, J. (2002). Knowledge seeking and location choice of foreign direct investment in the United States. *Management Science*, 48(12):1534–1554.

Coughlin, C. C. and Segev, E. (2000). Location Determinants of New Foreign-Owned Manufacturing Plants. *Journal of Regional Science*, 40(2):323–351.

Daude, C. and Fratzscher, M. (2008). The pecking order of cross-border investment. *Journal of International Economics*, 74(1):94–119.

De Simone, G. and Manchin, M. (2012). Outward Migration and Inward FDI: Factor Mobility between Eastern and Western Europe. *Review of International Economics*, 20(3):600–615.

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

Defever, F. (2012). The spatial organization of multinational firms. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 45(2):672–697.

Defever, F. and Toubal, F. (2013). Productivity, relationship-specific inputs and the sourcing modes of multinationals. *Journal of Economic Behavior & Organization*, 94:345–357.

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

Dow, D. and Karunaratna, A. (2006). Developing a multidimensional instrument to measure psychic distance stimuli. *Journal of International Business Studies*, 37(5):578–602.

Du, J., Lu, Y., and Tao, Z. (2008). Economic institutions and FDI location choice: Evidence from US multinationals in China. *Journal of Comparative Economics*, 36:412–429.

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

Duranton, G. and Puga, D. (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics*, 57(2):343–370.

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

Etzo, I. and Takaoka, S. (2018). The impact of migration on the cross-border M&A: Some evidence for Japan. *The World Economy*, 41(9):2464–2490.

Felbermayr, G., Grossmann, V., and Kohler, W. (2015). Migration, International Trade, and Capital Formation. *Handbook of the Economics of International Migration*, 1:913–1025.

Flisi, S. and Murat, M. (2011). The hub continent. Immigrant networks, emigrant diasporas and FDI. *The Journal of Socio-Economics*, 40:796–805.

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

Fujita, M. and Thisse, J.-F. (1996). Economics of agglomeration. *Journal of the Japanese and International Economies*, 10(4):339–378.

Gao, T. (2003). Ethnic Chinese networks and international investment: evidence from inward FDI in China. *Journal of Asian Economics*, 14:611–629.

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

Girma, S. and Yu, Z. (2002). The link between immigration and trade: Evidence from the United Kingdom. *Review of World Economics*, 138:115–30.

Goerzen, A., Asmussen, C. G., and Nielsen, B. B. (2013). Global cities and multinational enterprise location strategy. *Journal of International Business Studies*, 44(5):427–450.

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):302–316.

Head, C. K., Ries, J. C., and Swenson, D. L. (1999). Attracting foreign manufacturing: Investment promotion and agglomeration. *Regional Science and Urban Economics*, 29(2):197–218.

Head, K. and Mayer, T. (2004). Market potential and the location of japanese investment in the european union. *Review of Economics and Statistics*, 86(4):959–972.

Head, K., Ries, J., and Swenson, D. (1995). Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States. *Journal of International Economics*, 38(3-4):223–247.

Helpman, E., Melitz, M. J., and Yeaple, S. R. (2004). Export Versus FDI with Heterogeneous Firms. *American Economic Review*, 94(1):300–316.

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

Hornstein, A. S. and Greene, W. H. (2012). Usage of an estimated coefficient as a dependent variable. *Economics Letters*, 116(3):316 – 318.

Horstmann, I. and Markusen, J. (1992). Endogenous market structures in international trade. *Journal of International Economics*, 32:109–129.

Hymer, S. (1976). *The International Operations of National Firms: A Study of Direct Foreign Investment*. MIT Press, Cambridge, MA.

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.

Jayet, H. and Marchal, L. (2016). Migration and FDI: Reconciling the standard trade theory with empirical evidence. *Economic Modelling*, 59:46 – 66.

Kogut, B. and Chang, S. J. (1991). Technological Capabilities and Japanese Foreign Direct Investment in the United States. *The Review of Economics and Statistics*, 73(3):401–413.

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

Markusen, J. (2006). Modelling the Offshoring of White-Collar Services: from Comparative Advantage to the New Theories of Trade and FDI. In Brainard, L. and Collins, S., editors, *Offshoring White-Collar Work ? The Issues and the Implications*. Brookings Institution, Washington.

Marschak, J. (1974). *Binary-Choice Constraints and Random Utility Indicators*, pages 218–239. Springer Netherlands, Dordrecht.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Montresor, S., Perani, G., and Vezzani, A. (2013). How do companies "perceive" their intangibles? New statistical evidence from the Innobarometer 2013. Technical report, European Commission, Joint Research Centre – Institute for Prospective Technological Studies, Luxembourg, Publications Office of the European Union.

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

Nachum, L. (2003). Liability of foreignness in global competition? Financial service affiliates in the city of London. *Strategic Management Journal*, 24(12):1187–1208.

Nefussi, B. and Schwellnus, C. (2010). Does FDI in Manufacturing Cause FDI in Business Services? Evidence from French Firm-Level Data. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 43(1):180–203.

Nielsen, B. B., Asmussen, C. G., and Weatherall, C. D. (2017). The location choice of foreign direct investments: Empirical evidence and methodological challenges. *Journal of World Business*, 52(1):62–82.

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

Openshaw, S. (1983). *The modifiable areal unit problem.* Geo Books, Norwick.

Özden, Ç., Parsons, C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody? the evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, 25(1):12–56.

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):1433–1459.

Rauch, J. (2001). Business and Social Networks in International Trade. *Journal of Economic Literature*, 39(4):1177–1203.

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

Rauch, J. E. and Casella, A. (2003). Overcoming informational barriers to international resource allocation: prices and ties. *The Economic Journal*, 113(484):21–42.

Redding, G. (1995). Overseas Chinese networks: Understanding the enigma. *Long Range Planning*, 28(1):61–69.

Saxonhouse, G. R. (1976). Estimated parameters as dependent variables. *The American Economic Review*, 66(1):178–183.

Spies, J. (2010). Network and border effects: Where do foreign multinationals locate in Germany? *Regional Science and Urban Economics*, 40(1):20–32.

Tadesse, B. and White, R. (2008). Do immigrants counter the effect of cultural distance on trade? Evidence from US State-level exports. *Journal of Socio-Economics*, 37(6):2304–2318.

Tintelnot, F. (2017). Global Production with Export Platforms. *The Quarterly Journal of Economics*, 132(1):157–209.

Tong, S. Y. (2005). Ethnic Networks in FDI and the Impact of Institutional Development. *Review of Development Economics*, 9(4):563–580.

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

Wagner, D., Head, K., and Ries, J. (2002). Immigration and the trade of provinces. *Scottish Journal of Political Economy*, 49(5):507–525.

Weidenbaum, M. L. and Hughes, S. (1996). *The bamboo network: How expatriate Chinese entrepreneurs are creating a new economic superpower in Asia*. Simon and Schuster.

White, R. (2007). Immigrant-trade links, transplanted home bias and network effects. *Applied Economics*, 39(7):839–852.

Zaheer, S. (1995). Overcoming the Liability of Foreignness. *The Academy of Management Journal*, 38(2):341–363.

A Data Appendix

The data we use originate from the linkage of different data sources.

FDI flows data are drawn from the *fDI markets* database, a comprehensive and regularly updated online database of crossborder 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 into Italian provinces for which the destination city was available. These correspond to $1,147$ individual foreign direct investments into 85 Italian provinces (NUTS-3 level) occurred over the 2003-2015 period[19], i.e. a choice set of $97,495$ investment-province couples. 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, i.e. Barletta-Andria-Trani, Fermo and Monza-Brianza were founded in 2009 and are located in Apulia, Marche and Lombardy respectively. 25 provinces were never chosen as an investment location and had therefore to be excluded from the analysis. Of these, six are the newly-founded provinces located in the Centre-South[20]. The new province of Monza-Brianza, instead, was chosen as the destination for 5 investment ventures occurred after 2009, the year of its establishment. Employing the choice set as such would bear the paradoxical implication that Monza-Brianza was among the location options available to investors even in the years before it was constituted. Excluding Monza-Brianza from the alternatives available to the 769 investments occurred before 2010, the choice set reduces to a maximum of $96,726$ feasible alternatives.

Data on the main variables of interest, i.e. immigrant and emigrant stocks, are drawn respectively from the demography unit of the 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. Immigrants' data are available for a panel of 13 years, from 2002 to 2015. The ISTAT data lack information about immigrants originating from Hong Kong, as many of those hold British or, to a lesser extent, Chinese passports. Hong Kong, however, is a significant partner of Italian provinces, being home to 9 FDI occurred over the considered period

[19] 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.

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

(8 before 2010 and 1 after), and the missing data problem applies to immigrants but not to emigrants. Subtracting the 757 alternatives (i.e. $84 \times 8 + 85$) relating to Hong-Kong's choice of Italian provinces, the number of observations available for immigrants reduces to $95,696$. Emigrants' 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[21]. A limitation of both variables is that they refer to the regularly registered residents. Hence, they probably underestimate the actual stocks of both immigrants and emigrants. Notice that, as it 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, 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 were 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 was added to the estimated fixed effects.

We also included a set of control variables:

1. *Sectoral agglomeration.* Considering that agglomeration factors are likely to play an attractive role for FDI, we matched the sector of the investment with the corresponding agglomeration in each province. The province-level measures of agglomeration have been calculated based on the AIDA database, that includes the firms registered in Italy above a given turnover threshold.[22] Data cover the 2002-2014 period. Among the different measures of sectoral agglomeration (count of firms per sector, agglomeration based on value added, sales revenues, or employment) we opted to compute our agglomeration index using the count of firms in each sector and province due to the partial availabilty of the other variables. 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

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

[22]The version we use of AIDA is the largest available, the so-called "full" one, which covers firms above a fairly low turnover threshold (one million Euros).

NACE classification[23] could not be applied as such and the match was done manually. It is worth noticing 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 loss of $2,955$ observations.

In order to assess the relevance of Jacobian externalities in the location choices of FDI, we used 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 standardized in the empirical analysis.

2. *Bilateral (province-country) controls: FDI stocks up to 1997, 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 the same country into the same province between 1985 and 1997. Trade flows data are drawn from Italian international trade data publicly available at the province-country pair level (`https://www.coeweb.it`). Because the data downloading is an extremely time-consuming manual process, we opted to exclude minor remote islands from the analysis, a choice which did not affect the quality of the merge with the FDI data. The data cover both import and export flows over the 2002-2015 period; trade between specific country-province pairs is zero in 692 cases for the import data, and in 164 cases for the export data, leading to a corresponding reduction in the sample size for these variables when taking the log.

   The distances are calculated as great circle distances as in Bratti et al. (2014) based

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

on latitude and longitude (in decimal degrees) of provinces and partner countries[24]. A dummy variable for common border is equal to 1 if the province of destination is located in a region that is bordering the country of origin of the FDI, and to 0 otherwise.

3. *Province-level controls: population, per capita value added, aggregate value added, infrastructure endowment, education, count of patent applications per province, average wage, unemployment rate, share of residents with a tertiary degree.* As for per capita value added and 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[25]. The data about the resident population over the 2002-2015 period are drawn from the demography unit of the ISTAT[26]. Annual data on GDP and population are available, respectively, until 2014 and 2015. Instead, reportedly due to its limited time variation, the infrastructural endowment 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, however, could be performed for the province of Monza-Brianza, as the information about infrastructure endowment is not available for the new provinces even in the post-2009 years, which reduces the sample size by 378 observations. As regards the province-level shares of residents with a tertiary degree, meant to proxy for the human capital available in the province, the data are drawn from the 2011 Census and are publicly available from the ISTAT at `http://dati-censimentopopolazione.istat.it`. Tertiary education data were standardised in the empirical analysis.

   To add a measure of the R&D intensity of the province, the 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

[24]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`.

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

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

available at the province level for later years) and extrapolated for the later years. The extrapolation was performed similarly as in the other cases: the available patent data were first regressed on time with province fixed effects, then the out-of-sample prediction for the 2013-2015 period was added to the estimated fixed effects.

Finally, we included the province-level unemployment rate (because this statistics is only available for the 2004-2013 period due to changes in the computation rules at ISTAT, but it 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). In addition, annual wage data in Euro originating from the social security data of the Work Histories Italian Panel (WHIP) (Bena et al., 2012) were averaged by NUTS2 region to get a proxy for the regional (unfortunately not province-level due to limited information on firm location) labour costs.

4. *Co-location.* Recent studies (e.g. Defever, 2006; Castellani and Lavoratori, 2019) highlight the positive effect on locational choice of previous investments of the same parent company in a given province. Due to the limited number of observations in our data, we are unable to disentangle the function of the previous investment, nor to detail the number of previous investments; hence, we opted to construct a binary variable for co-location that is equal to 1 in the case that the same parent company has already invested in the same province, and zero otherwise.

   Overall, the observations for which we have complete information about our variables amount to $91,502$ excluding trade-related variables, and to $90,915$ including them.