



Doctoral School
of Social Sciences



Doctoral program in Development Economics and Local Systems

Curriculum in Development Economics

International Investments Flows:

The Role of Cultural Preferences and Migrants Networks

a dissertation submitted in partial fulfillment of the requirements for the Doctoral degree
(Ph.D.) in Development Economics and Local Systems

**Filippo
Santi**

XXXIth cycle
Academic year 2017/18

Supervisor: Giorgia Giovannetti

Doctoral Committee:

- Prof. Christine Oughton (SOAS University of London)
- Prof. Noemi Pace (Ca' Foscari University of Venice)
- Prof. Martina Menon (University of Verona)

Contents

Introduction	7
1 Asymmetric Cultural Preferences and Greenfield FDI	17
1.1 Introduction	17
1.1.1 Related literature	20
1.2 Asymmetric cultural proximity	21
1.2.1 Bilateral cultural trade as a proxy for CP	23
1.2.2 A detour on asymmetry	24
1.3 Econometric framework	28
1.3.1 Asymmetric CP and FDI gravity models	28
1.3.2 Baseline estimation, identification strategy and data	30
1.4 Results	32
1.4.1 Baseline results	33
1.4.2 Robustness checks	34
1.5 Extensions	42
1.5.1 Destination-side mechanisms	43
1.5.2 Heterogeneous impact of the asymmetric and time-dependent dimension of CP	46
1.6 Conclusions	47
Appendices	52
1.A Data: sources and general features	52
1.B Cultural trade as a proxy of the symmetric component of CP	57
1.C Extensions to the detour on asymmetry	60
1.D Further addressing the omitted variable bias	64
1.E Relevance of the instruments	67
1.F Further Robustness Tests	68
1.G Country excluded from the dataset	70
2 A network analysis of the migration-Investment nexus	72
2.1 Introduction	72
2.2 Literature Review	74
2.2.1 Migrants Networks and FDI: the Bilateral approach	75
2.2.2 Migrants (proper) Networks and FDI: the “complex” approach	77
2.3 Methodological Note	79
2.3.1 What is a network? Basic terminology and representation	79
2.3.2 From local to structural properties of nodes: using networks to analyze actors	81
2.3.3 Networks in Gravity	85
2.4 Topology and Descriptive Statistics	89
2.4.1 Data	89

2.4.2	Migrants Network and Greenfield FDI: a Network Description	91
2.5	Empirical Analysis	95
2.6	Conclusions	100
Appendices		109
2.A	Terminology Appendix	109
2.B	Mixed models vs FE gravity	110
2.C	Descriptive and Topological analysis Appendix	112
2.D	Estimates Appendix	115
3	The Heterogenous Impacts of Cultural Preferences on Bilateral M&A flows	123
3.1	Introduction	124
3.2	Cultural Proximity and Bilateral Economic Exchanges	125
3.3	The model	129
3.3.1	The Head and Ries model	130
3.3.2	Extending the model	133
3.4	Empirical Framework	135
3.4.1	Data Description	137
3.5	Results	138
3.6	Robustness Checks	141
3.6.1	Sensitivity to the Exclusion of Null Flows	142
3.6.2	Stability/Sensibility of the Symmetric Measures of Cultural Proximity . .	144
3.7	Conclusions	145
Appendices		150
3.A	Data Appendix	150
3.B	Strength and Weaknesses of Mean-Value Gravity Estimators	152
3.C	Extended Tables of Results	153
3.D	Potential non-linearities	154

Abstract

Foreign Direct Investments are the most complex form of internationalization. A large part of the recent international trade literature has focused on their determinants on the ground that they spur growth and have a positive impact on development. This thesis examines FDI along two different and understudied lines. The first line of research focuses on cultural factors promoting bilateral investments flows. In chapter 1 and chapter 3, I propose a novel definition of Cultural Proximity which separates the effect of cultural similarity from the role of perceptions and cultural affinity. I am able to innovate with respect to the existing literature by capturing the effect of time varying and possibly asymmetric patterns in the reciprocal cultural appreciation between two countries. In Chapter 1 I explicitly deal with the potential asymmetry in bilateral cultural appreciation, and test for the emergence of non reciprocal cultural patterns in the analysis of bilateral Greenfield FDI. An example clarifies what I mean: consider South Korea and Latin America. The so called Korean Wave, consisting of soap operas and Korean pop music has become extremely popular in Latin America since the mid 2000s, despite of geographic and cultural distance in terms of language and ethnicity. Yet, there is no evidence of a symmetric rise in popularity of Latin American culture in South Korea. The underlying idea is that the "new" positive perception of Korea enhances bilateral (trade and) FDI. In Chapter 3 I highlight the heterogeneity of FDI and the non-linearities that could emerge in the relationship between cultural affinity and bilateral M&A. In the empirical exercise, I use an econometric model that allows me to disentangle the impact of the different level on M&A. The second line of research explores the role of migrants' flows on bilateral FDI. Borrowing the tools from social network analysis, in Chapter 2 I investigate whether and how the position of a country in the International Migration Network affects a country's bilateral investment flows beyond the direct role of its local emi(immigrant)-grant population. The empirical application is on Greenfield FDI.

Introduction

Foreign Direct Investments (FDI) grew relentlessly in the last decade, even during the financial turmoil generated by the 2008 global financial crisis (UNCTAD, 2017). Despite such trend slightly reversed in 2017, their growth appears to be only temporarily suspended (UNCTAD, 2018). FDI constitute one of the most complex (and debated) mode of firm internationalization. Such complexity translates in the substantial heterogeneity of the existing economic literature.¹

This thesis contributes to the existing literature by focusing on the cultural determinants of bilateral investments decisions, adopting a macroeconomic partial equilibrium perspective. Understanding the factors promoting FDI is a matter of the utter importance for many countries worldwide, whose economic policies must face the increasing competition from abroad and the rising pressures from domestic firms which often demand for more protection or to relocate segments of the production abroad.

While the analysis of the institutional and political conditions on bilateral investments has been a object of enormous interest in the past, the role of different types of promoting factors gained substantial interest only in the last 10-15 years. In this thesis I focus on two specific drivers of bilateral investments. In Chapters 1 and 3 I explore the concept of Cultural Proximity (CP), and its role in promoting bilateral FDI. I begin by discussing the idea of CP and the limitations of the mainstream approach, which does not acknowledge the possibility that different countries might perceive differently the cultural aspects that classify them as similar. Then, I propose an empirical definition of CP which encompass the role of such subjectively perceived affinity as a distinct factor complementing the idea of CP in terms of similarity between cultures. I empirically test the implications of such an augmented definition of CP on both Greenfield FDI (Chapter 1) and on Mergers and Acquisitions (Chapter 3). Even though both chapters hinge on CP, they tackle different conceptual and methodological issues, to shed light on different aspects of the Investments-CP relationship. On the one hand, Chapter 1 insists on the properties of affinity itself, thanks to the use of trade flows in cultural goods as a proxy for the asymmetric and time varying component of CP. On the other hand, Chapter 3 focuses on the characteristics of the distribution of bilateral FDI data, to investigate how the asymmetric component introduced in Chapter 1 responds to the heterogeneity that derive from the investment distribution at world

¹Beyond the division between different sides of the existing debate, it is important to notice the pluralism that characterize the analysis of both the determinants and impacts of international investments worldwide: international business scholars, as much as international economists and economic geographers all contributed to shed light on the phenomenon as it is understood today, and the cross-field contributions are rapidly increasing (as they are in many other fields of research).

level. Differently, Chapter 2 detach from the concept of CP, to investigate the role of Migrants' Networks as a determinant of bilateral greenfield FDI. The literature insists on a substantially positive effect of a country's migrant population on FDI. Indeed, migrants are credited to favor the flow of information and to dispose of the right type of social capital which is necessary to make international investments profitable. However, the strict bilateral approach that is usually adopted in the literature is at odds with the fact that emigrants might not only maintain relationships with their motherland, but could also establish preferential connections with fellow nationals migrated to different destinations. Using a complex network approach, I investigate whether and how the position of a country in the International Migration Network (IMN) affects its bilateral investment position beyond the direct effect of its local emi(immi)-grant population.

More specifically, in Chapter 1, I explore the relationship between Greenfield FDI and *Explicit Cultural Preferences* (EP Hellmanzik and Schmitz, 2017) as a component of a broader concept of Cultural Proximity. Despite the fact that the notion of CP is not recent, it remained confined to the idea of the existence of some objective similarity between countries. Nonetheless, CP might be also affected by factors other than those defining similarity. The way such similarity is perceived (in terms of affinity and attractiveness) is nonetheless ignored by the largest majority of the literature, and still lack of a clear framing. The main contribution of this chapter lies in the provision of a new conceptual framework for analyzing such an extended concept of Cultural Proximity. The aim is to interpret the mechanisms linking similarity to cultural preferences, and to understand how the two contribute to explain bilateral greenfield investment flows at global level. Thus, I question the idea that the cultural relationship between countries has to be reciprocal, as it is implicit in the choice of the traditional measures of proximity/distance used in most of the existing international economic literature. As a matter of fact, there is no theoretical prior to assume that the intangible factors that concur to determine the concept of cultural proximity (which narrow the burden of otherwise defined remoteness between economic actors) have the same effect in both sides of a bilateral economic exchange. Interestingly, the scarce empirical evidence suggests that CP (and other types of intangible factors) is far from being undirected and symmetric. The definition of CP introduced in the chapter roots in the area of the cognitive traits of remoteness (Bergamo and Pizzi, 2014), of which the cultural component represents a fundamental building block.² In point of fact, considering CP as a cognitive process opens to the possibility for it to be asymmetric and not-objective. If this is the case, it follows that the traditional proxies of similarity cannot be able to capture the true impact of CP on economic exchanges. Limiting CP to the objective similarity between countries would exclude the possibility for a country to implement active (and relatively rapidly adjusting) investment

²As a small caveat to footnote , it is worth noticing here that international economists detach from the other economic disciplines dealing with cultural proximity, though they generally exchange notions and tend to examine similar phenomena from different angles. For what concerns the international economics approach, Bergstrand and Egger (2013) include CP within the class of factors that affect economic exchanges by either introducing or relieving what they define as the “unnatural” frictions to international exchanges (i.e., all those obstacles that do not refer directly to geography and natural impediments in general). Thus, the gravity approach is mostly interested in understanding how factors such as CP affect bilateral flows, and on how do they explain the effects of specific policies on the relationship between countries. In this sense, the concept of cultural proximity and the approach to attrition factors in economics I follow is also related to the economic geography approach, that developed a similar classification scheme (Boschma, 2005).

promotion policies³ based on cultural promotion initiatives (such as those carried on by the Chinese cultural promotion agency among others). Empirically, I assume CP as a composite construct, in which perceived cultural affinity co-determines cultural similarity.⁴ In order to capture the explicit cultural preferences (EP) component, which in turn captures the directed and potentially asymmetric elements of cultural proximity, I use *trade in cultural goods* as a proxy (UNESCO, 2005; UNCTAD, 2010; Disdier et al., 2010). Using trade data has the remarkable advantage of an almost global coverage, but requires in turn some important premises. Indeed, dealing with a notion of Cultural Proximity at global level requires a strong assumption on what should be considered as culturally valuable. The idea of CP was first used in the (relatively homogenous) developed context. For this reason, using international classifications to include developing and transition economies into the frame requires to re-think the definition of culture itself, of the measures that are better able to capture it, and of the channels through which culture can be transmitted and exchanged. On the one hand, the definition of what can be considered as culturally valuable is subject to substantial evolution over time as much as across disciplines. For instance, Capone and Lazeretti (2016) include heritage within the notion of cultural goods, while according to Meigs (1987), food too should be considered the product of a country’s cultural identity. On the other hand, a similar reasoning applies concerning the channels through which cultural products are traded and exchanged, which are also subject to constant evolution. Think for instance to iTunes, Youtube, Reddit, and Spotify: they represent just the tip of the iceberg of a new set of “cultural intermediaries” through which cultural products (mostly in the form of music and video media products) flow across users, and eventually countries. Yet, they are not included in official trade data statistics: ignoring such channels limits the capacity of cultural trade data to properly represent explicit cultural preferences between countries⁵. Therefore, a trade off exists between the adoption of a measure of cultural proximity applicable to a large set of highly heterogeneous countries, and the effective capability of that measure to capture the full extent of the cultural exchanges between them.

The empirical application turns around the investigation of the impact of explicit cultural preferences, and of proximity in general, on Greenfield FDI (Financial, 2017)⁶. Despite the same mechanisms might be valid for any type of economic exchange, I restrict the analysis to Greenfield FDI as they represent the perfect case study to test the extent of the asymmetric effect of EP. Indeed, CP is particularly effective at mitigating informative and trust-related barriers, which are particularly relevant for this type of investment (Harding and Javorcik, 2011; Sula and Willett, 2009). After discussing the issue of reciprocity in the reciprocal cultural appreciation at bilateral level, I discuss how the lack of a theoretical justification to treating CP as symmetric or time invariant finds confirmation in the data, which points in favor of a redefinition of the

³The same reasoning could apply to other types of flow

⁴This formalization opens up interesting policy scenarios, where governments may actively influence foreign perceptions through active trade and promotion policies.

⁵This limitation got increasingly relevant in the recent years, given the substantial increase in the diffusion of streaming platform for cultural products. Thus, to understand the processes of cultural transmission cannot disregard the search for a way to account for or estimate those new channels for cultural products transmission.

⁶Data are accessed in 2015, and include all Greenfield FDI recorded by the data provider for the period 2003-2014, excluding the investments in natural resources.

idea of cultural proximity⁷. The descriptive evidence and the resulting interpretative framework are finally plugged into an existing theoretical model of Greenfield FDI (de Sousa and Lochard, 2011), which is conveniently adapted to include my broader definition of proximity based on the dichotomy similarity/affinity.⁸ The model is estimated by mean of a poisson pseudo-maximum likelihood estimator (PPML) with high dimensional fixed effects (HDFE), able to account for the appropriate set of multilateral resistance terms (Anderson and Van Wincoop, 2003; Yotov et al., 2016; Larch et al., 2017).

The results reject the hypothesis that the culture-driven preference between two countries symmetrically affect bilateral FDI. Not only, they suggest that perceived affinity, as captured by the directed and time varying EP terms, may act as a bridge between otherwise culturally dissimilar countries. In particular, I identify a dominant effect of an investment's destination side cultural perceptions' channel over the origin's side perceived affinity. This finding is ultimately reasonable: nonetheless, the existing literature substantially ignored it. Thus, a firm located in a country might be encouraged to invest where it is more likely to be favourably welcome, either by the potential customers or by the presence of preferential treatments by the government. As the existing related literature on EP ignored so far the possibility for investments to be driven by the consideration of the recipient side, I introduce, discuss, and ultimately test two potential mechanisms that might explain the insurgence of asymmetric patterns in cultural proximity between countries. On the one hand, to capture the destinations' consumers' preference for an investing country, I split investments between those that are likely to explicitly target the destination market, and those which are more likely to constitute a platform for re-exporting production. I expect the destination's cultural preference for a country to be more relevant for the formers than for the latters. On the other hand, to capture the effect of the political economy adopted by a potential recipient country, I split the sample between countries with high government accountability and countries with lesser accountability. A government that is accountable in front of its citizen will be more inclined to fulfil its requests/aspiration. For this reason, cultural preferences at destination should be more relevant in presence of highly accountable governments. The analysis conducted on both mechanisms, renamed *Destination Consumers' Demand* channel and the *Destination Political Economy* channel respectively, validate the conceptual framework proposed in this paper, and demonstrate the importance of dealing with cultural preferences alongside the usual measures of cultural similarity.

Overall, the notion of cultural proximity defined and adopted in Chapter 1 allows us to understand more accurately the effect of culture and of cultural preferences on greenfield FDI. Nonetheless, other factors concur to reduce informative barriers and other intangible frictions between countries, facilitating in this way the decision of a firm to invest in a specific destination. In chapter 2 I tackle one of them investigating the role of migrants' networks on greenfield FDI. According to the economic literature, migrants can facilitate bilateral economic exchanges

⁷A similar conclusion is achieved by Boschma et al. (2016), who studied the role of asymmetry in various forms of proximity measure, applied to the Italian M&A case.

⁸The model is not redefined and reelaborated. Starting from its implications and its predictions, the final empirical equation is augmented consistently.

in many different ways. For instance, diaspora increase trade by fostering the foreign demand for goods produced in the country of origin or by increasing the demand for goods emigrants bring back home (White, 2007; Peri and Requena-Silvente, 2010; Giovannetti and Lanati, 2016). Similarly, migrants can also be beneficial for bilateral FDI: think for instance of their role in lifting the informative barriers that prevent entrepreneurs from taking advantage of opportunities abroad (Flisi and Murat, 2011; De Simone and Manchin, 2012). Not last, migrants favor cultural assimilation and can affect the distance that is perceived across countries, by signalling the national predisposition toward their host country, or by acting as intermediaries with their homeland productive system. Despite the fact that the relationship between international migration and economic exchanges has already been extensively investigated by the economic literature (especially at bilateral level), some aspects still remain to be fully understood. For instance, the way the economic performance of a country is affected by the global set of interactions its emigrant community maintains at global level (and that determine the position of a country within the international migration network) is still far from being clear. In this sense, the fact that emigrants communities tend to be highly transnational make this lack of clarity particularly relevant. In facts, diaspora are transnational in the sense that they tend to develop connections with the respective homelands as well as with their fellow nationals located elsewhere. In addition, migrants communities in host countries tend to integrate with each other, often overcoming national identity (Docquier and Lodigiani, 2010; Rapoport, 2018). Nonetheless, most of the existing literature linking FDI and Migration is still in a bilateral perspective (Bertoli and Moraga, 2013). In chapter 2 I abandon this pure bilateral approach by considering international migration as a *complex network* of global relationships. This means to shift the focus from the single bilateral channel toward considering how the position of a country in global international migration network (IMN) affects its economic relationships. The existing literature on migration and FDI as complex networks is quite limited: with the sole exception of Garas et al. (2106), most of the existing related studies focus on trade (see Fagiolo and Mastrorillo, 2014; Schiavo et al., 2010; Sgrignoli et al., 2015, among others). From an operational viewpoint, I approach migration and FDI flows as two layers of the same global economic network, to study how the structure of the migrants' network and the position of a country within it affects bilateral FDI flows. The strength of this approach consists of the possibility to analyze the direct, as much as the indirect effects of international migration on FDI (thus, taking into consideration the potential transnational effect of a country's migrant community). The chapter contributes to different aspects of the existing literature. First, it constitutes the first attempt to investigate the effect of skilled migration as captured by the structure of its network. The decision to focus on skilled migration as opposed to overall migration recognizes the relevance of the skill composition of the migrants community. Indeed, the economic impacts of migration flows in both the host and the home country is highly heterogeneous, and depends on the skill composition of the migrant community. While this evidence is widely accepted in the non-network based literature on the economic impacts of migration (see for instance Docquier and Lodigiani, 2010; Peri and Requena-Silvente, 2010; D'Agosto et al., 2013, among others), it fails to find a consistent application in network related applications.⁹ I depart from most of the related network literature by retaining

⁹Despite data availability of migration flows with skill intensity break-down implies a substantial restriction of

the directed nature of both the migrant and the investment networks (and avoiding to fall in what I define as the *undirectdness attribution problem*). This point is particularly interesting, as most of the literature estimates the impact of a country’s position in the network on economic exchanges by mean of a gravity model. However, obtaining consistent estimates of monadic (i.e. related to one country instead of being couple specific) terms is not immediate if the suitable set of FE (to control for multilateral resistance) is included. Existing studies either consider international flows as undirected networks (Fagiolo and Mastrorillo, 2014, among others), or limit the number of fixed effects in the empirical gravity estimation. This strategy may not be appropriate as it fails to consistently account for the multilateral resistance term (Anderson and Van Wincoop, 2003), even though it allows to retain the monadic, directed network characteristics that would otherwise be absorbed by the country-specific multilateral resistance terms (see for instance Garas et al., 2106). As recommended by Head and Mayer (2014), the correct procedure to estimate monadic variables in a conventional gravity framework would be to explicitly model the stochastic and deterministic component of the empirical model. This point leads to the third contribution. Instead of estimating a structural gravity equation via fixed effects, as done in the scant related literature, I apply a *multilevel regression model* (Rabe-Hesketh and Skrondal, 2012) with random intercept. This innovative methodology allows greater flexibility in the definition of the stochastic component of the empirical model, and avoids to “flatten” a large portion of variability between observations, as it is the case in fixed effects estimation.¹⁰

The results suggest a strong and consistent role of IMN centrality and prestige of a country on its bilateral greenfield FDI exchanges, and highlight the importance of third party network effects as magnifying factors for bilateral investments. The backbone of the chapter is represented by the empirical section, which is based on three steps: first, I propose an in-depth analysis of the co-evolutionary patterns between the International Migrants Network (IMN) and the Greenfield FDI Network (GFDIN) at world level. The clear patterns highlighted by the graphical comparison of the two networks suggests a positive correlation between the two. In order to explain the stylized facts emerging from the comparative analysis, I develop a conceptual framework to capture the mechanisms through which the position of a country in the migrants network might affect bilateral investments, beyond the established direct migrations channel. To test the hypotheses emerging from the conceptual model, I include a set of IMN-related measures of network centrality in the econometric analysis. Each measure reflects in turn a different aspect of a country’s relevance and position in the global network. In line with the existing evidence on migrants networks and FDI, I detect a positive and significant impact of the migrants’ network on greenfield FDI, with the position in the global network acting as a magnifying factor.

Finally, in the third and final chapter I examine go back to the link between Explicit Cultural

the estimation sample.

¹⁰Innovative refers to the field of application, not to the methodology itself. As a matter of fact, the multilevel regression is well established in many other economic applications, but it is an outsider of the gravity literature. Nonetheless, multilevel regression allows to retain monadic coefficient estimates without resorting to multiple steps estimation procedures, as it is necessary in fixed effects estimation. A two steps-Fixed Effects gravity equation is reported in the appendices, to benchmark the coefficients obtained via multilevel regression. Estimates hold across the different methodologies, scoring in favour of the econometric approach adopted.

Preferences and FDI, focusing this time on bilateral M&A. As a matter of fact, the analysis conducted in the first chapter does not address two main issues. On the one hand, while providing a mechanisms valid for Greenfield FDI, it cannot be immediately generalized. On the other hand, it does not take into consideration the possibility that Cultural Preferences and Cultural Proximity could impact differently different types and size of investment flows. In other words, it does not take into account the role of heterogeneity in the CP-FDI relationship. In Chapter 3 I extend the analysis conducted in the first chapter in both directions. To begin with, I investigate the effects of cultural proximity (as defined in the first chapter) on bilateral M&A, which are characterized by a higher degree of reversibility when compared to greenfields (Barba Navaretti et al., 2006); in addition, according to the data I use, their entity also appears to be smaller (on average).

Relying on (and expanding) the conceptual framework developed and tested in Chapter 1, I expect the asymmetric patterns identified in the EP terms when dealing with Greenfield FDI not to hold for M&A (despite the directed and time varying components of EP might remain significant).¹¹ As far as the methodological approach is concerned, I depart from analyzing the average effect of cultural proximity (and EP) on bilateral FDI (as implicitly done in the first chapter), to explicitly deal with the *high heterogeneity* of bilateral investment flows. Heterogeneous flows imply that their determinants are likely to have differentiated impacts according to a wide set of contingent factors, not last the size and intensity of the single bilateral channel (in terms of both number and value of the aggregate projects) between a country and a given economic partner. Thus, the presence of a large heterogeneity causes the traditional mean-value estimators (ppml or pooled OLS among others) to hide a substantial amount of information. To answer all those questions left open from Chapter 1, I extend a partial equilibrium model of affiliate sales developed by Head and Ries (2008) in two directions. First, I extend their reasoning to the longitudinal case. Then, I also extend the empirical panel gravity equation, to explicitly tackle the overdispersion issue that characterize bilateral M&A data. To do so, I apply a *Censored Quantile Regression* (CQreg) with High Dimensional Fixed Effects (HDFFE) (Powell, 1986; Canay, 2011; Figueiredo et al., 2014) to analyze the extent of the potential heterogeneity in the cultural determinants of M&A, controlling at the same time for the Multilateral Resistance Term by mean of Fixed effects and the large presence of null flows in the lower quantiles of the distribution of bilateral M&A through censoring. In contrast with the results of the first chapter, the findings not only suggest a smaller relevance of the economic and business environment at destination, but also that the asymmetric patterns highlighted in the case of greenfield FDI does not emerge for the case of M&A. This result is robust across estimators, confirming the initial concerns about the universality of the asymmetric impacts of cultural preferences. Despite both directions of EP, defined in terms of perceived affinity between countries, remain statistically significant

¹¹Nonetheless, there might be other mechanisms in play. Indeed M&A are generally looked at with great concern when they aim strategic or traditional companies. Think for instance to the recent (2018) strife between the Italian company Fincantieri and the French government for the acquisition of the majority share of a large French shipyard. Assuming that a greater consideration on the French side might have eased the acquisition (and that such fact applies universally to M&A), the role of the destination side appreciation mechanisms detected in Chapter 1 might play a crucial role for M&A too. This issue remains ultimately an empirical issue to be tested in the data.

both at mean level estimation (obtained via ppml) and across quantiles, there is no statistical evidence of the asymmetric patterns detected in the case of Greenfield. The comparison of the coefficients across quantiles finally suggests that if an asymmetric impact of EP on M&A exists in quantitative terms, it is not statistically relevant. Interestingly, both directions for EP remain significant across quantiles (that is, they are both relevant in quantitative terms), but their discordant trends suggest that the relative importance of either EP channel depends on the size of the bilateral flow between two countries.

Bibliography

- Anderson, James E. and Eric Van Wincoop**, “Gravity with gravitas: a solution to the border puzzle,” *American Economic Review*, 2003, *93* (1), 170–192.
- Barba Navaretti, Giorgio, Anthony J. Venables, and Frank Barry**, *Multinational firms in the World Economy*, Princeton University Press, 2006.
- Bergamo, Gianni and Claudio Pizzi**, “Foreign direct investment and psychic distance: a gravity model approach,” *RIEDS - Rivista Italiana di Economia, Demografia e Statistica - Italian Review of Economics, Demography and Statistics*, 2014, *68* (3-4), 167–174.
- Bergstrand, Jeffrey H. and Peter Egger**, “Gravity Equations and Economic Frictions in the World Economy,” in Daniel Bernhofen, Rod Falvey, David Greenaway, and Udo Kreickemeier, eds., *Palgrave Handbook of International Trade*, London: Palgrave Macmillan UK, 2013, pp. 532–570.
- Bertoli, Simone and Jesús Fernández-Huertas Moraga**, “Multilateral resistance to migration,” *Journal of Development Economics*, 2013, *102*, 79–100.
- Boschma, Ron**, “Proximity and Innovation: A Critical Assessment,” *Regional Studies*, 2005, *39* (1), 61–74.
- , **Emanuela Marrocu, and Raffaele Paci**, “Symmetric and asymmetric effects of proximities. The case of M&A deals in Italy,” *Journal of Economic Geography*, 2016, *16* (2), 505–535.
- Canay, Ivan A.**, “A simple approach to quantile regression for panel data,” *The Econometrics Journal*, 2011, *14* (3), 368–386.
- Capone, Francesco and Luciana Lazzeretti**, “Inter-organisational networks and proximity: an analysis of R&D networks for cultural goods,” *Sinergie*, 2016, *34* (101).
- D’Agosto, Elena, Nazaria Solferino, and Giovanni Tria**, “Does FDI affect migration flows? The role of human capital,” *Migration Letters*, 2013, *10* (3), 359.
- De Simone, Gianfranco and Miriam Manchin**, “Outward migration and inward FDI: factor mobility between eastern and western Europe,” *Review of International Economics*, 2012, *20* (3), 600–615.
- de Sousa, José and Julie Lochard**, “Does the Single Currency Affect Foreign Direct Investment?,” *The Scandinavian Journal of Economics*, 9 2011, *113* (3), 553–578.

- Disdier, Anne-Célia, Silvio H. T. Tai, Lionel Fontagné, and Thierry Mayer**, “Bilateral trade of cultural goods,” *Review of World Economics*, Jan 2010, *145* (4), 575–595.
- Docquier, Frédéric and Elisabetta Lodigiani**, “Skilled migration and business networks,” *Open Economies Review*, 2010, *21* (4), 565–588.
- Fagiolo, Giorgio and Marina Mastrorillo**, “Does human migration affect international trade? A complex-network perspective,” *PloS one*, 2014, *9* (5), e97331.
- Figueiredo, E., L.R. Lima, and G. Schaur**, “When elephants mate do ants get crushed? The WTO impact on trade dispersion,” Technical Report, University of Tennessee at Knoxville 2014.
- Financial, Times**, “FDI Market Database,” online, Financial Times Business Intelligence Division 2017.
- Flisi, Sara and Marina Murat**, “The hub continent. Immigrant networks, emigrant diasporas and FDI,” *The Journal of Socio-Economics*, 2011, *40* (6), 796–805.
- Garas, Antonios, Athanasios Lapatinas, and Konstantinos Poullos**, “The relation between migration and FDI in the OECD from a complex network perspective,” *Advances in Complex Systems*, 2106, *19* (06n07), 1650009.
- Giovannetti, Giorgia and Mauro Lanati**, “Do High-Skill Immigrants trigger High-Quality Trade?,” *The World Economy*, 2016.
- Harding, Torfinn and Beata Smarzynska Javorcik**, “Roll out the red carpet and they will come: Investment promotion and FDI inflows,” *The Economic Journal*, 2011, *121* (557), 1445–1476.
- Head, Keith and John Ries**, “FDI as an outcome of the market for corporate control: Theory and evidence,” *Journal of International Economics*, 2008, *74* (1), 2 – 20.
- **and Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Hellmanzik, Christiane and Martin Schmitz**, “Taking gravity online: The role of virtual proximity in international finance,” *Journal of International Money and Finance*, 2017, *77*, 164 – 179.
- Larch, Mario, Joschka Wanner, Yoto Yotov, and Thomas Zylkin**, “The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects,” 2017. Working Paper.
- Meigs, Anna**, “Food as a cultural construction,” *Food and Foodways*, 1987, *2* (1), 341–357.
- Peri, Giovanni and Francisco Requena-Silvente**, “The trade creation effect of immigrants: evidence from the remarkable case of Spain,” *Canadian Journal of Economics/Revue canadienne d'économique*, 2010, *43* (4), 1433–1459.

- Powell, James L.**, “Censored regression quantiles,” *Journal of Econometrics*, 1986, 32 (1), 143–155.
- Rabe-Hesketh, Sophia and Anders Skrondal**, *Multilevel and longitudinal modeling using Stata*, STATA press, 2012.
- Rapoport, Hillel**, “Diaspora externalities: A view from the South,” Technical Report, World Institute for Development Economic Research (UNU-WIDER) 2018.
- Schiavo, Stefano, Javier Reyes, and Giorgio Fagiolo**, “International trade and financial integration: a weighted network analysis,” *Quantitative Finance*, 2010, 10 (4), 389–399.
- Sgrignoli, Paolo, Rodolfo Metulini, Stefano Schiavo, and Massimo Riccaboni**, “The relation between global migration and trade networks,” *Physica A: Statistical Mechanics and its Applications*, 2015, 417, 245 – 260.
- Sula, Ozan and Thomas D. Willett**, “The reversibility of different types of capital flows to emerging markets,” *Emerging Markets Review*, 2009, 10 (4), 296–310.
- UNCTAD**, *Creative Economy Report. A Feasible Development Option*, UNDP, 2010.
- , *World Investment Report 2017 - Investment and the Digital Economy*, UNDP, 2017.
- , *World Investment Report 2018 - Investment and New Industrial Policies*, UNDP, 2018.
- UNESCO**, “International flows of selected cultural goods and services,” Technical Report, UNESCO Institute for Statistics 2005.
- White, Roger**, “Immigrant-trade links, transplanted home bias and network effects,” *Applied Economics*, 2007, 39 (7), 839–852.
- Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch**, “An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model,” *World Trade Organization, Geneva*, 2016.

Chapter 1

Asymmetric Cultural Preferences and Greenfield FDI

This paper investigates the role of asymmetric cultural proximity (CP) on greenfield foreign direct investment (FDI) from an origin to a destination country. We build a conceptual framework that explicitly accounts for cultural attractiveness as an asymmetric dimension within a broad notion of CP. We revisit the existing origin-side theories of bilateral FDI to derive a gravity equation suited for testing the impact of (i) the attractiveness of destination's culture for citizens in the origin country, and (ii) the attractiveness of origin's culture for individuals in the destination economy. While the role of the former direction of CP is well understood in the literature, we propose new mechanisms to rationalize that of the latter. We use exports and imports of cultural goods to proxy for the two directions of asymmetric and time-dependent CP in the same empirical specification. The econometric analysis confirms a positive role of asymmetric CP as a determinant of Greenfield FDI. Moreover, it suggests a stronger investment effect of the origin's culture attractiveness for the destination country. Finally, it provides support for the mechanisms proposed in the theoretical discussion.

Keywords: *Greenfield FDI, Gravity, ppml, Cultural Preferences.*

1.1 Introduction

The role of foreign direct investment (FDI) in generating net gains for both origin and destination countries is well documented. The growth-enhancing potential of FDI has spurred an in-depth analysis of its determinants. One of the most robust findings pertains to the cultural relationships between the investing and the receiving country: investment from origin to destination is relatively higher if the two countries share similar cultural traits, such as those embedded in language, religion, ethnicity or genetics (see for instance Blonigen and Piger, 2014). However, economically relevant dimensions of cultural relationships go well beyond the symmetric (and largely time-invariant) nature of proxies capturing the extent to which individuals in two coun-

tries speak the same language (Melitz and Toubal, 2014) or share similar genetic traits (Shenkar, 2001; Felbermayr and Toubal, 2010; Tung and Verbeke, 2010). In this sense, the recent study by Melitz and Toubal (2018) on the role of genetics and its impact on international economic exchanges is particularly interesting: the authors find co-ancestry to have a more reliable influence on bilateral trade than other cultural variables: yet, the possibility that cultural preferences shape international exchanges beyond the role of those measures of objective similarity still remains substantially neglected. This leads to the question of whether and how asymmetric (and time-dependent) cultural variables, such as preferences for cultural systems or bilateral trust, play out as determinants of investment patterns. The literature here offers only half of the answer. While the seminal contribution by Guiso et al. (2009) has shown that investment increases if individuals in the investing country trust the citizens of the receiving economy, the potential role of the opposite direction of trust is left unexplored. More generally, we lack a comprehensive assessment of the asymmetric dimensions in bilateral cultural relationships as determinants of FDI. Given the premise that the cultural relationship between two countries, say Kenya and the UK, features a potentially asymmetric element such as the appreciation of each other's cultural systems, it is a fairly safe assumption that the way individuals in Kenya appreciate British culture might be very different from how Kenyan culture is attractive for the UK. It is equally safe to expect that these patterns are likely to change over time. How do these two different and evolving forces affect British FDI in Kenya? Is one more relevant than the other? These are questions that motivate this paper, which represents a first attempt to assess the effect of cultural proximity (CP) on FDI, explicitly accounting for the asymmetric and time-dependent dimensions of CP, based on countries aggregate preferences. To this end we first provide a simple conceptual framework for the notion of CP. By encompassing contributions from international business scholars and economists, we present a workable definition of CP accounting for multiple dimensions of the cultural relationship between two countries. These include symmetric sharing of common cultural traits as well as asymmetric cultural attractiveness. The latter component is allowed to vary over time. I refer to such manifestation of cultural attractiveness as Explicit Cultural Preferences (EP) (Hellmanzik and Schmitz, 2017). The word preference denotes both "the fact that you something or someone more than another thing or person" and "an advantage that is given to a person or a group of people" (McIntosh, 2017). Therefore the term EP is linked to the possibility of an economic actor to signal a positive attitude toward a potential partner, a signal that might in turn imply the possibility for a preferential treatment, beyond the existence of an objective cultural proximity between the two actors¹.

In line with Disdier et al. (2010), we use bilateral trade in cultural goods as a proxy for asymmetric and time-dependent CP. Indeed, the value of imports of cultural goods reflects the attractiveness of the exporter's culture for the importer. Moreover, bilateral cultural trade is correlated with standard, symmetric and time-invariant measures of CP, showing the capacity of this proxy to capture all dimensions of CP. We provide some suggestive evidence of the asymmetry embedded in bilateral cultural relationships with a descriptive exercise, conducted on a broad sample of countries. The perspective on cultural asymmetry embedded in cultural trade data differs from

¹The definition of Explicit Cultural Preferences is resumed and expanded in Chapter 2 of this thesis.

and complements the seminal work by Guiso et al. (2009), where data on bilateral trust are analyzed on a sample of European countries. The variation in cultural relationships that can be captured with trade in cultural goods covers both developed and developing countries, an advantage with respect to other asymmetric measures which tend to be confined to EU countries. This is particularly relevant when greenfield FDI is the object of interest, as the scale and scope of South-South greenfield FDI is growing at fast pace (UNCTAD, 2017) and North-South and South-North greenfield has increased their size and relevance.

Equipped with a definition and an empirical measure of CP that account for asymmetry and time variation, we investigate the linkages between CP and greenfield FDI. The paper revisits the theories used in the literature to derive gravity equations of greenfield FDI. These are partial-equilibrium, supply-side models that subsume all gravity forces into monitoring and transaction costs which ultimately determine the investment decisions of the multi national enterprise (MNE). In this context we discuss the role played as determinants of investment decisions of both directions of asymmetric CP, i.e. the attractiveness of the culture in the origin country for individuals in the destination and the attractiveness of destination's culture for the origin. On the one hand, we argue that the cultural attractiveness of the destination country plausibly (and exhaustively) operates via the monitoring-transaction cost channel. On the other hand, the cultural attractiveness of the origin country for the destination is likely to play a role also through other channels. If the FDI project is conducted to serve consumers demand in the destination country (i.e. horizontal FDI), the attractiveness of the origin country's culture for (destination) consumers positively affects the value they put on the output of the origin's MNE and therefore increases the payoff of the FDI project. We denote this mechanisms as 'destination consumers demand' channel. Moreover, the realization of an FDI project can be facilitated (or opposed) by political pressures in the destination country. Under the assumption of political accountability, politicians in the destination country will allocate pressures to facilitate FDI projects also according to the degree by which the culture of the origin countries are attractive for the individuals (voters) in the destination (we call this the 'destination political economy' channel). All in all, the monitoring-transaction costs channels and the 'destination-side' mechanisms unambiguously imply a positive role of both directions of asymmetric CP in determining greenfield FDI from the origin to the destination country. However, the assessment of the relative importance of one direction over the other is an empirical matter.

A structural gravity equation, fully consistent with our theoretical discussion, is brought to the data. The primary source of information on bilateral greenfield FDI is the fDI Market Database, collected by the FDI Intelligence Unit of the Financial Times Ltd. The database contains detailed information on all the greenfield investment projects across more than 150 origin/destination countries for the period 2003-2014. Relying on the Poisson pseudo-maximum likelihood (PPML) estimation technique, our baseline results show a positive and significant effect of asymmetric CP on greenfield FDI. As for the relative importance of each direction of asymmetric CP, our findings suggest that investment projects from an origin to a destination country tend to increase more with the attractiveness of the origin for the destination. More precisely, the elasticities of the number of greenfield investment projects amount to 0.30 and 0.07 for (origin to destination) cultural exports and (origin from destination) imports, respectively. This baseline pattern holds

across a number of alternative specifications, including the addition of source-destination dyadic fixed effects and instrumentation of cultural trade. Moreover, results are robust to the use of total and average value of greenfield FDI as dependent variables and to different approaches in the definition of cultural trade.

Our findings shed new light on the mechanisms linking asymmetric CP and greenfield investment. In particular they suggest a stronger role of the ‘destination-side’ mechanisms. We extend the core analysis of the paper by conducting an empirical test of ‘destination consumers demand’ and the ‘destination political economy’ channels and find supportive evidence for the existence of these mechanisms. We also investigate whether and how the effect of the asymmetric and time-dependent dimension of CP varies at different levels of its symmetric and time-invariant components. We find that time-contingent positive shocks in the asymmetric component of CP increase greenfield FDI only at low levels of the time-invariant, symmetric dimension of CP. This is consistent with a relationship of substitutability between (i) time-contingent, asymmetric and (ii) time-invariant, symmetric dimensions of CP in triggering FDI, with the former operating as a bridgehead between otherwise culturally distant countries.

1.1.1 Related literature

Our paper speaks to the growing literature that considers culture as an important determinant of economic outcomes (see among others Guiso et al., 2006; Fernández, 2008, 2011; Alesina and Giuliano, 2015). We contribute in particular to the debate on whether and how the relationship between cultures affects exchanges and investment patterns across countries (see for instance Head and Mayer, 2014; Giuliano et al., 2014).

To the best of our knowledge this is the first analysis that explores the relationship between CP and FDI fully accounting for the asymmetric nature of CP.² This complements the seminal contribution by Guiso et al. (2009) that focus on the impact on international transactions of a related cultural variable: trust. While trust is inherently asymmetric these authors only focus in their FDI gravity regression on one direction of the cultural relationship: i.e. how much individuals in the FDI origin country trust on average individuals in the destination country. While CP and trust are two different cultural variables, their positive correlation (empirically assessed by the these authors in the same paper) and our results suggest that FDI could also positively respond to the trust of citizens in the destination country for those in the country where FDI is coming from.

Our paper is closely related to the two existing studies on the relationship between asymmetric CP and international trade: Disdier et al. (2010) and Felbermayr and Toubal (2010). The former introduces for the first time cultural trade as a proxy for asymmetric and time-dependent CP,

²There exist empirical studies of bilateral FDI that, while not centering their research question on the link between CP and FDI, include a symmetric (and often time-invariant) regressor to capture CP in an FDI gravity equation. These include Javorcik et al. (2011) and Blonigen and Piger (2014). They all find a positive relationship between CP and FDI. Similar symmetric and often time-invariant measures of CP have been used extensively in gravity equations for trade (see among others Anderson and Van Wincoop, 2003; Head and Mayer, 2014; Feenstra, 2015) as well as migration flows (Bertoli and Moraga, 2013; Beine et al., 2016).

the latter uses instead the Eurovision Song Contest voting results. They both find a positive role CP as determinant of trade patterns. Beside the focus on FDI, we contribute to this literature by providing a unifying conceptual framework for CP. In doing that we establish a connection with a related strand in the international business literature, where scholars have started to criticize the symmetric and time-invariant concept and measures of CP well before economists. We draw from the seminal work of Shenkar (2001) and propose a definition of CP which accounts for many of the critiques emerging from that literature. From the same strand in international business we acknowledge the recent contribution by Li et al. (2017). These authors focus on role of cultural attractiveness for FDI related outcomes. Differently from our approach, they construct a measure of cultural attractiveness using survey data from the GLOBE project covering 62 societies (House et al., 2004) and do not rely on a structural gravity econometric framework. Moreover, similarly to Guiso et al. (2009), while both directions of cultural attractiveness can potentially affect the same direction of the economic relationship, these authors only focus on the attractiveness of the destination's culture for the origin country, showing a positive role of attractiveness for FDI. Our finding of a strong role of the the origin's culture attractiveness for the destination country extends and complements their investigation.

Our conceptual framework speaks to the theoretical literature that provides micro-foundations to a structural gravity equation for FDI, notably Head and Ries (2008) and de Sousa and Lochard (2011). The 'destination-side' channels that explain the role of the origin's culture attractiveness for the destination country bring novel forces in the existing supply/origin-side gravity models, providing a rationale for the introduction of an additional term in the gravity equation to capture multilateral resistance from the side of the destination country. Our empirical results suggest that these forces are actually at work.

The rest of the paper is organized as follows. Section 1.2 builds a conceptual framework that explicitly accounts for the asymmetric dimension of CP and presents our proxy based on cultural trade. Section 1.3 discusses the various elements of the econometric framework proposed to assess the empirical role of CP as a determinant of Greenfield FDI. Baseline estimation results and robustness checks are discussed in Section 1.4 while Section 1.5 presents our extensions to the main analysis. Section 1.6 concludes.

1.2 Asymmetric cultural proximity

Economists and international business scholars have successfully used the concept of culture to identify factors that - in their cross-country variation - (i) explain international economic interactions and (ii) are not captured by relevant parameters such geographic distance or other forms of transaction costs.³ The definition of culture used in this paper is willingly broad and it accounts for the ideas (values, beliefs, norms) and practices (behavioral patterns) prevailing

³While not departing from this approach, we acknowledge that it is not uniformly adopted across social sciences. Indeed, many anthropologists tend to refuse the notion of cultures as bounded, essentialized and internally homogenous entities that can be used to classify, differentiate and compare groups of individuals (see for instance Abu-Lughod, 1996; Appadurai, 1996).

among respective groups of agents (Leung et al., 2005).

The characterisation of CP between two countries - i and n - as the degree by which the shared ideas and practices of one country tend to be similar to the ones of the other suffers from important limitations which have been highlighted in both the international business and the economic literature. Numerous studies including Shenkar (2001), Tung and Verbeke (2010) and Li et al. (2017) demonstrate how cultural relationships which are relevant in the context of international investment are far from being symmetric. For instance Shenkar (2001) relabels the assumption of symmetry in CP as the “illusion of symmetry”. One key element is that “symmetry between (1) the distance perceived by country n economic actors vis-à-vis country i and (2) the distance perceived by country i economic actors vis-à-vis country n , is often not warranted” (Tung and Verbeke, 2010). Ultimately, the behaviour of economic agents will be affected by their perceptions and therefore needs to be taken as a function of an asymmetric construct of CP. The analysis conducted by these papers provides empirical ground to support this critique. Using data from the GLOBE Project survey Li et al. (2017) find evidence of asymmetry in CP once cultural practices of a target country are mapped with values of an observer country. Practices records represent how a number of cultural elements (such as assertiveness, future orientation, gender egalitarianism) “are” according to the respondents in target while perceptions reflect how the same elements “should be” according to respondents in the observer country. Similar conclusions have been reached by economists. Felbermayr and Toubal (2010) state that “[a] country’s citizens can display respect and sympathy for the cultural, societal, and technological achievements of another country without this feeling necessarily being reciprocal”. They argue that such asymmetric assessment is relevant in determining bilateral economic interactions among countries and therefore call for a broad notion of CP capable of reflecting asymmetric affinity between two countries. Similar considerations can be found in Guiso et al. (2009) and Disdier et al. (2010) even though, because the empirical exercise in these papers involve only one focal country, the asymmetric aspect of CP is reduced to imply symmetry.

Consistently with these approaches, we assume cultural relationships to be asymmetric and we propose a notion of CP that accounts for that. We explicitly introduce cultural attractiveness as an element of CP. Indeed, individuals in country i can attribute desirable properties to the culture of country n independently on actual similarity between the two cultures.⁴ Overall, attractiveness is asymmetric and varies over time. For instance, certain historical events happening in a country could alter the degree by which foreigners find that country’s culture attractive. The election of a new president in the United States is likely to change the way countries around the world find American culture attractive as a function of the ideas and practices which are more represented by the elected candidate as well as the specific perceptions of each observer country. This alters the distribution of the US culture’s attractiveness across foreign countries, not necessarily having any effect on the way Americans find foreign cultures attractive.

The implication of this discussion is that the asymmetric dimension in the relationship between

⁴Li et al. (2017) derive the construct of cultural attractiveness from the interpersonal attraction framework introduced by the social psychology and sociology literature. The analysis in the present paper does not depart from that conceptualisation.

two cultures can potentially affect economic interactions, and therefore needs to be taken into account when investigating the role of CP for international trade or investment. Formally, we define CP between two countries i and n as

$$CP_{ni,t} = f(S_{ni}; A_{ni,t}) \quad (1.1)$$

where f is an increasing function on the unspecified support between minimum and maximum CP. S_{ni} denotes the actual similarity between i 's culture and n 's culture, with $S_{ni} = S_{in}$, while $A_{ni,t}$ is the attractiveness of the n 's culture for individuals in i . A is asymmetric as the identity $A_{ni,t} = A_{in,t}$ is potentially not verified. Finally, we allow $A_{ni,t}$ to vary over time.⁵ In practice S_{ni} can also be subject to time variation. Patterns of migration or geo-political design of national entities are two potential time dependent factors shaping religious, ethnic, linguistic similarity between two countries. We neglect this dimension for three reasons. First, its inclusion does not alter in any way the key results of our study. Second, changes in S_{ni} tend to take place in the long run while variations in the asymmetric component of CP can be relatively quick. This is because attractiveness might respond to a much broader set of events: from the changes of political representation (as in the case of the election example above), to the adoption of new communication technologies capable of better transmitting/accessing cultural contents across countries (for instance the development of machine learning translation algorithms), to the effectiveness of governments to promote the visibility of national cultures abroad, to the international diffusion of pop music from one particular country (e.g. the big success of pop music from South Korea in South America in 2016 and 2017). Third, a symmetric component of CP which is also time invariant represents the exact conceptual counterpart of the standard symmetric and time invariant empirical measures of CP and therefore will allow us for a more direct mapping between the theoretical constructs and the empirical measures (see Section 1.5.2).⁶

1.2.1 Bilateral cultural trade as a proxy for CP

We argue that bilateral trade flows in cultural goods can be used as meaningful proxies for CP. In particular, the value of i 's imports of cultural goods exported by n at time t - $CulIMP_{ni,t}$ - is an accurate proxy for $CP_{ni,t}$. As discussed by Disdier et al. (2010), $CulIMP_{ni,t}$ directly and intuitively accounts for n 's culture attractiveness for individuals in i . Similarly, the value of i 's exports of cultural goods imported by n - $CulEXP_{ni,t}$ - is an accurate proxy for $CP_{in,t}$. As for the capacity of cultural trade to capture the symmetric component of CP, our data shows that there exists a statistically significant empirical relationship between the two, indicating that

⁵This definition and the subsequent analysis do not rest on the assumption that cultures and perceptions are fixed over time and therefore avoid the "illusion of stability" (Shenkar, 2001).

⁶The definition given in (1.1) is silent on the potential relationships between S_{ni} and $A_{ni,t}$ or $A_{in,t}$. The theoretical discussion of these links remain to a large extent outside the scope of the current paper. However, on an empirical ground there exists a positive correlation between S_{ni} and $A_{ni,t}$ (see Appendix 1.B). Moreover, the subsequent empirical exercise allows us to assess the qualitative nature of the relationship between S_{ni} and $A_{ni,t}$ (whether they are complements or substitutes) as determinants of patterns of FDI.

attractiveness is positively correlated with similarity.⁷

Bilateral cultural trade flows are constructed from the BACI dataset by CEPII⁸ and cultural goods identified through the classification of proposed by UNCTAD (UNCTAD, 2010).⁹ Table 1.1 reports the products which are classified as cultural goods. The UNCTAD classification divides them into two categories, ‘core’ and ‘optional’ cultural goods, listed in the first and second column of Table 1.1 respectively. Each category has two headings, arts and media within the ‘core’ category and heritage and functional creation within the optional one. Core cultural goods generally embed a higher cultural content and they are listed across other available classification schemes such as the one developed by UNESCO.

Table 1.1: Categories of Goods with Cultural Content (UNCTAD, 2010)

Core Cultural Goods	Optional Cultural Goods
<i>Arts (Performing and Visual)</i>	<i>Heritage (Arts Crafts)</i>
Music (CD, Tapes), Printed Music, Painting, Photography, Sculpture and Antiques	Carpets, Celebration, Paperware, Wickerware, Yarn and Other
<i>Media (Publishing and Audio-Visual)</i>	<i>Functional Creations (Design and New-Media)</i>
Books, Newspaper, Other Printed Matter, Film	Architecture, Fashion, Interior, Glassware, Jewellery, Toys, Recorded Media and Video Games

Notes: Further information on the classification can be found in UNCTAD (2010). This table replicates Table 4.2, p. 112 of UNCTAD (2010).

Before the merging with FDI and other data the cultural trade database has a coverage of 176 countries on the period 2003-2014. On average across countries and over time trade in cultural goods accounts for 2.7% of total trade in this sample. As noted in Disdier et al. (2010), cultural trade is highly concentrated. Summing cultural trade flows across importers and over time, the top five exporters - China, Germany, USA, Italy and France - account for 55% of total cultural trade. When looking at all trade instead, the top 5 exporters - China, Germany, USA, Japan and France - account for 37% of the total.

1.2.2 A detour on asymmetry

Before turning to the main research question in the paper, we provide some descriptive evidence of the asymmetry embedded in the bilateral flows of cultural goods.

Stepping from the descriptive exercise proposed by Felbermayr and Toubal (2010), we construct an empirical measure of asymmetry in EP. The construction of the empirical asymmetry measure is done by replicating the simple descriptive exercise of Felbermayr and coauthor, and is made of two steps. First, we estimate a simple linear model where cultural trade $CuIMP_{ni,t}$ is regressed

⁷See Appendix 1.B.

⁸See http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=1 and Appendix 1.A for a detailed discussion of the data.

⁹The choice of the UNCTAD classification to define the relevant set of cultural goods serves the purpose of maximizing the country coverage of the resulting estimation sample. We depart from Disdier et al. (2010) that define cultural goods following a different scheme. The implications due to the adoption of a different classification scheme are discussed in Appendix 1.A.

on importer-time fixed effects $\delta_{i,t}$; country pair fixed effects γ_{ni} ; and an error term $\varepsilon_{ni,t}$. The empirical estimate $\hat{\gamma}_{ni}$ has a useful economic interpretation: it captures, on average over time, how much individuals in (importing) country i consider the culture of (exporting) country n attractive above or below the attractiveness of the average country (this is what Felbermayr and Toubal refer to as “*excess*”). Second, for each (undirected) pair of different countries we compute the absolute value of the difference between $\hat{\gamma}_{ni}$ and $\hat{\gamma}_{in}$. We interpret the result as a proxy for the degree of asymmetry in the CP between two countries. While the data - covering bilateral cultural trade for 176 countries - would in principle allow to estimate this measure for 15400 country pairs, due to the high number of zeros we are able to derive both $\hat{\gamma}_{ni}$ and $\hat{\gamma}_{in}$ only for 4137 pairs. However, despite they account for just less than one third of all potential combinations, these 4137 pairs account for 49.1% and 55.8% of total trade and total trade in cultural goods respectively. This exercise, counterposed to Felbermayr and Toubal’s estimates, provide an interesting insight on the importance of keeping the unit of analysis into consideration. Differently from their study, our sample is much larger and involves a far more heterogeneous set of actors: the result is a generally lower “*excess*” among European countries as opposed to the estimates obtained focusing on their restricted sample. The issue of overestimated asymmetry in a narrower and relatively homogeneous sample is made clearer by figure 1.1, where the distribution of asymmetry premia in our data is plotted against different samples used in the related literature. Despite the distribution is much more skewed in the total sample (the solid blue line) than in both the sample used by Felbermayr and Toubal (2010) and Guiso et al. (2009) (the green dash-and-dot line and the red short-dashed line respectively) it is worth noticing that both samples in the abovementioned studies are positioned in the increasing segment of the blue line. This fact suggests that the asymmetry in cultural premia identified for instance by Felbermayr and Toubal or by Guiso and coauthors is actually overestimated, since the asymmetry measure among similar countries exhibits much smaller magnitude when less homogeneous partners are taken into consideration. The discussion on asymmetry premia and sample selection is further explored in Appendix 1.C.

Kernel density distribution of asymmetry premia according to different sample selection.
 Full sample includes all 4137 country pairs for which both affinity premia were available;
 F&T 2010 refers to the sample used by Felbermayr and Toubal (2010); GSZ 2009 includes
 Guiso et al. (2009) sample.

To illustrate the scope of the asymmetry embedded in cultural trade, table 1.2 reports the country pairs with the highest and the lowest value of the asymmetry measure. For these two pairs we report the directed attractiveness premia and the resulting value of asymmetry implied by cultural trade.

Table 1.2: Max and Min Asymmetry

Country n	Country i	Attractiveness premium of i for n ($\hat{\gamma}_{ni}$)	Attractiveness premium of n for i ($\hat{\gamma}_{in}$)	Asymmetry ($ \hat{\gamma}_{ni} - \hat{\gamma}_{in} $)
China	Paraguay	7.211	-3.686	10.897
Morocco	Singapore	0.047	0.046	0.001

Notes: The table lists the two pairs showing respectively the higher (lower) asymmetry in attractiveness premia awarded to each other, according to the full sample of countries for which the estimated measure of asymmetry is available.

Figure 1.1: Distribution of Asymmetry Premia according to different Sample selections.

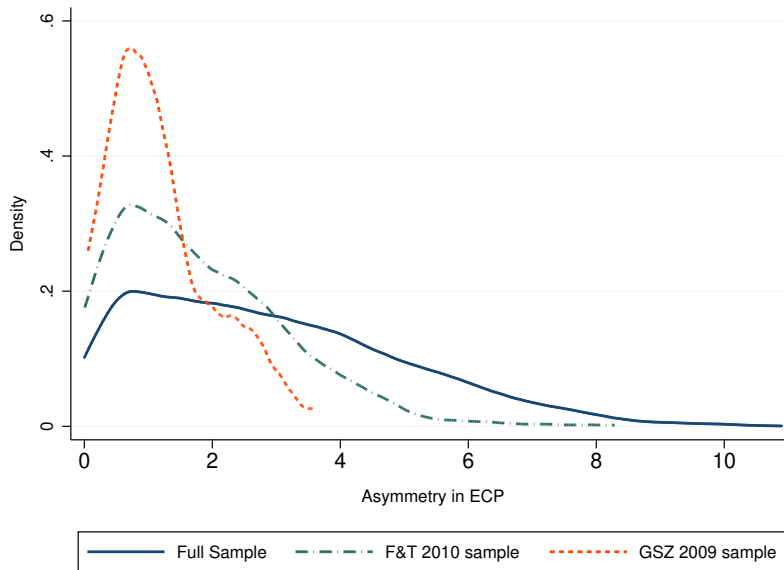


Table 1.2 shows the maximum and minimum values taken by the measure of asymmetry described above. The highest asymmetry estimated from our sample is between Paraguay (i) and China (n). In particular, China appears much more attractive for Paraguay relative to the average country ($\hat{\gamma}_{ni} = 7.211$). On the contrary the attractiveness of Paraguay's culture for China is lower than the average country's attractiveness ($\hat{\gamma}_{in} = -3.686$). In other words, individuals in Paraguay tend to put a positive attractiveness premium on Chinese culture while Chinese individuals tend to find Paraguay's culture less attractive than others. In order to get a more concrete understanding of this maximum asymmetry one can look at the actual value of the relevant cultural trade flows in the whole sample of bilateral cultural trade. In particular, the average value - across years and exporters - of Paraguay's imports of cultural goods is USD 2,087,000 while on average across years Paraguay imports from China USD 273,137,000 (almost 131 times the cross country average). On the other hand, the average Chinese imports of cultural good (across years and exporting countries) is USD 29,563,000 while its average yearly imports from Paraguay is just USD 23,000 (0.08% of the average value across exporters).¹⁰ Minimum asymmetry is found between Morocco and Singapore. In this case there exists a very balanced neutrality, with each country awarding the other with a very low attractiveness premium.

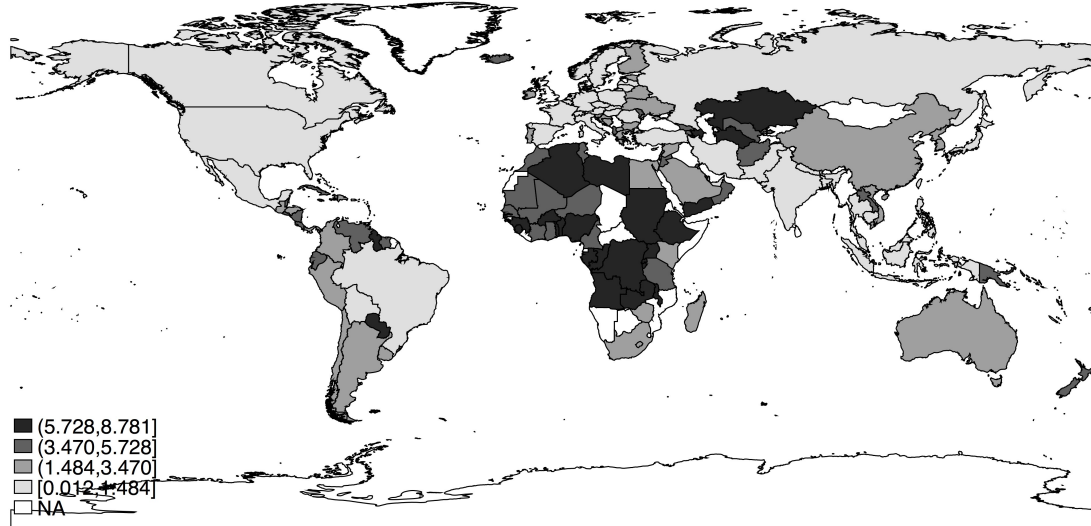
We complement the discussion of the extreme values of asymmetry by exploring the case of the UK and its bilateral cultural relationships with the other countries. The UK is the sixth biggest exporter and the second importer of cultural goods.¹¹ Because of the British Empire the legal, linguistic and cultural connections of the UK are many and relatively well known. For these reasons the UK represents a useful reference point for this exercise. Figure 1.2 provides a graphical

¹⁰This case seems to be suggestive of a potential correlation between asymmetry in export capacity and high asymmetry in cultural relationships: indeed, even if the table only shows the upper bound, this pattern finds support in the data. See Appendix 1.C for a simple assessment of this correlation. A comprehensive investigation of the determinants of asymmetry in CP goes beyond the scope of the preset paper.

¹¹This ranking is based on total trade flows for the period 2003-2014 across 176 countries.

representation of the distribution of asymmetry in the 156 available country pairs involving the UK. The colors denote the four quartiles of the distribution over these 156 observations: darker tones indicate higher asymmetry.

Figure 1.2: Asymmetry in CP Between the UK and the Rest of the World



A low degree of asymmetry in the cultural relationship reflected in cultural trade involving the UK is apparent for many European countries (with the notable exception of Ireland); for many economies in the South-East Asia region; for Russia; for the North American countries; and for some Latin American ones. High asymmetry emerges between the UK and countries in the African continent (with few exceptions below the median level of asymmetry including Madagascar and South Africa); countries in the Central Asia region; and few countries in Latin America. Relatively low asymmetry in the cultural relationships with European countries highlights the capacity of our empirical framework and of its wide country coverage to complement previous studies on the role of asymmetric cultural variables for economic transactions with a focus on European countries. Indeed, both Guiso et al. (2009) and Felbermayr and Toubal (2010) document the existence of a significant degree of asymmetry in patterns of trust and of affinity by using data on a relatively narrow and homogeneous set of countries. The case of the UK presented in Figure 1.2 suggests that intra Europe bilateral cultural relationships appear relatively more symmetric when extrapolated from a global empirical framework.

Finally, while the exercise in Figure 1.2 provides suggestive evidence for the distribution of the asymmetric component in cultural relationships, it remains largely uninformative regarding the type of asymmetry in each country pair. For instance, the relatively high asymmetry between the UK and Ireland (2.700) originates from a very high affinity premium placed by Ireland on the UK ($\hat{\gamma}_{GBR,IRL} = 8.677$) and only partly reciprocated by the still high affinity premium of the UK for Ireland ($\hat{\gamma}_{IRL,GBR} = 5.977$). On the contrary, the almost identical asymmetry score between the UK and Honduras reflects a low affinity premium of Honduras for the UK ($\hat{\gamma}_{GBR,HND} = 0.175$) to which the UK corresponds a negative one ($\hat{\gamma}_{HND,GBR} = -2.525$).

The descriptive detour proposed in this section served the purpose of illustrating the existence and scope of asymmetry in CP as an empirical phenomenon captured by bilateral cultural trade.

A focus on such asymmetry is central to our main research question, which we now turn to address.

1.3 Econometric framework

The econometric framework used to assess the empirical relationship between CP proxied by cultural trade and Greenfield FDI is constructed in several steps. First, we introduce a gravity model of bilateral FDI building on Head and Ries (2008) and de Sousa and Lochard (2011). Then, equipped with the definition of CP given in Section 1.2, we discuss theoretical mechanisms linking CP and greenfield FDI. Finally, the estimation strategy and data are presented.

1.3.1 Asymmetric CP and FDI gravity models

To assess how bilateral, asymmetric and time-varying CP affects bilateral patterns of greenfield FDI, we follow the theoretical model of greenfield FDI proposed by de Sousa and Lochard (2011) which is rooted in the seminal theory by Head and Ries (2008). Both models are characterized by a partial equilibrium, supply side perspective. Moreover, their gravity nature accounts for multilateral frictions, i.e. decisions made by MNEs to invest in a particular destination are not independent on their investment decisions into other countries.¹²

The theory is simple. Greenfield FDI projects are modelled as inspection games between the manager of a MNE (MM) and that of its foreign subsidiary (Sub). The payoff of the MM denoted by ν is a negative function of an inspection cost c and a transaction cost τ . The former reflects the standard costs of monitoring which can be implemented by the MM in order to detect a shirking behavior of Sub. The latter materializes whenever Sub exerts effort and adds value to the investment project. τ encompasses all types of costs associated with greenfield FDI beyond inspection costs. Examples includes the costs of dealing with “currency risks, exchange-rate transaction costs, trading- and liquidity-related costs as well as differentials of taxation, accounting, and legal standards in a broader interpretation” (de Sousa and Lochard, 2011, p. 554). Both c and τ are functions of a vector of formal investment policies, geographic and cultural proximity.

In a multi country framework with stochastic MNE’s payoff functions, MM chooses to invest in a country where the highest value of a project is higher than the highest value of projects in all other countries. The model allows to represent the number (or value) of greenfield FDI projects from origin country i into destination country n with a structural gravity equation of the kind

$$FDI_{ni} = K_i A_i^{-1} M_n T_{ni} \quad (1.2)$$

The term K_i is a function of the origin/parent country specific parameters, such as the total

¹²This approach differentiates these models from the knowledge-capital model of MNEs.

number of investment projects that can be financed (the total capital stock). A_i^{-1} is a multilateral resistance component, capturing the attractiveness of alternative locations for investors in country i . M_n is a function of the destination/host country specific parameters, which include the total number of potential investment projects and the average contribution of Sup across projects. Finally, T_{ni} is the bilateral component, a function of both monitoring and transaction costs, but also of the vector of formal investment policies, geographic proximity and CP. Intuitively, the model specifies T_{ni} as a decreasing function of c and τ . The qualitative relationship between these costs and formal investment policies as well as geographical distance parameters is taken from Head and Ries (2008) and de Sousa and Lochard (2011). The existence of FTAs (Free Trade Agreements) or BITs (Bilateral Investment Treaties) between i and n can potentially reduce both monitoring and transaction costs, which are also assumed to decrease with geographical proximity.

The way c and τ depend upon the symmetric component of CP is not new to the FDI gravity literature in economics: higher similarity between the two cultures implies lower monitoring as well as lower transaction costs. What has not been discussed is how monitoring and transaction costs react to the asymmetric component of CP. In what follows we address this in a broader discussion on how greenfield FDI from origin i to destination n depends upon both $CP_{ni,t}$ and $CP_{in,t}$.

Higher $CP_{ni,t}$ reduces the costs that the parent MNE has to pay to monitor the activities of its foreign subsidiary. This is intuitive if higher $CP_{ni,t}$ reflects higher S_{ni} . Indeed, for many symmetric dimensions of CP (common language, similar legal practices and contracting behaviour) clearly facilitate monitoring activities. However, $A_{ni,t}$, the degree of attractiveness for individuals in the origin country i of the ideas and practices which are prevalent among individuals in destination n , is also a determinant of lower monitoring costs. It minimizes assessment errors and facilitate the assessment processes themselves by making easier for i individuals (that have to evaluate the effort exerted by the subsidiary located in i) to establish an effective interaction with n agents, beyond a common language framework. By effective interaction we mean an interaction that favours a quicker and more precise understanding of what the other is saying as well as of what she is hiding. As for transactions costs, both S_{ni} and $A_{ni,t}$ minimize the costs to cope with different accounting/legal standards and in general with all corporate standards that might differ across the parent and the host country. Finally, from the point of view of country i parent personnel, if an inspection activity or the work needed to harmonize different corporate-related standards involves interaction with n 's individuals and/or business trips to country n , higher appreciation by country i individuals of the culture of country n reduces the costs associated with these activities.¹³ These mechanisms altogether unambiguously predict a positive effect of $CulIMP_{ni,t}$ on greenfield investment from i to n .

Let us now consider the role of $CP_{in,t}$ in explaining greenfield FDI from origin country i to destination n . Notice that our arguments on the role of S_{ni} apply to S_{in} as well due to the symmetric nature of S . Discussing the role of $CP_{in,t}$ therefore amounts to consider the role

¹³For a detailed review of the mechanisms that make destination's cultural attractiveness for the origin country a relevant driver of origin's MNEs' FDI decisions see Li et al. (2017).

of $A_{in,t}$, i.e. of the attractiveness of the i 's culture for individuals in n . From the point of view of the subsidiary personnel in the destination country n , the attractiveness of i 's culture for them results in a good attitude toward interactions with the parent's personnel. Smoother interactions reduce inspection as well as transaction costs for the MNE. But $A_{in,t}$ can be relevant for i 's investment in n beyond its effect on i 's MNE monitoring and transaction costs. First, in so far as the n subsidiary is intended to serve the n market, the value that consumers in n put on the output of i 's MNE increases the average payoff from a greenfield investment in country n . This preference value is likely to be a positive function of how much individuals (consumers) in n are attracted by i 's culture ($A_{in,t}$), also relatively to the cultures of other potential investors. This 'destination consumers demand' channel is likely to be particularly relevant (i) when the outcome of the FDI project is a final consumption good and (ii) in sectors where FDI is the prevailing mode of international provision, as it is still the case for many services sectors. Second, the realization of an FDI project by i can be facilitated or opposed by political pressures in the host country n . A plausible assumption is that political pressures to facilitate inward foreign investment will be allocated to i 's projects, also according to the degree by which individuals (voters) in n appreciate i 's culture with respect to those of other potential investors. We expect this 'destination political economy' channel to be more pronounced for destination countries with higher political accountability, i.e. where politicians tend to be less independent from voters preferences in their political and economic decisions.

These 'destination-side' mechanisms are not accounted for in the classical theoretical framework of de Sousa and Lochard (2011) and they call for an additional term in the gravity equation to capture multilateral resistance from the side of the destination country n . We rewrite (1.2) as

$$FDI_{ni} = K_i A_i^{-1} M_n B_n^{-1} T_{ni} \quad (1.3)$$

where B_n^{-1} is a function of the attractiveness of alternative investors for n 's consumers and/or voters.

The micro-foundation of the destination-side mechanisms by extending the theory of de Sousa and Lochard (2011) is a task that goes beyond the scope of the current paper: in fact they do not suggest any theoretical ambiguity about the sign of the relationship between $CP_{in,t}$ and i 's investment into n . All in all, the discussed mechanisms unambiguously imply a positive effect of $CP_{in,t}$ on greenfield investment from i to n .

1.3.2 Baseline estimation, identification strategy and data

The structural gravity model (1.3) can be brought to the data. Following Santos Silva and Tenreyro (2006) we rely on the Poisson Pseudo-Maximum Likelihood (PPML) as our workhorse estimator¹⁴. We are aware of the existence of some limitation of the PPML approach, limitations that call for additional considerations when choosing it among potential alternative estimators.

¹⁴See Santos Silva and Tenreyro (2011) and Martinez-Zarzoso (2017) for a detailed debate on the capacity of such estimator to handle large shares of null flows.

Concerning our framework, the more stringent issues to be considered are represented by the actual distribution of the error term, the effective capability to deal with a substantial fraction of zeroes, and the possibility to efficiently control for the Multilateral Resistance terms (MR, Anderson and Van Wincoop, 2003). Inherently to the first two observations, we follow Head and Mayer (2014), as we compare the results from different alternative estimators (each of them characterized by a different set of desirable features): our favourite PPML, the EK-Tobit (Eaton and Kortum, 2001), the Negative Binomial, the Gamma Pseudo-Maximum Likelihood (GPML), and the pooled OLS. The results of this preliminary exercise hold across different estimators/econometric techniques and are in line with the estimates presented in Section 1.4 and 1.5. Unfortunately, such an exercise is plagued by a substantial limitation: none of the proposed estimator (with OLS as the only exception) is able to properly control for the suitable set of fixed effects that would be necessary to control for MR¹⁵ (which should at least consider origin×year and destination×year FE, according to Baldwin and Taglioni, 2006). In short, the large amount of dummy variables controlling for FE triggers a form of “curse of dimensionality”, which prevents standard statistical packages from converging in large samples. Yet, this issue is particularly relevant, given the importance of the correct specification of the MR term to properly identify the effect of cultural preferences on FDI, net of initial condition and of the country specific unobservable factors which might affect the estimates. For this reason, we rely on PPML as our preferred technique, estimated via the recent *ppml_panel_sg* STATA package (developed by Tom Zylkin and introduced by Larch et al., 2017), which is able to absorb high-dimensional fixed effects (HDFE) without incurring in convergence issues. The results related to the other estimators, are available upon request to the corresponding authors.

The dependent variable used in the baseline estimation exercise is $C_{ni,t}$, the number of Greenfield FDI project from an origin country i to a destination country n at time t . The origin and destination specific components $K_{i,t}$ and $M_{n,t}$, as well as the multilateral resistances $A_{i,t}^{-1}$ and $B_{n,t}^{-1}$ are accounted for through origin-time and destination-time fixed effects. The elements of the bilateral component $T_{ni,t}$ are captured through (i) the log of the distance between origin and destination ($\ln \text{dist}_{ni}$); (ii) a dummy for geographical contiguity (contig_{ni}) as proxies for transportation costs; (iii) the number of FTAs and BITs involving i and n which are in force at time t ($\text{FTA}_{ni,t}$ and $\text{BIT}_{ni,t}$) as measures of formal investment policy. The elements of $T_{ni,t}$ which pertain to CP are proxied with both directions of cultural trade between i and n , ($\text{CulIMP}_{ni,t}$ and $\text{CulEXP}_{ni,t}$) at current time. One could expect a time lag between time-varying cultural attraction factors and FDI. The time lag is not taken into account in the baseline specification, as it is specifically dealt with when we discuss about the potential sources of reverse causality (in Table 1.6)¹⁶. Finally, in order to identify the specific role of the asymmetric component of CP ($A_{ni,t}$ and $A_{in,t}$) we control for its symmetric component ($S_{ni} = S_{in}$) by adding to our specification the standard symmetric and time-invariant measures of CP (a former colony dummy colony_{ni} , linguistic lang_{ni} , religious comrelig_{ni} , and institutional proximity comleg_{ni}). We acknowledge from the outset that our identification can be potentially undermined by endogeneity arising

¹⁵They only include dummies for year, origin, and destination country separately

¹⁶As reported in Table 1.6, the results of the specifications including lags do not experience a substantial change in the point estimates.

from omitted variable or reverse causality issues. We address this concern in Section 1.4.2.

The fDiMarket Database we use, collects information on greenfield FDI from January 2003 onward, and it is constantly updated. To the best of our knowledge, it constitutes the most reliable and complete existing source of greenfield investment data.¹⁷

In addition to Greenfield FDI information for the dependent variables and the data on cultural trade flows which constitute the main regressors of interest (see Section 1.2.1 above), we include in the gravity specification measures of linguistic proximity from Melitz and Toubal (2014) and Adsera and Pytlikova (2015). These indices integrate the standard bilateral linguistic measures adopted in the majority of gravity models that do not focus on CP. Data on bilateral investment treaties come from the UNCTAD Investment Policy Hub. All remaining gravity and distance related variables used throughout the empirical analysis come from the CEPII's *geodist* and *gravdata* datasets. See Appendix 1.A for a more thorough description of data sources and how the dataset is created.

The dataset used for the baseline estimation consists of an unbalanced panel of 87,448 observations. It features 144 origin and 178 destination countries over the 12 years period from 2003 to 2014. Summary statistics for the variables used in the baseline estimation are given in Table 1.3.

Table 1.3: Summary Statistics from Baseline Estimation Sample

Variable	Mean	Median	sd	Min	Max
$C_{ni,t}$	1.551	0	8.897	0	400
$\ln \text{dist}_{ni}$	8.482	8.747	0.910	4.107	9.892
colony_{ni}	0.032	0	0.177	0	1
lang_{ni}	0.157	0	0.364	0	1
comrelig_{ni}	0.173	0.033	0.266	0	0.989
contig_{ni}	0.038	0	0.190	0	1
comleg_{ni}	0.293	0	0.455	0	1
$\text{FTA}_{ni,t}$	0.269	0	0.444	0	1
$\text{BIT}_{ni,t}$	0.393	0	0.488	0	1
$\ln \text{CultIMP}_{ni,t}$	-0.454	-0.429	3.273	-6.908	10.644
$\ln \text{CultEXP}_{ni,t}$	-0.145	-0.086	3.114	-6.908	10.644

Notes: This table reports summary statistics for the variables used in the baseline estimation exercise (see Table 1.4). The related estimation sample consists of 87,448 observations.

1.4 Results

In this section we present the results of the empirical analysis. We discuss the baseline estimation results in Section 1.4.1 and then the main robustness tests in Section 1.4.2. Further extensions

¹⁷Completeness does not exclude misreporting or missing data, but such missing data are likely to be very limited and continuously revised by the dataset provider (<http://www.fdiintelligence.com/fDi-Tools/fDi-Markets>).

to the core analysis of the paper are discussed separately in Section 1.5.

1.4.1 Baseline results

Table 1.4 below presents the main results of our empirical exercise. The positive and statistically significant coefficient of $\ln \text{CultIMP}_{ni,t}$ in column (1) shows that the attractiveness of the n 's culture for individuals in country i ($A_{ni,t}$) is a determinant of the number of greenfield FDI projects from i to n . In particular, the number of investments from an origin country to a destination economy increases with $A_{ni,t}$ as captured by the value of i 's cultural imports from n . Analogously, the estimated coefficient of $\ln \text{CultEXP}_{ni,t}$ in column (2) is positive and statistically significant, showing that the number of greenfield FDI projects from origin i to destination n is higher for stronger attractiveness of the i 's culture for individuals in the in n ($A_{in,t}$). Finally, both bilateral flows of cultural goods between the origin i and the destination n are included in the specification reported in column (3) of Table 1.4. Their estimated coefficients remain positive and highly significant but the magnitude of the point estimate for $\ln \text{CultIMP}_{ni,t}$ is more than halved. Overall, the impact of trade in cultural goods on the number of greenfield FDI projects is identified beyond the role of the other gravity variables and of the standard proxies for CP. This shows that the asymmetric component of CP plays a role above and beyond its symmetric elements.

These results suggest that investment projects from i to n tend to increase more with the attractiveness of the origin's culture for individuals in the destination - $A_{in,t}$ - rather than with $A_{ni,t}$. Relying on the point estimates in column (3) of Table 1.4, the elasticities of cultural trade on the number of greenfield investment projects amount to 0.30 and 0.07 for (source to destination) exports and (source from destination) imports respectively. This finding sheds some light on the relative importance of the theoretical mechanisms linking asymmetric CP and greenfield investment. In particular it points to a relatively stronger role of those mechanisms discussed in Section 1.3.1 that explain greenfield FDI of i into n with the attractiveness of the culture of the origin country i for individuals in the destination country n . Our results confirm that it certainly matters how much the manager of the i MNE appreciates the culture in the country where the company invests, as this would imply expectations of lower monitoring and transaction costs. However, it matters more how much individuals in the destination economy appreciate the culture represented by the affiliate of the MNE in their country. Our conceptual framework (see Section 1.3.1) suggests that this too can be due to the MNE manager's expectations of lower monitoring and transaction costs (because of smoother interaction with agents that appreciate the culture represented by the MNE) but also to destination-specific channels. These are a higher propensity of the individuals in the destination country to buy the output of the MNE affiliate in their country ('destination consumers demand' channel) as well as to approve political (and economic) support toward the FDI project by their government ('destination political economy' channel). Both channels increase the profitability of the FDI project and therefore stimulate greenfield investment.¹⁸

¹⁸In Section 1.5 we present a more detailed test of the 'destination consumers demand' and the 'destination

Table 1.4: Impact of CP on Greenfield FDI (Number of Projects)

Dep. Var.	Count $C_{ni,t}$		
	(1)	(2)	(3)
$\ln \text{CultIMP}_{ni,t}$	0.165*** (11.87)		0.0690*** (5.90)
$\ln \text{CultEXP}_{ni,t}$		0.330*** (23.71)	0.305*** (21.91)
$\ln \text{dist}_{ni}$	-0.407*** (-11.60)	-0.214*** (-6.19)	-0.179*** (-5.13)
colony_{ni}	0.478*** (7.89)	0.387*** (6.95)	0.366*** (6.85)
lang_{ni}	0.254*** (4.20)	0.189*** (3.73)	0.181** (3.53)
comrelig_{ni}	1.002*** (9.47)	0.893*** (9.51)	0.883*** (9.21)
contig_{ni}	-0.114 (-1.71)	0.0752 (-1.21)	-0.0977 (-1.61)
comleg_{ni}	0.253*** (6.01)	0.170*** (4.59)	0.153*** (4.06)
$\text{FTA}_{ni,t}$	0.172** (3.02)	0.135* (2.49)	0.118* (2.19)
$\text{BIT}_{ni,t}$	0.0398 (0.93)	0.0119 (0.29)	0.0115 (0.29)
Imp×Year FE	✓	✓	✓
Exp×Year FE	✓	✓	✓
Obs	87448	87448	87448
% Zeros	0.749	0.749	0.749
R ²	0.9056	0.9216	0.9221
Test 1	-	-	585.19
Test 2	-	-	141.81
Estimator	PPML	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows.

The estimates are obtained with PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair-wise as well as origin-by-time and destination-by-time FEs (see Larch et al., 2017). The model includes origin×time and destination×time FEs. The sample size in this table is invariant to the number of covariates included and refers to the regression which features both imports and exports of cultural goods. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

1.4.2 Robustness checks

In this section we present some robustness checks dealing with the potential endogeneity of our proxy for CP - i.e. trade in cultural goods. We argue that all sources of endogeneity - namely

political economy’ channels. Table F-1 in appendix 1.F replicates the same specification as above with the coefficients of interests expressed in terms of share over aggregate imports (exports) respectively. Such an exercise allows to clean the coefficients accounting for Cultural Preferences from potential shocks affecting bilateral cultural trade. The estimates from this additional robustness check confirm the results reported in the main text.

omitted variables, *reverse causality* and *measurement error* - may potentially contribute to the bias of our parameters of interest. In what follows we discuss and address each of these sources.

Controlling for time-invariant unobserved factors and reverse causality

The correlation of the error term with CP and the determinants of Greenfield FDI in (1.3) may arise primarily because of the omission of dyadic specific unobserved factors. In particular, as noted by Felbermayr and Toubal (2010) and Disdier et al. (2010) these unobserved elements are often related to initial conditions, since the mutual learning due to strong pre-existing ties may favor convergence of cultural characteristics which in turn can trigger even more intense FDI flows. Furthermore, the relationship between CP and FDI can also be subject to reverse causality as there might be determinants of FDI that drive both economic outcomes as well as cultural attractiveness, making it difficult to establish a clear direction of causation (see Felbermayr and Toubal, 2010; Guiso et al., 2009). Indeed, positive FDI shocks may increase the interactions with foreign partners which in turn could lead to mutual learning and further cultural convergence and appreciation. We deal with these first two sources of endogeneity - namely omitted variables and reverse causality - through the inclusion of asymmetric dyadic fixed effects and by adopting an instrumental variable (IV) approach, respectively.¹⁹

We start discussing the inclusion of dyadic fixed effects. Table 1.5 compares our benchmark results with the fully specified model. The inclusion of dyadic fixed effects absorbs all the cross section variability in our sample, so that the impact of CP depends solely upon time contingent cultural factors. To allow for comparison of the results, the sample size is identical in all columns as we maintain the same sample for the fully specified model across all specifications. The models with country×year fixed effects (columns 1-3) deliver roughly the same results as Table 1.4, so the reduction of the sample size does not significantly alter our benchmark estimates. On the other hand, similarly to Felbermayr and Toubal (2010) and Disdier et al. (2010), the inclusion of dyadic fixed effects in column (4) substantially affects our parameters of interest. Trade in cultural goods retains a positive impact on FDI, but the magnitude of both the elasticities of cultural imports and exports is much lower with respect to the benchmark equation, indicating that CP is partly captured by an unobservable time invariant component. In addition, only the impact of exports remain statistically significant, which suggests that only the time variation of attractiveness of the origin's culture for the individuals in the destination economy plays a role in the MNE decision to invest.

We now move to the issue of reverse causality. In the literature the simultaneity problem has been commonly addressed with an IV strategy where current levels of CP are instrumented with their past values (see for instance Felbermayr and Toubal (2010)). This strategy hinges on the assumptions that (i) lagged bilateral values CP predict their current levels sufficiently well and that (ii) current shocks in the gravity equation are uncorrelated to past cultural relationships.

¹⁹In Appendix 1.D we further test the consistency of our benchmark results by augmenting the specification with the inclusion of observable variables of dimension *nit* that might capture (part) of the unobserved time-varying dyadic factors.

Table 1.5: Impact of Cultural Proximity on Greenfield FDI: Adding Country Pair FE

Dep. Var.	Count $C_{ni,t}$			
	(1)	(2)	(3)	(4)
$\ln \text{CultIMP}_{ni,t}$	0.145*** (10.35)		0.0522*** (4.43)	0.00677 (0.78)
$\ln \text{CultEXP}_{ni,t}$		0.314*** (22.57)	0.295*** (21.04)	0.0499*** (3.72)
$\ln \text{dist}_{ni}$	-0.404*** (-11.94)	-0.208*** (-6.27)	-0.181*** (-5.42)	
colony_{ni}	0.481*** (8.04)	0.388*** (7.14)	0.372*** (7.08)	
lang_{ni}	0.244*** (4.06)	0.180*** (3.58)	0.173*** (3.43)	
comrelig_{ni}	0.957*** (9.04)	0.855*** (9.06)	0.847*** (8.84)	
contig_{ni}	-0.0905 (-1.40)	-0.0578 (-0.96)	-0.0754 (-1.28)	
comleg_{ni}	0.246*** (5.90)	0.164*** (4.43)	0.151*** (4.03)	
$\text{FTA}_{ni,t}$	0.147** (2.62)	0.109* (2.09)	0.0976 (1.87)	0.0499 (1.12)
$\text{BIT}_{ni,t}$	-0.0145 (-0.34)	-0.0368 (-0.93)	-0.0358 (-0.92)	0.117 (1.41)
Imp×Year FE	✓	✓	✓	✓
Exp×Year FE	✓	✓	✓	✓
Country Pair FE				✓
Obs	49702	49702	49702	49027
% Zeros	55.99	55.99	55.99	55.99
R ²	0.9053	0.9222	0.9224	0.9686
Test 1	-	-	526.13	14.85
Test 2	-	-	146.33	6.92
Estimator	PPML	PPML	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the value of the aggregated bilateral flow of greenfield investments from country i to country n , including zero flows. The estimates are obtained with PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair-wise as well as origin-by-time and destination-by-time FEs. The columns (1) to (3) replicate table 1.4 results, and include origin×time and destination×time FEs only. Column (4) includes Country Pair FE, to address multilateral resistance, Baldwin and Taglioni (2006), Baier and Bergstrand (2007), Head and Mayer (2014) and Piermartini and Yotov (2016) among the others, suggest to include country×time dummy and trading pair dummies.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

While we find the first validity condition plausible, the latter which refers to the exogeneity of the instrument is neither obvious, nor easy to demonstrate. For instance, it could be argued that part of the current variation of FDI is associated to the evolution of cross-country cultural relationships and therefore depends on past shocks of CP. Indeed, FDI normally requires a long-term focus and the MNEs decision to invest is likely to depend even more on past than current levels of CP. In our conceptual framework an alternative way to address the issue of reverse causality is to adopt a completely different approach by replacing current levels of cultural trade with their lagged

values as the main variable of interest. The advantage of this strategy is that trade flows are predetermined with respect to FDI, a condition which attenuates the issue of reverse causality, without (ii) being a binding/necessary condition for the consistency of the estimator. Although they are based on somewhat contrasting assumptions, in our robustness analysis we propose both strategies - the IV and the lagged approach - to address the simultaneity problem. In the first two columns of Table 1.6 we estimate our baseline specification with the predetermined values of cultural trade at $t-2$ and $t-5$ in columns 1 and 2, respectively. The point estimates of our parameters of interest in both regressions are very close to the baseline results, which we find as reassuring. In addition, the very limited variation over time of the impact of trade in cultural goods suggests a persistence in bilateral cultural tastes or, alternatively, a very similar variation in CP over time for all country pairs. This finding corroborates the relatively low impact of the time variation of CP on FDI obtained by introducing country pair fixed effects in Table 1.5.

The IV strategy reported in the remaining two columns of Table 1.6 builds on Combes et al. (2005), Briant et al. (2014) and Felbermayr and Toubal (2010) and exploits the longitudinal nature of the BACI dataset by instrumenting current levels of cultural trade flows with lagged values of the same variables ($t - 12$).²⁰ Columns 3 and 4 compare the PPML estimates with the correspondent coefficients obtained with IVPPML using the reduced sample of Felbermayr and Toubal (2010). Concerning our parameters of interest, controlling for endogeneity leads to results that are in line with the literature and consistent with the estimates of the fully specified model. The elasticity of imports of cultural goods roughly maintain the same magnitude as in the PPML model, but becomes statistically not significant. As for exports, when instrumented their coefficient remains statistically significant at the 1% confidence level, and substantially increases in magnitude. Hence, once we control for reverse causality, we find that only cultural attractiveness of the origin country for potential destinations have an impact on greenfield investment. Furthermore, the instrumented exports' elasticity is more than twice as large, suggesting a downward bias in the impact of exports of cultural goods. However, the resulting downward bias is substantially smaller compared to previous studies (see Guiso et al. (2009) and Felbermayr and Toubal (2010)), as in our gravity specification the elasticities are far closer across estimators.²¹

On the Measurement Error

In our econometric analysis the issue of measurement error is particularly compelling as the accuracy of our results may be severely affected by imprecise measures of both Asymmetric CP and Greenfield FDI. More specifically, while asymmetric CP may not be fully reflected by the

²⁰The earliest year available from BACI dataset is 1995: this forces us to reduce the time span (2007-2014) in our IV analysis. The time varying lagged instrument is relevant as it is strongly correlated to the endogenous variable as showed in Appendix 1.E. The IV strategy is performed with the Stata command IVPOISSON which doesn't allow for the inclusion of high dimensional fixed effects. In order to include a comprehensive set of fixed effects which account for time varying importer and exporter heterogeneity, our strategy is to reduce the sample size to ensure convergence in the estimation.

²¹In Felbermayr and Toubal (2010) the impact of cultural proximity on trade is more than ten times higher when instrumented. The gap between OLS and 2SLS estimates is even higher in the analysis of Guiso et al. (2009) when the dependent variable is FDI.

Table 1.6: Impact of Instrumented Cultural Proximity on Greenfield FDI

Dep. Var.	Count $C_{ni,t}$			
	2 year lag (1)	5 year lag (2)	Baseline (3)	IV (4)
$\ln \text{CultIMP}_{ni,t}$			0.0658** (2.96)	0.0736 (1.35)
$\ln \text{CultEXP}_{ni,t}$			0.247*** (9.43)	0.619*** (6.54)
$\ln \text{lagged CultIMP}_{ni,t-2}$	0.0740*** (6.32)			
$\ln \text{lagged CultEXP}_{ni,t-2}$	0.296*** (21.27)			
$\ln \text{lagged CultIMP}_{ni,t-5}$		0.0784*** (6.59)		
$\ln \text{lagged CultEXP}_{ni,t-5}$		0.286*** (19.51)		
$\ln \text{dist}_{ni}$	0.179*** (5.08)	0.182*** (5.17)	0.806*** (11.26)	0.350** (2.70)
colony_{ni}	0.380*** (7.14)	0.385*** (7.23)	0.0193 (0.23)	0.0177 (0.18)
lang_{ni}	0.167** (3.26)	0.152** (2.99)	0.0723 (0.70)	0.0436 (0.30)
comrelig_{ni}	0.877*** (9.02)	0.872*** (8.99)	0.118 (0.95)	0.206 (1.49)
contig_{ni}	0.106 (1.75)	0.117 (1.92)	0.147* (2.36)	0.283*** (3.93)
comleg_{ni}	0.155*** (4.07)	0.157*** (4.17)	0.330*** (5.89)	0.219** (3.20)
$\text{FTA}_{ni,t}$	0.127* (2.34)	0.133* (2.45)	0.394*** (3.49)	0.0725 (0.48)
$\text{BIT}_{ni,t}$	0.00909 (0.23)	0.0311 (0.78)	0.172* (2.23)	0.0757 (0.83)
Imp×Year FE	✓	✓	✓	✓
Exp×Year FE	✓	✓	✓	✓
Obs	84568	80057	10596	10040
Estimator	PPML	PPML	PPML	IV PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z statistics in parentheses. Standard errors are clustered by trading pair. The dependent variable “Count” $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows. The estimates in columns (1) to (3) are obtained by PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair wise as well as origin by time and destination by time FEs. The model includes origin×time and destination×time FEs. Estimates in column (4) are computed via IVPML using the *ivpoisson* command built in *STATA 13*. Due to convergence reasons, in column (3) and (4) the sample is reduced to the subset of importing and exporting countries as in Felbermayr and Toubal (2010). A drawback of IVPOISSON command is that it cannot handle high dimensional FE. Nonetheless, the estimates are consistent to a broader sample estimated with a reduced set of fixed effects (available upon request to the authors), suggesting that they are robust to different specifications.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultEXP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultEXP}_{ni,t} = 0$).

intensity of bilateral exchanges in cultural trade - either for the gross nature of cultural trade and/or because of the issue of global value chain - also the data on Greenfield FDI from the FT

dataset include estimates for capital investment (derived from algorithms) when a company does not release the information (see Desbordes and Wei, 2017; Lee and Ries, 2016). Here we address these two sources of measurement error in turn.

Asymmetric CP: Specific characteristics of cultural goods may fail to adequately represent local cultural identity and therefore, as a result, the intensity of their cross-country exchanges may not appropriately reflect the actual patterns of asymmetric CP. For instance, facing a world trading system where global supply chains are prevalent, one may argue that Chinese exports of fashion products or toys (included in the category of optional cultural goods) to an import country not only (and not necessarily) reflect Chinese cultural content, and therefore the cultural attractiveness of China for the importer, but also some third country's cultural content embedded in the fashion or pottery design performed in that country before actual manufacturing happening in China. This concern is legitimate as long as few countries in our sample have a comparative advantage in the manufacture of a number of cultural products, fostering a disproportionate concentration of production in (and export from) these countries of cultural goods embedding foreign cultural value added. This might actually be the case for several Asian countries and for some of the products included in the sub-category of optional cultural goods (see Table 1.1). It is well known that countries in the so called Factory Asia have an international specialisation in the manufacturing of low tech goods, including for instance toys (see Baldwin and Lopez-Gonzalez, 2015). The average revealed comparative advantage (RCA) across optional cultural goods for the period of our analysis is equal to 1.2 for China and above the threshold value of 1 also for India, Indonesia, Malaysia, Thailand and Vietnam.²² These arguments intuitively point to a downward bias of the impact of CP estimated in our baseline analysis, as the error in measuring actual CP is likely to be positively associated with both the extensive and intensive margin of cultural trade. This conjecture is also supported theoretically (see for instance Kukush et al., 2004) for the Poisson estimator, although to the best of our knowledge the conclusions have not been extended to models with multiple regressors and with non-classical measurement error. Columns 1-3 of Table 1.7 compares our benchmark results of Table 1.4 with the estimates obtained with only core and optional cultural goods, respectively. The distinction between core and optional hinges on the cultural content embodied in these types of products: hence, it is reasonable to expect the impact of CP as mostly driven by the trade (in either direction) of core cultural goods as they are likely to better capture proximity in cultural tastes.²³ However, optional cultural goods represent the lion share of cultural trade from and between developing countries: failing to account for these flows would exclude many South countries from the analysis, limiting the impact of CP on specific FDI channels (especially North-North). The pattern of results is stable across different measures of cultural trade, showing the capacity of both types of cultural goods to reflect the same underlying forces. The stability of our results across core and optional cultural goods also suggests that this potential source of measurement error is not biasing our results as the average RCA of the listed Asian countries across core cultural goods is always well

²²RCA is computed following the Balassa index.

²³The distinction between core and optional cultural goods is described in detail in the Appendix 1.A.

below one (for instance it is equal to 0.378 of China and 0.165 for Vietnam).²⁴ In Column 4 we restrict our analysis to *Newspapers* which is arguably a category less subject to GVC bias as the papers are produced locally and plausibly reflect more strongly the cultural identity of the country where these goods are purchased from. The estimates confirm the asymmetric nature of CP and the predominant role of $\ln \text{CultEXP}_{ni,t}$ in influencing Greenfield FDI. Finally, we test the robustness of our analysis by including both directions of bilateral *trust* as an alternative measure of CP. Data on *trust* are from Table I of Guiso et al. (2009) and measure the average level of trust among selected EU countries from citizens of country of origin to citizens of country of destination. While the use of this alternative proxy imposes obvious limitations in terms of sample size and composition, we believe it is the best available proxy to compare the validity of our conclusions with. Indeed, *trust* is time varying, can be safely assumed to be strongly dependent on (at least some elements of) cultural proximity (see for a discussion Guiso et al., 2009) and - most importantly - allows to test for the role of the asymmetric nature of CP. Column 5 shows that the elasticity of *Trust* from the destination to the origin of FDI is the most important determinant of the MNE's decision to invest: this finding corroborates the soundness of our conclusions on the stronger investment effect of the origin's culture attractiveness for the destination country and at the same time substantiates the validity of cultural trade as a valid proxy for CP.

Measure of FDI: The focus on the number of projects (count) as opposed to their total or average value has the advantage of minimizing the potential distortions induced by the imputation techniques used in the construction of the value-related variables,²⁵ but has its own limitations: for instance it is equivalent to imposing to all projects the same weight in terms of economic relevance, without discriminating them for their actual size. For instance, an investment in a legal consultant office (the business sector with the lowest average capital investment in our sample) is implicitly evaluated as an investment in a plant for oil refinery, which is roughly 257 times larger (5.344 millions US\$ against more than 1.372 billions US\$ on average for the two types of investments respectively). Beyond these measurement related considerations, the size of bilateral FDI and the number of investments may (or may not) react differently to variation in CP as they capture different aspects of internationalization. This is ultimately an empirical question. The reported results in Table 1.8 (column 1) show that while the impact of CP is still positive but generally lower when considering the value ($V_{ni,t}$) as dependent variable, imports of cultural goods become statistically not significant. These combined findings suggest that the destination side mechanisms are relevant across different measures of bilateral volume of FDI, and that the decision on whether or not to invest is more sensitive to the asymmetric components of CP than the actual size of bilateral FDI. Similar conclusions apply when we investigate the impact of asymmetric CP on the intensive margin of investment as captured by the average value of investment ($\bar{V}_{ni,t}$). The estimates reported in Column 2 indicate that, despite being halved in their magnitude, the coefficients of both $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultEXP}_{ni,t}$ remain statistically

²⁴A better test of the implications of relying on gross cultural trade would require the use of value added trade data. Unfortunately available sources such as the OECD/WTO TiVA database fail to match the country coverage and product desegregation required by the research design of the present study.

²⁵See Table A-3 in Appendix 1.A for a more precise assessment of the scope of imputation.

Table 1.7: Different Measures of CP: Core VS Optional Cultural Trade

Dep. Var.	Count $C_{ni,t}$				
	Total Cultural	Core Cultural	Optional Cultural	Newspapers	Trust
	(1)	(2)	(3)	(4)	(5)
$\ln \text{CultIMP}_{ni,t}$	0.0690*** (5.90)	0.0925*** (8.22)	0.0525*** (4.34)	0.0468*** (5.59)	
$\ln \text{CultEXP}_{ni,t}$	0.305*** (21.91)	0.285*** (20.18)	0.249*** (19.43)	0.112*** (10.23)	
$\ln \text{trust}_{ni,t}$					0.975 (1.74)
$\ln \text{trust}_{in,t}$					1.379* (2.48)
Imp×Year FE	✓	✓	✓	✓	✓
Exp×Year FE	✓	✓	✓	✓	✓
Obs	87448	67192	76951	19022	172
% Zeros	75%	53%	64%	8%	-
R ²	0.922	0.920	0.913	0.925	0.949
Estimator	PPML	PPML	PPML	PPML	OLS

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows. All estimates but in the last column are obtained by PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair-wise as well as origin-by-time and destination-by-time FEs. The model includes origin×time and destination×time FEs; the last column is computed using the OLS equivalent of ppml, *reghdfe*, developed by Sergio Correia.

The first column replicates column (3) of table 1.4. The second column refers to the effect on greenfield FDI of ‘core’ cultural trade, while the third refers to ‘optional’ cultural trade, as defined by UNCTAD (2010). The fourth perform the same exercise with newspaper trade only, while the final column perform a similar exercise with the measure of trust as in Guiso et al. (2009). NOTICE: The sample is substantially reduced in column (4) and column (5), due to the large number of null values that are dropped when taken in logarithmic form. The absence of a coefficient estimate for contiguity, FTA, BIT, and colony in the last column is due to the very small sample of countries for which the *Eurobarometer* Survey is available.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

significant at least at the 5% confidence level.

Table 1.8: Impact of Cultural Proximity on Value of Greenfield FDI

Dep. Var.	Value $\bar{V}_{ni,t}$	Average Value $\bar{V}_{ni,t}$
	(1)	(2)
$\ln \text{CultIMP}_{ni,t}$	0.0221 (1.07)	0.0390* (2.11)
$\ln \text{CultEXP}_{ni,t}$	0.269*** (11.44)	0.137*** (6.11)
$\ln \text{dist}_{ni}$	-0.237*** (-4.44)	-0.166** (-3.20)
colony_{ni}	0.364*** (4.76)	0.0290 (0.25)
lang_{ni}	0.109 (1.20)	0.0222 (0.24)
comrelig_{ni}	1.210*** (8.42)	0.750*** (5.09)
contig_{ni}	-0.0952 (-0.94)	0.0874 (0.66)
comleg_{ni}	0.0724 (1.28)	0.0215 (0.32)
$\text{FTA}_{ni,t}$	0.260*** (3.52)	0.120 (1.34)
$\text{BIT}_{ni,t}$	-0.0443 (-0.74)	0.284*** (4.33)
Imp×Year FE	✓	✓
Exp×Year FE	✓	✓
Obs	87448	87448
% Zeros	0.749	0.749
R ²	0.9221	0.4961
Test 1	196.27	59.70
Test 2	41.24	8.33
Estimator	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair.

The dependent variable $C_{ni,t}$ is the value of the aggregated bilateral flow of greenfield investments from country i to country n , including zero flows, while $\bar{V}_{ni,t}$ represents the average value of bilateral greenfield investments.

The estimates are obtained with PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair-wise as well as origin-by-time and destination-by-time FEs. The model includes origin×time and destination×time FEs. The sample size in this table is invariant to the number of covariates included and refers to the regression which features both imports and exports of cultural goods. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed.

fDIMarket database provides information on the value of each greenfield. When no official figures are provided by the parent company, the value is estimated by FDIIntelligence unit. Information about the estimation algorithm can be found on fDIMarket website. The χ^2 refers to the cultural trade (trust) coefficients equivalence. Rejecting the null implies the two direction to be significantly different from each other.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

1.5 Extensions

This section proposes three extensions to the analysis conducted so far. First, we propose two empirical tests of the ‘destination-side’ mechanisms as introduced in the conceptual framework laid out in Section 1.3.1. Then, we test whether the role of asymmetric and time-dependent component of CP changes at different levels of its symmetric and time-invariant component. Finally,

we test whether similar conclusions about the stronger role of the destination side mechanisms apply to different forms of economic exchanges such as overall trade and M&A.

1.5.1 Destination-side mechanisms

The empirical analysis so far has established the relative importance of the two directions of asymmetric CP in explaining Greenfield FDI from an origin country i to a destination n . In particular the attractiveness of the i 's culture for individuals in n - $A_{in,t}$ proxied by $\text{CultEXP}_{ni,t}$ - seems to play a much stronger role than the attractiveness of the destination for the origin, $A_{ni,t}$ proxied by $\text{CultIMP}_{ni,t}$. This is somehow at odds with the standard theories of bilateral FDI which tend to focus on 'origin-side' mechanisms and calls for a more careful consideration of 'destination-side' mechanisms. In this section we propose an empirical test of the 'destination consumers demand' and the 'destination political economy' channels introduced in Section 1.3.1.

According to the 'destination consumers demand' channel, $A_{in,t}$ can be relevant to explain FDI from i to n because the preferences of consumers in n for the the affiliate's production in their country would be a positive function of i 's cultural attractiveness for them. This leads us to expect $A_{in,t}$ to be more relevant with respect to $A_{ni,t}$ when the FDI projects are intended to target consumer demand in the destination country rather than to serve as an intermediary step in a global supply chain type of production. In the case of horizontal FDI the attractiveness of the origin's culture for consumers in the destination country could be a stronger driver of the investment decision as it might positively affect the expected revenues of the FDI project. This is confirmed empirically by the estimation results presented in Table 1.9.

Both columns replicate results as in column (3) of Table 1.4 on two different subsamples. Column (1) includes only FDI projects in those sectors that are more likely to target the consumers demand in the destination country, i.e. that include consumption (final) goods and services. Conversely, the estimation sample used to derive the results presented in column (2) is restricted to those sectors where the importance of local consumption is lower compared to the location advantages of different kind: such sectors include mainly intermediate goods.²⁶ Taking the ratio between the point estimates of the coefficients for $\ln \text{CultEXP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ as a measure of the relative importance of $A_{in,t}$ with respect to $A_{ni,t}$ in explaining $C_{ni,t}$ we notice that this ratio is higher when the estimation sample is restricted to those sectors that are more likely to target the consumers demand in the destination country. We take this a suggestive evidence of the existence of the hypothesized 'destination consumers demand' channel in determining the role of CP for FDI.

The 'destination political economy' channel, on the other hand, rationalizes the role of $A_{in,t}$ in determining greenfield FDI from i to n , through the potential political and economic support

²⁶The estimation sample in the first column includes only FDI projects classified in the following sectors: beverages, consumer electronics, consumer product, financial services, food and tobacco, leisure and entertainment, software and ICT devices, and transportation. The estimation sample in the second column instead includes only the following sectors: automotive components, biotech, building and construction material, ceramics, glasses, chemical, coal, oil gas, electronic component, engines and turbines, industrial machinery, metals, minerals, plastic, rubber, semiconductors.

Table 1.9: Destination Consumers Demand Channel

Dep. Var.	Count $C_{ni,t}$	
	More likely	Less likely
FDI targeting consumers in n	(1)	(2)
$\ln \text{CultIMP}_{ni,t}$	0.0768*** (5.85)	0.0731*** (4.12)
$\ln \text{CultEXP}_{ni,t}$	0.317*** (20.12)	0.255*** (14.70)
$\ln \text{dist}_{ni}$	-0.258*** (-7.34)	-0.0730 (-1.42)
colony_{ni}	0.315*** (4.48)	0.369*** (5.50)
lang_{ni}	0.244*** (3.97)	0.0386 (0.46)
comrelig_{ni}	1.047*** (9.60)	0.872*** (6.50)
contig_{ni}	-0.153* (-2.21)	-0.0963 (-1.13)
comleg_{ni}	0.204*** (4.64)	0.0174 (0.31)
$\text{FTA}_{ni,t}$	0.0138 (0.24)	0.171* (2.15)
$\text{BIT}_{ni,t}$	0.0467 (1.10)	-0.0522 (-0.83)
Imp×Year FE	✓	✓
Exp×Year FE	✓	✓
Obs	78697	62989
% Zeros	0.82	0.83
R ²	0.90	0.88
Test 1	5389.02	2310.47
Test 2	874.19	331.26
Test Imp		0.05
Test Exp		13.49
Estimator	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. Both columns replicate results as in column (3) of Table 1.4. The estimation sample in the first column includes only FDI projects classified in the following sectors: beverages, consumer electronics, consumer product, financial services, food and tobacco, leisure and entertainment, software and ICT devices, and transportation. the estimation sample in the second column instead includes only the following sectors: automotive components, biotech, building and construction material, ceramics, glasses, chemical, coal, oil gas, electronic component, engines and turbines, industrial machinery, metals, minerals, plastic, rubber, semiconductors.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultEXP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultEXP}_{ni,t} = 0$). *Test Imp* and *Test Exp* reports the result of a χ^2 test on the equality of the coefficients across specifications (Low vs High) for cultural imports and exports respectively. For practical reasons, the last two tests are performed on a reduced set of fixed effects and approaching the estimates via GLM instead of HDFE absorbing routines, and comparing the two equation via *Seemingly Unrelated Estimation* (SUEST). Formally, *Test Imp* = $H_0 : \ln \text{CultIMP}_{ni,t}^H = \ln \text{CultIMP}_{ni,t}^L = 0$; *Test Exp* = $H_0 : \ln \text{CultEXP}_{ni,t}^H = \ln \text{CultEXP}_{ni,t}^L = 0$.

granted by the government in n to an FDI project coming from i . In a political economy model this would need to respond - at least to some extent - to the preferences of voters in n , affected by their appreciation of the culture in i . This mechanism implies a stronger relative importance

of the origin's cultural attractiveness for the destination when politicians in the destination country are subject to a higher degree of accountability with respect to their citizens, i.e. when their allocation of support across projects coming from different sources is likely to more closely reflect voters' preferences. The estimates reported in Table 1.10 represent an empirical test of this implication.

Table 1.10: Destination Political Economy Channel

Dep. Var. Accountability in n	Count $C_{ni,t}$	
	Low (1)	High (2)
$\ln \text{CultIMP}_{ni,t}$	0.0966*** (5.00)	0.0466 (1.65)
$\ln \text{CultEXP}_{ni,t}$	0.229*** (10.07)	0.486*** (16.23)
$\ln \text{dist}_{ni}$	-0.635*** (-6.37)	-0.509*** (-4.33)
colony_{ni}	0.814*** (5.89)	0.771*** (4.59)
lang_{ni}	0.176 (1.72)	-0.133 (-0.79)
comrelig_{ni}	0.144 (0.63)	-0.279 (-0.56)
contig_{ni}	0.167 (1.27)	0.0688 (0.32)
comleg_{ni}	-0.00287 (-0.03)	-0.111 (-0.80)
$\text{FTA}_{ni,t}$	0.0613 (0.49)	1.249*** (6.40)
$\text{BIT}_{ni,t}$	0.0938 (1.13)	-0.0934 (-0.87)
Imp×Year FE	✓	✓
Exp×Year FE	✓	✓
Obs	3755	2376
% Zeros	0.76	0.68
R ²	0.85	0.99
Test 1	174.09	270.38
Test 2	15.02	107.60
Test Imp		1.10
Test Exp		17.01
Estimator	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. Both columns replicate the specification as in column (3) of Table 1.4. The estimation sample used to derive the estimates reported in the first (second) column is restricted to destination countries in the first (fourth) quartile of the accountability score below. Accountability is measured with by the World Bank CPIA indicator, which reports a country's perception on Corruption, Accountability and Transparency.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultEXP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultEXP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultEXP}_{ni,t} = 0$). *Test Imp* and *Test Exp* reports the result of a χ^2 test on the equality of the coefficients across specifications (Low vs High) for cultural imports and exports respectively. For practical reasons, the last two tests are performed on a reduced set of fixed effects and approaching the estimates via GLM instead of HDFE absorbing routines, and comparing the two equation via *Seemingly Unrelated Estimation* (SUEST). Formally, *Test Imp* = $H_0 : \ln \text{CultIMP}_{ni,t}^H = \ln \text{CultIMP}_{ni,t}^L = 0$; *Test Exp* = $H_0 : \ln \text{CultEXP}_{ni,t}^H = \ln \text{CultEXP}_{ni,t}^L = 0$.

Both columns replicate the specification as in column (3) of Table 1.4. The estimation sample used to derive the estimates reported in the first (second) column is restricted to destination countries with an accountability score below (above) the sample median. Accountability is measured with the accountability index, from the World Bank CPIA indicators on Corruption, Accountability and Transparency perception. The ratio between the point estimates of the coefficients for $\ln \text{CultEXP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ is higher for the subsample of high accountability destination countries, suggesting a relative higher importance of $A_{in,t}$ when politicians in the destination country are more accountable vis-à-vis their citizens and therefore providing empirical evidence for the existence of the hypothesized ‘destination political economy’ channel²⁷.

1.5.2 Heterogeneous impact of the asymmetric and time-dependent dimension of CP

This section tests how the asymmetric and time-dependent component of CP affects bilateral investment flows at different degree of the symmetric and time-invariant component of CP. In order to do so, we explore the effect of trade in cultural goods at different values (above and below the median value) of three symmetric, and time-invariant measures of cultural proximity previously used in the literature: religious proximity, the Melitz and Toubal (2014) “Common Spoken Language” (CSL) measure of linguistic proximity, and the composite index of linguistic proximity (AP Index) by Adsera and Pytlikova (2015).²⁸ Moreover, to identify the impact of time-contingent shocks in CP all regressions include a full set of fixed effects as in Table 1.5. The inclusion of dyadic fixed effects absorbs all the cross section variability in our sample, a necessary feature if we are interested in exploring the time-varying dimension of cultural trade. Results are reported in Table 1.11 below.

Consistently with the results presented in Table 1.5 the reported estimates suggest that only time contingent shocks in terms of cultural attractiveness of the origin country for the destination seem to trigger investments. However, it seems that those results are mainly driven by pairs characterized by low level of time-invariant and symmetric CP: time contingent shocks to the cultural attractiveness of the origin country for the destination only play a role when the level of pre-existing or historical cultural ties is relatively weak. This is consistent with a relationship of substitutability between time-contingent, asymmetric and time-invariant, symmetric dimensions of CP in triggering FDI, with the former operating as a bridgehead between otherwise culturally distant countries.

²⁷To test this assumption, replicating the baseline specification with both cultural export- and import-related interactions would be necessary. However, the large heterogeneity that characterizes FDI flow data does not allow to draw statistically meaningful conclusions. In addition, collinearity causes the coefficients of the export-related interaction terms to be dropped systematically.

²⁸The choice of these measures is constrained by our intention to split the estimation sample. The majority of the “traditional” measures used in existing literature have a binary structure and for this reason they are not suitable to split our sample in a simple and effective way.

Table 1.11: Heterogeneous impact of the asymmetric and time-dependent dimension of CP

Dep. Var.	Count $C_{ni,t}$					
	Religion ¹		CSL ²		AP index ³	
	(1-50 pct)	(51-100 pct)	(1-50 pct)	(51-100 pct)	(1-50 pct)	(51-100 pct)
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln \text{CultIMP}_{ni,t}$	0.00639 (0.53)	-0.000994 (-0.07)	0.00920 (0.82)	-0.0151 (-1.03)	-0.00908 (-0.57)	-0.0434 (-0.92)
$\ln \text{CultEXP}_{ni,t}$	0.0554*** (3.34)	0.0122 (0.75)	0.0604*** (3.59)	0.00995 (0.66)	0.0713*** (3.51)	-0.0779 (-1.26)
$\text{FTA}_{ni,t}$	0.136* (2.06)	-0.0640 (-1.09)	0.0315 (0.50)	-0.0336 (-0.66)	0.0130 (0.14)	-0.0475 (-0.55)
$\text{BIT}_{ni,t}$	0.0273 (0.27)	0.0754 (0.65)	0.223* (2.32)	0.0187 (0.19)	0.0859 (0.64)	0.289 (0.77)
Imp×Year FE	✓	✓	✓	✓	✓	✓
Exp×year FE	✓	✓	✓	✓	✓	✓
Country Pair FE	✓	✓	✓	✓	✓	✓
Obs	23209	23916	22657	23465	12487	23465
% Zeros	59.78%	55.25%	64.04%	51.00%	45.77%	4.47%
R ²	0.9687	0.9770	0.9721	0.9791	0.9730	0.9895
Test 1	11.52	0.56	14.59	1.3905609	12.43	2.46
Test 2	5.53	0.35	5.71	1.31	9.09	0.19
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the number of the aggregated bilateral flow of greenfield investments from country i to country n , including null flows. The estimates are obtained with PPML using the *PPML panel sg* command written by Thomas Zylkin which simultaneously allows to absorb pair-wise as well as origin-by-time and destination-by-time FEs.

¹ Division along the median of the distribution of religious proximity between country i and country n .

² Division along the median of the distribution of Common Spoken Language as in Melitz and Toubal (2014) between country i and country n .

³ Division along the Composite Index of Linguistic Proximity as in Adsera and Pytlikova (2015) between country i and country n .

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

1.6 Conclusions

Many countries are pursuing policies to attract foreign direct investments because they reckon FDI will contribute to their economic growth by creating a more competitive business environment, triggering technology spillovers, increasing capital accumulation and generating more job opportunities. The growth-enhancing role of FDI is well documented in the literature and is particularly evident for developing countries. Over the last 15 years the share of FDI originating from developing countries over total flows has increased from 8% to 26% while recent research has showed that much of this investment takes place between developing economies (Gold et al., 2017)

The overall economic benefits of FDI have motivated a thorough investigation of its determinants and CP has been established as an important driver of the firm’s decision to invest abroad. However, the definition of CP used assumed that it was symmetric and stable over time. The resulting standard measures of CP - including the composite indexes (as the one proposed by Kogut and Singh, 1988, based on Hofstede, 2003’s cultural dimensions) - employed in the existing empirical studies are therefore inadequate to capture a broader and more refined notion of CP.

In this paper we have assessed the effect of CP on greenfield FDI explicitly accounting for its asymmetric and time-dependent dimensions. In line with Disdier et al. (2010), we used bilateral trade in cultural goods as a proxy for asymmetric and time-dependent CP. The exercise contributes to the literature as the effects of asymmetric bilateral cultural measures remain largely understudied and the few papers that include FDI as outcome variable as well as an asymmetric measure of bilateral cultural relationship have been confined mainly to samples of OECD economies. The use of two comprehensive datasets on trade and greenfield FDI - namely BACI (CEPII) and Financial Times FDI Market dataset, respectively - allows the present study to feature a very extended country coverage which also includes South-South FDI, for which CP may be particularly relevant.

Relying on the PPML estimation technique with high-dimensional fixed effects our results have shown that asymmetry in cross-country cultural proximity matters for FDI flows: more precisely, investment projects from a source to a destination country tend to increase more with cultural exports from source to destination rather than with imports. In other words, the evidence points to a stronger role of the cultural attractiveness of the country where the investment is coming from for individuals in the destination economy. This result suggests that higher relevance in explaining patterns of FDI should be attributed to the cultural preferences of the individuals in the destination country, both as consumers potentially buying the outcome produced by the subsidiary as well as voters, affecting the allocation of political pressures across competing investment projects.

Our analysis leaves at least two interesting questions open to future research. First, while the study of asymmetry in CP is limited in the context of the present paper to a descriptive assessment, it undoubtedly proves that such phenomenon exists in the data, namely that cultural relationships are indeed asymmetric. More can be done to identify a statistically robust and convincing measure of the degree of asymmetry in cultural relationships and to study its determinants and effects in the realm of economic phenomena. Second, our findings shed new light on the role played by individuals in the destination country to trigger inward FDI. While this paper focuses on the cultural dimension of these preferences, further theoretical investigation can be conducted to broadly assess their contribution within a fully micro-founded general equilibrium model of bilateral FDI.

Bibliography

- Abu-Lughod, Lila**, "Writing against culture," in Richard G. Fox, ed., *Recapturing Anthropology: Working in the Present*, School of American Research Press., 1996, pp. 137–162.
- Adsera, Alicia and Mariola Pytlikova**, "The role of language in shaping international migration," *The Economic Journal*, 2015, 125 (586), F49–F81.
- Alesina, Alberto and Paola Giuliano**, "Culture and Institutions," *Journal of Economic Literature*, December 2015, 53 (4), 898–944.

- Anderson, James E. and Eric Van Wincoop**, “Gravity with gravitas: a solution to the border puzzle,” *American Economic Review*, 2003, *93* (1), 170–192.
- Appadurai, Arjun**, *Modernity at Large: Cultural Dimensions of Globalization*, Minneapolis-London: University of Minnesota Press, 1996.
- Baldwin, Richard and Daria Taglioni**, “Gravity for dummies and dummies for gravity equations,” Technical Report, National Bureau of Economic Research 2006.
- **and Javier Lopez-Gonzalez**, “Supply-chain Trade: A Portrait of Global Patterns and Several Testable Hypotheses,” *The World Economy*, 2015, *38* (11), 1682–1721.
- Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga**, “A practitioners’ guide to gravity models of international migration,” *The World Economy*, 2016, *39* (4), 496–512.
- Bertoli, Simone and Jesús Fernández-Huertas Moraga**, “Multilateral resistance to migration,” *Journal of Development Economics*, 2013, *102*, 79–100.
- Blonigen, Bruce A. and Jeremy Piger**, “Determinants of Foreign Direct Investment,” *Canadian Journal of Economics*, 2014, *47* (3), 775–812.
- Briant, Anthony, Pierre-Philippe Combes, and Miren Lafourcade**, “Product complexity, quality of institutions and the protrade effect of immigrants,” *The World Economy*, 2014, *37* (1), 63–85.
- Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer**, “The trade-creating effects of business and social networks: evidence from France,” *Journal of international Economics*, 2005, *66* (1), 1–29.
- de Sousa, José and Julie Lochard**, “Does the Single Currency Affect Foreign Direct Investment?,” *The Scandinavian Journal of Economics*, 9 2011, *113* (3), 553–578.
- Desbordes, Rodolphe and Shang-Jin Wei**, “The effects of financial development on foreign direct investment,” *Journal of Development Economics*, 2017, *127*, 153–168.
- Disdier, Anne-Célia, Silvio H. T. Tai, Lionel Fontagné, and Thierry Mayer**, “Bilateral trade of cultural goods,” *Review of World Economics*, 2010, *145* (4), 575–595.
- Eaton, Jonathan and Samuel Kortum**, “Trade in capital goods,” *European Economic Review*, 2001, *45* (7), 1195 – 1235. International Seminar On Macroeconomics.
- Feenstra, Robert C.**, *Advanced international trade: theory and evidence*, Princeton university press, 2015.
- Felbermayr, Gabriel J. and Farid Toubal**, “Cultural proximity and trade,” *European Economic Review*, 2010, *54* (2), 279 – 293.
- Fernández, Raquel**, “Culture and Economics,” in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, Basingstoke: Palgrave Macmillan, 2008.

- , “Does Culture Matters?” in Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds., *Handbook of Social Economics*, Vol. 1A, Elsevier, Elsevier, North-Holland, 2011, pp. 481–510.
- Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon**, “Genetic distance, transportation costs, and trade,” *Journal of Economic Geography*, 2014, *14* (1), 179–198.
- Gold, Robert, Holger Görg, Aoife Hanley, and Adnan Seric**, “South–South FDI: is it really different?,” *Review of World Economics*, 2017, pp. 1–17.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, “Does Culture Affect Economic Outcomes?,” *Journal of Economic Perspectives*, June 2006, *20* (2), 23–48.
- , – , and – , “Cultural biases in economic exchange?,” *The Quarterly Journal of Economics*, 2009, *124* (3), 1095–1131.
- Head, Keith and John Ries**, “FDI as an outcome of the market for corporate control: Theory and evidence,” *Journal of International Economics*, 2008, *74* (1), 2 – 20.
- and **Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Hellmanzik, Christiane and Martin Schmitz**, “Taking gravity online: The role of virtual proximity in international finance,” *Journal of International Money and Finance*, 2017, *77*, 164 – 179.
- Hofstede, Geert**, *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*, Sage publications, 2003.
- House, Robert J., Paul J. Hanges, Mansour Javidan, Peter W. Dorfman, and Vipin Gupta**, *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*, Sage publications, 2004.
- Javorcik, Beata S., Çağlar Özden, Mariana Spatareanu, and Cristina Neagu**, “Migrant networks and foreign direct investment,” *Journal of Development Economics*, 2011, *94* (2), 231–241.
- Kogut, Bruce and Harbir Singh**, “The effect of national culture on the choice of entry mode,” *Journal of international business studies*, 1988, pp. 411–432.
- Kukush, Alexander, Hans Schneeweis, and Roland Wolf**, “Three estimators for the Poisson regression model with measurement errors,” *Statistical Papers*, 2004, *45* (3), 351–368.
- Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin**, “The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects,” 2017. Working Paper.
- Lee, Hyun-Hoon and John Ries**, “Aid for Trade and Greenfield Investment,” *World Development*, 2016, *84*, 206–218.

- Leung, Kwok, Rabi S. Bhagat, Nancy R. Buchan, Miriam Erez, and Cristina B. Gibson**, “Culture and International Business: Recent Advances and Their Implications for Future Research,” *Journal of International Business Studies*, 2005, 36 (4), 357–378.
- Li, Chengguang, Felix C. Brodbeck, Oded Shenkar, Leonard J. Ponzi, and Jan Hendrik Fisch**, “Embracing the foreign: Cultural attractiveness and international strategy,” *Strategic Management Journal*, 2017, 38 (4), 950–971.
- McIntosh, Colin**, *Cambridge Advanced Learner’s Dictionary and Thesaurus*, Cambridge University Press, 2017.
- Melitz, Jacques and Farid Toubal**, “Native language, spoken language, translation and trade,” *Journal of International Economics*, 2014, 93 (2), 351–363.
- **and** –, “Somatic Distance, Trust and Trade,” Working Papers 2018-11, CEPPII September 2018.
- Santos Silva, João M C and Silvana Tenreyro**, “The log of gravity,” *The Review of Economics and statistics*, 2006, 88 (4), 641–658.
- **and** –, “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator,” *Economics Letters*, 2011, 112 (2), 220–222.
- Shenkar, Oded**, “Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences,” *Journal of international business studies*, 2001, 32 (3), 519–535.
- Tung, Rosalie L. and Alain Verbeke**, “Beyond Hofstede and GLOBE: Improving the quality of cross-cultural research,” *Journal of International Business Studies*, Oct 2010, 41 (8), 1259–1274.
- UNCTAD**, *Creative Economy Report. A Feasible Development Option*, UNDP, 2010.
- , “51 World Investment Report,” Technical Report, UNCTAD 2017.

Appendix

1.A Data: sources and general features

The data used throughout both the descriptive and the analytical parts of the paper come from a variety of sources. Table A-1 displays the major sources and related links where additional information on the different databases used to create our final dataset: most of the other data come from sources that are well known in empirical gravity literature.

The focus of the analysis is on testing the role and the extent of the non-reciprocal component of CP on international economic flows, with the specific focus on greenfield FDI. For this reason we aggregate the projects according to the country of origin, destination and year in which the investment has been made. Then, we label missing dyadic flows at this stage as null investment channels, to obtain a square bilateral FDI matrix accounting for 184×185 countries of origin and destination. Cultural Trade data are then merged accordingly. Given that some territorial units in fDIMarket are not matched in BACI, some countries are dropped throughout the empirical analysis (see Table G-1 in Appendix 1.B with the complete list of unmatched and excluded countries). In this respect, our strategy is similar to the one adopted by Aubry et al. (2014), Desbordes and Wei (2017), and Lee and Ries (2016) among the others. As a consequence, our FDI data reveals a pattern that is consistent with the findings from the recent theoretical and empirical literature in international economics (see for instance Mayer and Ottaviano, 2008), i.e. that only few firms are able to undertake FDI as a form of internationalization.²⁹

However, the databases related to our variables of interest, cultural trade and greenfield FDI respectively, present some peculiarities that demand for some crucial choices in terms of data aggregation and classification, in order to obtain the least distortionary measures possible. In the remaining of this section, we explore the main issues related to cultural trade (that constitute our main variable of interest) and greenfield FDI respectively.

Data on trade in cultural goods Trade data come from the BACI dataset by CEPII³⁰, a proper workhorse in empirical gravity analysis in international trade. It is not the purpose of this appendix to describe the features of the BACI dataset as it is, for which we suggest the interested reader to check directly on the web link provided in Table A-1 above. Much more interesting for the purpose of this paper is to define what can be labelled as *Cultural Good* and what classification scheme is better able to fit to the purpose of this paper that is, to investigate the role of imperfect reciprocity in cultural proximity in international economic flows.

Many countries and international organizations developed their own classification scheme, based

²⁹As this is particularly true for greenfield FDI, the result is that null bilateral flows account for more than 94% of the possible bilateral channels in our dataset. See Table A-2 below for a detailed report concerning the incidence of null flows in our dataset.

³⁰http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=1

Table A-1: Main Sources of Data used in the Empirical Section

Variables	Dataset / Source / Website / Reference and Accessibility
FDI Variables	FDIMarket / FDI Intelligence Unit, The Financial Times / http://www.fdiintelligence.com/ / FDI Market License
Trade Variables	BACI / CEPII / http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=1 / UN COMTRADE access required
Gravity Variables	Gravdata / CEPII / http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8 / Free
Bilateral Distance	Geodist / CEPII / http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6 / Free
Migrant Stock	WB Global Bilateral Migration Dataset / The World Bank / http://data.worldbank.org/data-catalog/global-bilateral-migration-database / Artuç et al. (2015) / Free
Language I	Lingweb / CEPII / http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=19 / Melitz and Toubal (2014) / Free
Language II	Data S1 / The Economic Journal / http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12231/abstract / Adsera and Pytlikova (2015) / Free
Cultural Distance	Hofstede Index / The Journal of Population Economics / https://link.springer.com/article/10.1007/s00148-011-0356-x / Belot and Ederveen (2012) / Free
BITs	UNCTAD Investment Policy Hub / http://investmentpolicyhub.unctad.org/IIA / Free
CPIA	Country Policy and Institutional Assessment / The World Bank / https://data.worldbank.org/data-catalog/CPIA / Free

Notes: This table lists the main sources in the data used throughout the dataset. Additional information are available upon request to the corresponding author. Concerning the sources of the single variables referring to a particular dataset used in this paper, the authors encourage to search directly in the websites provided.

on precise principles and content of the single class of product: for this reason, identifying the most suitable scheme for the sample considered in this paper is not an easy task. Yet, the choice of the classification is particularly sensible. Given the world coverage of our analysis, we restricted our search to two alternative classifications for cultural goods promoted by United Nations agencies, the UNESCO and the UNCTAD,³¹ each of them based on different criteria and different categories of goods to be included in the count. Disdier et al. (2010) classified cultural goods using the definition proposed by UNESCO. Despite we build upon their seminal work, we depart from that approach and adopt the scheme proposed by UNCTAD (2010). There exist two main reasons for this choice: (i) a technicality related to the *time coverage* of the data, and (ii) a more substantial issue concerning the *sample selection*.

As for time coverage, the decision to prefer the UNCTAD classification leans on the different coding system adopted by the two different classifications. With respect to this point, UNESCO adopts the 2007 Harmonized Commodity Description and Coding System (HS 2007), that would call for the adoption of a conversion table to arrange the data along our time period. Conversely, UNCTAD (2010) adopts the HS 2002 coding system, that is more suitable for the time period at stake, as it allows not to convert the trade flows prior of 2007.³² The conversion may distort the data, since the way they are collected is not always consistent across different coding systems: for this reason, the adoption of the UNCTAD (2010) classification could turn out to be not only less burdensome from a computational point of view, but also less prone to distortions.

Much more relevant for the choice of the classification scheme is the the sample coverage issue. The dataset used throughout this paper has global coverage,³³ with a large number of developing and transition economies in addition to developed ones. Conversely, Disdier et al. (2010) confine

³¹Other criteria can be found in the classification schemes developed by national and smaller international institutions (see UNCTAD, 2010, for a review).

³²Nonetheless, as we adopted lag values of cultural trade as instruments in our IV analysis, we could not eventually avoid the burden of converting trade data prior to 2002. See Section 1.4.2.

³³See Appendix 1.G for the list of excluded countries.

their analysis to a much more homogeneous group of OECD countries. This could not seem a major concern, but it is important to acknowledge that cultural goods are neither homogeneous nor equally produced worldwide. Both UNESCO and UNCTAD classifications uphold this fact by splitting cultural goods into “**core**” and “**optional**” cultural goods, with the former generally dominated by developed economies. By construction, in both classifications “optional” cultural goods encompass a wide range of products that are more likely to be produced in, and traded by developing countries too.³⁴

A potential drawback of the wider conception of what can be considered as cultural good is that the UNCTAD classification has a much more diluted cultural content when compared to the UNESCO’s. In fact, despite the latter encompasses a narrower set of traded goods, they are the ones with the larger cultural content. Nonetheless, given the world coverage of our sample, developed countries account for less than 30% of the whole set of countries included. For this reason, in order to balance the cultural composition of trade flows, and to construct a comparable measure of cultural trade across different development stages, the classification that is able to guarantee a relatively higher weight to those goods more evenly distributed across developed, developing and least developed economies should be preferred. This problem was not relevant in Disdier et al. (2010) because of the relative homogeneity of the sampled countries. Comparing the two classifications suggests that “core” goods account for 60% of total cultural goods in UNESCO’s classification; barely 20% in UNCTAD’s. For this reason, “[...]the UNESCO classification is better at capturing the experience of countries in the global North, while UNCTAD’s better reflects opportunities for countries in the South.[...]” (UNCTAD, 2010, p. 111). This issue is more explicitly tackled in Section 1.4.2, where separate regressions on “core” and “optional” trade are run separately and compared to the results of our benchmark specification, where cultural goods encompasses both groups of goods.

Greenfield FDI data Data on FDI (that constitute the dependent variable in our empirical analysis) come from the *fDIMarket* database, that includes a detailed collection of all (and only) greenfield investments occurred worldwide in the period 2003-2014 (the first available year for greenfield FDI and the last year available for cultural trade data - our variable of interest - respectively). In figures, *fDIMarket* contains more than 169,000 investment projects, carried on by roughly 67700 different companies worldwide in the period considered. The dataset include a large amount of information related to each recorded investment, included the the declared capital expenditure and the estimated number of jobs created at the moment the investment is carried out. Beyond the “quantitative” information, the dataset includes several additional investment-level entries such as location (up to NUTS 3 level of disaggregation), economic activity of the parent company as well as the (broad) sector in which such activity can be associated to in the host country. The high level of detail would ideally allow a much finer aggregation than the broad national-sectoral unit most of the more common datasets allow, but this type of analysis goes beyond the scope of the current research.

However, despite the exceptionally wide coverage of the dataset and its reliability in terms of missing records,³⁵ *fDIMarket* data present some important issues that worth to be introduced.

³⁴The definition of “core” goods made in UN agencies’ and sovra-national organizations’ classifications in general derive from this consideration, since most of the minor classification tend to include those “high cultural content” goods in their schemes. Conversely, “optional” goods refers to those goods that are included by certain countries or agencies’ classification, but not by others (the inclusion of a class of goods depends on the productive system of the country that develop the classification). However, since all those schemes refer to developed countries, they tend to mirror goods prevalently traded by advanced economies, leaving apart those goods that may have a diluted cultural content.

³⁵UNCTAD itself bases part of its investments’ reports on *fDIMarket*’s figures. Not only, the database constitutes one of the sources of the UNCTADSTAT’s FDI dedicated section.

The first issue relates to the cross sectional dimension: Table A-2 shows the incidence of null flows over the full set of potential country pairs in the dataset, at a yearly break down. The estimation via OLS is therefore excluded by the zero-inflated structure of the full dataset, that would distort the estimates downward (see for instance Head and Mayer, 2014, for a thorough discussion on the choice of the correct estimator for gravity analysis in the context of zero-inflation). To the best of our knowledge, the incidence of null flows in the full dataset is larger than any other previous study: nonetheless, in the empirical section the sample is reduced by the estimation routines to those observations for which the FDI flow is non-zero in at least one year out of 12. This refinement substantially reduces the amount of zeroes to slightly less than 70%, allowing us to obtain consistent estimates via PPML (See Santos Silva and Tenreyro, 2011, for a comprehensive proof of the consistency of PPML estimator in presence of both over-dispersion of the data and over-inflation of null values in the dependent variable.).

Table A-2: Percentage of “Zeroes” by Year

Year	Null	Non-Null	Total	Incidence
2003	32,453	1,587	34,040	95.34%
2004	32,442	1,598	34,040	95.31%
2005	32,405	1,635	34,040	95.20%
2006	32,289	1,751	34,040	94.86%
2007	32,151	1,889	34,040	94.45%
2008	31,751	2,289	34,040	93.28%
2009	31,960	2,080	34,040	93.89%
2010	31,931	2,109	34,040	93.80%
2011	31,833	2,207	34,040	93.52%
2012	31,916	2,124	34,040	93.76%
2013	31,756	2,284	34,040	93.29%
2014	31,901	2,139	34,040	93.72%
Total	384,788	23,692	408,480	94.20%

Notes: This table breaks down the incidence of null flows by year. It becomes apparent that the issue of null flows is pervasive in the *FDIMarket* dataset as we constructed it. The high incidence of zeroes and the data over-dispersion in the sample prevent us from using OLS. We resort to use a PPML estimation technique as suggested by Santos Silva and Tenreyro, and raised to workhorse strategy by authors (see for instance Head and Mayer (2014), Yotov et al. (2016) among the others.

The second issue concerns the reliability of the “quantitative” information available, namely the *Capital Expenditure* (CAPEX) and *Job Creation* entries. Section 1.3.1 provides a theoretical justification for the use of count instead of the value of FDI flows as dependent variable: nonetheless, being able to test the theoretical prediction about the role of asymmetry in CP would call for a comparison across different measures of bilateral FDI. *fdIMarket* database is one of the few existing datasets that could potentially allow for this issue. Nonetheless, such an exercise calls for additional prudence: as stressed by both Desbordes and Wei (2017) and Lee and Ries (2016), *fdIMarket* collects information on all existing greenfield FDI *projects* as they are officially released by the respective investing companies. Unfortunately, in most of the cases no communication is made about the true CAPEX value. In all those cases, CAPEX is imputed according to an algorithm summarily described on the *DIMarket*’s website. Such imputation is likely to introduce non-trivial distortions in the data, the more relevant (a) the higher is the percentage of estimated projects over the total number of projects in a given bilateral corridor; (b) the lower the number of projects from the country of origin. Table A-3 below provides the tabulation of the projects for which only the *imputed* CAPEX was available, broken down by year. Given the incidence of estimated observations, we suggest a particular care when handling

estimates obtained using value related dimensions as dependent variables, though the picture they provide may be particularly interesting. In Section 1.4.2, those results are presented and commented in light of our measure of CP.

Table A-3: Percentage of Imputed Values by Year

Year	Imputed	Real Value	Observations	Incidence
2003	6,325	3,182	9,507	67%
2004	7,270	3,143	10,413	70%
2005	7,849	2,883	10,732	73%
2006	9,534	3,301	12,835	74%
2007	8,968	4,006	12,974	69%
2008	13,416	3,794	17,210	78%
2009	12,063	2,723	14,786	82%
2010	12,843	2,629	15,472	83%
2011	14,101	2,757	16,858	84%
2012	13,088	2,181	15,269	86%
2013	14,319	2,399	16,718	86%
2014	13,044	2,344	15,388	85%
Total	132,820	35,342	168,162	79%

Notes: The table report the percentage of estimated capital investment. The number of observations refers to the number of single projects collected by FDI Market for the period 2003-2014. The large incidence of estimated values makes the estimates obtained using values as dependent variables not fully reliable: as a matter of facts, in addition to the lack of clarity in the imputation technique, imputation brings in a component of uncertainty per se.

1.B Cultural trade as a proxy of the symmetric component of CP

Building upon Disdier et al. (2010), we identified the exchange of cultural goods as classified by UNCTAD (2010) as a good proxy of CP. In this Appendix we show how trade in cultural goods strongly relates to the symmetric component of CP as defined in Section 1.2. In other words, we provide a rough indication of the dependency of cultural attractiveness on cultural similarities. To that end we regress cultural trade on various conventional symmetric (and time invariant) proxies for cultural distance such as a dummy for *common border* $contig_{ni}$, the *log of weighted distance* $\ln dist_{ni}$, a measure of *religious proximity* $relig_{ni}$, a dummy rta_{ni} , which takes the value of 1 if both countries belong to a *regional trade agreement*, 0 otherwise, a binary variables for *common legal origin* $comleg_{ni}$, and finally a binary variable for *past colonial relationship* $colony_{ni}$ which takes the value of 1 if the two countries have ever been in a colonial relationship, 0 otherwise. All these variables are sourced from CEPII databases. Among the covariates the regression also includes a time varying component ($\ln mig_{ni,t}$), namely the *stock of bilateral immigrants* resident in the exporting country (Source: World Bank). Because data are available every 10 years (with the notable exception of the year 2013), our empirical exercise is a Pooled regression for the years 2010 and 2013 only, which nonetheless guarantees a still reasonably high number of observations. Furthermore, as in Felbermayr and Toubal (2010) we enrich the number of proxies by adding more refined measures of linguistic proximity obtained from Melitz and Toubal (2014): along with the standard dummy that equals 1 if a two countries share the same official language and 0 otherwise (COL_{ni} “*common official language*”), we include CSL_{ni} “*common spoken language*” as the probability that a pair of people at random from two countries understands one another in some language and CNL_{ni} “*common native language*” as the probability that a random pair from two countries speak the same native language. Lastly, we employ a comprehensive measure of cultural distance widely used in the literature, namely the *Hofstede Index* $Hofstede_{ni}$ (Hofstede, 1991). This composite Index has been one of the main workhorses for the empirical of test the impact of cultural proximity on economic exchanges such as trade and FDI (see for instance Du et al. (2012)), but other than being at the same time pre-determined and symmetric, has the drawback of covering a fairly limited sample (see for a discussion Shenkar (2001)). The data are from Belot and Ederveen (2012).³⁶ Results for this exercise are reported in Table B-1.

The estimates in Table B-1 indicate that trade of cultural goods relates to almost all the proxies of CP we included, whose impacts have the expected sign. The first column reports the OLS results with log of imports of cultural goods as dependent variable. The coefficients are all statistically significant with the exception of CNL_{ni} : this is likely to be imputed to an high degree of collinearity between linguistic distance measures. The loss of information of zero bilateral trade due to the logarithmic specification could be a serious concern in our case, as the zeros in trade of cultural goods stand for a large share of the total available information. The main issue with the elimination of the zeros is a possible selection bias. Indeed, it might be that proxies for cultural proximity are associated with the intensity of trade in cultural goods only in the instances of positive trade and have no role in explaining the cases of the zeros. To address this issue we report the PPML results in Column 2. Despite the change in the sample size, almost all the effects retain the expected sign. The only exceptions are the measures of language proximity and the RTA dummy that, in any case, maintain a pairwise correlation coefficient that is positive and statistically different from zero (see Table B-2 below). Lastly, the inclusion of the Hofstede Index in the third column causes a considerable loss of information as the sample reduces to 19 OECD countries. The Index seems to be capturing most of the effect of religious and linguistic proximity and - most importantly for our purposes - is negatively related to the

³⁶See Belot and Ederveen (2012) for the details related to the construction of the Hofstede Index. See Section 1.A for a more thorough description of the data and the complete list of sources and data accessibility.

Table B-1: Testing the Validity of Cultural Trade as a Proxy of CP

Dep. Var.	$\ln \text{CultIMP}_{ni,t}$	$\ln \text{CultIMP}_{ni,t}$	$\ln \text{CultIMP}_{ni,t}$
	(1)	(2)	(3)
$\ln \text{mig}_{ni,t}$	0.115*** (20.83)	0.0761*** (4.30)	0.0880** (2.89)
$\ln \text{dist}_{ni}$	-1.225*** (-49.15)	-0.695*** (-10.61)	-0.921*** (-6.77)
contig_{ni}	0.317*** (3.74)	0.260** (2.86)	0.440* (2.34)
$\text{FTA}_{ni,t}$	0.266*** (6.24)	0.0807 (0.77)	0.683** (2.96)
comrelig_{ni}	0.236*** (3.55)	0.440* (2.28)	0.235 (1.26)
comleg_{ni}	0.281*** (8.66)	0.303*** (4.43)	0.411** (2.68)
colony_{ni}	0.500*** (5.67)	0.383*** (3.65)	0.763*** (3.45)
COL_{ni}	0.374*** (6.13)	0.0786 (0.55)	-0.0000199 (-0.00)
CSL_{ni}	0.683*** (6.52)	-0.350 (-1.45)	-0.394 (-0.74)
CNL_{ni}	0.0691 (0.48)	0.209 (0.71)	-0.402 (-0.92)
Hofstede_{ni}			-1.034*** (-4.01)
Imp×Year FE	✓	✓	✓
Exp×year FE	✓	✓	✓
Sample	Full	Full	Reduced
Obs	24620	54525	684
% Zeros	-	0.5485	-
R ²	0.7476	0.8993	0.9118
Estimator	OLS	PPML	OLS

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t (z) -statistics in parentheses. Standard errors are clustered by trading-pair. The model includes importer×time and exporter×time FEs. The first and third columns' estimates are estimated with OLS. The sample size in this table reflect the way the different estimators deal with null flows as well as the sample size. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed. The sample in the third column is reduced due to those countries for which the Hofstede Index of Cultural Proximity is available (see Belot and Ederveen, 2012).

imports of cultural goods.

Table B-2: Testing Validity of Cultural Trade as a Proxy of CP - Correlations

Correlation with:	cult.trade $T_{ni,t}$	
	Baseline Covariates Set	Linguistic and CP proxies
	(1)	(2)
$\ln mig_{ni}$	0.0955* (0.0000)	0.0955* (0.0000)
$\ln dist_{ni}$	-0.0218* (0.0000)	-0.0218* (0.0000)
$contig_{ni}$	0.0771* (0.0000)	0.0771* (0.0000)
$FTA_{ni,t}$	0.0363* (0.0000)	0.0363* (0.0000)
$comrelig_{ni}$	-0.0049 (0.2433)	-0.0049 (0.2433)
$comleg_{ni}$	-0.0037 (0.3691)	-0.0037 (0.3691)
$colony_{ni}$	0.0265* (0.0000)	0.0265* (0.0000)
$lang_{ni}$	0.0130* (0.0018)	
COL_{ni}		0.0101* (0.015)
CSL_{ni}		0.0359* (0.0000)
CNL_{ni}		0.0275* (0.0000)
$Hofstede_{ni}$		-0.2507* (0.0000)
Obs	57672	703

Notes: * $p < 0.01$. SE in parentheses are clustered by trading-pair. The table show pairwise correlation coefficients between trade in cultural goods and all standard coefficients of proximity. Coefficients in the first column refers to the whole sample for which all variables are available. This means that it is limited to just year 2010 and year 2013 because of bilateral stock of migrants availability. Coefficients in the second column refers instead to the reduced sample for which the Hofstede index is available.

1.C Extensions to the detour on asymmetry

Asymmetry in CP and export capacity This Appendix investigates the correlation between the degree of asymmetry in CP and the relative cultural export capacity between trading partners. This is done by dividing the set of countries which appear in at least one pair for which a value of asymmetry is available into four classes, depending on the value of their exports of cultural good with respect to the 3 quartiles of the distribution of cultural exports. The first class consists of countries below the first quartile of cultural exports, the second class of those between the first and the second quartile, the third class of those between the second and third quartile, and finally the fourth class of those countries above the third quartile of the distribution. The set of country pairs are then partitioned according to all possible combinations of two elements with repetitions from the four classes defined above. One pair could be classified either as containing two first class countries (both at the bottom of the cultural export distribution), one first and one fourth class country (the former at the bottom and the latter at the top of the cultural trade distribution) and so on and so forth for all 10 possible combinations. Finally, the value of asymmetry is regressed on the ten dummies identifying the elements of this partition (First-First, Second-Second, . . . , First-Second, . . .), taking those pairs with two bottom cultural exporters (First-First) as the base group. Results are reported in Table C-1.

Looking at the first column of Table C-1, we notice that on average across all pairs including two bottom cultural exporters the value of asymmetry is equal to 2.078, below both the mean and median values of asymmetry, equal to 2.932 and 2.614 respectively. Less asymmetry appears to be present in the CP between countries with a similar but higher value of cultural exports, and also between a country in the fourth class (top cultural exporter) and one in the third (quasi-top cultural exporter). Higher levels of asymmetry in CP instead are expected among countries which are relatively more heterogeneous in terms on cultural export capacity. Higher asymmetry in bilateral CP is associated with wider heterogeneity in export capacity and, to a lesser extent, with average export capacity within the pair. These patterns are generally confirmed when restricting the analysis to bilateral cultural relationships characterized by attractiveness premia with the same sign (both positive and negative) as well as with different sign. These results are presented in the second, third and fourth columns of Table C-1.

Asymmetry across different samples The motivation of this extension is to show how the width and degree of homogeneity within the sample of countries may be crucial when the impact of CP on the economy is investigated. We argue that the empirical assessment of the role of asymmetric CP for economic transactions needs to be conducted with the widest possible country coverage. A empirical analysis conducted on a narrow and homogeneous set of countries could potentially overestimate the degree of asymmetry embedded in cultural relationship and therefore undermine the assessment of the role of such asymmetric component in determining economic outcomes. In order to show this we replicate the construction of our empirical measure of asymmetry in CP starting from a sample with the same country coverage of the database used by Felbermayr and Toubal (2010) to construct their asymmetric measure of CP based on Eurovision Song Contest scores.³⁷ This subsample includes only European countries, that can be considered as a relatively homogeneous group under many respects, and especially when compared with the rest of the World. We denote by $|\hat{\gamma}_{ni}^{full} - \hat{\gamma}_{in}|^{FT}$ the resulting measure of asymmetry. $|\hat{\gamma}_{ni} - \hat{\gamma}_{in}|^{full}$ indicates instead the asymmetry whose components have been estimated on the whole sample. Table C-2 reports both measures of asymmetry and their difference for a number of selected country pairs. The + and - signs below the first two columns reflect the sign of the attractiveness premium exerted by country i and country n on each other. Take

³⁷The country coverage is identical with the exception of Yugoslavia due to availability of cultural trade data.

Table C-1: Asymmetry Across Different Types of Cultural Traders

Dep. Var.	Asymmetry $ \hat{\gamma}_{ni} - \hat{\gamma}_{in} $			
	All types	Both positive	Both negative	Opposite sign
Attractiveness premia	(1)	(2)	(3)	(4)
Second-Second	-0.400** (-3.13)	-0.279 (-1.35)	-0.0767 (-0.54)	-0.561** (-2.75)
Third-Third	-0.610*** (-4.90)	-0.143 (-0.74)	-0.946*** (-5.45)	-0.399 (-1.60)
Fourth-Fourth	-0.828*** (-5.59)	-0.172 (-0.82)	- -	- -
First-Second	1.048*** (6.77)	1.104*** (3.96)	0.299* (2.10)	1.532*** (7.12)
Second-Third	0.188 (1.56)	0.00573 (0.03)	0.110 (0.79)	-0.00420 (-0.02)
Third-Fourth	-0.586*** (-4.75)	0.0328 (0.17)	- -	0.973 (1.18)
First-Third	1.682*** (12.21)	1.380*** (4.79)	0.889*** (6.50)	1.721*** (9.06)
Second-Fourth	0.779*** (5.97)	0.607** (3.07)	1.093 (1.12)	0.889*** (4.61)
First-Fourth	2.690*** (21.84)	1.270*** (5.07)	1.651*** (10.23)	2.043*** (11.96)
Constant (First-First)	2.078*** (19.70)	1.423*** (7.76)	1.392*** (12.86)	3.194*** (20.20)
Obs	4137	1486	793	1858
R ²	0.3424	0.1274	0.2285	0.2421
Estimator	OLS	OLS	OLS	OLS

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. In this table the proxy for asymmetry ($|\hat{\gamma}_{ni} - \hat{\gamma}_{in}|$) is regressed on a constant and 9 dummies. As an illustration, the dummy "Fourth-Fourth" takes value one for those country pairs where both countries have a value of cultural exports above the third quartile of the distribution of cultural exports. As a further illustration the dummy "First-Fourth" takes value one for those country pairs where one country is a bottom exporter of cultural goods (below the first quartile of the cultural exports distribution) and the other is a top cultural exporter (above the third quartile). When point estimates and t statistics are not reported it is because the respective dummy coefficient has no variability (always equal to 0) in the corresponding estimation sample. The case in which both countries in the pair are bottom exporters (below the first quartile of the cultural exports distribution) is set as base level and the related dummy variable is omitted from the regression.

for instance the UK and France. The asymmetry computed from the whole sample is very low and equal to 0.17. The first + sign below the asymmetry score indicates that the attractiveness premium that France exerts on the UK with respect to the average country is positive. The same is true the other way round, as indicated by the second + sign. When computed on a smaller sample featuring only European countries, the value of asymmetry increases by more than 180% and becomes equal to 0.48 (still relatively small compared to the average asymmetry over the whole sample).

The last column of the table shows the extent of the bias induced by considering only a subsample of (relatively) homogenous countries: a negative sign in the difference between $|\hat{\gamma}_{ni} - \hat{\gamma}_{in}|^{full}$ and $|\hat{\gamma}_{ni} - \hat{\gamma}_{in}|^{FT}$ means that the degree of asymmetry in the country pair under consideration decreases when other, more heterogeneous countries are considered. Failing to consider the role of the rest of the world within the system of cultural affinity could result in a severe bias in cultural relationship between countries.

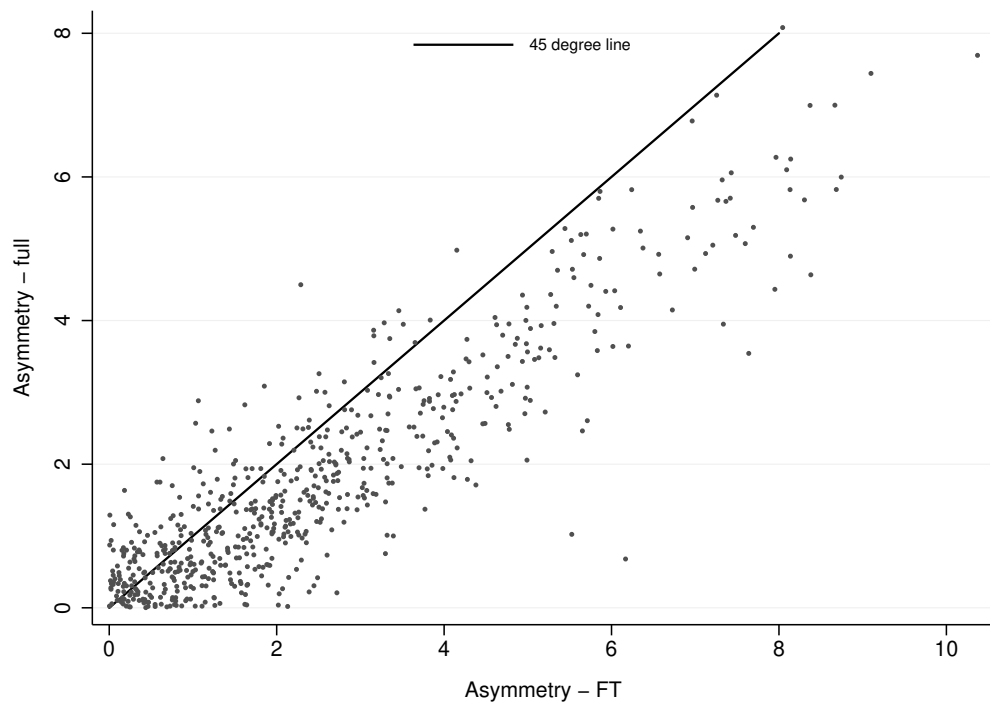
Table C-2: Asymmetry Across Different Samples

Country n	Country i	Asymmetry - full $ \hat{\gamma}_{ni} - \hat{\gamma}_{in} ^{\text{full}}$	Asymmetry - FT $ \hat{\gamma}_{ni} - \hat{\gamma}_{in} ^{\text{FT}}$	Differential $ \hat{\gamma}_{ni} - \hat{\gamma}_{in} ^{\text{full}} - \hat{\gamma}_{ni} - \hat{\gamma}_{in} ^{\text{FT}}$
Finland	Italy	1.16 + +	2.35 + +	-1.19
United Kingdom	France	0.17 + +	0.48 + +	-0.31
Russian	United Kingdom	0.95 + +	1.60 + +	-0.65
Germany	Turkey	0.33 + +	1.46 + +	-1.13
Spain	Russian	2.19 + +	2.20 - +	-0.01
Norway	Sweden	1.49 + +	1.95 + +	-0.46
Croatia	Sweden	0.31 + +	1.89 + -	-1.58
Belgium	Malta	2.88 + +	5.02 + -	-2.14
Ireland	United Kingdom	2.70 + +	3.32 + +	-0.62
Ukraine	Ireland	3.04 + -	3.45 - -	-0.41

Notes: The table lists a selection of country pairs and shows the extent of the bias in the empirical assessment of asymmetry due to adopting a sample of relatively homogeneous countries. A positive (negative) value of the differential across the full sample and the restricted one implies that the restriction is actually over-(under-) estimating the true extent of CP. The sample of countries used in Felbermayr and Toubal (2010), which only includes European countries is taken as the restricted set of relatively homogeneous countries. The + and - signs below the two columns of symmetry report the sign of the attractiveness premium exerted by country i and country n on each other.

Beyond the few examples reported in Table C-2, Figure gives a sense of the sign of the bias on all the country pairs generated from the restricted sample for which both measures of asymmetry are estimated. This is done by plotting, for each pair the value of asymmetry coming from the full sample (on the vertical axis) against the value of the asymmetry generated from the restricted sample (on the horizontal axis). With the bulk of the observations below the 45 degree line, especially moving away from the origin, we conclude that the overestimation of asymmetry in CP implied by an empirical framework with limited country coverage can be highly widespread.

Figure C-1: Asymmetry Full Sample VS Asymmetry Felbermayr and Toubal (2010) Sample



1.D Further addressing the omitted variable bias

An important econometric issue in our regressions is the potential endogeneity of our proxy for CP which mainly arises because of the potential omission of unobserved factors that might be correlated both with the error term (and thus FDI) and with trade in cultural goods. Both the proposed IV analysis and the inclusion of dyadic FEs in Section add robustness to our estimates and confirm our main conclusions. Here we further test the consistency of our benchmark results by augmenting the specification with the inclusion of observable variables of dimension nit that might capture (part) of these unobserved time-varying dyadic factors.

A variable which potentially shapes both cultural trade as well as FDI is represented by the migrants' networks. Migrants are able to form important linkages between the country of origin and the one of destination. To this regard, the literature identified a positive impact of migrants' networks on both FDI and international trade (see for instance Javorcik et al., 2011; Gould, 1994; Giovannetti and Lanati, 2016), which is predominantly imputed to the "insider knowledge" that migrants provide to reduce information costs in international transactions. The time varying impact of migrants' networks on FDI cannot be entirely absorbed through our comprehensive set of fixed effects and its exclusion from the list of regressors may introduce an omitted variable bias.³⁸ The results are reported in Table D-1 below, that replicates the specifications of Table 1.4, but comprises the bilateral stocks of immigrants from both n to i and i to n as additional regressors.

Including the stocks of immigrants does not alter our overall conclusions. In particular, the positive impact of exports in cultural goods which proxy for the destination side mechanisms driving the firm's decision to invest is always statistically significant and does not vary in magnitude as we control for the network effect (column 1-3). In a nutshell, the destination side mechanisms driving FDI seem to be independent from the network channel. This points to the goodness of our proxy in capturing the role of cultural proximity and score in favor of its robustness to the inclusion of alternative measures of time varying CP.

Finally, in Table D-2 we augment our baseline specification with the total volume of bilateral non cultural trade. In particular, $\ln \text{bil_trade_NC}$ captures the effect of the sum of bilateral non cultural imports and exports between origin and destination at time t on FDI. The evidence from Disdier et al. (2010) shows that bilateral flows of cultural products can be highly related to the overall flows of bilateral trade, while at the same time bilateral economic exchanges such as aggregate trade are likely to be positively associated with FDI. The statistics indicate that the volume of bilateral non-cultural trade does not impact FDI and its inclusion substantially leaves our results unaffected, which we find as reassuring.

³⁸Their inclusion, however, reduces the explanatory power of our econometric exercise, as data on bilateral migrants' stocks with a global country coverage are generally only available with a 10 year interval between observation (Source: The World Bank). Therefore, we only include the migrants' stock as a robustness check.

Table D-1: Addressing Omitted Variable Bias: Including Migration

Dep. Var.	Count $C_{ni,t}$		
	(1)	(2)	(3)
$\ln \text{migstock}_{ni,t}$	0.0810*** (5.13)		0.0579** (2.63)
$\ln \text{migstock}_{in,t}$		0.0788*** (4.29)	0.0293 (1.33)
$\ln \text{CultIMP}_{ni,t}$	0.0507** (3.27)	0.0368 (1.90)	0.0204 (0.93)
$\ln \text{CultEXP}_{ni,t}$	0.290*** (15.12)	0.296*** (12.94)	0.290*** (11.37)
$\ln \text{dist}_{ni}$	-0.0566 (-1.25)	-0.0693 (-1.46)	-0.0574 (-1.13)
colony_{ni}	0.283*** (4.26)	0.308*** (4.41)	0.292*** (3.87)
lang_{ni}	0.117* (2.01)	0.0704 (1.11)	0.0725 (1.08)
comrelig_{ni}	0.930*** (7.48)	0.910*** (7.04)	0.960*** (6.82)
contig_{ni}	-0.0391 (-0.55)	-0.0447 (-0.60)	-0.0140 (-0.18)
comleg_{ni}	0.156*** (3.45)	0.189*** (3.84)	0.187*** (3.61)
$\text{FTA}_{ni,t}$	0.129 (1.94)	0.144* (2.10)	0.138 (1.84)
$\text{BIT}_{ni,t}$	0.0277 (0.51)	-0.0154 (-0.26)	-0.0315 (-0.93)
Imp×Year FE	✓	✓	✓
Exp×year FE	✓	✓	✓
Obs	9619	8756	5853
% Zeros	67%	67%	60%
R ²	0.91	0.92	0.92
Test 1	278.59	179.89	140.92
Test 2	76.53	66.75	53.26
Estimator	PPML	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows. This table replicates the baseline specification adding the bilateral stock of migrants from n to i as additional regressors. The reduced number of observations is due to the availability of the migration data, that allow to use only two points in time (2010 and 2013) for the period covered in the analysis (Source: The World Bank). TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

Table D-2: Addressing Omitted Variable Bias: Including Non-Cultural Trade

Dep. Var.	Count $C_{ni,t}$	
	(1)	(2)
$\ln \text{CultIMP}_{ni,t}$	0.0690*** (5.90)	0.0838*** (6.01)
$\ln \text{CultEXP}_{ni,t}$	0.305*** (21.91)	0.324*** (14.64)
$\ln \text{bil_trade_NC}_{ni,t}$		-0.0352 (-1.24)
$\ln \text{dist}_{ni}$	-0.179*** (-5.13)	-0.176*** (-5.08)
colony_{ni}	0.366*** (6.85)	0.367*** (6.90)
lang_{ni}	0.181*** (3.53)	0.176*** (3.50)
comrelig_{ni}	0.883*** (9.21)	0.876*** (9.21)
contig_{ni}	-0.0977 (-1.61)	-0.0947 (-1.56)
comleg_{ni}	0.153*** (4.06)	0.154*** (4.08)
$\text{FTA}_{ni,t}$	0.118* (2.19)	0.117* (2.17)
$\text{BIT}_{ni,t}$	0.0115 (0.29)	0.00749 (0.19)
Imp×Year FE	✓	✓
Exp×year FE	✓	✓
Obs	87448	87448
% Zeros	0.749	0.749
R ²	0.9221	0.9221
Test 1	585.19	214.46
Test 2	141.81	130.19
Estimator	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable “Count” $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows. The estimates in column (1) replicates column (3) in our baseline results in Table 1.4; column (2) provides the result of the same equation, augmented to include total bilateral trade of non-cultural goods. TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

1.E Relevance of the instruments

Table E-1 below mimics a first stage regression for the IV analysis, by showing the relevance of the instruments in explaining the endogenous variables to our analysis. Since the *IVPPML* command does not compute first stage regression, we regressed the endogenous variables on all the instruments as well as on the covariates of the second stage.

Table E-1: Relevance of the Instrument: First Stage Endogenous Variables on Instruments

Dep. Var.	Cult.Import _{ni,t}	Cult.Export _{ni,t}
	(1)	(2)
lnCultIMP _{ni,t-8}	0.560*** (14.73)	
lnCultEXP _{ni,8}		0.560*** (14.74)
ln dist _{ni}	-0.664*** (-9.15)	-0.663*** (-9.14)
colony _{ni}	-0.116 (-1.37)	-0.116 (-1.37)
lang _{ni}	0.123 (0.90)	0.124 (0.91)
comrelig _{ni}	0.0534 (0.44)	0.0539 (0.44)
contig _{ni}	0.0773 (1.13)	0.0776 (1.14)
comleg _{ni}	0.0481 (0.78)	0.0479 (0.78)
FTA _{ni,t}	0.324** (2.94)	0.325** (2.95)
BIT _{ni,t}	0.0485 (0.59)	0.0484 (0.58)
Imp×Year FE	✓	✓
Exp×Year FE	✓	✓
Obs	11117	11117
% Zeros	12.2%	12.2%
R ²	0.9502	0.9502
Estimator	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. This table shows the relevance of the selected instruments on the endogenous variables. The decision to adopt lagged values of the endogenous variables builds on Card (2001).

The estimates are obtained with PPML using the *PPML* command by Santos Silva and Tenreyro (2006) and Santos Silva and Tenreyro (2011) which perfectly deals with the reduced set of FE we are going to include in the instrumental analysis. Column (1) shows the correlation of the lagged value of import in cultural goods on current imports. Column (2) performs the same exercise on export. The sample is reduced in this specification, because of data availability for the lagged instruments. Time coverage: 2007-2014

1.F Further Robustness Tests

In this section, some additional robustness tests are presented. The first set of results replicates our baseline specification reported in Table 1.4, by considering cultural trade in terms of share with respect to aggregate trade. Considering shares instead of absolute values allows to clean the coefficients from potential shocks affecting an economy as a whole, that might affect a country’s total trade. Considering cultural trade in absolute terms might hide the impact of such shocks. Coefficients in table F-1 come from a transformation of the coefficients of interest in Table 1.4, where $\text{CultIMP}_{ni,t}^{\text{sh}} = \text{CultIMP}_{ni,t} / \sum_{n=1}^n \text{AggrIMP}_{ni}$. Despite the coefficients for the investing side appreciation channels turn insignificant in column (3), the results further confirm the conclusions of the empirical analysis conducted in this chapter: the cultural preference awarded by a destination to a potential investor proves to dominate the origin-side one.

The second set of results, presented in table F-2 analogously replicates our benchmark estimations on two distinct reduced samples. as mentioned in appendix 1.A, using trade in cultural goods does not guarantee that all goods containing cultural content of a certain country to be actually attributed to that country alone. Think for instance to a Canadian singer whose record company is actually located in the US: in this case, the cultural content would be Canadian, while the cultural trade would accrue to the US. Analogously, many cultural goods include intermediate inputs which are likely to be produced in some part of the world other than the country which assemble and finally trade those goods. The trade of intermediate cultural products is partly recorded as cultural trade on its own, even though it has keeps no track of a cultural content on its own. For this reason, panel A of table F-2 reports the estimates after excluding China and the USA, while panel B perform a similar exercise excluding China and the rest of the Manufacturing Asia. The results hold across samples, confirming the validity of the conclusions discussed in section 1.4.

Table F-1: Impact of CP on Greenfield FDI - Shares of Cultural Trade

Dep. Var.	Count $C_{ni,t}$		
	(1)	(2)	(3)
$\ln \text{CultIMP}_{ni,t}^{\text{sh}}$	0.0286* (2.32)		0.00793 (0.68)
$\ln \text{CultEXP}_{ni,t}^{\text{sh}}$		0.159*** (10.78)	0.157*** (10.88)
Imp×Year FE	✓	✓	✓
Exp×Year FE	✓	✓	✓
Obs.	87448	87448	87448
% Zeros	0.749	0.749	0.749
R ²	.9	.9	.9
Test 1	-	-	118.4
Test 2	-	-	67.78
Estimator	PPML	PPML	PPML

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. The dependent variable Count $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultIMP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultIMP}_{ni,t} = 0$).

Table F-2: Impact of CP on Greenfield FDI - Removing Big Traders

Dep. Var.	Count $C_{ni,t}$		
	(1)	(2)	(3)
Panel A: excluding US and China			
$\ln \text{CultIMP}_{ni,t}$	0.184*** (14.63)		0.0973*** (8.92)
$\ln \text{CultEXP}_{ni,t}$		0.316*** (23.13)	0.283*** (20.56)
Imp×Year FE	✓	✓	✓
Exp×Year FE	✓	✓	✓
Obs.	80641	80641	80641
% Zeros	.77	.77	.77
R ²	0.810	0.820	0.830
Test 1	-	-	633.8
Test 2	-	-	89.64
Panel B: excluding China and Manufacturing Asia			
$\ln \text{CultIMP}_{ni,t}$	0.179*** (13.29)		0.0733*** (5.98)
$\ln \text{CultIMP}_{ni,t}$		0.351*** (23.28)	0.322*** (20.48)
Imp×Year FE	✓	✓	✓
Exp×Year FE	✓	✓	✓
Obs.	69533	69533	69533
% Zeros	.75	.75	.75
R ²	0.900	0.920	0.920
Test 1	-	-	591.8
Test 2	-	-	119.4

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors are clustered by trading-pair. Panel A replicates baseline specification of table 1.4 excluding the USA and China. Panel B excludes all Manufacturing Asia and China.

The dependent variable Count $C_{ni,t}$ is the bilateral number of Greenfield FDI projects from country i to country n . It includes the zero flows.

TESTS: *Test 1* refers to the joint significance χ^2 test over the two coefficients for explicit preferences ($H_0 : \ln \text{CultIMP}_{ni,t} = \ln \text{CultIMP}_{ni,t} = 0$). *Test 2* reports instead the χ^2 test inherent to the statistical difference between $\ln \text{CultIMP}_{ni,t}$ and $\ln \text{CultEXP}_{ni,t}$ ($H_0 : \ln \text{CultIMP}_{ni,t} - \ln \text{CultEXP}_{ni,t} = 0$).

1.G Country excluded from the dataset

Table G-1: List of Countries Excluded from the Analysis

In both direction: no flows of greenfield FDI (in or out) over the entire period
Anguilla, Netherland Antilles, Cocos and Keeling Islands, Cook Islands, Christmas Islands, Western Sahara, Falkland Islands, Faeroe Islands, Gibraltar, French Guiana, Kiribati, Marshall Islands, Northern Mariana Islands, Montserrat, Norfolk Islands, Niue, Nauru, Pitcairn, Palau, Saint Helena and Tristan da Cunha, San Marino, Saint Pierre et Miquelon, Tokelau, Tonga, Tuvalu, British Virgin Islands, Vanuatu, Wallis and Futuna
No outward flows over the whole period (excluded as source countries)
Aruba, Benin, Bhutan, Cape Verde, Central African Republic, Chad, Comoros, Republic of the Congo, Dominica, Eritrea, Grenada, Guinea, Guinea-Bissau, PRD Korea, Liberia, Maldives, Mauritania, New Caledonia, Niger, Paraguay, Sao Tome and Principe, Seychelles, Sierra Leone, Somalia, Saint Kitts and Nevis, Sain Lucia, Saint Vincent and the Grenadines, Suriname, Timor Leste, Turkmenistan
Countries excluded or aggregated for inconsistencies between CEPII and fDIMarket
Serbia and Montenegro (both excluded)
Belgium and Luxembourg (both excluded)
Sudan and Sud Sudan (South Sudan is Excluded)
Switzerland and Liechtenstein (Aggregated)
France and Monaco (Aggregated)
<small>Notes: The result of the exclusion of these countries is a rectangular dataset of $n \times m$ countries. In addition to these countries - excluded for data inconsistencies - other dyadic flows are excluded when no investment occurs between two countries during the period analyzed. This explains the discrepancy between the size of the dataset and the number of observations used in the estimation</small>

Bibliography

- Adsera, Alicia and Mariola Pytlikova**, “The role of language in shaping international migration,” *The Economic Journal*, 2015, 125 (586), F49–F81.
- Artuç, Erhan, Frédéric Docquier, Çağlar Özden, and Christopher Parsons**, “A global assessment of human capital mobility: the role of non-OECD destinations,” *World Development*, 2015, 65, 6–26.
- Aubry, Amandine, Maurice Kugler, and Hillel Rapoport**, “Migration, FDI and the Margins of Trade,” *Boston: Center for International Development, Harvard University*, 2014.
- Belot, Michèle and Sjef Ederveen**, “Cultural barriers in migration between OECD countries,” *Journal of Population Economics*, 2012, 25 (3), 1077–1105.
- Desbordes, Rodolphe and Shang-Jin Wei**, “The effects of financial development on foreign direct investment,” *Journal of Development Economics*, 2017, 127, 153–168.
- Disdier, Anne-Célia, Silvio H. T. Tai, Lionel Fontagné, and Thierry Mayer**, “Bilateral trade of cultural goods,” *Review of World Economics*, 2010, 145 (4), 575–595.
- Du, Julan, Yi Lu, and Zhigang Tao**, “Institutions and FDI location choice: The role of cultural distances,” *Journal of Asian Economics*, 2012, 23 (3), 210–223.
- Felbermayr, Gabriel J. and Farid Toubal**, “Cultural proximity and trade,” *European Economic Review*, 2010, 54 (2), 279 – 293.
- Giovannetti, Giorgia and Mauro Lanati**, “Do High-Skill Immigrants trigger High-Quality Trade?,” *The World Economy*, 2016.

- Gould, David M.**, “Immigrant links to the home country: empirical implications for US bilateral trade flows,” *The Review of Economics and Statistics*, 1994, pp. 302–316.
- Head, Keith and Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Hofstede, Geert**, *Cultures and organizations. Intercultural cooperation and its importance for survival. Software of the mind*, Mc Graw-Hill, 1991.
- Javorcik, Beata S., Çağlar Özden, Mariana Spatareanu, and Cristina Neagu**, “Migrant networks and foreign direct investment,” *Journal of Development Economics*, 2011, *94* (2), 231–241.
- Lee, Hyun-Hoon and John Ries**, “Aid for Trade and Greenfield Investment,” *World Development*, 2016, *84*, 206–218.
- Mayer, Thierry and Gianmarco IP Ottaviano**, “The happy few: The internationalisation of european firms,” *Intereconomics*, 2008, *43* (3), 135–148.
- Melitz, Jacques and Farid Toubal**, “Native language, spoken language, translation and trade,” *Journal of International Economics*, 2014, *93* (2), 351–363.
- Santos Silva, João M C and Silvana Tenreyro**, “The log of gravity,” *The Review of Economics and statistics*, 2006, *88* (4), 641–658.
- and —, “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator,” *Economics Letters*, 2011, *112* (2), 220–222.
- Shenkar, Oded**, “Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences,” *Journal of international business studies*, 2001, *32* (3), 519–535.
- UNCTAD**, *Creative Economy Report. A Feasible Development Option*, UNDP, 2010.
- Yotov, Yoto V, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch**, “An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model,” *World Trade Organization, Geneva*, 2016.

Chapter 2

A network analysis of the migration-Investment nexus

In an increasingly interconnected world, the extent to which the complex web of relationship a country entertains with the rest of the world is still far from being fully understood. This chapter asks whether and to what extent the progressive integration and expansion of the Global Migration Network can be accounted for explaining the patterns of Greenfield Foreign Direct Investment exchanges at bilateral level. After constructing and comparing the two networks, I econometrically test the relationship between bilateral migration flows and FDI, controlling for the embeddedness of both the investing and the recipient country in the global migrants' network. To exploit the complex structure of the dataset to its full extent, I estimate a gravity-like equation of bilateral FDI departing from the usual fixed effect estimation to embrace a more flexible multilevel mixed estimator, that I maintain to be better deal with the potential hidden hierarchical structure of the dataset. Results confirm the positive direct relationship between migrants diaspora and FDI at bilateral level: the inclusion of network level statistics allows to spot a much larger degree of complexity in the migration-investment relationship. The emergence of third-party network effects constitute the most interesting finding of all, and might help to better understand the mechanisms triggered by migration flows beyond the pure bilateral perspective. Such interesting insight would not have been detected with a non-network approach.

Keywords: *Greenfield FDI, Multilevel, Networks, Migration.*

2.1 Introduction

During the last decades, the share of migrant population worldwide grew remarkably in absolute terms, in spite of a global trend of imposing limitations to human mobility, especially along the South-to-North trajectory (McKenzie, 2007). Nonetheless, not considering forced displacements, cross border movements only account for slightly more than 3% of total world population (Öz-

den et al., 2011; UNDESA, 2015). At the same time, the number of countries involved in the International Migration Network (Fagiolo and Mastrorillo, 2013, IMN) grew remarkably in the last decades. Despite this fact might be partially related to the better data coverage available in recent years, it motivates the efforts spent to understand the economic impacts of international migration at world level. Notwithstanding the simultaneous growth recorded in the circulation of goods, people, and capital, most of the literature studying the impact of migration on economic exchanges focuses on trade. Not much attention has been dedicated to the relationship between investments and migration, despite the relevance of people movements as a conveyor of information, able to mitigate uncertainty and to reduce the transaction costs that limit investments more than other types of flows.

This chapter explores the role played by migrants' networks as determinants of bilateral FDI, considering the two flows as overlapping layers of the same global macroeconomic network. Approaching international migration data as a network offers the advantage of taking the complexity of the world system into account, beyond the direct bilateral relationship between two countries. This study explores the relationship between the migrants' network and two countries' bilateral investment position, using a sample of 20 OECD countries as reference. Since highly educated and skilled migrants are more likely than their less educated counterpart to dispose of the social capital (Burchardi et al., 2018) or of the informative capacity (White, 2007) needed to trigger FDI (Flisi and Murat, 2011), the choice of the sample is led by the availability of education-based bilateral migration flows (Brucker et al., 2013). Following Fagiolo and Mastrorillo (2014); Sgrignoli et al. (2015), and Garas et al. (2016), I explore the correlation between the international migrants network (IMN) and the global Greenfield FDI network (GFDIN) under a complex network perspective. After describing and comparing the structure of the two networks, and their topological characteristics in order to identify potential co-evolution patterns, I econometrically test the relationship between the two, applying a multilevel mixed regression modeling (Rabe-Hesketh and Skrondal, 2012).

The original contribution of this chapter is essentially threefold. First, no study so far has analyzed the impacts of a country's position in the migrants network on its bilateral FDI "strategy": thus, the complexity of the relationship between migration and FDI flows remains largely unexplored. Existing evidence is confined to basic measures of connectivity and does not explore how the structure of a network ultimately affects nodes connectivity on a different type of flow.

Second, even if I do investigate bilateral flows, I depart from the typical fixed effects (FE) gravity estimation. In fact, the complex structure of the data, which are characterized by different levels of information (country's network specific, dyad specific, and contextual), is likely to imply a hidden hierarchical structure. For this reason, Multilevel Mixed Regression analysis might represent an appropriate alternative to the analysis of bilateral (network) data.¹

Third, conversely from the earlier literature, I explicitly address the attribution problem which affects most of the existing related studies. Such attribution problem makes impossible to distin-

¹In Appendix 2.B I also report the estimates obtained from a two-step fixed-effects estimation as a benchmark for my studies.

guish the network position of observations characterized by very different attributes. I solve this issue by retaining the direct structure of both networks.²

The findings are consistent with the existing evidence on the migrants-FDI relationship: traditional socioeconomic measures are confirmed as crucial drivers of bilateral greenfield FDI flows, which are also boosted by the presence of highly educated migrants.³ Also, consistently with the literature on complex network and macroeconomic flows, bilateral FDI are both directly and indirectly affected by migrants' network topological features. Thus, as the presence of an established bilateral migration corridor directly affects the economic exchanges between two countries, bilateral FDI are also positively related to both the *size* of a country's specific migration network and the *quality* of the position of such country in that network, which capture the indirect effect of the IMN.⁴

The remainder of the chapter is as follows. Section 2.2 presents a brief review of the related literature, divided between "traditional" studies on migrants flows/stocks and the new stream of evidence based on complex networks. Section 2.3 introduces to the methodological tools used to describe the network topology and the interconnection between flows, which are first described and then analyzed by means of a multilevel gravity-like model in Sections 2.4 and 2.5 respectively. Section 2.6 finally concludes and set the stage for future research.

2.2 Literature Review

It is worth to begin with a little notation. Throughout this chapter, I generically refer to migration flows in terms of network. Such definition may be open to some misinterpretation, since the idea of migrants networks has evolved in the recent years. The earlier definition of migrants' network indicate the set of connections that a migrant community may activate both at home and at destination, able to trigger a certain economic outcome (be it to facilitate job search and hiring, or to facilitate trade, capital movements, etc.): in other words, the social capital that is mobilized in the migration process (Munshi, 2003). The most important contributions in this approach are reviewed in section 2.2.1. More recently, economists began to question the pure bilateral domain of migration. Complementary to this, the research turned with increasing interest to the

²An example of *attribution problem* is the following. Suppose two dyads, composed on the one hand of a country experiencing small immigration flows at front of massive capital outflows and on the other hand, by a country characterized by the opposite features, ususally result equal to dyads in which both countries experience average inflows of migrants and average outflows of capital. Summed together, the two types of dyads cannot be distinguished. Such attribution problem allows to identify only the existence of correlation patterns between the flows considered, while it does not allow drawing any conclusion concerning the direction of the IMN-Investment nexus.

³To be more precise, they are positively related to the presence of migrants in general. Yet, the role of the highly educated diaspora confirms the claims of Flisi and Murat (2011) concerning a larger impact of the better educated migrants with respect to the total migration flow.

⁴With "size" I intend the number of contacts centered on a country in the IMN, measured in terms of the number of partners a country links to (outward connectivity), or it is linked by (inward connectivity), both in *extensive* (existence of a link) and in *intensive* (strength of that link) terms. With "quality", I refer instead to the importance of a country's individual network in terms of size (extensive or intensive) of the partners it is connected to.

complex web of relationships that migrants' flows trigger among countries worldwide. The focus shifted from the mechanisms dominating a dyadic relationship between two countries to how the set of global interconnections affects those same mechanisms. Section 2.2.2 reviews these latter contributions.

2.2.1 Migrants Networks and FDI: the Bilateral approach

The effect of migrants' networks as informal institutions, which affect individual as much as collective outcomes has been thoroughly explored in the last decades. For instance, they proved to be fundamental in job seeking (Montgomery, 1991) and in driving both working and location choices (Massey et al., 1993; Munshi, 2003). But migrants network are highly effective in favouring international economic exchanges too. Gould (1994) first explored the relationship between the presence of immigrants communities in the US and trade flows with their country of origin. Interestingly, migrants' flows affect economic exchanges both from the supply and the demand side. Concerning the former, two main factors help explaining the usually positive effect of migration on trade. On the one hand, migrants' communities can exploit alternative informative channels, which favor in turn the establishment of privileged forms of information exchanges. (Leblang, 2010, refers to this in terms of the information potential). On the other hand, migrants dispose of the social capital that helps reducing the cost of contract enforcement (which heavily affect the transaction costs).⁵ According to Rauch (2001) these two mechanisms defines migrants' flows' "Business and Social Network effect".

On the demand side, migrants networks can encourage trade by increasing the demand for goods produced back home (White, 2007, referred to this effect in terms of "transplanted home bias"), and may change the consumption pattern of their home country via "tastes contamination". Many different theoretical frameworks have been developed concerning the way these two effects operate and how they participate in shaping trade patterns in the real world (Giovannetti and Lanati, 2016).

In the last 10-to-15 years, the study of the impacts of migrants networks on bilateral economic exchanges widened to encompass cross-border financial flows, and particularly FDI (Gheasi and Nijkamp, 2017). However, while the same mechanisms facilitating trade seem to be much more relevant for cross border investments, the empirical and theoretical evidence is much less clear cut. Kugler and Rapoport (2007) investigates the causes of this lack of clarity, highlighting two main opposite theoretical mechanisms linking migration to investments. On the one hand, FDI and migrants outflows may be substitutes: outmigration, especially of the better educated, could reduce the incentives to invest in the country of origin because of the lower labor productivity (which is often the main driver of vertical FDI). This may be the case for countries that are subject to *brain drain* (Beine et al., 2008), which is often favored by the existence of selective immigration policies and cherry-picking strategies in developed countries. Thus, emigration could negatively affect the economic performance of a country (Checchi et al., 2007), and could explain why some

⁵Interestingly, both mechanisms appear to be stronger for highly educated migrants flows.(Barro and Lee, 2013)

studies found migration and FDI to be substitutable. For instance, Aroca and Mohoney (2005) detect a negative correlation between FDI and Emigration from Mexico to the US, concluding that the two flows could be considered as substitute. On the other hand, FDI and migrants might be characterized by a complementarity relationship à la Docquier and Lodigiani (2010). This is closely related to the idea of business and social network effect, introduced by Rauch and Trindade (2002) and further expanded by White (2007) and Simone and Manchin (2012).⁶ Accordingly, migrants networks may act as a bridge across countries, favoring investments by lifting informational barriers, as well as signaling opportunities to potential economic partners in both their home and host countries (Cuadros et al., 2016). This mechanism seems to be also related to the idea that investments are affected by the *skill composition* of a country, whose effect is arguably stronger for investment than for trade. In short, migration might affect FDI flows by determining the factor endowment of a country. This drives in turn the type of investments that can be attracted (see Feenstra and Hanson, 2008; Zhu and Treffer, 2005; Khalifa and Mengova, 2015). Thus, countries with higher human capital should be able to attract more skill-intensive investments from abroad. The larger opportunities offered by better paid jobs should favor additional human capital accumulation, by making more profitable to invest in education and providing at the same time an incentive not to migrate abroad (to work in the better paid industry back home). This argument extends the *Brain Gain* theory, first proposed by Stark et al., 1997 for migration, but remains mostly a theoretical mechanism that only holds in the long run.⁷ The empirical evidence too is still scarce and divided in two streams. The first one focuses on the dichotomy “substitutability vs complementarity”, while the second (and to some extent more recent one) studies the circumstances which determines which of the two effects (complementarity or substitutability) prevails. Taking explicitly into account the skill composition of the migrants network, Kugler and Rapoport (2005) conclude that future FDI inflows depend positively on the stock of expected future highly educated migrants, but that a substitution effect is also at play when secondary school educated migrants outflows are considered. Similarly, Flisi and Murat (2011) detect a positive correlation between bilateral immigration flows on investment outflows, suggesting that the information channel triggered by the presence of ethnic ties between the migrant community and their homeland might be beneficial to those companies wishing to invest there. Checchi et al. (2007) too detect a virtuous relationship between skilled migration and FDI, despite the evidence of a negative impact on overall human capital accumulation. Instead, Kugler and Rapoport (2007) analyze the FDI-migration nexus taking the US alone as destination country and analyzing the flows of FDI toward a large set of migrants’ origin countries: by differentiating between a static and a dynamic perspective, they conclude that there is evidence of substitution effect when data are analyzed statically, but that the relationship reverses in favor of complementarity when a dynamic approach is preferred.

⁶To be more precise, Rauch talks about “Business Network externality”, whilst de Simone and coauthors refers to it as “Diaspora Externality”.

⁷The same mechanism applied to the domain of migration implies that having the chance to migrate may encourage human capital accumulation. Since not all the better educated workers are actually able to migrate, they nurture the stock of nationally available skilled workforce that in turn attracts investments: in the end, the incentives to migrate shrinks due to the availability of better jobs back home, encouraging further FDI inflows and possibly an upgrade along the value chain. For instance, this is what happened in certain provinces of Mexico following the set up of aircraft plants by Bombardier after the 1990s (Baldwin, 2016).

Acknowledging the presence of some inconsistencies in the existing evidence, Federici and Giannetti (2010) develop a theoretical model to take into account both substitutability and complementarity in the FDI-Migration relationship. Their target is to model what they define as the “Information-revealing effect” of migration (Leblang, 2010), i.e. the role of the information channeled by migrants which is able to trigger complementarity. Javorcik et al. (2011) study the impact of US immigrants networks on the destination of US outward investments, finding evidence of complementarity.⁸ D’Agosto et al. (2013) explore the FDI-Migration nexus across the north-south trajectories, finding evidence of both substitution and complementarity effect, with the latter prevailing in magnitude. Kugler et al. (2017) study the relationship between migrants networks and financial flows other than investments, starting from the idea that the diaspora from a country reduces the effect of the “informational frictions”, stimulating those flows that are more sensitive to the availability of reliable information. Their findings reinforce the previous results by Kugler and Rapoport (2007): migrants’ flows positively affect international capital flows, especially when the informational frictions are more acute, i.e. when there is very low cultural proximity between countries. Not surprisingly, the impact of the highly educated diaspora is stronger than the impact of the rest of the migrants’ stock. Burchardi et al. (2018), by focusing more specifically on the historical migrant communities and their distribution within the US, reach similar conclusions. Despite the debate on which effect between complementarity and substitutability dominates theoretically, the former does prevail empirically, as diaspora appears to be highly effective in reducing those informational asymmetries affecting bilateral FDI (much more than trade and bilateral equity flows, according to Portes and Rey, 2005; Wang, 2017). Evidence of complementarity also comes from the research of Cuadros et al. (2016), which explicitly explores the possibility for migrants communities to be particularly effective in driving FDI when the differences in terms of financial development between two countries are large: the so-called *bridge effect* of bilateral diaspora over bilateral FDI proved to be at work, though only at low levels of cross country interaction (in terms of bilateral reciprocal FDI).

2.2.2 Migrants (proper) Networks and FDI: the “complex” approach

The major limitation in the works reviewed so far is that they limit the analysis to the pure bilateral perspective. However, as much as a country cannot be considered independent from the set of partners it interacts with, the same is true when the focus shifts to a bilateral level. The decision to interact and the way two countries eventually do so should be considered as a response to the overall set of stimuli generated by all the existing and potential connections in which these two countries engage. In fact, economic exchanges are rarely the result of pure bilateral dynamics: often, they can be interpreted as a form of *relational data*, resulting from the complex patterns in which two actors/countries are embedded. (Social) Network Analysis represents a crucial tool to understand both global and local dynamics, and has been applied

⁸This last work is particularly interesting as it is one of the first to explicitly tackle the issue of the heterogeneous impacts of skilled bilateral migration on investments, detecting a much stronger role of the highly skilled migration as opposed to the less educated one, addressing at the same time the potential endogeneity of migration flows and FDI.

to a wide set of relational data in many fields: from criminology, to finance (to understand contagion and herd behavior), election patterns, business intelligence, industrial clusters, and economics (Jackson, 2010). In international economics and finance, SNA focuses on the study of *Complex Networks*, and on the way their structure affects their evolution over time.⁹ Most of the related international economic literature either focuses on, or is related to, trade networks (see for instance Li et al., 2003; Fagiolo et al., 2008, 2009; De Benedictis and Tajoli, 2011). Currently, the world trade web is the only extensively investigated large network. The only attempts to map in a similar fashion other international flows at the global level focused on international migration flows (Fagiolo and Mastrorillo, 2013; Abel and Sander, 2014; Sciabolazza, 2018; Garas et al., 2106, the latter limited to OECD destinations only). Conversely, a comprehensive mapping of the international investments network is still missing, and the few related analyses in this sense are limited to a few geographically circumscribed studies. De Masi et al. (2013) map the internationalization strategy of Italian firms, while Joyez (2017) proposes a similar application to French multinationals; De Masi and Ricchiuti (2018) investigate the structure of the EU outward FDI network. Only recently, Metulini et al. (2017) mapped the network of corporate control.

More recently, econo-physicists began to study the economic relationship between countries in terms of complex networks, by overlapping different flows as “layers” (or levels) of a wider, global macroeconomic network. Fagiolo and Mastrorillo (2014) investigate the presence of a potential common structure between the world trade web (ITN) and the International Migration Network (IMN), finding evidence of important interconnections between the two flows. Sgrignoli et al. (2015) performed a similar exercise, extending the analysis of (Rauch, 2001) on the impact of migrants networks on different trading sectors, while Metulini et al. (2017) studied the interrelation between trade and capital flows. These studies represent just a part of the literature that developed in the last few years, which focused on the juxtaposition between different macroeconomic networks in order to better understand how countries ultimately connect to each other, and the degree of permeability across the different types of flow. While the network analysis of migration and trade is a well established application, disregarding of the network approach (see Docquier and Lodigiani, 2010; Peri and Requena-Silvente, 2010; Giovannetti and Lanati, 2016; Sgrignoli et al., 2015, among the others), nothing similar has been done to exploit the potentialities of network analysis in the analysis of the migration-investment nexus. The only attempt in this sense is represented by the analysis of Garas et al. (2106). However, they limited their analysis to the search of a causal link between the connectivity of a country in the IMN, and the amount of bilateral investments. Understanding what node-level characteristics prove to be relevant in driving bilateral FDI remained outside the scope of their analysis. To current date, no further study succeeded in mapping the global investment network at world level, and in analyzing it in a multilayer framework as it has been done for trade.

⁹Complex Networks refer to graphs with non-trivial topological features (as summarized by Lü et al., 2013).

2.3 Methodological Note

In this section I introduce a few definitions to get acquainted with SNA tools, and the differences across the various elements that define the structure of a network. I also introduce some measures of network centrality. These measures relate to the position of a country within the network. Despite all of them describe the importance of an economic actor with respect to a type of relationship, they also convey different information. It is therefore necessary to explain how they are built, and the interpretation when applied to both the International Migration Network (IMN) and the GFDIN (Greenfield Investment Network). For reason of generalizability, in the following pages I denote the nodes of the network (i.e. the subjects whose relationships constitute the network itself) as actors or vertices rather than as countries.¹⁰

2.3.1 What is a network? Basic terminology and representation

Networks represent perhaps the most natural way to represent relational data (such as international economic flows) retaining the systemic perspective. In Social Network Analysis (SNA), a network (or graph) is represented by a set of (un)ordered dyads $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. The number of actors \mathcal{N} (also called nodes or vertices) and the number of links \mathcal{E} define the size and the density of the network. Two nodes are said to be connected if it is possible to find a path between them, “walking” through their connections. This point is particularly interesting and describe in a single sentence how working with networks offers the possibility to focus on the way an actor (country) interacts with its partners keeping in sight the global structure in which such relationship is embedded.

Analytically, a network is usually represented in matrix form as an *adjacency matrix* \mathcal{A}_n , a $\mathcal{N} \times \mathcal{N}$ square matrix with rows and column containing the whole ordered list of vertices in \mathcal{G} : all entries $a_{i,j} \neq 0$ indicate the existence of a direct link between node i and node j .¹¹ Graphically, vertices are represented as dots, linked together with lines connecting the different actors according to the structure of the adjacency matrix.¹²

Graphs can be either *directed* or *undirected*: such difference is crucial to define the way actors connects with each other. Undirected means that no direction is associated to the edges linking the vertices: for instance, free trade agreements can be modeled as an undirected graph (Soprancetti, 2017) where a treaty is considered as binding for both sides reciprocally. On the

¹⁰Also, I may refer to a network as “graph”, borrowing from the branch of mathematics known as *graph theory*. Since the use of graph in place of network helps to figure out the set of relationships in a graphical way, I refer to graphs whenever this mental representation can be helpful, or when I explicitly talk about the graphical representation of the network.

¹¹A network can also be represented as an edgelist $\mathcal{E}_{n,e}$, where n represents the number of vertices, and e the number of active links between them. Edgelists are usually more tractable when the size of the network is very large, as they only contain information about the existing links, not about all the potential connections, generally marked with a 0 in the standard adjacency matrix.

¹²The representation in graph form is particularly effective when dealing with small graphs. However, larger networks might result confusing, and might be better represented by different forms of plot, such as circular plots, dendograms, and so on.

contrary, economic flows cannot be considered symmetric: trade flows (Fagiolo et al., 2008, 2009; Schiavo et al., 2010; De Benedictis and Tajoli, 2011), FDI (Vitali et al., 2011; De Masi et al., 2013; Joyez, 2017), portfolio investments (Chuluun, 2017), and migration (Fagiolo and Mastro-riillo, 2013; Abel and Sander, 2014; Sciabolazza, 2018) are all characterized by an asymmetric component and by non-reciprocity. In all these cases, the presence of a positive flow from an node to another neither implies the presence of a counter flow in the other direction, nor that (if existing) such counter flow equals the former: there may be an actor that links to another one which does not links back to the former. When this happens, a network is said to be directed.¹³

In addition, a network can be weighted or binary, depending on whether the size of a certain flow is taken into consideration. Back to the example of FTA, their network is both undirected and binary, where each edge can be seen as a logical operator such as

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ directly connects to } j \\ 0 & \text{if } i \text{ does not directly connects to } j \end{cases} \quad (2.1)$$

While the binary dimension gives the idea of the extension of a network (similar to the *extensive margin* in economics), international economists are often interested in the intensive margin of a bilateral relationship.¹⁴ Thus, the intensity of the flow may be included into the analysis, turning equation (2.1) into

$$a_{ij} = \begin{cases} v_{ij} & \text{if } i \text{ directly connects to } j \\ 0 & \text{if } i \text{ does not directly connects to } j \end{cases} \quad (2.2)$$

, where v_{ij} represents the value of the bilateral exchange between country i to country j . Graphically, weights are included by changing the thickness of the edges/arcs, or by changing the intensity of the colors. For completeness, a network that is directed or weighted (or both) can be represented also in its undirected and binary structure.¹⁵

Finally a network may be simple or multiplex. The first one takes into consideration a single “environment” or “socio-economic context”, and analyzes the relationships occurring between actors on that level: the only structure that counts is the one of interest, and all other dimensions are excluded. The second one instead represents one of the most promising developments in SNA (and in network science in general) as it takes into consideration several levels of relationships (≥ 2) or layers of the potential interactions between actors. Thus, the development of a link or its intensity is analyzed both in function of the specific referring context and in light of the other potential relationships that may occur across layers. (See for instance Kivelä et al., 2013, for an introduction to multilevel and multiplex network analysis).

¹³The terminology also changes a bit: edges become *arcs* and are graphically represented by an arrow.

¹⁴The FTA network might turn to be weighted if the number of treaties in which two countries are simultaneously involved is considered.

¹⁵The opposite is not feasible.

This chapter analyzes the impacts of migrants networks on bilateral (greenfield) FDI. Given the nature of migration flows, their network is better characterized as **weighted** and **directed**. Also the time dimension is interesting: observing both the international migration and the FDI network over time offers an interesting perspective of their evolution (Fagiolo and Mastrorillo, 2014; Sgrignoli et al., 2015; Garas et al., 2106). By explicitly linking the migrants network to bilateral FDI, this chapter joins the small but growing literature on the interrelation between different networks. Even though it does not refer to a proper multilayer network, it constitutes a substantial improvement of the existing FDI literature.¹⁶

2.3.2 From local to structural properties of nodes: using networks to analyze actors

A graph provides a huge load of information, in terms of its global structure as well as in terms of the behavior of its actors, which affected by it. Indeed, the decision of two vertices to link (and the way they do so) is unlikely to be taken irrespectively from the overall context in which those agents operate. For instance, there may be some node that link to many others: newcomers may find convenient to establish a connection with those highly influential nodes, in order to take advantage of the channels they may offer toward the rest of the network. This phenomenon, known as *Preferential Attachment* (Barabási and Albert, 1999), characterizes many real world international networks, from the FTA (Sopranzetti, 2017) to trade (Fagiolo et al., 2008), investments (Garas et al., 2106), and migration (Fagiolo and Mastrorillo, 2013; Sciabolazza, 2018) networks (as well as the Internet - see for instance Brin and Page, 1998; Kleinberg, 1999). Understanding the effect of actors' position with respect to the overall network may be very informative about their economic potential and the influence that a single actor may exert on the system as a whole. Since the position of an actor can be defined in several ways, it is important to understand what is the meaning of the different alternatives. According to Borgatti (2005), each measure of centrality conveys a specific set of information about the role of a node within the network. In the empirical section, I include and discuss several measures of network position. I introduce them in this section, to give an idea of the type of information they add to the analysis, their relative advantage and their shortcomings.¹⁷

1. **(Weighted) Degree Centrality.** It is the most immediate measure of network centrality. In its binary formulation, degree centrality counts the number of connections of a node

$$k_i = \sum_{j=1}^n a_{ij} \quad , \quad \forall i \in \mathcal{G} \quad (2.3)$$

where a_{ij} represents the ij^{th} entry of the binary adjacency matrix. Intuitively, it is associated to the extensive margin of network, i.e. to the number of (new) connections that

¹⁶This brief introduction is not complete: for instance, the distinction between uni- and bi(multi)-modal networks (where actors are connected to each other indirectly, by sharing the same set of occurrences) remains untackled, as well as the possibility of complex structures generated, for instance, by self-loop edges (typical in expectation and contagion network models).

¹⁷Additional terminology can be found in appendix 2.A

an actor establishes.¹⁸ For the purpose of this chapter, I am also interested in the analysis of the intensive margin of a network. Weighted degree centrality (often referred to as *Strength Centrality*), analogously to degree centrality, sums all entries of the weighted adjacency matrix referring to a certain node i . The difference is that it takes the size of the flow into consideration, not just the presence of a link.

$$s_i = \sum_{j=1}^n v_{ij} \quad , \quad \forall i \in \mathcal{G}. \quad (2.4)$$

Both Degree and Strength can be normalized in order to obtain a measure ranging between 0 and 1. Interestingly, despite both measures are based on a very similar type of information, they may not perfectly coincide in complex networks (even if their correlation is usually very large). For instance, actors with a lot of small connections may coexist with nodes with a small number of extremely large links. Measures of purely local connectivity fail to capture this possibility.¹⁹

2. **Average Nearest Neighbour (ANN) Centrality** (Barrat et al., 2007). Measures of local network connectivity may be informative about the size (and possibly, the intensity) of a node’s relational web. Taken alone, such measures offer a very limited perspective about the context an actor operates in. A perhaps related concept, though slightly wider, consists in the average size and intensity of a node’s neighbors connectivity, which represents a form of “indirect” measure of centrality. ANN Degree (ANND) counts the average number of links accruing to/from the whole direct neighborhood of a node; analogously, ANN Strength (ANNS) measures the average intensity that characterizes its proximate countries. ANN centrality (Barrat et al., 2007) is defined as

$$ANNC = \frac{1}{s_i} \sum_{j=1}^n (a_{ij} w_{ij} k_j) \quad (2.5)$$

, where s_i represents the strength of node i and constitutes the normalizing factor (in the binary version, s_i is replaced by k_i). a_{ij} represents the ij^{th} entry of the adjacency matrix \mathcal{A} , and w_{ij} the weight of the link (that is conveniently set to 1 in the binary case); k_j indicates node j ’s degree centrality. Comparing Degree (strength) and ANNC measures is particularly interesting under a topological perspective, as it allows identifying whether a network presents an *assortative* (nodes connect to others with similar connectivity score) or a *dissortative* (nodes connect to others having very different connectivity scores) mixing patterns.²⁰

3. **Overlapping and Complementary Networks.** The reasoning underlying ANN centrality is that the dimension of the partners’ self-centered networks may be relevant in explaining the outcomes at the bilateral level. Nonetheless, these measures could reflect

¹⁸In directed graph, a distinction can be made between arcs departing from that vertex and those reaching it: in this case we talk about in-degree and out-degree centrality

¹⁹Also, these measures ignore all other structural characteristics of the neighbors, or of the network itself.

²⁰Assortativity should not be confused with homophily, which identify the tendency of actors to connect to each other according with certain individual characteristics.

some factors that do not really deal with the bilateral relationship between two countries, not last because they refer to the importance of actors other than those in a specific dyad. For instance, large ANND may lead to a large noise in the bilateral informative channels, resulting in a diversion of resources from a given bilateral relationship. This is more likely to happen if the overlapping networks between two countries is large. On the other side, a large complementary network would constitute a bridge between two otherwise unrelated nodes.²¹ For these reasons, the impact of such measures are likely to be non-trivial and constitute an ultimately empirical issue. For instance, Fagiolo and Mastrorillo (2014) account for such “third party” effect by including both migrants overlapping and complementary network measures to study the impact of migrants network on trade. They detected a magnifying impact of both, with a stronger role for the overlapping share. Overlapping and non-overlapping *inward* networks between two nodes (countries) i, j are defined respectively as the set of edges that are shared or that characterize only one of the two vertices.

$$\text{overlapping}_{ij} = \mathcal{I} = (1, \dots, n), \mathcal{J} = (1, \dots, m) \quad \mathcal{I} \cap \mathcal{J} = 0 \quad (2.6)$$

$$\text{overlapping}_{ij} = \mathcal{I} = (1, \dots, n), \mathcal{J} = (1, \dots, m) \quad \mathcal{I} \cap \mathcal{J} \neq 0 \quad (2.7)$$

where $\mathcal{I} = (1, \dots, n)$ and $\mathcal{J} = (1, \dots, m)$ represent the set of immigrants in county i and j respectively, while n and m represent the sets of countries that send migrants to i and j respectively.²² Dealing with both shared and non-shared network linkages represents a further improvement with respect to ANN centrality, since they allow to decompose the effect of partners’ networks in the role of direct competitors (the shared links) and that of indirect competitors (the non-overlapping connections).

4. **Eigenvector and Bonachich Centrality.** Degree and Strength centrality offer an insight over the size of an actor’s self-centered network. Unfortunately, they do not say anything about the position of the actor within the network. The “prestige” of a node within the network may be a much more interesting information, which refers to the importance (centrality) of a node as determined by the importance of its neighbors. An interesting measure of prestige is constituted by the *Eigenvector Centrality* (Bonacich, 1972, 1987), which defines a node’s importance as proportional to the sum of the importance of its own neighbors, which in turn depends on the importance of their own neighbors and so on and so forth in a recursive fashion. Thus, the prestige of a node no longer depends just on the size of its neighborhood, but also on the size of the neighborhood of the actors it is connected to (this time negatively). Despite the eigenvector centrality works fine for binary, unweighted graphs, it can also be generalized to the weighted case (The so called Katz prestige. See Newman, 2003; Jackson, 2010, for a review). Building upon eigenvector centrality and Katz prestige, Bonacich (1987) proposed a measure to take into consideration the “distance decay” effect of the indirect connections, to discount neighbors’ influence by

²¹Overlapping refers to the set of shared neighbours between a pair of nodes. Conversely, complementarity refers to the share of a country’s neighbors that have no relationship with the economic partner that is being considered

²²The reverse holds to compute the overlapping and complementary *outward* network

the distance on the adjacency matrix. Bonachich Centrality depends on two parameters, which are defined to privilege either the local centrality of a node (degree) or the decay effect due to distance. It can be expressed in matrix form as

$$Ce^B(\mathcal{A}, a, b) = (\mathcal{I} - bA)^{-1}agl \quad (2.8)$$

where \mathcal{A} is the adjacency matrix, $a, b > 0$ with b small captures the role of distance decay, while a captures the importance of a node's base value.²³ A high Bonachich and Eigenvector centralities imply that a node is important as it is able to link with (many) other important nodes, i.e., as long as it is located in the middle of a well established community, not just at its boundaries.

5. **PageRank Centrality.** It is another measure of prestige based on the eigenvector of the adjacency matrix. Though very similar to Bonachich centrality (it weight both the number of incoming links and the centrality of the connected nodes), it also takes their *link propensity* to send/receive connections from third parties into account. Thus the pagerank centrality of a node, is proportional to the number of links *received* from other important nodes, and to the degree of parsimony with which it links to other actors (or if a node is heavily linked *by* others). In a simplified formulation, PageRank centrality (PR) of a given node i is defined as

$$PR_i = \alpha \sum_k \frac{a_{k,i}}{d_k} x_k + \beta \quad (2.9)$$

α and β are constants, while k_i^{out} is the out-degree of node i . In the case a link does not link to any other nodes, this measure is set to 1, rather than to null. Despite having been developed to study the world wide web, PageRank presents some features that make it relevant also for the application of this research.²⁴ According to the existing theories linking migration to bilateral economic exchanges, migrants' networks operate as to solve informational problems, and as a signal for market and social conditions in both the country of origin and of destination. Such diaspora externality (Leblang, 2010; Kugler et al., 2017) implies that information flow in a privileged way between countries with a well established migration corridor between them. But when a country receives migrants from too many countries or send migrants to too many countries, it is possible that a noise is introduced, and the informative power of a certain channel is diluted by the competition of other countries/diaspora. As in the case of the indirect measures of centrality presented above, whether a dilution or a reinforcing effect is in play remains an ultimately empirical matter.

6. **Hub and Authority Scores.** Kleinberg (1999) proposed two alternative measures of centrality, the one specular to the other. The Kleinberg's HITS algorithm assigns both an Authority (prestige as linker) and a Hub (prestige as recipient) score to each node, according to a mutual recursive process in which the node is evaluated for how many links it receives from authorities (hub score) and for how many links it sends to hubs. In

²³Notice that all eigenvector-based measures, because of the recursive computational process, require both a starting base centrality value (often the node degree) and a condition to stop iterations.

²⁴Pagerank is also the algorithm which defines the ranking of Google's search results

simplified terms, the hub and the authority scores of a given actor i can be defined as

$$hub_i = \sum_{j=1}^n authority_j \quad (2.10)$$

$$authority_i = \sum_{j=1}^n hub_j \quad (2.11)$$

The HITS algorithm, analogously to PageRank, was originally developed to rank web pages on the web. HITS measures may be highly informative for the application of this study. First of all, being a hub in the IMN means that a country receives (many) migrants from many countries characterized by high emigration rates toward many other important destinations. The opposite is true for being an authority in the IMN. Thus, a positive correlation between bilateral migration and FDI would rule out any evidence of a potential “dilution effect” of the migrants informative channel (that is, the fact that many immigrant communities at a certain destination end up competing with each other in order to attract investments to their homeland).²⁵

Beyond node-level statistics, SNA offer a wide range of “system-level” statistics, related to the overall structure of the network. Such structure affects not only the probability of actors to establish a link, but also the resilience of a relational system, the rapidity of a potential contagion (financial or epidemic), the emergence of a core-periphery structure (Wallerstein, 1987; Borgatti and Everett, 2000), the emergence of rich clubs or even “riches get richer” patterns (the so called *preferential attachment* mechanism described at the beginning of this section), as well as many other real-world phenomena. For instance, transitivity (also known as clustering coefficient), network density, the diameter and the average path length (APL) of a network provide other useful information about the relational content of a network itself.²⁶

2.3.3 Networks in Gravity

So far, I offered an overview of what a network and the metrics defining it are, both in terms of size and in terms of the different definitions of centrality/prestige. I now turn to discuss how network features can be included in the econometric analysis of bilateral data (usually addressed in with ad hoc extensions of the gravity model. See for instance Head and Mayer, 2014).

The underlying limitation of the usual gravity estimation (implemented by mean of FE or pooled estimation) is that it considers the single bilateral relationships as independent from the whole set of relationships a country may be involved in (which are usually very numerous in international economic networks). Nonetheless, the behavior of economic actors cannot be considered as independent from the complex web of relationships they are embedded into. To control for this issue, Anderson and Van Wincoop (2003) introduced the idea of *Multilateral Resistance*, to

²⁵Garas et al. (2106) excluded this possibility, but made no effort to include this class of measures into the analysis).

²⁶These statistics are briefly described in Appendix 2.C proposes a review similar to the one above. Given the purpose of this research, they are mostly suitable to describe the networks, rather than to be included into the analysis

take into account the relative attractive force exerted by potential economic partners outside the bilateral relationship between two countries.²⁷ Graham (2015) maintain that the empirical counterpart of the theoretical multilateral resistance limits the control of interdependence to the structure of the error term. Despite this strategy fit satisficingly the structure of the error term in the cross-section case, a much more complex mathematical validation process would be needed to extend the same idea to the longitudinal case.

Approaching complex longitudinal dyadic data in terms of complex networks helps showing how the structure of interdependence between economic actors (countries) changes over time, and how it may ultimately “shape multiple structures of interdependence over time” (Sciabolazza, 2018, p. 243). Yet, the practical estimation of the position of migrants networks on bilateral FDI flows remains an issue to be carefully considered. In the econometric exercise, I depart from the traditional fixed effects estimation, to move towards a more flexible approach based on a multilevel mixed effects model.

Empirical Approach and Estimation Methodology

In the econometric analysis I depart from most of the existing literature on complex networks and bilateral economic exchanges in at least three respects.

First, as opposed to Fagiolo and Mastrorillo (2014); Sgrignoli et al. (2015); Garas et al. (2106) I do not use total bilateral flows as dependent variable. Indeed, taking the reciprocal flows (stocks) between two countries, and working with the resulting undirected network generally solves many computational issues (not last, the possibility to estimate in a single step the bilateral network effect on the dependent variable). Nonetheless, it does not allow to understand the channels through which the IMN topology affects bilateral FDI.²⁸ Collapsing both the IMN and the GFDIN by taking the sum of the respective flows at bilateral level may perfectly fit the purpose of the researcher looking for a general causal relationship between the IMN and FDI (eventually, that is what Fagiolo and Mastrorillo found in the ITN-IMN relationship), but does not allow to investigate further how the migrants network affects such mechanism. This whoul require both the IMN and the GFDIN to be kept directed.²⁹ In this sense, my approach is closer to Metulini et al. (2017) and to the second part of Garas et al. (2106).³⁰

²⁷Bertoli and Moraga (2013) provided the theoretical foundation of multilatera resistance in the context of migration.

²⁸Suppose to take the aggregate bilateral flow between two countries (denoted i and n). This implies that the values of the flows (stocks) going in either direction between them ($ni = i \rightarrow n + n \rightarrow i$) are summed together. Suppose that i receives a modest number of migrants from n , despite a large number of investments flowing the other direction; on the other side, n could invest very little in i . While receiving a huge number of migrants from it. Summing up migrants and FDI flows (stocks) between those two countries would make impossible to distinguish them from another pair c,d which is characterized by average flows (stocks) of FDI and migrants in both direction.

²⁹Remind from above that keeping the network directed implies that each node is characterized (at bilateral level) by two arcs instead of just one edge.

³⁰This latter study is particularly relevant. To the best of my knowledge, it represents the only existing study of the impacts of the structure of migrants networks on FDI bilateral exchanges, despite their empirical analysis is plagued by misspecification issues. Indeed, applying a poisson pseudo-maximum likelihood (ppml) estimator, the authors consider the dependent variable in log form, causing the coefficients not to be correctly specified

Second, I do not limit the analysis to a single direction in the IMN, as done in most of the existing literature: in most studies, bilateral network statistics are built summing up the reciprocal inward migration flows of the countries in each dyad and the reciprocal outward trade/investment flows between them.³¹ Since I am interested in the directed effect of migration from an hypothetical country n to an OECD country i and in the FDI flow in the opposite direction, I expect the outward connectivity of the emigration country to be as statistically relevant as the inward connectivity of the destination (with respect to the IMN). As a matter of facts, both measures capture different aspects of the openness of the two countries, and may affect the bilateral FDI channel differently. Indeed migrants also provide information about the work ethics and labor potentiality of a certain country. To consider only the joint bilateral flow means to ignore the possibility that the signal from a country's diaspora in a specific destination may suffer from the competition arising from all other migrants communities in different destinations (and vice versa). Thus, the competing communities abroad may suffer from the higher efficacy of a certain group in directing investments or exploiting connections in the motherland. Considering the inward connectivity in the IMN of the investing country as opposed to the outward connectivity of the recipient economy is a way to take into account such potential "dilution effect". Existing evidence seems to point in favor of a virtuous impact of IMN centrality and bilateral FDI, as they may both indicate overall openness for both countries. Given that no existing study takes both sides of the IMN into account, the potential existence of a dilution effect remains an ultimately empirical issue, that has to be tested in the data.

Third and last, under the estimation perspective I depart from the traditional gravity estimation. Indeed, the hierarchical structure of the dataset I construct (by merging together bilateral data as well as contextual information and network variables) implies that a large amount of information, as well as all the heterogeneity across observations, would be controlled out by standard FE estimation. As a matter of facts, the inclusion of network variables into the frame adds a new level to the structure of the data, in which a country is located into the bilateral relationships with its partners and, at the same time, within the (more or less thick) web of relationships defined globally within the IMN. For this reason, I detach from the existing international trade and investment literature, to adopt a multilevel mixed modeling, similar to the one proposed by Drzewoszewska (2014) and Giovannetti et al. (2018), which better control for the structure of the data when in presence of such complex structures (Rabe-Hesketh and Skrondal, 2012). The strength of this class of models is that they endogenize the hidden hierarchical structure (for the purpose of this chapter, I assume bilateral flows nested at country pair level, nested within countries individual networks, nested within time), while allowing the error term to be determined at the level of the single dyad, varying across levels. A mixed model constitutes an alternative approach to gravity analysis (that I applied in the first two chapters) that tends to flatten data heterogeneity by focusing on the sole within-group variability (Giovannetti et al.,

(Santos Silva and Tenreyro, 2006). Replicating their analysis (especially the one taking the aggregate bilateral flows as dependent and explanatory variables) with the data used in this chapter confirms such concerns: i.e. their results are only robust when the dependent variable is taken in logarithmic terms, loosing consistence when the dependent variable is correctly kept in levels.

³¹This make sense in the bilateral approach, but looses of justification when incoming and outgoing directed arcs are maintained separated, as it is in this case

2018).³² Differently, a mixed model does not merely control out the structure of the data: it explicitly models it, allowing for instance both the intercept and the slope of certain coefficients to vary across dyads (by letting the variance of such coefficients to have a stochastic component). For this reason, according to Bell and Jones (2015), mixed models are generally more elastic than their FE counterpart, which should be interpreted as a special case of mixed model where the entire between group variability is controlled out. Since the between group variability (in this case, the difference across country pairs) is generally larger than the within group variability (that is, within the same group of observations), the adoption of a mixed model allows exploiting a much larger amount of information.

Such hierarchical data structure, together with the assumptions of non-homogeneity and non-constant correlation in the structure of the error term across the different levels considered (as well as across dyads), implies the estimated equation to resemble

$$y_{ni,t} = \sum_{r=1}^R \beta_r X_{ni,t}^r + \nu_i + e_{ni} + \nu_t \quad (2.12)$$

$r = (1, \dots, R)$ is the number of regressors included, while i, n denote the country of origin and destination of the flow of interest respectively. Finally, ν_i, e_{ni}, ν_t explicit the three levels of the error term. Accordingly, the empirical multilevel gravity equation is defined as

$$\ln FDIcount_{ni,t}^{5y} = \ln mig_{in,t}^{high,l5} + \sum_{net=1}^{NW} \gamma_{net} X_{ni,t}^{net,l5} + \sum_{r=1}^R \beta_r X_{ni,t}^r \quad (2.13)$$

In equation (2.13), the dependent variable $\ln FDIcount_{ni,t}^{5y}$ represents the 5-year cumulative sum of the *count* of greenfield FDI from country i to country n . $\ln mig_{in,t}^{high,l5}$ is the 5-years lagged stock of tertiary educated migrants (which proxies for highly skilled migration) from country n to country i , while $\sum_{net=1}^{NW} \gamma_{net} X_{ni,t}^{net,l5}$ is the vector of both direct and indirect (lagged) network characteristics included in the regression, as described in section 2.3. I use lags to reduce reverse causality, which might affect the relationship between migration and FDI, while at the same time acknowledging that migrants' networks (whatever defined) need time to structure and become able to establish those channels for information to flow. Finally, $\sum_{r=1}^R \beta_r X_{ni,t}^r$ includes all the usual bilateral control variables, such as the log bilateral distance between i and n , and the logs of i and n 's real per capita GDP. The estimates are reported in section 2.5, while more on debate between Mixed and Fixed effects estimation is reported in Appendix 2.B, which also reports and compare the coefficients for some centrality measure when estimated in either way.

³²With mixed model I refer to multilevel regression composed of a fixed part as a base level, and by one or more random levels defined according to the assumptions about the hidden hierarchical structure of the data. See appendix 2.B for a comparison with FE estimation.

2.4 Topology and Descriptive Statistics

Below, I introduce the data used in the empirical section and provide some basic descriptive statistics (section 2.4.1). Then, I describe the correlation between the structure of the IMN and that of the GFDIN, in a fashion similar to the one proposed by Fagiolo et al. (2008) and Garas et al. (2106). The objective of section 2.4.2 is to describe the relationship between a country openness in terms of capital flows and in terms of migration (inward and outward), by means of the statistics presented in section 2.3.2.

2.4.1 Data

Data on bilateral FDI statistics, which constitute our dependent variable, come from the *FDI Market* database, provided by Financial Times (2017). *FDI Market* constitutes one of the two main sources of transactional investment data available (i.e. data at level of single project or transaction). For a discussion about pros and cons of using this kind of data as compared to traditional balance of payment records, see the data appendix in chapter 1.³³

Figures on total bilateral migrants statistics come from UNDESA-population statistics division (UNDESA, 2015), which collects data for 202 countries and autonomous territories around the world with a 5 year span between 1990 to 2015. I compute network statistics from it, not to lose the structural information which derive from a global coverage. Differently, high skill migration data comes from the IAB Brain Drain Dataset (Brucker et al., 2013), which records census based immigration flows data in 20 OECD countries.³⁴ Using both sources together allows to identify the effect of bilateral migration once the systemic network effect is taken into account.³⁵ Other data used throughout the current and the next section come from open access sources: distance, gdp data, and other usual gravity variables come from CEPII database (gravdata and geodist dataset freely available from CEPII, 2017). Language data come from Melitz and Toubal (2014). Table 2.1 above reports the summary statistics of the main variables included in the econometric

³³The dataset contains information about all Greenfield investments which took place between January 2003 and December 2015 (last available year at the time the database was accessed, in June 2016). As recalled in the appendix to chapter 1, this statement is not entirely true: being based on personal interviews to MNEs' representatives, this type of data suffer from both recall bias, voluntary omission, and missed response issues. These three issues are more severe for some countries than they are for others, and some flows tend to be systematically underreported. Nonetheless, transactional data still represent the best source of FDI records broken by sector and type.

³⁴These countries are Australia, Austria, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States.

³⁵Focusing on the 20 OECD countries included in the IAB dataset drops a good number of large investors, such as Italy, Japan, and China. The latter in particular (given its fundamental role in many developing countries) constitutes a particularly relevant loss. Results have been checked to test the sensitivity of the coefficients to the exclusion of those countries that recorded a substantial outflow of FDI in the period considered, but that are de facto excluded in the reduced sample. Relaxing the skill requirement by including total migration flows did not lead to significant changes in terms of sign, even though the numerical magnitude of the coefficients is reduced (in line with the existing evidence reviewed in section 2.2.1). Table c-3 in appendix 2.C reports basic descriptive for both bilateral investment and migration flows, divided by groups of countries, to better explain the sample representativity issue.

Table 2.1: Regressors Description: Estimation Sample summary statistics

	Obs	Mean	SD	min	MAX
Standard Gravity Controls					
Count ^{5y} _{ni,t}	9,225	0.998978	1.39583	0	7.6587
ln Mig ^{tot,15} _{ni,t}	9,225	6.716067	2.929089	0	16.26366
ln Mig ^{high,15} _{ni,t}	8,039	5.944147	2.638684	0	14.09002
ln dist _{ni}	9,225	8.580736	0.838096	4.708416	9.880192
ln GDP _{c_i}	9,225	10.59282	0.47766	8.937813	11.54109
ln GDP _{c_n}	9,225	8.43646	1.5681	4.968309	11.54109
contiguity _{ni}	9,225	0.019079	0.136809	0	1
colony _{ni}	9,225	0.050623	0.219239	0	1
comlegal _{ni}	9,225	0.303848	0.459943	0	1
Measures of Local Connectivity (centrality)					
ln INDEG _i ¹⁵	9,225	5.151201	0.226307	4.110874	5.356586
ln OUTDEG _n ¹⁵	9,145	3.954678	0.446999	1.791759	5.036952
ln INSTR _i ¹⁵	9,225	14.42818	1.351787	11.92099	17.58938
ln OUTSTR _n ¹⁵	9,145	13.02374	1.590762	7.005789	16.56104
(NON)-Overlapping Network					
ln INcommon _i ^{BIN,15}	9,225	3.405737	1.112924	0	5.313206
ln INcomple _i ^{BIN,15}	9,225	3.699064	0.521604	1.386294	5.141664
ln OUTcommon _n ^{BIN,15}	9,225	3.701753	0.434784	1.098612	4.875197
ln OUTcomple _n ^{BIN,15}	9,225	4.810991	0.259271	3.89182	5.236442
ln INcommon _i ^{W,15}	9,225	12.91427	2.089016	1.94591	17.58648
ln INcomple _i ^{W,15}	9,225	13.2609	2.116861	3.912023	17.58717
ln OUTcommon _n ^{W,15}	9,225	12.51571	1.733822	4.158883	16.32769
ln OUTcomple _n ^{W,15}	9,225	10.87305	2.296394	0	16.30458
Eigenvector-based measures of Network Prestige					
ln Eigen _i ^{w,15}	9,225	-3.46662	1.359396	-6.45196	-0.41093
ln Eigen _n ^{w,15}	9,145	-6.91571	3.24157	-21.9757	0
ln PageRank _i ¹⁵	9,225	-4.33001	1.145174	-6.23998	-1.86192
ln PageRank _n ¹⁵	9,145	-6.09213	1.067524	-7.2866	-1.86192
ln Bonacich _i ¹⁵	1,576	-1.40723	0.455734	-1.8529	-0.5696
ln Bonacich _n ¹⁵	1,956	-1.16428	1.102108	-4.35237	1.064278
Kleinberg's centrality measures					
ln Hub _i ^{BIN,15}	9,225	-0.19265	0.107514	-0.51965	0
ln Auth _n ^{BIN,15}	9,145	-1.62914	1.111284	-4.57818	0
ln Hub _i ^{W,15}	9,225	-4.76347	1.217841	-8.21987	-2.52916
ln Auth _n ^{W,15}	9,145	-7.93241	2.419418	-17.8871	0
Barrat et al.'s ANNC Measures					
ln ANND _i ^{ININ,15}	9,225	3.862308	0.222935	3.58575	4.849128
ln ANND _i ^{INOUT,15}	9,225	3.405628	0.277202	3.021282	4.326693
ln ANND _n ^{OUTIN,15}	9,225	4.762068	0.176364	4.157443	5.204007
ln ANND _n ^{OUTOUT,15}	9,225	4.214646	0.088371	3.928725	4.504848

Summary Statistics based on the estimation sample, computed on the observations for which data on highly educated migration is available. This sample includes flows from the *IAB-20* to the rest of the world (including *IAB-20* countries, as in the second group in Panel B of table c-3 in the appendix).

Variables names are reported as acronyms.

analysis, reported in the same order in which they are analyzed and discussed in section 2.5.

2.4.2 Migrants Network and Greenfield FDI: a Network Description

Following Newman (2018), I compare the basic topological structure of the the IMN and the GFDIN over time, to spot some potential co-evolution patterns between the two, as well as to identify the background mechanisms that drive the evolution of each network separately. Since migration data are only available on a 5-year span, table 2.2 reports the main topological features for the years in which they overlap (that is: 2005, 2010, and 2015).³⁶ Complemented with table c-2 in appendix 2.C, they provide an overview of the size and of the system-level connectivity of the two layers (as well as their evolution).

Table 2.2: Network Comparison: General Connectivity (I)

PANEL A: Network Size						
	2005		2010		2015	
	Mig	FDI	Mig	FDI	Mig	FDI
Nodes Count	220.00	228.00	220.00	228.00	220.00	228.00
Edges Count	10534.00	1802.00	10685.00	2399.00	10688.00	2347.00
Mean In-degree	47.88	7.90	48.57	10.52	48.58	10.29
Min. In-degree	0.00	0.00	2.00	0.00	2.00	0.00
Max. In-Degree	212.00	47.00	204.00	57.00	204.00	64.00
Mean Out-degree	47.88	7.90	48.57	10.52	48.58	10.29
Min. Out-degree	5.00	0.00	4.00	0.00	3.00	0.00
Max. Out-Degree	153.00	107.00	154.00	120.00	154.00	115.00
PANEL B: General Connectivity						
	2005		2010		2015	
	Mig	FDI	Mig	FDI	Mig	FDI
Density	0.22	0.03	0.22	0.05	0.22	0.05
APL	1.67	2.17	1.67	2.08	1.67	2.08
Diameter	4.00	6.00	4.00	5.00	4.00	5.00
Assortativity	-0.29	-0.23	-0.28	-0.19	-0.28	-0.25
Transitivity	0.56	0.46	0.57	0.45	0.57	0.45

Network Topological Comparison. APL = Average Path Length. Other measures of general connectivity are reported in table c-2 in appendix 2.C. All the networks statistics reported in this table that have not been discussed in section 2.3.2 are briefly described in appendix 2.A.

SOURCES: FDI data come from fDIMarket database (Financial Times ltd). Migration data from UNDESA Population Division. The years indicated on top of the table refers to the GFDIN. Migration statistics refers to the network 5 years earlier. (UNDESA, 2015)

The first point to be noticed relates to the size of the two networks: as a matter of fact, the IMN

³⁶Since I include the 5 year lagged migration network into the econometric analysis, the plot too compare the GFDIN in a given year against the IMN as it was 5 years earlier

is substantially larger than the GFDIN (Greenfield FDI Network), at least in terms of number of edges (the \mathcal{E} dimension discussed at the beginning of section 2.3.1). The size differential is also reflected in the remaining IN- and OUT-degree statistics shown in Panel A. It emerges that while the IMN size remains substantially unchanged over time, the GFDIN grew substantially between 2005 and 2010, arresting its growth in the following five years.³⁷ These trends emerge clearly in Panel B as well, where general connectivity is reported: again, the IMN did not change substantially over the period considered, at least not as much as it did in the earlier decades (see Fagiolo and Mastrorillo, 2013, 2014; Sciabolazza, 2018, for a thorough analysis of the migrants network over a longer time span); the GFDIN on the other hand experienced a slight decrease in both its average path length (APL) and diameter. Assortativity and transitivity statistics remained substantially stable, notwithstanding the decrease in the dissortative behavior recorded in the 2005-2010 window. Overall, despite the sustained growth, the GFDIN remains far smaller than its migration counterpart (especially in terms of its intensity). Both networks share the tendency to link countries with different size and local connectivity structure (as highlighted by the negative assortativity score) and an average clustering coefficient (as captured by a transitivity score around 0.5). Nonetheless, the IMN appears closer small world configuration (Watts and Strogatz, 1998) with respect to the GFDIN.³⁸

Figure 2.1 plots the two networks against each other, comparing the link weights (value of the stock of migrants and flow of FDI respectively). Every coordinate $(x, y) = (i, j)$ constitutes an ordered pair of countries, with flow going from country i to country j . Coloring (from warmer to cooler) and size (small to large) reflect the product of i and j 's populations and per capita GDPs respectively, for all the three time spans considered. All in all, larger country pairs in terms of both GDPc and population are also characterized by larger linkages in the two layers, and this trend reinforces over time. This rough graphical comparison suggests that the IMN and the GFDIN co-evolved over time, at least in terms of flows intensity (this is different from the size statistics presented in Panel A of table 2.2, where the binary structures were compared, instead of their weighted counterpart.).

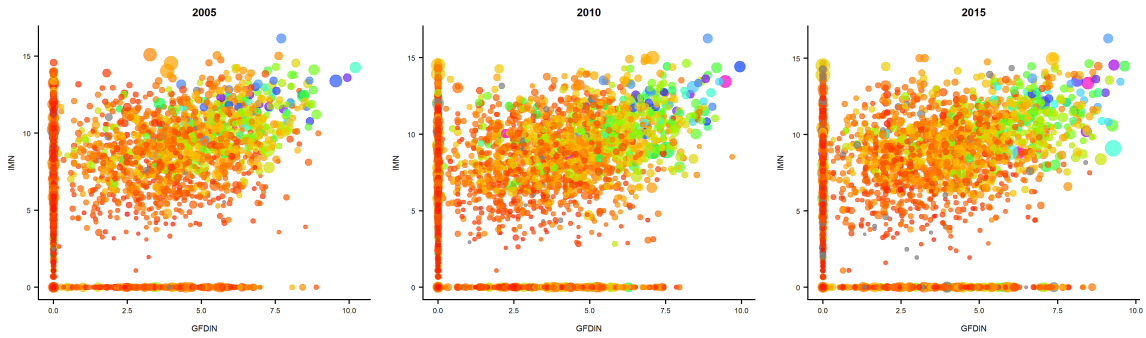
Such correlation pattern is confirmed when I look to the node-level centralities, both weighted and binary. Figure 2.2 shows node level statistics for the year 2010 only.³⁹ This time, each point

³⁷This is not entirely true: the slowdown was mostly due to the outbreak of the global financial crisis and its aftermath. Yearly data allow to have a clearer picture of the evolution of the GFDIN in that period. However, given the purpose of this exercise i.e., to compare FDI and Migrants networks, yearly network statistics for the former are not reported here. They remain available upon request to the author

³⁸Concerning these two last measures, it is worth noticing the discrepancies between such estimates and those obtained by Garas et al. (2106), which represent the only other work who compare the iMN against an investment network. For both networks, they detect a much higher (negative) assortative score. This fact could be due to the different types of data used to build the FDI and the migrants' networks (For instance, they did not distinguish FDI stocks between Greenfield and M&A. Conversely, I only focus on Greenfield flows). The major difference is represented by the migrants dataset. Whilst I use migrants stocks, they adopt flow data: yet, the choice of building a migrants stock network as opposed to a flow network is not trivial, and relates to the nature of the mechanisms I want to test. As a matter of fact, the informative potential of migrants networks and their pro-investment effects are more likely to emerge when a migrant has the time to get acquainted to its new destination, in order to establish contacts and collect knowledge about the economic tissue there. Flows are more likely to reflect short run synergies that may share the same drivers of FDI, raising additional concerns about a potential reverse causality.

³⁹Same plots for 2005 and 2010 are not shown for reason of space. They are still available upon request.

Figure 2.1: Link weight comparison over time

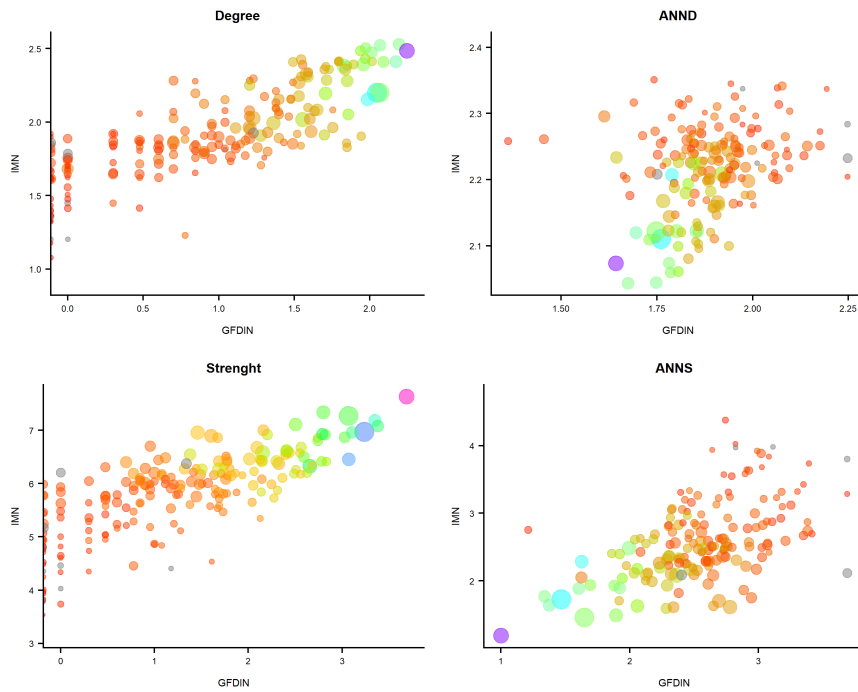


(a) Log-Log scale. Markers size is proportional to the product of the population of country i and country j . Colors from red to light blue reflect the product between per capita GDP of country i and country j

on the graph represents a single country, while size and coloring maintain the same criteria as in Fig 2.1. Every point (x, y) represents an ordinate couple $(centr_i^{gfdin}, centr_i^{imn})$. As it was in the dyadic case reported in figure 2.1, larger nodes (in terms of both GDPc and population) are characterized by a larger connectivity in both binary and weighted terms (left hand side plots), as well as a lower average connectivity of the neighbors they link to/are linked from (plots on the right). Not only these patterns confirm the relatively marked correlation between the two networks, but also give a preliminary graphical representation of the dissortative patterns highlighted in Panel B of table 2.2. Figure 2.3 further explores the dissortative patterns in both networks by plotting both binary and weighted node centrality against the respective average neighbor counterparts. Even though the trend is much better defined for the weighted networks (bottom panels), the binary structure too presents a clear dissortative trend (as suggested by Fagiolo and Mastrorillo, 2013; Garas et al., 2106, among the others).

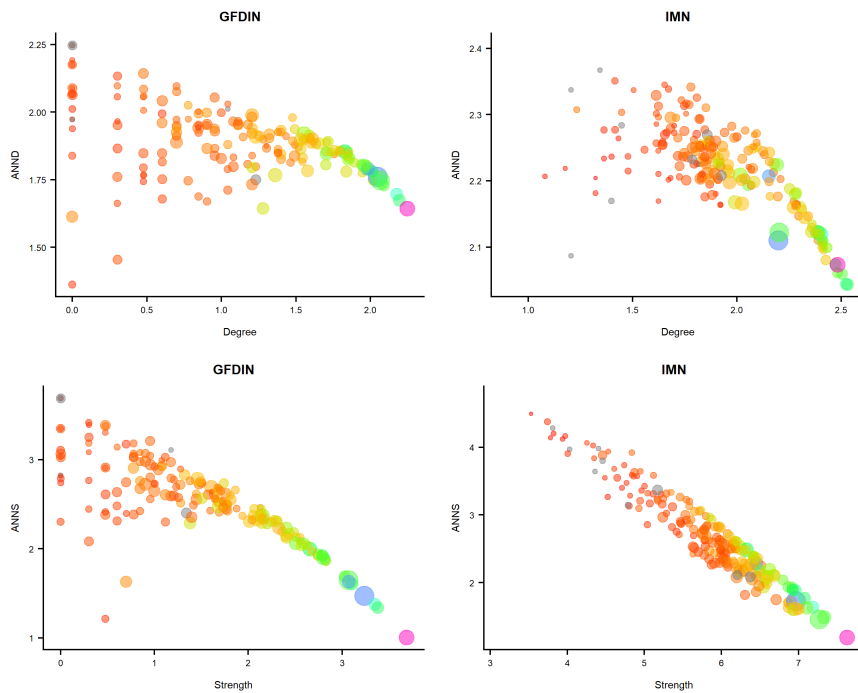
Finally, figure 2.4.2 plots the degree distribution (Total, IN, and OUT) for both the IMN and the GFDIN. While the structure of the IMN has already been investigated properly (Fagiolo and Mastrorillo, 2013; Sciabolazza, 2018), the degree distribution of greenfield network has not been reported yet. As expected, both networks share a heavily skewed distribution, with the vast majority of nodes with little or no connections and few large hubs. Similarly for both the GFDIN and the IMN, these figures suggest a power law distribution that may imply a preferential attachment structure (Barabási and Albert, 1999). Nonetheless, an in depth analysis of the GFDIN is beyond the purpose of this chapter, and is left for future analysis.

Figure 2.2: Node Centrality across layers: Year 2010



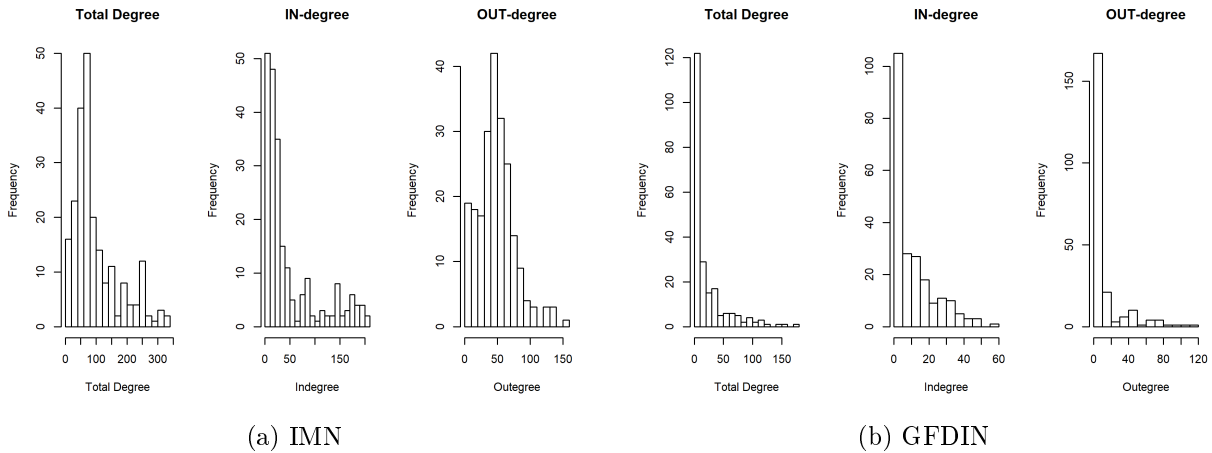
(a) Log-Log scale. Markers size is proportional the population of the country, while colors from red to light blue reflect the per capita GDP. ANN(D,S) = Average Nearest Neighbour Degree or Strength (Barrat et al., 2007)

Figure 2.3: Dissortative Patterns: Centrality vs ANNC



(a) Log-Log scale. Markers size is proportional the population of the country, while colors from red to light blue reflect the per capita GDP. ANNC = Average Nearest Neighbor Centrality (generic for ANND and ANNS)

Figure 2.4: Degree Distribution (Year: 2010)



2.5 Empirical Analysis

Despite the clear correlation pattern that emerges from the simple graphical comparison of the two networks, no conclusion can be drawn on the nature of such relationship. For this reason, I fit a gravity-like model of FDI including those measures of network centrality into the estimated equation. The purpose of this section is to provide an insight into how the position of a country in IMN ultimately affect bilateral greenfield FDI flows. Five separate exercises are conducted, with the objective of highlighting different aspects of the migration network-FDI relationship. Tables 2.1 to 2.5 are all based on equation (2.13). Each column refers to a different measure of network centrality/privilege, or to a consistent set of them, and do not report the coefficients for the set of controls included in the empirical specification (distance, colonial relationship, contiguity, and legal system, per capita GDP, etc.), to focus on the migration related variables only.⁴⁰ $\ln \text{Mig}_{ni,t}^{\text{high},l5}$ captures the direct effect of bilateral (highly educated) migration flows. It maintains a positively stable and statistically significant coefficient, with a numerical value ranging between 0.09 (4th column of table 2.2) and 0.323 (column 5 of table 2.3).⁴¹ All in all, it suggests that migrants and investments are complementary rather than substitutes.

I now turn to consider the position of each country in the IMN, which constitute the innovative part of the analysis. Table 2.1 controls for the effect of the local centrality position. Conversely

⁴⁰Since they are consistent in their sign across all specifications and maintain a stable magnitude (when significant), they are only reported in the appendices. Per capita GDPs, which account for the size effect of a given bilateral channel, also maintain a significant and positive role throughout the whole battery of results. Distance maintains a negative and significant coefficient, ranging between 0.08 to 0.16. The remaining dummy variables, indicating geographic, historical, and institutional proximity respectively maintain a positive, though not always significant value. Only common legal system and the colonial ties dummy (1 means both countries were part of the same colonial rule) show a bit of noise, probably due to the high correlation among the two. Appendix 2.D reports the full specification tables (including coefficients that have been excluded from the main text for reasons of better legibility) and the pairwise correlation between network measures included in each of them.

⁴¹These two bounds also represent the only two “irregular” estimates: indeed, taking them apart implies the coefficient for the highly educated bilateral migrants channel to remain always between 0.204 and 0.263, statistically indistinguishable across specifications.

Table 2.1: Binary and Weighted Local Connectivity

	Dep. Var. : FDI Count ^{5y} _{ni,t}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Mig ^{high,15} _{ni,t}	0.263*** (16.97)	0.264*** (15.65)	0.207*** (38.08)	0.246*** (14.95)	0.209*** (32.04)	0.204*** (31.54)	0.121*** (105.72)
ln INDEG ¹⁵ _i		-0.0343 (-0.51)				0.0823 ⁺ (1.87)	
ln OUTDEG ¹⁵ _n			0.612*** (5.42)			0.621*** (5.72)	
ln INSTR ¹⁵ _i				0.0598*** (19.75)			0.209*** (11.57)
ln OUTSTR ¹⁵ _n					0.156*** (5.96)		0.250*** (7.55)
LRtest	1545.63***	1479.55***	1305.26***	1386.03***	1363.26***	1233.10***	1219.51***
Obs	9308	9308	9225	9308	9225	9225	9225

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable “FDI” Count_{ni,t} is the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is risible: The selection of the investing subset i left only 20 OECD countries, all among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time. Table b-1 in Appendix 2.B compares local network coefficients estimates obtained via Mixed Multilevel Regression with those obtained via a two-steps FE estimator.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and common legal system; and a constant. All coefficients are shown in the full table in the empirical appendix.

from previous related studies, I distinguish between the outward connectivity of the emigration country (which captures the size and the extent of its diaspora) from the inward connectivity of the destination/investing country - which controls for the openness with respect to migration and for the presence of different migrants communities. On average, the higher connectivity of both origin and destination countries appears to have a positive effect on the bilateral number of investments from the 20 OECD countries considered as investors (i) toward the rest of the world (n), with the sole i in-degree (column 2) to have a non-statistically significant effect. This result confirms the previous evidence focused on the IMN’s effect on both trade (Fagiolo and Mastrorillo, 2014; Sgrignoli et al., 2015) and FDI (Garas et al., 2106).⁴²

Degree and Strength centralities are simple to compute and extremely easy to handle. Nonetheless, they ignore the complex structure of the network, to focus exclusively on the size of the network centered in a given country. Thus, they do not say anything about who the neighbors are, how large or how central they are, or where a node is located within the overall network.

To overcome this limitation, following Fagiolo and Mastrorillo (2014), I first try to expand the analysis by dividing the the IMN between overlapping or complementary networks. It could be that a country’s (local) centrality on the IMN captures a generic openness to international flows (which constitute different layers of the same global economic network). Indeed, splitting

⁴²A possible explanation may be that larger connectivity on the IMN implies a larger degree of integration in the global economy in general: the lack of evidence about a potential “dilution effect” of the bilateral signal could be explained with a better capability to deal with differentiated flows of information coming from different emigrant communities abroad when the countries operate in a (fairly) open economy.

global centrality between connectivity in the common destination/origin countries and connectivity outside the common set of destination/origin countries may help understanding whether the complexity of the IMN triggers third party effects. In short, third party effects in the overlapping network would imply that the connection in the emigrants/immigrants community does not extinguish its role at the bilateral level, but it also works as a bridge between their origin (or destination) country and other partners, by simply connecting fellow nationals abroad. Similarly, a large non-overlapping network could imply that two countries are embedded in different communities.⁴³ Thus, the complementary IMN subgraph may either be negatively correlated with the bilateral FDI flow (suggesting that the two countries belong to different economic spheres), or positively - in case the ethnic ties between different migrants communities “team up” in establishing international connections that may turn beneficial for bilateral FDI.⁴⁴ Table 2.2 reports the estimates for both the binary (first 4 columns) and the weighted (columns 5 to 8) IMN subnetworks, constructed to favor i 's inward connectivity as opposed to n 's outward connectivity. Results neither confirm nor exclude any of the expected mechanism discussed above. In the binary network, the outward connectivity of the migrants country of origin n points in favour of such mechanism. The inward connectivity of the investing country (i) rather shows complementarity effects, suggesting that high integration in a different community is still beneficial for bilateral investments between two countries (this finding is analogous to Fagiolo and Mastrorillo, 2014; Garas et al., 2106). Nonetheless, this pattern is reversed when considering the weighted network. This inconsistency calls for a more thorough analysis of the dualistic nature of international network when complementary and overlapping connections are considered separately.

Table 2.2 complements the results of table 2.1. It also says a little more on the effect (at least in terms of bilateral FDI) of *who* a country connects to. The two tables combined seem to rule out the possibility of a “dilution effect”, that could have emerged if being central in the IMN would have constituted a source of noise for the bilateral informative channel set up by the bilateral migration.

Tables 2.3 and 2.4 make an additional step forward and explore the role of a set of measures of centrality/prestige which takes the prestige of a neighbour into account in the definition of a node's centrality itself. These measures, which are based on the eigenvector of the adjacency matrix, take into account not just the connectivity of a certain node in the network, but also the importance of the connections themselves, resulting in a recursive definition of centrality. Thus, a country is considered as central as long as its own connections are central in the network. This is similar to distinguishing the size of a country's network to the quality of its connections. With the only exclusion of n 's Bonacich centrality and weighted hubness score of country i when taken alone (columns 6 of table 2.3 and column 4 of table 2.4 respectively) all the proposed measures show a positive and statistically significant effect on bilateral FDI. These results confirm the idea that being highly connected in the network (as defined by tables 2.1 and 2.2) is not the

⁴³Communities in SNA refer to groups of actors/nodes characterized by high density between them and little connectivity between them

⁴⁴The evidence on the pro-trade effect of the IMN connectivity actually detected this last effect to be at play. See Fagiolo and Mastrorillo (2014)

Table 2.2: Overlapping and non-Overlapping Network Effects

	Dep. Var. : FDI Count ^{5y} _{ni,t}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Mig ^{high} _{ni,t} ¹⁵	0.245*** (16.71)	0.252*** (16.83)	0.245*** (16.71)	0.0965*** (273.93)	0.243*** (14.27)	0.244*** (14.21)	0.219*** (29.93)	0.208*** (37.00)
ln INcommon _i ^{BIN,15}	0.153*** (13.42)	0.178*** (12.27)						
ln INcomple _i ^{BIN,15}		0.160*** (4.94)						
ln OUTcommon _n ^{BIN,15}			0.153*** (13.42)	0.966*** (10.46)				
ln OUTcomple _n ^{BIN,15}				-1.261*** (-10.34)				
ln INcommon _i ^{W,15}					0.0474*** (8.06)	0.0505*** (8.38)		
ln INcomple _i ^{W,15}						-0.00914** (-3.19)		
ln OUTcommon _n ^{W,15}							0.128*** (5.20)	0.0820*** (4.14)
ln OUTcomple _n ^{W,15}								0.0785*** (9.19)
LRtest	1590.11***	1588.19***	262.82***	965.41***	1584.26***	1516.20***	1425.06***	1354.23***
Obs	9308	9308	9308	9308	9308	9308	9308	9308

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable “FDI” Count_{ni,t} is the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is risible: The selection of the investing subset i left only 20 OECD countries, all among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and common legal system; and a constant. All coefficients are shown in the full tables in the empirical appendix.

only thing that matters: rather, the importance of who a country links to (is linked from) is also relevant to explain its bilateral investments patterns.

Finally, table 2.5 explores another aspect so far neglected in the related literature: the average connectivity of a country’s closest neighbors.⁴⁵ A country that is small in the network (i.e. has a reduced number of connections on its own) may take advantage from linking to highly connected partners. Such small country may exploit the ethnic ties of its small migrant community (in or out) and of its connections with those large emi-/immigration countries, via non-nationality related ethnic ties (Sgrignoli et al., 2015; Sciabolazza, 2018).⁴⁶ Consistently, the average centrality in the outward neighborhood of the investment recipient country (i.e., the average centrality in the set of countries toward which n sends migrants) always appears to have a detrimental effect on bilateral FDI, whereas the average centrality in i ’s inward neighborhood (the set of countries that send migrants to i) positively affects the bilateral FDI flows toward a specific destination. Results are (partially) in line with Fagiolo and Mastrorillo (2014), and points toward the existence of a potential pro-investment effect of common ethnic ties when the country of destination

⁴⁵Table 2.5 reports the estimates for the binary network only.

⁴⁶This is related to the idea of *preferential attachment*, proposed by Barabási and Albert (1999) and introduced in section 2.3

Table 2.3: Baseline Results: Eigenvector Based Prestige Measures

	Dep. Var. : FDI Count ^{5y} _{ni,t}					
	(1)	(2)	(3)	(4)	(5)	(6)
ln Mig ^{high,15} _{ni,t}	0.246*** (15.40)	0.235*** (18.04)	0.255*** (14.11)	0.221*** (28.40)	0.323*** (13.22)	0.259*** (8.44)
ln Eigen ^{w,15} _i	0.0649*** (14.65)					
ln Eigen ^{w,15} _n		0.0750*** (8.25)				
ln PageRank ¹⁵ _i			0.0342** (3.10)			
ln PageRank ¹⁵ _n				0.352*** (5.08)		
ln Bonacich ¹⁵ _i					0.224*** (12.32)	
ln Bonacich ¹⁵ _n						0.0357 (1.15)
LRtest	1431.60***	1475.05***	1497.73***	1246.31***	121.43***	156.87***
Obs	9308	9225	9308	9225	1514	1972

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable “FDI” Count_{ni,t} is the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is risible: The selection of the investing subset i left only 20 OECD countries, all among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and common legal system; and a constant. All coefficients are shown in the full table in the empirical appendix.

Table 2.4: Baseline Results: Kleinberg’s Hubness and Authority Scores

	Dep. Var. : FDI Count ^{5y} _{ni,t}					
	(1)	(2)	(3)	(4)	(5)	(6)
ln Mig ^{high,15} _{ni,t}	0.246*** (14.08)	0.244*** (17.45)	0.215*** (13.43)	0.265*** (15.52)	0.224*** (23.66)	0.221*** (22.84)
ln Hub ^{BIN,15} _i	0.764*** (7.64)		1.211*** (13.08)			
ln Auth ^{BIN,15} _n		0.153*** (12.39)	0.183*** (15.35)			
ln Hub ^{w,15} _i				-0.0173 (-1.41)		0.0198*** (8.67)
ln Auth ^{w,15} _n					0.138*** (5.93)	0.140*** (6.02)
LRtest	1404.52***	1554.88***	1456.09***	1452.04***	1307.76***	1219.01***
Obs	9308	9225	9225	9308	9225	9225

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable “FDI” Count_{ni,t} is the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is risible: The selection of the investing subset i left only 20 OECD countries, all among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and common legal system; and a constant. All coefficients are shown in the full table in the empirical appendix.

is involved. In other words, immigrants communities in a certain country i may “team up” to exploit ethnic ties, with the result of ultimately foster investments toward a specific country n .⁴⁷

Table 2.5: Baseline Results: Barrat et al ANNC

	Dep. Var. : FDI Count $_{ni,t}^{5y}$			
	(1)	(2)	(3)	(4)
$\ln \text{Mig}_{ni,t}^{\text{high},l5}$	0.263*** (16.79)	0.259*** (17.22)	0.214*** (26.53)	0.234*** (20.15)
$\ln \text{ANND}_i^{\text{ININ},l5}$	0.115+ (1.94)			
$\ln \text{ANND}_i^{\text{INOUT},l5}$		0.177*** (4.63)		
$\ln \text{ANND}_n^{\text{OUTIN},l5}$			-1.782*** (-5.75)	
$\ln \text{ANND}_n^{\text{OUTOUT},l5}$				-2.604*** (-6.53)
LRtest	1524.86***	1540.36***	1218.00***	1382.94***
Obs	9308	9308	9308	9308

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable “FDI” Count $_{ni,t}$ is the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is risible: The selection of the investing subset i left only 20 OECD countries, all among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

** Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and common legal system; and a constant. All coefficients are shown in the full table in the empirical appendix.

2.6 Conclusions

Migration flows are increasing worldwide, despite several attempts of many developed countries to limit the incoming of people (usually, of the less skilled) (McKenzie, 2007; UNDESA, 2015). However, their impact on economic exchanges is potentially very high, especially for those flows demanding for better coordination across actors, access to better information, or that may be subject to more severe contract enforcement issues. Thus, understanding the impact of diaspora on bilateral exchanges represents an important step in the understanding of the mechanisms that dominate an increasingly integrated global economy.

⁴⁷ At the same time, the negative impact on the origin side partially indicates that a sort of “dilution effect” is at play. Nonetheless, since the earlier results do not point in this direction, further analysis would be needed to clear this point beyond any reasonable doubt.

This chapter analyzes the impact of the international migrants' network (particularly, the role of highly educated migrants) on bilateral greenfield FDI exchanges between 20 OECD countries and the world.⁴⁸ Despite the relatively limited size of the dataset, which accounts for less than 2 percent of global bilateral migration channels, the results are representative for more than 40 percent of total non-null greenfield FDI channels recorded worldwide.⁴⁹ The descriptive evidence collected examining the correlation patterns between the IMN and the GFDIN (discussed in section 2.4) shows the existence of co-evolution trends between the two different networks, that should be considered as layers of the same global economic network rather than separate entities.

Building upon such exploratory evidence, I set up an econometric exercise focused on (a) the identification of the impacts of the IMN structural properties on bilateral greenfield FDI exchanges between two countries, beyond the role of traditional bilateral factors shaping bilateral investment flows; and (b) the discussion of the potential mechanisms that may be at play.

In the econometric exercise I apply an estimation technique (multilevel mixed model Rabe-Hesketh and Skrondal, 2012) not yet used in this context, in order to relax the strict assumptions over the correlation in the error term that is proper of traditional fixed effect estimation (and of longitudinal gravity models in particular: see Head and Mayer, 2014; Graham, 2015), as well as to control for the potential hidden hierarchical structure of bilateral exchange data (Giovannetti et al., 2018). Five separated sets of results are presented, each of them exploring a different network-related aspects of a country's connectivity. Bilateral FDI from country i to country n grow with the number of connections that both countries have on the IMN, with n outward connectivity being much more relevant than i 's inward connectivity, both in terms of the weighted and the binary definition of network. Consistently with the earlier literature on the relationship between IMN and Trade (Fagiolo and Mastroiello, 2014; Sgrignoli et al., 2015), I explored to what extent the diversity in the network, split between overlapping (shared) and complementary (not shared) connections in the IMN affects bilateral FDI flows between i and n . Ex ante, I could expect the size and the variety of a migrant community in/from a country to be affected by a trade-off, in terms of openness against competition between migrants' communities. Yet, the results of such exercise are not conclusive: while the overlapping network proved to be always positively associated to bilateral greenfield FDI, the role of the exclusive connections (i.e. those links that are not shared by both i and n) provided mixed evidence. The results of the Eigenvector-based measures of centrality/prestige are particularly interesting. The reason is that they do not only take into account the quantitative aspect of the positioning in the network, but also the "quality" of such position (measured in terms of the relevance of countries neighborhood) with respect to the whole network. This is in line with the idea of centrality not only as a matter of *who you are*, but also of *who you know* (Jackson, 2010). These findings are to be ultimately complemented with the analysis of the indirect network effect, captured by the average size of a node's neighbouring connectivity (Barrat et al., 2007). What emerges is that, while average neighbors inward connectivity positively affects bilateral FDI from the investing country, the same does not hold for the investment recipient economy, where a certain

⁴⁸The full list of countries is available in section 2.3

⁴⁹figures based on the values reported in table c-2

competition might be detected when its average neighbors' inward connectivity is considered. In conclusion, results appear to be robust, and confirm the importance of the network systemic dimensions of the International Migration Network (IMN) on bilateral FDI, which increase with a better positioning (of both countries of a given dyad) in it.

Despite this chapter add substantially to the understanding of the impacts of diaspora on bilateral FDI flows, a few limitations set the stage for the work beyond this thesis. The first one concerns the sample selection: in an unreported set of results I replicate the same set of regressions, first retaining all OECD countries as migrants destinations, then on the entire migrants network (as a result, I could only use total migration to proxy for the direct bilateral migrants effect, instead of limiting the analysis to the highly educated diaspora only).⁵⁰ While in the first case the results did not change substantially, they radically shifted downward when the sample includes all countries as investors rather than reducing them to 20/23/35, becoming much less clear.⁵¹ The second issue concerns the potential endogeneity of the results. In fact, I have not been able to identify a proper instrument for implementing a sound IV strategy. Even the two-steps strategy proposed in Ortega and Peri (2014) and Fagiolo and Mastrorillo (2014) did not lead to satisfactory results in terms of fit and correlation. Further effort should be put in this direction. Third, as highlighted by Fagiolo and Mastrorillo (2014) and later received by Sgrignoli et al. (2015), standard econometric techniques do not easily allow to control for the potential spatial correlation at dyadic and extra-dyadic level (Krisztin and Fischer, 2015). A spatial filtering approach could help disentangling the severe problems related to potential correlation between investment flows when serial spatial auto-correlation cannot be excluded. The fourth and final consideration refers to the scope of the analysis: indeed, this study considered the IMN and the GFDIN as two layers of the same global macroeconomic system. Previous authors focused on the relationship between the IMN and the world trade web, or between the Corporate Control Network and the ITN. However, as suggested by many scholars (Helpman et al., 2008; Aubry et al., 2012, see for instance), the movements of goods, people, and capital are likely to be co-determined: excluding one of the flow may limit the analysis, with the risk of obtaining results which are not correct, or conclusions plagued by potential endogeneity issues. Addressing this issue by means of complex networks seems to me a promising, yet under-explored field of application.

Bibliography

- Abel, Guy J. and Nikola Sander**, “Quantifying Global International Migration Flows,” *Science*, 2014, *343* (6178), 1520–1522.
- Anderson, James E. and Eric Van Wincoop**, “Gravity with gravitas: a solution to the border puzzle,” *American Economic Review*, 2003, *93* (1), 170–192.

⁵⁰Results of regressions on different samples are available upon request.

⁵¹These figures refers to the Brucker et al. (2013)'s sample; the same sample with the addition of Italy, Japan, and China; and the full OECD group respectively.

- Aubry, Amandine, Maurice Kugler, and Hillel Rapoport**, “Migration, FDI and the Margins of Trade,” *Boston: Center for International Development, Harvard University*, 2012.
- Baldwin, Richard**, *The Great Convergence: Information Technology and the New Globalization*, Harvard University Press, 2016.
- Barabási, Albert-László and Réka Albert**, “Emergence of Scaling in Random Networks,” *Science*, 1999, *286* (5439), 509–512.
- Barrat, Alain, Marc Barthelemy, and Alessandro Vespignani**, “The Architecture of Complex Weighted Networks: Measurements and Models,” in “Large Scale Structure And Dynamics Of Complex Networks: From Information Technology to Finance and Natural Science,” World Scientific, 2007, pp. 67–92.
- Barro, Robert J. and Jong Wha Lee**, “A new data set of educational attainment in the world, 1950–2010,” *Journal of development economics*, 2013, *104*, 184–198.
- Beine, Michel, Frédéric Docquier, and Hillel Rapoport**, “Brain drain and human capital formation in developing countries: winners and losers,” *The Economic Journal*, 2008, pp. 631–652.
- Bell, Andrew and Kelvyn Jones**, “Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data,” *Political Science Research and Methods*, 2015, *3* (1), 133–153.
- Bertoli, Simone and Jesús Fernández-Huertas Moraga**, “Multilateral resistance to migration,” *Journal of Development Economics*, 2013, *102*, 79–100.
- Bonacich, Phillip**, “Factoring and weighting approaches to status scores and clique identification,” *The Journal of Mathematical Sociology*, 1972, *2* (1), 113–120.
- , “Power and Centrality: A Family of Measures,” *American Journal of Sociology*, 1987, *92* (5), 1170–1182.
- Borgatti, Stephen P.**, “Centrality and network flow,” *Social Networks*, 2005, *27* (1), 55 – 71.
- **and Martin G. Everett**, “Models of core/periphery structures,” *Social Networks*, 2000, *21* (4), 375 – 395.
- Brin, Sergey and Lawrence Page**, “The anatomy of a large-scale hypertextual Web search engine,” *Computer Networks and ISDN Systems*, 1998, *30* (1), 107 – 117. Proceedings of the Seventh International World Wide Web Conference.
- Brucker, Herbert, Stella Capuano, and Abdeslam Marfouk**, “Education, gender and international migration: insights from a panel-dataset 1980-2010,” Technical Report 2013.
- Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan**, “Migrants, Ancestors, and Foreign Investments,” *The Review of Economic Studies*, 08 2018.
- CEPII**, “CEPII Gravity Database,” online, CEPII 2017.

- Checchi, Daniele, Gianfranco De Simone, and Riccardo Faini**, “Skilled migration, FDI and human capital investment,” 2007. Working Paper.
- Chuluun, Tuugi**, “Global portfolio investment network and stock market comovement,” *Global Finance Journal*, 2017, *33*, 51 – 68.
- Cuadros, Ana, Joan Martín-Montaner, and Jordi Paniagua**, “Homeward bound FDI: Are migrants a bridge over troubled finance?,” *Economic Modelling*, 2016, *58*, 454 – 465.
- D’Agosto, Elena, Nazaria Solferino, and Giovanni Tria**, “Does FDI affect migration flows? The role of human capital,” *Migration Letters*, 2013, *10* (3), 359.
- De Benedictis, Luca and Lucia Tajoli**, “The World Trade Network,” *The World Economy*, 2011, *34* (8), 1417–1454.
- De Masi, Giulia and Giorgio Ricchiuti**, “The network of European outward foreign direct investments,” in A. Amighini, S. Gorgoni, and M. Smith, eds., *Networks of International Trade and Investment*, 2018.
- , **Giorgia Giovannetti, and Giorgio Ricchiuti**, “Network analysis to detect common strategies in Italian foreign direct investment,” *Physica A: Statistical Mechanics and its Applications*, 2013, *392* (5), 1202–1214.
- Docquier, Frédéric and Elisabetta Lodigiani**, “Skilled migration and business networks,” *Open Economies Review*, 2010, *21* (4), 565–588.
- Drzewoszevska, Natalia**, “Multilevel Modelling of Bilateral Trade Flows between European Union Countries,” *Przegląd Statystyczny*, 2014, *61* (3), 283–305.
- Fagiolo, Giorgio and Marina Mastrorillo**, “International migration network: Topology and modeling,” *Physical Review E*, 2013, *88* (1), 012812.
- and – , “Does human migration affect international trade? A complex-network perspective,” *PloS one*, 2014, *9* (5), e97331.
- , **Javier Reyes, and Stefano Schiavo**, “On the topological properties of the world trade web: A weighted network analysis,” *Physica A: Statistical Mechanics and its Applications*, 2008, *387* (15), 3868–3873.
- , – , and – , “World-trade web: Topological properties, dynamics, and evolution,” *Physical Review E*, 2009, *79* (3), 036115.
- Federici, Daniela and Marilena Giannetti**, “Temporary migration and foreign direct investment,” *Open Economies Review*, 2010, *21* (2), 293–308.
- Feenstra, Robert. C. and Gordon. H. Hanson**, “Foreign investment, outsourcing and relative wages,” Working Paper 2008.
- Financial Times**, “fDI Market Database,” online, Financial Times Business Intelligence Division 2017.

- Flisi, Sara and Marina Murat**, “The hub continent. Immigrant networks, emigrant diasporas and FDI,” *The Journal of Socio-Economics*, 2011, 40 (6), 796–805.
- Garas, Antonios, Athanasios Lapatinas, and Konstantinos Poullos**, “The relation between migration and FDI in the OECD from a complex network perspective,” *Advances in Complex Systems*, 2106, 19 (06n07), 1650009.
- Gheasi, Masood and Peter Nijkamp**, “A brief overview of international migration motives and impacts, with specific reference to FDI,” *Economies*, 2017, 5 (3), 31.
- Giovannetti, Giorgia and Mauro Lanati**, “Do High-Skill Immigrants trigger High-Quality Trade?,” *The World Economy*, 2016.
- , **Marco Sanfilippo, and Margherita Velucchi**, “Diverse twins: analysing China’s impact on Italian and German exports using a multilevel quantile regressions approach,” *Applied Economics*, 2018, 50 (28), 3051–3065.
- Gould, David M.**, “Immigrant links to the home country: empirical implications for US bilateral trade flows,” *The Review of Economics and Statistics*, 1994, pp. 302–316.
- Graham, Bryan S.**, “Methods of Identification in Social Networks,” *Annual Review of Economics*, 2015, 7 (1), 465–485.
- Head, Keith and Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Helpman, Elhanan, Marc Melitz, and Yona Rubinstein**, “Estimating trade flows: Trading partners and trading volumes,” *The Quarterly Journal of Economics*, 2008, 123 (2), 441–487.
- Jackson, Matthew O.**, *Social and economic networks*, Princeton University Press, 2010.
- Javorcik, Beata S., Çağlar Özden, Mariana Spatareanu, and Cristina Neagu**, “Migrant networks and foreign direct investment,” *Journal of Development Economics*, 2011, 94 (2), 231–241.
- Joyez, Charlie**, “On the topological structure of multinationals network,” *Physica A: Statistical Mechanics and its Applications*, 2017, 473, 578–588.
- Khalifa, Sherif and Evelina Mengova**, “Trading Tasks and Skill Premia,” *Journal of Finance and Economics*, 2015, pp. 46–71.
- Kivelä, Mikko, Alexander Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter**, “Multilayer Networks,” *ArXiv e-prints*, 2013.
- Kleinberg, Jon M.**, “Authoritative Sources in a Hyperlinked Environment,” *J. ACM*, September 1999, 46 (5), 604–632.
- Krisztin, Tamás and Manfred M Fischer**, “The gravity model for international trade: Specification and estimation issues,” *Spatial Economic Analysis*, 2015, 10 (4), 451–470.

- Kugler, Maurice and Hillel Rapoport**, “Skilled emigration, business networks and foreign direct investment,” 2005. Working Paper.
- **and** —, “International labor and capital flows: Complements or substitutes?,” *Economics Letters*, 2007, *94* (2), 155–162.
- , **Oren Levintal, and Hillel Rapoport**, “Migration and cross-border financial flows,” *The World Bank Economic Review*, 2017.
- Leblang, David**, “Familiarity Breeds Investment: Diaspora Network and International Investment,” *American Political Science Review*, 2010, *104* (3), 584–600.
- Li, Xiang, Yu Ying Jin, and Guanrong Chen**, “Complexity and synchronization of the world trade web,” *Physica A: Statistical Mechanics and its Applications*, 2003, *328* (1), 287–296.
- Lü, Jinhua, Guanrong Chen, Maciej J. Ogorzalek, and Ljiljana Trajković**, “Theory and applications of complex networks: Advances and challenges,” in “Circuits and Systems (ISCAS), 2013 IEEE International Symposium on” IEEE 2013, pp. 2291–2294.
- Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and Edward J. Taylor**, “Theories of international migration: A review and appraisal,” *Population and Development Review*, 1993, pp. 431–466.
- McKenzie, David**, “Paper walls are easier to tear down: Passport costs and legal barriers to emigration,” *World Development*, 2007, *35* (11), 2026–2039.
- Melitz, Jacques and Farid Toubal**, “Native language, spoken language, translation and trade,” *Journal of International Economics*, 2014, *93* (2), 351–363.
- Metulini, Rodolfo, Massimo Riccaboni, Paolo Sgrignoli, and Zhen Zhu**, “The Indirect Effects of FDI on Trade: A Network Perspective,” Working Paper 2017.
- Montgomery, James D.**, “Social networks and labor-market outcomes: Toward an economic analysis,” *The American economic review*, 1991, pp. 1408–1418.
- Munshi, Kaivan**, “Networks in the modern economy: Mexican migrants in the US labor market,” *The Quarterly Journal of Economics*, 2003, pp. 549–599.
- Newman, Mark**, “The Structure and Function of Complex Networks,” *SIAM Review*, jan 2003, *45*, 167–256.
- Newman, Mark**, *Networks*, Oxford university press, 2018.
- Ortega, Francesc and Giovanni Peri**, “Openness and income: The roles of trade and migration,” *Journal of International Economics*, 2014, *92* (2), 231 – 251.
- Özden, Çağlar, Christopher R. Parsons, Maurice Schiff, and Terrie L. Walmsley**, “Where on earth is everybody? The evolution of global bilateral migration 1960–2000,” *The World Bank Economic Review*, 2011, *25* (1), 12–56.

- Peri, Giovanni and Francisco Requena-Silvente**, “The trade creation effect of immigrants: evidence from the remarkable case of Spain,” *Canadian Journal of Economics/Revue canadienne d’économique*, 2010, 43 (4), 1433–1459.
- Portes, Richard and H el ene Rey**, “The determinants of cross-border equity flows,” *Journal of International Economics*, 2005, 65 (2), 269 – 296.
- Rabe-Hesketh, Sophia and Anders Skrondal**, *Multilevel and longitudinal modeling using Stata*, STATA press, 2012.
- Rauch, James E.**, “Business and social networks in international trade,” *Journal of economic literature*, 2001, pp. 1177–1203.
- **and Vitor Trindade**, “Ethnic Chinese networks in international trade,” *Review of Economics and Statistics*, 2002, 84 (1), 116–130.
- Santos Silva, Jo ao M. C. and Silvana Tenreyro**, “The log of gravity,” *The Review of Economics and statistics*, 2006, 88 (4), 641–658.
- Schiavo, Stefano, Javier Reyes, and Giorgio Fagiolo**, “International trade and financial integration: a weighted network analysis,” *Quantitative Finance*, 2010, 10 (4), 389–399.
- Sciabolazza, Valerio Leone**, “A Net of Moving People: Network Analysis of International Migration Flows,” in A. Amighini, S. Gorgoni, and M. Smith, eds., *Networks of International Trade and Investment*, 2018.
- Sgrignoli, Paolo, Rodolfo Metulini, Stefano Schiavo, and Massimo Riccaboni**, “The relation between global migration and trade networks,” *Physica A: Statistical Mechanics and its Applications*, 2015, 417, 245 – 260.
- Simone, Gianfranco De and Miriam Manchin**, “Outward migration and inward FDI: factor mobility between eastern and western Europe,” *Review of International Economics*, 2012, 20 (3), 600–615.
- Sopranzetti, Silvia**, “Overlapping free trade agreements and international trade: A network approach,” *The World Economy*, 2017, 41 (6), 1549–1566.
- UNDESA, Population Division**, “Trends in International Migrant Stock: Migrants by Destination and Origin,” online 2015.
- Vitali, Stefania, James B. Glattfelder, and Stefano Battiston**, “The Network of Global Corporate Control,” *PLOS ONE*, 10 2011, 6 (10), 1–6.
- Wallerstein, Immanuel**, *World-systems analysis* 1987.
- Wang, Li-Hsun**, “Accounting quality and information asymmetry of foreign direct investment firms,” *Research in International Business and Finance*, 2017, 42, 950 – 958.
- Watts, Duncan J. and Steven H. Strogatz**, “Collective dynamics of "small-world" networks,” *nature*, 1998, 393 (6684), 440.

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

Zhu, S. C. and D. Trefler, “Trade and inequality in developing countries: a general equilibrium analysis,” *Journal of International Economics*, 2005, pp. 21–48.

Appendix

2.A Terminology Appendix

The network statistics introduced in section ?? are by no mean sufficient to describe the network itself: indeed, in section ?? as well as in appendix 2.C, I use several other measures to describe both the IMN and the GFDIN. While many of them are quite intuitive in what they represent, others deserve a clarification. Below, I briefly review some additional notation.

- (i) **Triad** or triplet: it represents the simplest structure which network characteristics can be identified. As the name suggests, is composed of 3 nodes or actors. Triads can be of different kinds, reflecting different types of basic social structures. Mapping the emergence of each possible type of triads (the so called *triad census*) gives important information over the basic structure of a network.
- (ii) **Density**: mathematically, it is the ratio between the number of existing edges/arcs in a network and the number of potential links that might exist in a graph of a specific size N . It gives an idea of the overall network connectivity. Density in real world networks can be considered a two edge blade: the denser the network, the highest its resilience (resistance to a node's dropout); at the same time, stronger density implies also faster contagion and faster information circulation.
- (iii) **Walks and Paths**: A walk is a sequence of links that allow moving from a node toward another, no matter whether such sequence includes repeated links. A path is a walk that does not pass through the same link twice. Both walks and paths can be either directed or not, so that they are defined in both directed and undirected graphs. A path that starts from a node and finishes at the same node is called **cycle**. Two nodes can be connected by more than one walk/path/cycle.
- (iv) **Geodesic, Diameter and APL**: the geodesic is the shortest path connecting a node to another (in terms of number of links between the two nodes). The diameter is then the largest geodesic in the network, and gives an idea of the size and the connectivity of a graph. The APL (or, Average Path Length) is the average geodesic (shortest distance) between any couple of nodes (or dyad) in the network. Small APL and limited diameter identify *small world* patterns (Watts and Strogatz, 1998) .
- (v) **Degree Distribution** (Figure 2.4.2): is the distribution of the relative frequencies of nodes having different degree. Degree distribution is a fundamental characteristic of a network, as it allows to detect interesting feature, such as the existence of a *power law*. The identification of Barabási and Albert (1999)'s' preferential attachment behavior (or scale-free network) is based on degree distribution.
- (vi) **Clustering Coefficient**: Clustering refers to the tendency of nodes to form closed triads. It is often referred to as *transitivity* since it reflects the tendency of nodes sharing a common

neighbor to link together. Real networks exhibit higher clustering than expected by a random generating process.

- (vii) **Assortativity and Homophily**: two similar though unrelated concepts. Assortativity is the tendency of nodes to link with neighbors that are similar to them in terms of network connectivity: when nodes with different size are more likely to link together, a network is said to follow a *dissortative* pattern (that is another typical feature of scale free networks). Homophily on the other hand reflects the tendency of nodes to link to their similar in terms of a specific shared characteristics (sex, status, ethnicity among others).

2.B Mixed models vs FE gravity

Bilateral data such as trade, often rely on variations of the gravity equation. This class of model is generally estimated by means of fixed effects (FE) or pooled data estimation. However, when the data structure is highly complex and can be ranked according to a hidden hierarchy, FE estimation may not be the best strategy to follow. In this chapter, I overlap both bilateral and country specific information, by including migrants' network variables into the frame. Network variables represent higher order data that operate at different levels compared to the usual bilateral perspective included into gravity analysis: thus, applying multilevel regression seems appropriate. Nonetheless, there are also other reasons to prefer a mixed approach to a FE.

The use of FE estimation and pooled regression in gravity analysis is usually justified by two features. First, FE reduce the incidence of omitted variable bias. Second, the inclusion of the appropriate set of effects allows to control for the multilateral resistance term (MRT, Anderson and Van Wincoop, 2003). As a matter of fact, the exclusion of MRT, in case they are correlated with the gravity dimensions included in the model, would substantially bias the coefficients of the estimated structural gravity (Baier and Bergstrand, 2007).⁵² Provided the model to be correctly specified, FE (within) estimation allows to think of the estimated coefficients as causal effects (According to Bell and Jones, 2015, this constitute the stronger factor in favor of FE modeling. Indeed, the authors express skepticism about the effective capability of FE modeling to identify a causal effect, especially when the model falls into overfitting problems.). However, FE gravity estimation is not free from criticalities. First, it implies a strong analytical assumption concerning the structure of the error term and the degree of interconnection between units. Indeed, by considering the correlation in the error term to be constant across observations (country pairs), FE models ignore the specificity of each bilateral relationship and the resulting average variation that occur between country pairs. This might be particularly relevant in the case of historical, geographical (Egger, 2000), as well as relational features. When data are characterized by a highly complex structure, the assumption of homogeneity across units could generate a bias in the coefficients.⁵³ Second, the inclusion of the proper set of FE to control for the MRT complicates the estimation of all those variables that are time invariant or monadic, as they would require the variable of interest to be regressed on the estimated fixed effects in a second step (Head and Mayer, 2014; Head and Ries, 2008). Fagiolo and Mastrorillo (2014) and many others solved this issue by summing up the trade and migration flows in either directions at country-pair level: this strategy allows to maintain the variability of the variables

⁵²In addition, estimating structural gravity by means of ppml, allows the FE included in the equation to precisely represent the empirical counterpart of the theoretical MRT (Yotov et al. (2016) and Fally, 2012)

⁵³This issue is central in empirical gravity estimation. Among the others, Cameron and Miller (2015) developed a methodology to consistently cluster SE when dealing with dyadic data, made operational in the STATA suite by Belotti et al. (2018)

Table b-1: Binary and Weighted Local Connectivity

Panel A: FE second stage monadic estimation			
ln INDEG _{<i>i</i>}	0.298*** (47.46)		
ln OUTDEG _{<i>n</i>}		0.804*** (82.76)	
ln INSTR _{<i>i</i>}			0.172*** (117.69)
ln OUTSTR _{<i>n</i>}			0.171*** (60.01)
Panel B: Mixed regression (Tab 2.1 estimates)			
ln INDEG _{<i>i</i>}	-0.0343 (-0.51)		
ln OUTDEG _{<i>n</i>}		0.612*** (5.42)	
ln INSTR _{<i>i</i>}			0.0598*** (19.75)
ln OUTSTR _{<i>n</i>}			0.156*** (5.96)

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses.

Panel A reports the estimates from the second stage of the FE estimation as in Head and Mayer (2014). The first stage (available upon request) includes all the controls as in equation (??) with country×year FE to control for MR and it is estimated via PPML, using the user-written command *ppml_panel_sg* (Larch et al., 2017). Standard Errors are clustered at country pair. The dependent variable in the second step depends on which country’s specific characteristic is estimated. Thus, when we are interested in origin (destination) side centrality, the dependent variable is represented by the first stage’s ln (origin×year) (ln (destination×year)) FE.

Panel B reports the estimates from the 3-level regression with random intercepts as reported in Table 2.1. Dyads are nested in networks, all nested in time.

of interest beyond the reach of the FE, but does not allow to differentiate the relative impacts of the two coupled channels. Garas et al. (2106) “solved” this issue by applying a reduced set of FE (including separately origin, destination, and time): nonetheless, their approach is likely to incur in what Baldwin and Taglioni (2006) defined as bronze, silver, and gold medal specification errors.

As pointed out in section 2.3.3, I detach from proper gravity analysis, to embrace a multilevel mixed estimator. This methodology can be applied whenever data are suspected to be hierarchically structured, and the observed units have been monitored over time. In short, mixed models merge together a fixed component with one or more random parts that constitute the levels of the hidden hierarchy. This class of models relaxes the strict homogeneity assumption over the error term (by specifying the structure of the dataset instead of controlling out the heterogeneity across the observed dyads, it avoids to incur in both misspecification and overfitting issues, providing in this way an equally efficient estimator when compared to FE). See ??.

Table b-1 replicates the result in table 2.1, but compares the results of FE estimation against the results of the baseline mixed estimator. Notice that I report the estimate for the second stage of a two-steps procedure. In the first step I fit a full gravity equation with all the necessary bilateral controls and the suitable set of FE. The estimated country×year FE (country alone in the cross sectional case) is then regressed on the full set of monadic characteristics that are de

facto absorbed by such term.

Since the second stage regresses the country×year FE on that country specific monadic feature, it is not possible to estimate an equation including monadic information that is specific to the partner. For this reason, table b-1 does not reports the replication of columns (6) and (7) from table 2.1, where both in-degree(-strength) of country i and out-degree(-strength) of country n were included. Beyond the greater flexibility, this last issue reinforces our argument in favour of adopting a multilevel mixed modelling. Also coefficients estimates remain relatively similar across estimators, with the small differences that might be ascribed to the different way FE and Mixed models handle the stochastic component. Only the coefficient for $\ln \text{INDEG}_i$ turns significant in FE estimation when it is included alone⁵⁴.

2.C Descriptive and Topological analysis Appendix

Table c-3 reports the structure of the whole dataset, splitting the world into gorups of countries. The purpose of the table is to provide an overview of the overall dataset, and to show the representativity of the data used throughout the empirical analysis. Table c-1 and c-2 complement the description of the network proposed in table 2.2 and commented in section 2.4.2.

Table c-1: Network Comparison: Centralization Scores

	2000		2010		2015	
	Mig	FDI	Mig	FDI	Mig	FDI
Degree Centralization	0.56	0.29	0.56	0.35	0.56	0.35
Closeness Centralization	0.74	0.01	0.71	0.01	0.71	0.01
Betweenness Centralization	0.08	0.05	0.08	0.06	0.08	0.06
Eigenvector Centralization	0.71	0.81	0.70	0.79	0.70	0.80

Network Topological Comparison over the years included in the empirical analysis.

The normalizaed centralization indices proposed above include only those measures for which a theoretical maximum can be estimated with no loss of accuracy: this feature allows the two networks to be compared to each other. Unfortunately, other measures of interest such as PageRank centrality, Bonachich, ANND and the weighted variant of Degree Centrality cannot be estimated normalized. They are therefore of little interest at this stage, as they could not be compared. SOURCES: FDI data come from fDIMarket database (Financial Times ltd). Migration data from UNDESA Population Division (UNDESA, 2015)

⁵⁴The FE estimates replicating Tables d-2 to d-5 maintain the same trends as those reported here, and are available upon request.

Table c-2: Network Comparison: Connectivity

	2000		2010		2015	
	Mig	FDI	Mig	FDI	Mig	FDI
Mutual Dyads	2677.00	337.00	2747.00	484.00	2749.00	486.00
Asymm. Dyads	5180.00	1128.00	5191.00	1431.00	5190.00	1375.00
Null Dyads	16233.00	24413.00	16152.00	23963.00	16151.00	24017.00
Weak Components	1.00	66.00	1.00	50.00	1.00	54.00
Average Clique	48.00	20.00	48.00	21.00	48.00	22.00
Average Coreness	55.56	9.09	56.56	12.08	56.80	11.84
Modularity	0.09	0.13	0.11	0.12	0.12	0.08

Network Topological Comparison over the years included in the empirical analysis.

Mutual and Asymmetric Dyads refer to the share of reciprocated and not reciprocated edges respectively. Weak Components represents the number of weakly connected components intended as the maximal subgraph that would be connected if the direction of the arcs is ignored. Modularity, Coreness and Clcquishness report similar information about the networks' internal "grouping", as characterized by larger density within than between them.

SOURCES: FDI data come from fDIMarket database (Financial Times ltd). Migration data from UNDESA Population Division (UNDESA, 2015)

Table c-3: Looking to the data: Summary by flow direction

	Obs	Non-missing	Mean	SD	min	MAX
Panel A - Total Sample						
NUM Greenfield (5 years)	145860	11,685	1.050343	15.204	0	1635
High Skill Migration (IAB)	11460	9,613	5968.422	37818.17	0	1315891
Total Migration (UNDESA)	145860	32,054	4278.237	75891.03	0	12100000
Panel B: Number of Observations by Direction in the sample						
IAB-20 to IAB-20						
NUM Greenfield (5 years)	1140	926	36.71754	108.0213	0	1559
High Skill Migration (IAB)	1140	1135	15247.4	49524.84	0	533808
Total Migration (UNDESA)	1140	1122	44918.89	116738.5	0	1289396
IAB-20 to ROW (Including IAB-20)						
NUM Greenfield (5 years)	13,260	4786	7.989668	46.68598	0	1635
High Skill Migration (IAB)	11,460	9613	5968.422	37818.17	0	1315891
Total Migration (UNDESA)	13,260	5356	5390.537	39408.96	0	1289396
IAB-20 to ROW						
NUM Greenfield (5 years)	12120	3860	5.287541	34.68435	0	1635
High Skill Migration (IAB)	10320	8478	4943.419	36151.46	0	1315891
Total Migration (UNDESA)	12120	4234	1672.523	16048.16	0	876528
ROW to ROW						
NUM Greenfield (5 years)	120600	5438	0.286194	5.019805	0	826
High Skill Migration (IAB)	-	-	-	-	-	-
Total Migration (UNDESA)	120600	17756	2626.872	51541.64	0	3584076
ROW to IAB-20						
NUM Greenfield (5 years)	12000	1461	1.062083	9.443657	0	420
High Skill Migration (IAB)	-	-	-	-	-	-
Total Migration (UNDESA)	12000	8942	19645.36	203298.1	0	12100000

Total Dataset description by direction of the flow.

IAB-20 refers to the countries available in the IAB Brain Drain database Brucker et al. (2013), and includes Australia, Austria, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States.

2.D Estimates Appendix

In this appendix to the empirical section I report the full regression tables, excluded from the main text to maintain the focus on the IMN. Each regression table is also complemented by a pairwise correlation table. I do so to show how the usually high correlation between network centrality measures (Jackson, 2010; Newman, 2018) requires to carefully choose the combination of measures to be included simultaneously, in order to avoid multicollinearity.

Full Estimates

Table d-1: Full Coefficients Replica of table 2.1

	Dep. Var. : FDI Count ^{5y} _{ni,t}						
ln Mig ^{high} _{ni,t}	0.263*** (16.97)	0.264*** (15.65)	0.207*** (38.08)	0.246*** (14.95)	0.209*** (32.04)	0.204*** (31.54)	0.121*** (105.72)
ln dist _{ni}	-0.121*** (-4.39)	-0.122*** (-4.62)	-0.108*** (-3.95)	-0.140*** (-5.03)	-0.106*** (-3.88)	-0.104*** (-4.00)	-0.162*** (-5.61)
ln GDPc _i	0.350*** (5.25)	0.346*** (5.79)	0.364*** (4.99)	0.342*** (5.06)	0.359*** (5.31)	0.373*** (5.45)	0.341*** (5.02)
ln GDPc _n	0.197*** (5.56)	0.196*** (5.67)	0.187*** (6.00)	0.201*** (5.67)	0.244*** (5.81)	0.188*** (6.11)	0.285*** (6.32)
contig _{ni}	0.846*** (18.65)	0.843*** (16.04)	1.033*** (24.67)	0.843*** (17.85)	1.060*** (26.70)	1.042*** (22.14)	1.108*** (23.82)
colony _{ni}	0.159 ⁺ (1.71)	0.163 ⁺ (1.66)	0.375** (2.79)	0.141 (1.52)	0.364** (2.81)	0.369** (2.70)	0.418** (3.20)
comleg _{ni}	-0.0585*** (-3.81)	-0.0618*** (-6.04)	-0.0103 (-0.45)	-0.0538*** (-3.65)	-0.0148 (-0.72)	-0.00177 (-0.09)	0.0261 (1.08)
ln INDEG _i		-0.0343 (-0.51)				0.0823 ⁺ (1.87)	
ln OUTDEG _n			0.612*** (5.42)			0.621*** (5.72)	
ln INSTR _i				0.0598*** (19.75)			0.209*** (11.57)
ln OUTSTR _n					0.156*** (5.96)		0.250*** (7.55)
Constant	-4.782*** (-4.58)	-4.551*** (-7.34)	-7.081*** (-4.74)	-5.351*** (-5.22)	-7.170*** (-5.07)	-7.673*** (-6.32)	-10.62*** (-6.29)
ln s _{1,1}							
Constant	-3.068*** (-10.04)	-3.062*** (-10.04)	-2.855*** (-9.59)	-3.123*** (-10.17)	-3.404*** (-9.67)	-2.864*** (-9.59)	-3.910*** (-9.25)
ln s _{2,1}							
Constant	-0.395*** (-5.16)	-0.395*** (-5.18)	-0.447*** (-4.79)	-0.410*** (-5.66)	-0.418*** (-5.14)	-0.446*** (-4.74)	-0.456*** (-7.26)
ln s _{3,1}							
Constant	-0.325*** (-7.46)	-0.325*** (-7.51)	-0.335*** (-21.35)	-0.318*** (-6.42)	-0.353*** (-13.71)	-0.337*** (-21.28)	-0.371*** (-8.98)
LRtest	1545.63	1479.55	1305.26	1386.03	1363.26	1233.10	1219.51
Obs	9308	9308	9225	9308	9225	9225	9225

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable Count_{ni,t} represents the 5-year cumulate count of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is riible, due to the selection of i country, that only include 20 OECD countries that where among the top investor in the period considered (Japan, China, and Italy are excluded). ull list of country can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i an n ; three dummies for contiguity, colonial history, and common legal system; and a constant. Al coefficients are shown in the full table in the empirical appendix.

Table d-2: Full Coefficients Replica of table 2.2

	Dep. Var. : FDI Count ^{5y} _{ni,t}							
ln Mig ^h _{ni,t} ^{igh}	0.245*** (16.71)	0.252*** (16.83)	0.245*** (16.71)	0.0965*** (273.93)	0.243*** (14.27)	0.244*** (14.21)	0.219*** (29.93)	0.208*** (37.00)
ln dist _{ni}	-0.0928*** (-3.65)	-0.0850** (-3.18)	-0.0928*** (-3.65)	-0.168*** (-6.06)	-0.128*** (-4.58)	-0.123*** (-4.55)	-0.106*** (-3.90)	-0.110*** (-4.15)
ln GDPc _i	0.365*** (5.38)	0.317*** (5.05)	0.365*** (5.38)	0.342*** (4.24)	0.333*** (4.83)	0.334*** (4.84)	0.355*** (5.20)	0.346*** (5.24)
ln GDPc _n	0.142*** (4.62)	0.139*** (4.66)	0.142*** (4.62)	0.171*** (5.86)	0.177*** (4.96)	0.171*** (5.07)	0.207*** (5.74)	0.260*** (6.18)
contig _{ni}	0.874*** (20.52)	0.871*** (22.20)	0.874*** (20.52)	0.904*** (18.13)	0.865*** (18.15)	0.869*** (18.56)	0.909*** (22.69)	0.848*** (19.81)
colony _{ni}	0.201* (2.10)	0.242* (2.27)	0.201* (2.10)	0.303** (2.83)	0.153 ⁺ (1.66)	0.158 ⁺ (1.69)	0.320** (2.60)	0.339** (2.72)
comleg _{ni}	-0.0302* (-1.98)	-0.0414** (-3.04)	-0.0302* (-1.98)	0.0451 ⁺ (1.91)	-0.0563*** (-3.85)	-0.0556*** (-3.81)	-0.0201 (-0.94)	-0.02645 (-1.25)
ln INcommon ^{BIN} _i	0.153*** (13.42)	0.178*** (12.27)						
ln INcomple ^{BIN} _i		0.160*** (4.94)						
ln OUTcommon ^{BIN} _n			0.153*** (13.42)	0.966*** (10.46)				
ln OUTcomple ^{BIN} _n				-1.261*** (-10.34)				
ln INcommon ^W _i					0.0474*** (8.06)	0.0505*** (8.38)		
ln INcomple ^W _i						-0.00914** (-3.19)		
ln OUTcommon ^W _n							0.128*** (5.20)	0.0820*** (4.14)
ln OUTcomple ^W _n								0.0785*** (9.19)
Constant	-5.147*** (-4.88)	-5.418*** (-4.65)	-5.147*** (-4.88)	-0.728 (-0.91)	-4.888*** (-4.69)	-4.814*** (-4.71)	-6.432*** (-4.80)	-6.961*** (-5.04)
ln s _{1,1}								
Constant	-2.810*** (-9.86)	-2.818*** (-9.87)	-2.810*** (-9.86)	-2.580*** (-9.34)	-2.958*** (-10.06)	-2.926*** (-10.02)	-3.202*** (-9.94)	-3.480*** (-9.57)
ln s _{2,1}								
Constant	-0.384*** (-5.14)	-0.405*** (-5.32)	-0.384*** (-5.14)	-0.506*** (-5.84)	-0.400*** (-5.45) (-5.41)	-0.397*** (-4.88)	-0.399*** (-4.95)	-0.445***
ln s _{3,1}								
Constant	-0.351*** (-7.70)	-0.340*** (-7.79)	-0.351*** (-7.70)	-0.362*** (-11.23)	-0.327*** (-6.72)	-0.329*** (-6.82)	-0.353*** (-12.77)	-0.337*** (-16.27)
LRtest	1590.11	1588.19	262.82	965.41	1584.26	1516.20	1425.06	1354.23
Obs	9308	9308	9308	9308	9308	9308	9308	9308

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable $\text{Count}_{ni,t}^{5y}$ represents the 5-year cumulate *count* of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is riible, due to the selection of i country, that only include 20 OECD countries that where among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i an n ; three dummies for contiguity, colonial history, and common legal system; and a constant. Al coefficients are shown in the full table in the empirical appendix.

Table d-3: Full Coefficients Replica of table 2.3

	Dep. Var. : FDI Count ^{5y} _{ni,t}					
ln Mig ^{high} _{ni,t}	0.246*** (15.40)	0.235*** (18.04)	0.255*** (14.11)	0.221*** (28.40)	0.323*** (13.22)	0.259*** (8.44)
ln dist _{ni}	-0.133*** (-4.74)	-0.0960*** (-3.67)	-0.134*** (-4.72)	-0.158*** (-5.26)	-0.0128 (-0.15)	-0.119 ⁺ (-1.76)
ln GDPc _i	0.327*** (4.79)	0.367*** (5.50)	0.340*** (4.94)	0.339*** (4.65)	0.643*** (8.20)	0.278*** (3.31)
ln GDPc _n	0.202*** (5.71)	0.134*** (5.75)	0.198*** (5.64)	0.105*** (8.34)	0.287*** (37.14)	0.213*** (16.00)
contig _{ni}	0.839*** (17.93)	1.063*** (29.05)	0.844*** (18.50)	0.791*** (9.34)	0.990*** (13.19)	1.190*** (13.80)
colony _{ni}	0.143 (1.54)	0.282** (2.66)	0.149 (1.56)	0.278* (2.46)	0.119 (1.18)	0.237* (2.19)
comleg _{ni}	-0.0428*** (-3.58)	-0.0280* (-2.06)	-0.0567*** (-3.93)	0.0343 ⁺ (-1.74)	-0.0541 ⁺ (-1.73)	-0.168** (-2.70)
ln Eigen _i ^w	0.0649*** (14.65)					
ln Eigen _n ^w		0.0750*** (8.25)				
ln PageRank _i			0.0342** (3.10)			
ln PageRank _n				0.352*** (5.08)		
ln Bonacich _i					0.224*** (12.32)	
ln Bonacich _n						0.0357 (1.15)
Constant	-4.172*** (-3.90)	-3.991*** (-4.31)	-4.397*** (-3.79)	-1.197* (-2.33)	-9.650*** (-5.37)	-4.057*** (-7.80)
ln s _{1,1}						
Constant	-3.330*** (-9.93)	-3.062*** (-9.87)	-3.025*** (-10.06)	-2.492*** (-8.81)	-21.74 (-0.01)	-2.651*** (-4.02)
ln s _{2,1}						
Constant	-0.404*** (-5.63)	-0.420*** (-5.00)	-0.403*** (-5.36)	-0.577*** (-5.28)	-0.0682 (-0.01)	-0.206* (-2.17)
ln s _{3,1}						
Constant	-0.324*** (-6.33)	-0.344*** (-10.79)	-0.320*** (-6.87)	-0.292*** (-38.36)	-0.535 (-0.02)	-0.603*** (-5.41)
LRtest	1431.60	1475.05	1497.73	1246.31	121.43	156.87
Obs	9308	9225	9308	9225	1514	1972

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable Count_{ni,t} represents the 5-year cumulate count of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is riible, due to the selection of i country, that only include 20 OECD countries that where among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i an n ; three dummies for contiguity, colonial history, and common legal system; and a constant. Al coefficients are shown in the full table in the empirical appendix.

Table d-4: Full Coefficients Replica of table 2.4

	Dep. Var. : FDI Count ^{5y} _{ni,t}					
ln Mig ^{high} _{ni,t}	0.246*** (14.08)	0.244*** (17.45)	0.215*** (13.43)	0.265*** (15.52)	0.224*** (23.66)	0.221*** (22.84)
ln dist _{ni}	-0.136*** (-4.82)	-0.111*** (-4.14)	-0.133*** (-4.84)	-0.122*** (-4.39)	-0.160*** (-5.10)	-0.160*** (-5.07)
ln GDPc _i	0.354*** (5.16)	0.357*** (5.26)	0.364*** (5.09)	0.328*** (6.53)	0.342*** (4.86)	0.367*** (5.32)
ln GDPc _n	0.202*** (5.74)	0.144*** (4.68)	0.141*** (4.53)	0.196*** (5.64)	0.109*** (8.14)	0.110*** (8.25)
ln contig _{ni}	0.853*** (17.66)	1.034*** (30.30)	1.063*** (27.40)	0.840*** (17.38)	0.968*** (20.55)	0.977*** (19.82)
ln colony _{ni}	0.119 (1.22)	0.214* (2.16)	0.161 (1.57)	0.173 [†] (1.71)	0.325** (8.14)	0.311* (2.56)
ln comple _{ni}	-0.0384** (-3.11)	-0.0381** (-2.64)	0.00323 (-0.30)	-0.0630*** (-4.52)	-0.0255 (-1.38)	-0.0200 (-1.10)
ln Hub _i ^{BIN}	0.764*** (7.64)		1.211*** (13.08)			
ln Auth _n ^{BIN}		0.153*** (12.39)	0.183*** (15.35)			
ln Hub _i ^w				-0.0173 (-1.41)		0.0198*** (8.67)
ln Auth _n ^w					0.138*** (5.93)	0.140*** (6.02)
Constant	-4.506*** (-4.10)	-4.155*** (-4.03)	-3.580** (-3.24)	-4.625*** (-4.94)	-2.315*** (-3.38)	-2.468*** (-3.69)
ln s _{1,1}						
Constant	-3.070*** (-10.13)	-2.806*** (-9.85)	-2.755*** (-9.86)	-3.041*** (-10.02)	-2.611*** (-9.20)	-2.627*** (-9.22)
ln s _{2,1}						
Constant	-0.400*** (-5.51)	-0.391*** (-5.42)	-0.401*** (-6.12)	-0.396*** (-5.15)	-0.478*** (-4.67)	-0.475*** (-4.66)
ln s _{3,1}						
Constant	-0.326*** (-6.25)	-0.349*** (-7.72)	-0.354*** (-5.88)	-0.324*** (-7.68)	-0.338*** (-57.17)	-0.340*** (-52.20)
LRtest	1404.52	1554.88	1456.09	1452.04	1307.76	1219.01
Obs	9308	9225	9225	9308	9225	9225

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable Count_{ni,t} represents the 5-year cumulate count of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is riible, due to the selection of i country, that only include 20 OECD countries that where among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i and n ; three dummies for contiguity, colonial history, and commonegal system; and a constant. All coefficients are shown in the full table in the empirical appendix.

Table d-5: Full Coefficients Replica of table 2.5

	Dep. Var. : FDI Count _{ni,t} ^{5y}			
ln Mig _{ni,t} ^{high}	0.263*** (16.79)	0.259*** (17.22)	0.214*** (26.53)	0.234*** (20.15)
ln dist _{ni}	-0.129*** (-4.94)	-0.138*** (-5.37)	-0.151*** (-5.18)	-0.144*** (-5.11)
ln GDPc _i	0.338*** (5.48)	0.333*** (5.16)	0.336*** (4.73)	0.341*** (4.82)
ln GDPc _n	0.196*** (5.59)	0.196*** (5.55)	0.187*** (5.85)	0.174*** (5.50)
ln contig _{ni}	0.835*** (16.35)	0.829*** (16.83)	0.770*** (13.40)	0.787*** (14.19)
ln colony _{ni}	0.167 ⁺ (1.75)	0.161 ⁺ (1.73)	0.296** (2.75)	0.218* (2.26)
ln comleg _{ni}	-0.0669*** (-5.97)	-0.0687*** (-5.51)	-0.0273 (-1.33)	-0.0434* (-2.35)
ln ANND _i ^{ININ}	0.115 ⁺ (1.94)			
ln ANND _i ^{INOVT}		0.177*** (4.63)		
ln ANND _n ^{OUTIN}			-1.782*** (-5.75)	
ln ANND _n ^{OUTOVT}				-2.604*** (-6.53)
Constant	-5.030*** (-4.30)	-5.049*** (-4.56)	4.463*** (6.03)	6.837*** (8.39)
ln s _{1,1}				
Constant	-3.069*** (-10.06)	-3.080*** (-10.09)	-2.830*** (-9.27)	-2.543*** (-9.02)
ln s _{2,1}				
Constant	-0.398*** (-5.28)	-0.403*** (-5.47)	-0.578*** (-4.31)	-0.468*** (-4.71)
ln s _{3,1}				
Constant	-0.323*** (-7.42)	-0.321*** (-7.02)	-0.273*** (-42.65)	-0.311*** (-17.71)
LRtest	1524.86	1540.36	1218.00	1382.94
Obs	9308	9308	9308	9308

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t-statistics in parentheses. Robust Standard Errors.

The dependent variable Count_{ni,t} represents the 5-year cumulate count of greenfield FDI from country i to country n . Zero flows are excluded by the log-linearization, even if their occurrence is riible, due to the selection of i country, that only include 20 OECD countries that where among the top investor in the period considered (Japan, China, and Italy are excluded). Full country list can be found in the main text.

Estimates refer to a 3-level regression with random intercepts. Diads are nested in networks, all nested in time.

LRtest refers to the *log-likelihood Ratio* test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012): rejecting the null hypothesis of no difference between the standard OLS/FE estimates suggests that the multilevel model is suitable to the data.

Notice: all equations include distance (in logs); per capita GDP (in logs) of both i an n ; three dummies for contiguity, colonial history, and common legal system; and a constant. Al coefficients are shown in the full table in the empirical appendix.

Pairwise correlation

Table d-6: Local Centrality: Network Pairwise Correlation Annex to table 2.1

	$Mig_{ni,t}^{high}$	INDEG _i	OUTDEG _n	INSTR _i	OUTSTR _i
$Mig_{ni,t}^{high}$	1				
INDEG _i	-0.0431*	1			
OUTDEG _n	0.2110*	-0.1654*	1		
INSTR _i	0.3737*	-0.1302*	0.0136	1	
OUTSTR _i	0.2170*	-0.0934*	0.5672*	0.0122	1

* $p = 0.05$

Pairwise correlations between network dimensions, computed on the estimation sample (including flows from the *IAB-20* to the rest of the world, including *IAB-20* countries).

Table d-7: Overlapping Links: Network Pairwise Correlation Annex to table 2.2

	$Mig_{ni,t}^{high}$	INcommon _i ^{BIN}	INcomple _i ^{BIN}	OUTcommon _n ^{BIN}	OUTcomple _n ^{BIN}
$Mig_{ni,t}^{high}$	1				
INcommon _i ^{BIN}	0.0869*	1			
INcomple _i ^{BIN}	0.0295*	-0.2533*	1		
OUTcommon _n ^{BIN}	0.3229*	0.5614*	-0.0377*	1	
OUTcomple _n ^{BIN}	-0.2664*	-0.0986*	0.0901*	-0.4289*	1
INcommon _i ^W	0.3718*	0.2991*	0.0245*	0.3475*	-0.4864*
INcomple _i ^W	0.2128*	-0.2549*	0.1867*	0.0011	-0.5402*
OUTcommon _n ^W	0.1628*	0.1773*	0.0468*	0.4923*	-0.2002*
OUTcomple _n ^W	0.1890*	-0.0516*	0.0864*	0.2297*	-0.1491*
	INcommon _i ^W	INcomple _i ^W	OUTcommon _n ^W	OUTcomple _n ^W	
INcommon _i ^W	1				
INcomple _i ^W	0.1768*	1			
OUTcommon _n ^W	0.0580*	-0.0282*	1		
OUTcomple _n ^W	-0.0042	0.01	0.2742*	1	

* $p = 0.05$

Pairwise correlations between network dimensions, computed on the estimation sample (including flows from the *IAB-20* to the rest of the world, including *IAB-20* countries).

Table d-8: Prestige: Network Pairwise Correlation Annex to table 2.3

	$Mig_{ni,t}^{high}$	Eigen _i ^w	Eigen _n ^w	PageRank _i	PageRank _n	Bonacich _i	Bonacich _n
$Mig_{ni,t}^{high}$	1						
Eigen _i ^w	0.3058*	1					
Eigen _n ^w	0.0573*	0.005	1				
PageRank _i	0.3568*	0.7889*	-0.0029	1			
PageRank _n	0.1312*	-0.0035	0.4333*	-0.0153	1		
Bonacich _i	0.009	0.0909*	-0.0071	0.0626*	-0.0051	1	
Bonacich _n	0.0021	-0.0121	-0.0084	-0.0008	0.0199	0.0890*	1

* $p = 0.05$

Pairwise correlations between network dimensions, computed on the estimation sample (including flows from the *IAB-20* to the rest of the world, including *IAB-20* countries).

Table d-9: Authority: Network Pairwise Correlation Annex to table 2.4

	$Mig_{ni,t}^{high}$	Hub_i^{BIN}	$Auth_n^{BIN}$	Hub_i^W	$Auth_n^W$
$Mig_{ni,t}^{high}$	1				
Hub_i^{BIN}	0.2250*	1			
$Auth_n^{BIN}$	0.0880*	-0.0377*	1		
Hub_i^W	0.0338*	0.5556*	-0.0117	1	
$Auth_n^W$	0.0418*	-0.015	0.1421*	-0.0009	1

* $p = 0.05$

Pairwise correlations between network dimensions, computed on the estimation sample (including flows from the *IAB-20* to the rest of the world, including *IAB-20* countries).

Table d-10: Neighbors Importance: Network Pairwise Correlation Annex to table 2.5

	$Mig_{ni,t}^{high}$	$ANND_i^{ININ}$	$ANND_i^{INOUT}$	$ANND_n^{OUTIN}$	$ANND_n^{OUTOUT}$
$Mig_{ni,t}^{high}$	1				
$ANND_i^{ININ}$	0.1023*	1			
$ANND_i^{INOUT}$	0.2014*	0.9022*	1		
$ANND_n^{OUTIN}$	-0.1861*	-0.1456*	-0.1177*	1	
$ANND_n^{OUTOUT}$	-0.1729*	-0.1078*	-0.0790*	0.8474*	1

* $p = 0.05$

Pairwise correlations between network dimensions, computed on the estimation sample (including flows from the *IAB-20* to the rest of the world, including *IAB-20* countries).

Bibliography

- Anderson, James E. and Eric Van Wincoop**, “Gravity with gravitas: a solution to the border puzzle,” *American Economic Review*, 2003, *93* (1), 170–192.
- Baier, Scott L. and Jeffrey H. Bergstrand**, “Do free trade agreements actually increase members’ international trade?,” *Journal of International Economics*, 2007, *71* (1), 72–95.
- Baldwin, Richard and Daria Taglioni**, “Gravity for Dummies and Dummies for Gravity Equations,” Working Paper 12516, National Bureau of Economic Research 2006.
- Barabási, Albert-László and Réka Albert**, “Emergence of Scaling in Random Networks,” *Science*, 1999, *286* (5439), 509–512.
- Bell, Andrew and Kelvyn Jones**, “Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data,” *Political Science Research and Methods*, 2015, *3* (1), 133–153.
- Belotti, Federico, Edoardo Di Porto, and Gianluca Santoni**, “Spatial Differencing: Estimation and Inference,” *CESifo Economic Studies*, 2018, *64* (2), 241–254.
- Brucker, Herbert, Stella Capuano, and Abdeslam Marfouk**, “Education, gender and international migration: insights from a panel-dataset 1980-2010,” Technical Report 2013.
- Cameron, Colin A. and Douglas L. Miller**, “A practitioner’s guide to cluster-robust inference,” *Journal of Human Resources*, 2015, *50* (2), 317–372.
- Egger, Peter**, “A note on the proper econometric specification of the gravity equation,” *Economics Letters*, 2000, *66* (1), 25 – 31.
- Fagiolo, Giorgio and Marina Mastrorillo**, “Does human migration affect international trade? A complex-network perspective,” *PloS one*, 2014, *9* (5), e97331.

- Garas, Antonios, Athanasios Lapatinas, and Konstantinos Poullos**, “The relation between migration and FDI in the OECD from a complex network perspective,” *Advances in Complex Systems*, 2106, 19 (06n07), 1650009.
- Head, Keith and John Ries**, “FDI as an Outcome of the Market for Corporate Control: Theory and Evidence,” *Journal of International Economics*, 2008, 74 (1), 2–20.
- **and Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Jackson, Matthew O.**, *Social and economic networks*, Princeton University Press, 2010.
- Larch, Mario, Joschka Wanner, Yoto Yotov, and Thomas Zylkin**, “The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects,” 2017. Working Paper.
- Newman, Mark**, *Networks*, Oxford university press, 2018.
- Rabe-Hesketh, Sophia and Anders Skrondal**, *Multilevel and longitudinal modeling using Stata*, STATA press, 2012.
- UNDESA, Population Division**, “Trends in International Migrant Stock: Migrants by Destination and Origin,” online 2015.
- Watts, Duncan J. and Steven H. Strogatz**, “Collective dynamics of "small-world" networks,” *nature*, 1998, 393 (6684), 440.
- Yotov, Yoto V, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch**, “An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model,” *World Trade Organization, Geneva*, 2016.

Chapter 3

The Heterogenous Impacts of Cultural Preferences on Bilateral M&A flows

Foreign Direct Investments (FDI) and Mergers and Acquisitions (M&A) constitute an increasingly important factor for economic growth. Thus, understanding the mechanisms that regulate cross border financial flows is a matter of the utter importance. This chapter contributes to the existing literature by focusing on the existence of quantitative heterogenous patterns in cross-border investment flows. Indeed, with very few notable exceptions, none of the studies reviewed considered the possibility that bilateral FDI might respond to economic, institutional, and cultural stimuli in differentiated ways according to their size. This chapter explores the extent of such heterogeneity, posing particular attention to the analysis of the non constant Explicit Cultural Preferences and of Cultural Proximity on bilateral M&A flows. By applying a Longitudinal Censored Quantile regression with high dimensional fixed effects to estimate a fully consistent gravity model of M&A, I explore the extent of such form of non-linearity, and discuss its potential policy implications. The results suggest that different stimuli do affect M&A flows in a heterogenous way. It emerges that such heterogeneity is stronger when the asymmetric and time varying components of Cultrual Proximity are considered, in whose respect FDI flows appear to differentiate both qualitatively and quantitatively. While qualitative heterogeneity (which refers to the differences across different modes of internationalization) is well acknowledged and extensively discussed in the existing literature, the possibility that the different drivers of FDI might not to be stable according to the size of the bilateral flow has remained substantially unexplored.

Keywords: *Mergers and Acquisitions, Gravity, Heterogeneity, Censored Quantile, Cultural Preferences.*

3.1 Introduction

The active promotion of cross-border investments (either inward or outward) is an often high-order priority in the development strategy of many countries worldwide. For this reason, understanding the determinants of Foreign Direct Investments (FDI), as well as the potentially differentiated impacts those factors have on the various types of investment is highly relevant under an economic policy perspective. Indeed, the entry mode of foreign firms crucially determines the favour with which a recipient economy receive them. While greenfield FDI are generally favourably welcome by the public opinion, the acquisition of a existing firms often raises perplexities in recipient countries.¹ This thing is made even more relevant considering that Mergers and Acquisitions (henceforth M&A) constitute the lion's share of overall Foreign Direct Investments. In 1997, M&A added up to roughly 60% of total investment flows worldwide (UNCTAD, 1998), topping more than two third of total FDI flows in the early 2000s. Indeed, M&A sustained FDI figures despite the substantial drop recorded by their greenfield counterpart, as a consequence of the global financial crisis (UNCTAD, 2017).

Also qualitatively speaking, M&A differ from greenfield FDI (analyzed in the first chapter of this thesis). Not only M&A appears to be “more reversible” than greenfields (Sula and Willett, 2009), but also usually reflect different investment strategies: for instance, they constitute an important source of technological acquisition.²

This chapter resumes the analysis of the first one, focusing on the role of Cultural Proximity (CP) on M&A. Despite a growing interest on the impact of common cultural traits on economic exchanges, the way CP and its different components ultimately affect bilateral exchanges in general, and bilateral FDI in particular is not yet entirely clear. Indeed, while most of the recent theoretical contributions acknowledge the time varying, directed, and potentially asymmetric effects of CP (intended as a social and cognitive construct) on bilateral economic exchanges, the existing empirical studies fail (with few notable exceptions) to consistently include those dimensions into the frame. As a result, there is a shortage of empirical evidence of the relative impact of the several dimensions of CP on economic exchanges and investment decisions. I extend the analysis conducted in Chapter 1 in at least two main directions.

On the one side, it investigates the role of CP on a different type of investments (Mergers and Acquisitions - M&A), which are likely to be affected differently by the directed and time varying components of CP (defined in terms of *Explicit Cultural Preferences* - EP) as captured by trade in cultural goods (Disdier et al., 2010).³ On the other side, this chapter acknowledges the

¹For instance, UNCTAD (2016) reports several examples of domestic economic policies aimed at defending strategic sectors from foreign acquisition. Even wealthier economies such as France, the US, and Germany to mention a few impose limitations under a national economic strategy perspective.

²This point explains why M&A are particularly relevant for developing countries, as well as why they are often perceived as a threat in developed countries. Partial confirmation of such differences can be found in some recent estimates by *Global Finance*. According to them, M&A from developing countries exceeded those coming from the richer economies in 2012 (197,003 millions US\$ versus 151,752 millions US\$ respectively). See <https://www.gfmag.com/global-data/economic-data/value-of-cross-border-maa-by-region-country>

³With explicit cultural preferences, I refer to the possibility of an economic agent to signal a positive attitude toward a potential partner, which might imply the possibility for a preferential economic treatment. Think for instance of bilateral trade and the related consumption of foreign goods: by consuming those goods which are

quantitative heterogeneity of bilateral investments data, and adopts an alternative analytical approach to take it into consideration within a structural gravity framework. Indeed, heterogeneity exists in at least two forms. Investment flows differentiate in terms of composition (qualitative heterogeneity): investments in different sectors are likely to respond differently to economic and bilateral relationship between two countries, as much as different types of investments have different impacts in a recipient economies. But bilateral FDI flows might also be quantitatively heterogeneous (in terms of size of the single investment as much as in terms of aggregate bilateral flows): in this respect, the same factors and conditions could affect differently bilateral corridors depending on their duration and size. Nonetheless, as asserted by Baltagi and Egger (2016), bilateral FDI channels are characterized by very different dimension and maturity: for this reason, it is likely that not all bilateral economic partnerships are dominated by the same mechanisms (or more precisely, they are likely to respond differently to the same mechanisms). While qualitative heterogeneity is well acknowledged by the existing literature (for instance Chan and Zheng, 2017, discuss the type-specific impact of migrants network on FDI), its quantitative counterpart remains largely unexplored, despite it could lead to severe data issues, and very few empirical studies paid attention to such issue. Starting from the theoretical and empirical contribution of Head and Ries (2008), which I conveniently adapt to deal with the potential nonlinearities brought in by heterogeneity, I apply a High Dimensional Fixed Effects (HDFE) Censored Quantile estimator (CQReg) (Powell, 1984, 1986; Chernozhukov and Hong, 2003; Canay, 2011; Figueiredo et al., 2014, 2015, among others) to investigate the effect EP on bilateral M&A. The advantage with respect to common mean value estimator (such as PPML, Negative Binomial, or Pooled OLS) lies in the fact that HDFE-CQReg makes it possible to explore the non-constancy of the explanatory variables along the distribution on the response one, allowing at the same time their identification and estimate in presence of both a large set of FE, and of a large fraction of null flows. Controlling for the overdispersion of bilateral M&A data, as much as for the large share of null flows and the multilateral resistance terms, EP appear to maintain a significant and quantitatively important effect on bilateral M&A. However the asymmetric patterns advocated by Shenkar (2001) and confirmed by the analysis in the first chapter for greenfield FDI is not statistically significant for M&A. The remainder of this chapter is structured as follows. Section 3.2 reviews the state of the art of both (a) the evolution of gravity analysis as applied to cross-border investment flows; and (b) the summary of the empirical evidence about CP as a determinant of bilateral economic exchanges. Section 3.3 presents the conceptual framework. The empirical methodology, the description of the data and the discussion of the results are reported in sections 3.4, 3.5, and 3.6 respectively. Finally, section 3.7 concludes.

3.2 Cultural Proximity and Bilateral Economic Exchanges

The economic interactions between countries are highly affected by distance (either physical or not), which is a crucial determinant of transaction costs. Until the advent of both the trans-

likely to be chosen from a basket of potential alternative and relatively homogeneous products, a certain country may be signaling a preference for a precise economic partner (everything else equal).

portation and the ICT revolutions, geography constituted the main obstacle to economic exchanges (Tinbergen, 1962), but it progressively lost importance in favor of other forms of remoteness/proximity. Such forms of “unnatural” frictions (Bergstrand and Egger, 2013) proved to affect economic exchanges with a magnitude comparable to that of geography. Cultural Proximity (CP) belongs to this broad category of intangible frictions. Nonetheless, unlike other forms of non-physical construct, the economic literature is still far from reaching an agreement of the dimensions that define it. Economists explored the impact of several measures of cultural similarity on economic exchanges. Linguistic similarity is one of the first and more explored proxy of CP. Speaking the same language is an important mean of cultural transmission and generally implies a certain degree of historical co-evolution between peoples.⁴ For instance, it reduces the cost for collecting information while facilitating at the same time the transmission of non-coded knowledge across countries. Melitz and Toubal (2014) thoroughly explored the relationship between linguistic similarity and trade, by constructing a set of measures of linguistic similarity between countries. Controlling for many potential confounding factors, they find a positive and statistically significant impact of language on both trade and migration flows. Interestingly, they detect a stronger role of common spoken idioms as opposed to official language. Similarly, Adsera and Pytlikova (2015) find that migration is increasing in the degree of linguistic proximity: after differentiating between official and common spoken language, they conclude the latter to be much more closely related to CP with respect to the former. To the best of my knowledge no existing study explicitly focuses on the effect of linguistic proximity on international investment flows, despite all models of bilateral FDI exchange include a measure of linguistic similarity (that usually maintains a positive and statistically significant coefficient. See Guiso et al., 2009; Aggarwal et al., 2012; Chakraborty et al., 2017, among others.)⁵. Religion constitute another milestone of a country’s national identity (and therefore of its culture): beliefs, social norms and even legislation are often affected (in some cases directly, think of the Sharia, in other cases indirectly) by it. Helble (2007) and Lewer and Van den Berg (2007) investigate the impacts of different religious beliefs on economic growth, and the role of religious mixing and similarity on bilateral trade across countries. They both detect a positive and statistically significant effect of religious proximity on economic exchanges.⁶ Lee and Park (2015) explore the effects of religious beliefs on trade in services, reaching the same conclusions. CP is also positively associated with the presence of ethnicity-based bonds. Among others, Felbermayr and Toubal (2010) include a measure of ethnic proximity to control for a potential “ethnicity bias”. Hofstede (1991, 2003) considers ethnic and genetic similarity (see for instance Cavalli-Sforza et al., 1988) while building his multidimensional index of CP, while Melitz and Toubal (2018) study the relationship

⁴Though the spread of colonial empires in the past loosened the relationship between language and historical co-evolution: thus, using national official languages often captures different mechanisms, unrelated to cultural proximity.

⁵Language is a mean through which information circulate, and it may increase reciprocal trust. However, it does not provide any tool to interpret the way people with different cultural background think or behave. For this reason, despite the high correlation between language and culture, the former should be better considered as a consequence of cultural proximity, not as a component of it.

⁶They also find that economic outcomes are positively associated to religious variety within a country, while the existence of a dominant religion, once controlled for any other factors, deters economic growth and exchanges with countries that do not share the same beliefs

between somatic distance and trust on trade. Still, genetic similarity does not directly affect CP (even if it might denote geographical proximity). Rather, it implies a certain degree of physical similarity, which in turn favors trust and appreciation between people that visually recognize each other as close. However, the mechanisms linking genetic similarity to cultural proximity are not clear cut: it is plausible that people from the same ethnic group share the same cultural background, but they are likely to derive from a broader “national identity”. Genetic distance is more likely to capture other trust-related mechanisms, different from those triggered by (properly defined) culture. Despite both Language, religion, and ethnicity capture different relevant aspects of the cultural identity of a country. In general, there is a substantial lack of clarity on the nature of CP itself, and neither language nor religion or ethnicity is able to fully account for culture and its impact on bilateral exchanges (in particular, bilateral greenfield FDI). To begin with, both the international business (Shenkar, 2001; Tung and Verbeke, 2010, among the others) and economic geography literature (Boschma et al., 2016) began to question the idea of reciprocity, which is still largely (implicitly) accepted in the majority of the empirical economic literature. Second, there is little awareness regarding the speed to which cultural proximity evolves. This in turn translates into the fact that most of the common empirical proxies of bilateral CP present a very reduced variability over time. Despite the fact that certain phenomena (think of the progressive shift in the religious and linguistic composition of a society in response to migratory inflows from abroad) are characterized by slowly adapting processes, these cannot be considered the only channels through which cultural affinity evolves over time. For instance, there may be events generating sudden shifts in perceived proximity that cannot be explained in terms of the cultural dimensions usually considered by international economists. Finally, the existing evidence is quite vague on the possibility for cultural affinity to be reversible in the long term, especially when some “culturally disruptive” event (such as a fashion) takes place. To put it differently, there is no assessed evidence on whether events leading to sudden changes in the perceived attractiveness of a country for a potential economic partner generate irreversible, self-reinforcing shifts in perceived CP, rather than a temporary deviation from its long term trends⁷. Thus, the usual definitions of CP are substantially limited. First, they appears to be more closely related to trust rather than to cultural perceptions. For instance, speaking the same language lowers the barriers to gather and exploit information, and at the same time it helps communicating and sharing non-tacit knowledge. Sharing the same (or similar) religious practices defines the system of beliefs and the domain of what is considered as socially acceptable: for this reason, sharing the same set of moral rules helps understand each other, and makes it easier for an economic actor to anticipate the behavior of a potential (economic) partner. Second, the proxies of CP discussed so far tend to be perfectly reciprocal, and only define whether two countries are objectively similar between each other. But, economic actors may respond to such similarity in different ways, making it more or less effective at determining economic exchanges. Recently, the issue of imperfect reciprocity has been tackled resorting to directed measures of

⁷Dixit (1989), Baldwin (1988), and Dixit and Pindyck (1994) among others extended the idea of hysteresis, formerly related to the analysis of unemployment in the labor market (Blanchard, 1986), to capture this kind of phenomena in international trade.

explicit affinity between countries⁸. Guiso et al. (2009) analyze the impact of declared reciprocal trust on economic exchanges within (a restricted number of) European countries. They show how trust is strongly related to all those dimensions of “objective” cultural similarity, and how it ultimately affects bilateral economic exchanges across countries (Spring and Grossmann, 2016, extended the same framework to human mobility).⁹ The idea of using declared reciprocal trust between countries as a proxy for the asymmetric component of CP constitutes the first attempt to overcome the reduction of CP to a measure of pure similarity. In fact, similarity represents just one of the building block of cultural proximity. Dealing with such dimension alone does not allow to consider neither the way a country “perceives” and evaluate such similarity, nor the existence of events affecting those perceptions.¹⁰ In the definition of CP that I adopt here (and in Chapter 1), similarity represents a component of a more complex construct that considers also the role of the culture-based affinity between countries. I define the latter component in terms of *explicit cultural preferences* (EP), which capture the way culture *is perceived* as opposed to how it *is*.¹¹ In this sense, such definition encompasses the idea of *psychic distance/proximity* (Dow and Ferencikova, 2010), which complements the idea of cultural similarity usually adopted in the related economic literature. As a matter of fact, psychic distance relates to the set of factors “preventing or disturbing the flow of information between firms and the market. Examples of such factors are differences in language, culture, political systems, level of education, level of industrial development[...]”. According to (Tung and Verbeke, 2010) there is no evidence for these factors to be time invariant, symmetric, nor undirected: there might be some occurrences, events, innovations, fashions, etc., which change (temporarily or permanently) the way a country is perceived by its counterparts, with no need for any change to occur in the way such country perceives the others. Take for instance the election of a particularly “meaningful” candidate to a political rally. Despite the absence of any remarkable changes under any other observable (and possibly unobservable) cultural dimension, the electoral turnover might trigger a major change in the way a country is perceived abroad. Capturing such directed and time varying phenomenon requires something more than the traditional bilateral and symmetric measures of CP. To the best of my knowledge, there have been only three other attempt to extend the definition of CP beyond the pure cultural similarity (apart from declared trust). Disdier et al. (2010) use trade in cultural goods to investigate patterns in bilateral trade across OECD countries. Adopting a different perspective, Felbermayr and Toubal (2010) use the average Eurovision Song Contest scores awarded in each country to the competitors from abroad: they found evidence of imperfect reciprocity in the cultural relationship, and investigated the effects brought by that asymmetry on bilateral trade. Finally, Hellmanzik and Schmitz (2017) developed the idea of *virtual proximity* as a proxy for cultural preferences. Building upon the database by Woo et al. (2011),

⁸See the first chapter of this thesis for a more detailed description of the relationship between similarity and affinity.

⁹Objective similarity refers to cultural traits that are observable, whose identification is (relatively) immediate: in this sense, the language spoken in a country, its dominant religion or the religious composition of its population represent objective dimensions.

¹⁰And hence, excludes the possibility for a government to adopt active cultural promotion policies with the aim to foster its economic exchanges.

¹¹A similar paradigm of EP/affinity is well established in the International Business (IB) literature, that widely accepts the idea of people and entrepreneurs taking decisions according to their personal scale of values.

who collected data on bilateral hyperlink connections for a sample of nearly 90 countries in two points in time, they test the relevance of the informative channel created by the Internet as a driver for cross border bilateral portfolio investments. All these studies, in line with Guiso et al. (2009), state that “proximate cultural tastes” boost bilateral trade and cross border portfolio investments beyond the role of traditional measures of CP. Finally, in chapter 1 I extend the analysis of Felbermayr and Toubal (2010) using trade in cultural goods as proxy for explicit cultural preferences (as proposed by Disdier et al., 2010). I find evidence that both direction and time variability play an important role in shaping the mechanisms through which CP affects bilateral Greenfield FDI flows. In what follows, similarly to what I do in Chapter 1, I refer to cultural proximity as the system of shared practices and norms able to reduce both the costs of communication and the effort required to source and interpret information about potential economic partners, beyond the tangible aspects of geographical distance and institutional similarities. In this sense, cultural proximity operates facilitating the flow of information and the formation of trust between countries, by providing a key to *understand and interpret* such information. This is translated operatively into a definition of CP as a complex construct, where both the aspects defining cultural similarity and those affecting the way such similarity is perceived by different economic actors coexist.¹²

3.3 The model

Gravity equations constitute a true workhorse of empirical economic analysis. Nonetheless, their application has long being subject to an important limitation, inherently the absence of a sound theory to refer to. This point has been highlighted by Kleinert and Toubal (2010), which demonstrated how very different theoretical foundations (and therefore, very different interpretations) can originate almost identical empirical gravity equations.¹³ To rule out any ambiguity, I borrow the theoretical foundation from the structural gravity equation of bilateral M&A proposed by Head and Ries (2008) (hereafter H&R2008). I expand the original model extending their definition of distance in order to explicitly focus on CP, to distinguish between cultural similarity and affinity. While the former refers to the existence of an objective shared cultural trait, the latter alludes to the possibility that a people appreciates a foreign economic partner beyond the existence of an observable cultural similarity between them. The empirical gravity equation is then re-adapted to explicitly deal with the potential heterogeneity (via quantile estimation) of bilateral

¹²See Chapter 1 for a detailed discussion on the mechanisms.

¹³This fact explains why just in 1984, Deardoff claimed that the lack of a compelling theoretical ground for the interpretation of the results was a sufficient reason not to use this class of models for predictive, as well as for analytical purposes. Indeed, the importance of a sound theoretical mechanism for the empirical analysis is straightforward. In a proximity-concentration trade off (Brainard, 1997), the role played by geographic and institutional factors is crucial in driving the decision to invest as an alternative to export directly. As geographic distance captures part of the transportation costs associated to bilateral trade, we could expect it to have a negative coefficient. However, the role of distance becomes less obvious when the decision is no longer whether to export or not, but whether to export or the set up/purchase a productive plant abroad. Everything else equal, a MNE planning to serve a larger portion of the global market may decide to set up a new productive plant abroad (export-platform hypothesis) in order to reduce the average transportation costs toward all the potential destination markets. In this case, we can expect distance to positively affect bilateral FDI (it is interesting that despite few notable exceptions, this has rarely been the case in empirical FDI gravity.)

aggregate M&A flows and the multilateral resistance terms (Anderson and Van Wincoop, 2003; Baier and Bergstrand, 2007; Baldwin and Taglioni, 2006) in the context of longitudinal data. To the best of my knowledge, this study represents the first application of a High Dimensional Fixed Effects (HDFFE) Censored Quantile Estimator (CQreg) (Powell, 1984; Chernozhukov and Hong, 2003; Canay, 2011) to the analysis of investment flows in general, and to the study of the cultural drivers of bilateral M&A in particular.

3.3.1 The Head and Ries model

The model hinges on the trade-off between the remunerativity of an asset and the geographic remoteness of with respect to the head quarter (HQ). The model steps from an inspection game framework, played between a multinational corporation (MNE)’s HQ and its subsidiary (SUB).

Both players simultaneously choose their behavior anticipating strategically their opponents potential decision. The HQ produces a fixed output $a \gg 0$ with no uncertainty. The SUB’s local manager has to choose between playing fair with the HQ ($e > 0$) and shirking: in the first case, it produces an output b , which sums up to the HQ total production. For this, the HQ compensates the SUB for her effort with w ($b > w > e$). Since the two players decide simultaneously, there is the possibility for the SUB to shirk, in order to get compensated without actually spending any effort. Knowing that the local manager may find convenient to shirk provided she can go away with it, the HQ decides whether to monitor (spending $c > 0$) or to trust the SUB’s manager. The payoff structure is such that no pure strategy could credibly represent an equilibrium: should the HQ always verify, the SUB would anticipate her behavior and would cooperate. Conversely, should the HQ always trusts, the SUB would have no incentive to apply any effort. Table 3.1 summarizes the payoff structure

Table 3.1: Inspection Game’s Payoff Structure

		HQ		
SUB	w, a-w	0, a-c	Shirk (x)	
	w-e; a+b-w	w-e, a+b-w-c	Work (1-x)	
	Trust (1-y)	Verify (y)		

Structure of the payoff in the original game. By construction, $a > b > w > e > c$: this setting does not allow any pure strategy to be pursued, so that no pure strategy represents a Nash equilibrium. The table replicates table 1 in H&R2008.

The optimal equilibrium consists of a mixed strategy Nash equilibrium, where the SUB shirks with probability x and the HQ verifies with probability y .¹⁴ The game clears solving the expected payoff functions of both player

$$E(v)^{hq} = a + b(1 - x) - cy - w(1 - xy) \tag{3.1}$$

¹⁴As a consequence, the only condition for the SUB not to compensated consists in a verification process that confirm a flawed behavior on her side, which takes place with probability $xy > 0$

$$E(v)^{sub} = w(1 - xy) - e(1 - x) \quad (3.2)$$

Both HQ and SUB maximize the expected payoff taking each other's strategy (and therefore, the probability of taking a given action) as fixed. Therefore, the FOC for the optimal probabilities x and y implies the maximization of (3.1) and (3.2) with respect to y and x respectively. Setting the derivatives for (3.1) and (3.2) equal to zero ($E(v)_{hq,y} = -c + wx = 0$ and $E(v)_{sub,x} = -wy + e = 0$) allows to find the equilibrium mixing probabilities for y and x to be equal to c/w and e/w respectively. Substituting $y = c/w$ and $x = e/w$ into (3.1) yields

$$E(v)_{hq} = a + b(1 - c/w) - w \quad (3.3)$$

The derivative of equation (3.3) with respect to w (the wage HQ is expected to SUB) allows to identify the optimal remuneration w

$$E[w] = \begin{cases} w = \sqrt{bc} & \text{if cooperates} \\ w = 0 & \text{if shirks} \end{cases}$$

, to be substituted back into equation (3.1) to find the expected HQ's payoff

$$E(v)_{sub} = a + b - 2\sqrt{bc} \quad (3.4)$$

At least two observations can be made on equation (3.4). First, the interest of a HQ for a SUB in a certain country is negatively correlated with the expected monitoring and verification cost c . Second, the value of the output b produced by the SUB enters the expected payoff function in two opposite directions. On the one side, it increases the gain a HQ expects from an acquisition. On the other side, it acts as a magnifying factor of the monitoring cost c . Assuming two similar HQ (1 and 2 respectively) producing respectively the same output $a_1 = a_2 = a$, it derives that the one facing the lower verification cost c will be also able to offer the higher bid. The cost parameter c (which determines the return of the investment and therefore the optimal allocation of investments in a country) can be assumed to be an increasing function of both geographic and cultural factors, that sum up in a distance vector D_{in} , included within c 's "remoteness function".

¹⁵ The remoteness function takes the form

$$c_{ni} = [r(D_{ni})/2]^2, \quad r' > 0 \quad (3.5)$$

Substituting equation (3.5) into (3.4) shows the existence of an ability/proximity trade-off, which

¹⁵Though the formulation is elastic enough to allow for alternative definitions of distance: indeed, it could be extended to include many other different measures of institutional similarity (Aleksynska and Havrylchyk, 2013), financial development (Desbordes and Wei, 2017), etc.

determines the profitability for a HQ in a given country i to invest in a SUB in any other given country n .

$$E(v)_{ni} = a + b - \sqrt{br}(D_{ni}) \quad (3.6)$$

Thus, more distant HQs have to compensate greater costs with a higher output a (which proxies for the ability of the HQ), in order to compete for more distant assets.

The HQ individual payoffs can be aggregated. Let us a global economy where MNE from all over the world bid to gain control over potential affiliates through stylized auctions. In equation (3.6) a MNE anticipates the expected return from a given investment depending on her personal ability/proximity trade off. The expected bilateral stock of M&A owned by all HQs from a given country i in a generic country n can be defined as

$$FDI_{ni} = \pi_{ni}K_n \quad (3.7)$$

where π_{ni} reflects the probability for a HQ from country i to take over a random SUB in a given country n , while K_n represents the total stock of existing assets in country n ¹⁶. π_{ni} is assumed to be distributed as a type-1 extreme value function (Gumbel), which allows to formulate π_{ni} as depending on the largest potential bid (m_i^{max}) that could be offered by a HQ from a given country i , without changing the functional form. For the same property, the probability of the highest bid for a random target SUB in a given country n to come from a specific country i , depends on the probability for the maximum bid m_i^{max} to exceed the maximum bid m_j^{max} , $\forall j \neq i$. π_{ni} can be rewritten¹⁷ as

$$\pi_{ni} = \frac{\exp[\mu_i/\sigma + \ln m_i - (\sqrt{b}\sigma)rD_{ni}]}{\sum_l \exp[\mu_l/\sigma + \ln m_l - (\sqrt{b}\sigma)rD_{ni}]} \quad (3.8)$$

Thus π is a function of the distribution of i 's HQ over the cumulative distribution of value added a , that also determines HQs' heterogeneity. Plugging equation (3.8) into equation (3.7) defines the *impeding effect of distance*, obtained by re-parametrizing the expected cost function $r(D_{ni}) \equiv \delta\sqrt{b}/\sigma$ with $\Theta \equiv \delta\sqrt{b}/\sigma$. In this way Θ represents a compound parameter capturing the role of distance as a function of the inspection costs (proxied by δ , increasing in remoteness) and the potential output value of the SUB.

The expected value of all investments for country i in the recipient economy n ($E[FDI_{ni}]$) depends on the share of n 's assets owned by all HQs from i . This defines the *bid competition* between all potential HQs worldwide to take over n 's assets. Let $s_{in} \equiv m_i/(\sum_l m_l)$ be the share of world "bidders" from country i , the bid competition for a given country n 's asset is represented

¹⁶The authors notice how expected bilateral M&A stocks may differ from the actual bilateral figures because of the discrete distribution of potential targets in a random country n . They refer to this process as *lumpiness*

¹⁷This is made possible by Gumbel's properties: in particular, the maximum m Gumbel draw maintain the same distribution, shifted by parameter $\mu = \sigma \ln(m)$. Thus, π_{ni} depends on m_i^{max}

by

$$B_n \equiv \exp^{[\mu/\sigma - D_{ni}\Theta]} s_i^m K_n B_n^{-1} \quad (3.9)$$

Equation (3.9) tells that the bid competition for country n 's assets grows in the bidders concentration nearby (proximity effect) and in the productivity of the competitors (ability effect): the consequence of a higher competition implies that a higher share of assets will likely accrue in the portfolio of foreign competitors of a given country i (lower π_{ni} for a given i). Thus, the expected bilateral FDI position $E[FDI_{ni}]$ can be defined as

$$E[FDI_{ni}] = \exp[\mu/\sigma - D_{ni}\Theta] s_i^m K_n B_n^{-1} \quad (3.10)$$

Equation (3.10) resembles a gravity equation, which originates from the initial inspection game, where expected bilateral FDI's are positively related to i and n 's size (s_i^m and K_n respectively), and decreasing in the bilateral distance vector $D_{in}\Theta$. The bid competition term B_n^{-1} , i.e. the multilateral resistance term, captures the difficulty on the side of i to take control of assets in n due to the large number of competitive rivals from other countries, and is therefore expected to reduce the bilateral FDI position of country i in n .

Equation (3.10) can be finally rewritten (as in Eaton and Kortum, 2001, 2002) as

$$E[FDI_{ni}] = \exp(\mu/\sigma + \ln(s_i^m) + \ln(K_n) - \ln(B_n) - D_{ni}^{-1}) \quad (3.11)$$

, to separate origin's and destination's specific terms, to be estimated via countries' specific fixed effects (FE) to better control for the multilateral resistance terms.

3.3.2 Extending the model

Including Cultural Preferences

To explicitly address the analysis of distance and to isolate the role of explicit cultural preferences (EP) I include all those characteristics that defines cultural similarity, which are likely to co-determine bilateral exchanges.¹⁸ To control for institutional proximity, I consider the degree of similarity of the legal environments in both countries, the existence of a free trade agreement

¹⁸H&R2008 assume inspection costs c to be increasing in distance, they are mostly concerned about the role of multilateral investments, so that their analysis over the role of distance was merely instrumental to include attrition in the model. They include geographic distance (symmetric), two indicators for common colonial ties (directed), and they control for common official language (COL, symmetric) as a measure of cultural proximity. However, both measures capture very similar aspects. On the one hand, they are highly correlated: in most cases, the language of the hegemonic power has been maintained in the form of official language by most of the newly independent colonies. On the other hand, despite Common Official Language (COL) does affect international flows, Common Spoken Language (CSL, computed as the percentage of a certain language speakers over total population) is much more relevant in cultural terms, as demonstrated by Melitz and Toubal (2014) and Adsera and Pytlikova (2015)

(FTA), and the existence of colonial ties. These measures control for the capability of an entrepreneur/CEO to orient herself into the bureaucratic environment of the recipient country. I also control for the physical remoteness, including a dummy for contiguity in addition to the usual geodesic distance.¹⁹ Finally, to control for both cultural similarity and affinity I include a variable for common religion and the two directions of bilateral cultural trade, which capture the role of reciprocal perceived affinity (in terms of explicit cultural preferences). Including the two terms of reciprocal affinity implies that I can test whether cultural proximity is actually symmetric or not. In chapter 1 I dealt with this issue focusing on greenfield FDI. However, there is no theoretical justification for the same mechanism to affect all types of bilateral investments. More specifically, M&A are likely to be less subject to asymmetric cultural determinants: the size of the investment is usually smaller, and the risk is shared among different actors (unless the acquisition leads to a 100% ownership). In addition, the fact that the potential affiliate is already operating (no matter how well it is performing) offers to a potential investor the possibility to evaluate strengths and weaknesses of the deal, allowing to make better informed investments. For this reason, I expect the two terms of perceived affinity to remain significant, but with no evidence of the same asymmetric patterns detected in Chapter 1.

Flows, Time, and Heterogeneity

With respect to the empirical equation, I extend H&R2008 model in three directions.

First , the original model was intended for stock data.²⁰ As pointed out by the same authors, “A model of flows requires accounting for divestitures of assets in a specification of the adjustment costs associated with convergence to desired FDI levels” (Head and Ries, 2008, p. 6). Nonetheless, I investigate whether explicit cultural preferences (EP) play (or do not play) a role in shaping M&A transactions and, in case they do, whether the asymmetric patterns detected for Greenfield FDI hold for M&A too. In this sense, the model allows to extend the gravity equation to flows without any loss in terms of theoretical interpretation of the coefficients.

Second , I introduce the temporal dimension, a fundamental aspect for EP. The way a country perceives the culture of its economic partners may be driven not only by the objective/observable cultural traits (such as language, religion, historical co-evolution, etc.), but also by temporary shocks that may either reduce or increase bilateral exchanges. Given the importance of the time variability in EP, I extend the gravity equation to take such temporal dimension into account.

Third, the original model does not offer any insight about the potential heterogeneity and the related role of EP with respect to (a) the different levels of the bilateral M&A; and (b) the relative

¹⁹Irrespectively of distance, contiguity may facilitate cultural transmission and assimilation. Its omission could therefore introduce an upward bias into the estimates for both the similarity and the affinity related terms.

²⁰H&R2008 collapse flow data in order to obtain an approximation of bilateral FDI stocks for one point in time. Indeed, equation (3.10) is a stock equation that makes no prediction about the annual flow level needed to reach the observed level of FDI stock.

importance of the EP components as opposed to other measures of distance, at different levels of bilateral M&A. Despite the high heterogeneity of bilateral M&A, very little effort has been spent so far to identify the implications of such high dispersion (Paniagua et al., 2015; Cuadros et al., 2016; Desbordes and Wei, 2017). I use Quantile Regression to explore how cultural proximity (and its time varying and potentially asymmetric component, EP) shapes bilateral M&A taking their quantitative heterogeneity into account.

3.4 Empirical Framework

I adopt a censored quantile estimator with high dimensional fixed effects. Before discussing the results, it is worth to discuss the appropriateness of such methodology and to review the main challenges in the empirical estimation of gravity models.

The effect of distance on bilateral M&A (and on economic exchange in general) is crucially affected by their heterogeneity. An investing MNE weights differently the economic conditions of a given destination when it has some experience operating in that country with respect to a situation in which it moves there for the first time. In aggregate bilateral terms, this is likely to apply as well. Economic, institutional, and cultural conditions are likely to affect differently large and well established investment corridors as opposed to small ones.

The fact that over-dispersion might hide the possibility that smaller channels are *qualitatively* different from larger ones in terms of the underlying ruling mechanisms has only been recently acknowledged in the gravity literature. Despite the rising concern for the heterogeneity of the determinants of bilateral economic exchanges, the related empirical evidence is still scant. With respect to mean value estimation, quantile regression presents some interesting advantages. Conversely from the log-linear transformation, quantile regression is not subject to the Jensen's Inequality. With respect to PPML, it also relaxes the strict assumption on the distribution of the error term, which raised several concerns in the critical assessment of Martínez-Zarzoso (2013) and Figueiredo et al. (2014). Finally, quantile decomposition generally returns robust estimates even when the outcome of interest is highly dispersed.²¹ Cairns and Ker (2013) and Baltagi and Egger (2016) estimate a structural gravity model by means of quantile regression, to assess whether observable trade costs are ultimately constant across different country pairs. As a matter of facts, the assumption of homogeneous effects is implicit in the empirical estimation of structural gravity models. Since the estimated coefficients differed not only quantitatively, but also statistically, they both concluded that the elasticity of trade to trade costs might actually be heterogeneous. Similarly, Paniagua et al. (2015) estimated a quantile gravity equation to take into account the different weight a firm attach to any potential attraction factor at country level, depending on the size of the expected investment. They detect a substantial instability in the coefficients along the distribution of the firms size. Cuadros et al. (2016, 2019) proposed two analogous application to study the role of migrants' network over bilateral FDI flows.

²¹Santos Silva and Tenreyro (2006) consider the error term to be distributed as $E[\eta_i|x] = 1$. Thus, it may return inconsistent estimates whenever this assumption fails. Figueiredo et al. (2014) proved quantile regression to perform better than PPML when this happens, while performing the same when the assumption over η holds.

Consistently with the assumption of heterogeneous returns, they found that only less mature (i.e. small) bilateral FDI channels are positively and significantly effected by migration (with the effect to be more pronounced in case of skilled migration). This is likely to occur because null and small FDI investment channels are more likely to be sensitive to information frictions, which tend to be alleviated by migrants network. All those applications, the use of mean value estimators would have prevented to detect the quantitative importance of heterogeneity. All the studies reviewed above applied a conventional quantile estimator, which remains highly sensible to the presence of null flows. While both PPML and GPML perform relatively well in presence of a large number of zeroes, provided the error term to be correctly specified (Santos Silva and Tenreyro, 2011), the same does not apply to quantile regression, which is based the log-linear transformation of the dependent variable. Since log-linearization drops null values, the lower quantiles would be actually composed by observations which are actually ranked higher in the real distribution of the response variable. All the studies mentioned so far ignored the presence of null flows. Such limitation can be solved via censoring techniques (Powell, 1984, 1986; Chernozhukov and Hong, 2003). Similarly to simple quantile regression, the outcome variable is log-linearized. However, null flows are considered as left-censored. Applying CQReg (in order to retain null flows), Figueiredo et al. (2014) estimate the impact of WTO membership on bilateral trade. They conclude that WTO effectively had a positive impact on worldwide trade. In particular, they conclude that WTO has been particularly effective at promoting small and new trading partnerships, i.e. between countries that did not use to trade much before entering the WTO, which benefited more than well established (and large) trading channels.

I apply a CQreg model with high dimensional fixed effect (HDFE, Canay, 2011; Figueiredo et al., 2014, 2015) to study the effect of the different components of CP on bilateral M&A.²²

To account for censoring, I transform the dependent variable as in equation (3.12), where C_{it} represents the minimum uncensored value of the dependent variable for country i in the t^{th} period.²³

$$\ln Y_{ijt}^{C_{it}} = \ln(\max(C_{it}; \ln Y_{ijt})) \quad (3.12)$$

Coherently with the discussion in Section 3.3.2, the empirical gravity equation takes the form

$$\begin{aligned} \log(\text{FDI})_{ijt}^{C_{it}, \tau} = & \beta_1^\tau \log(\text{CulImp})_{ijt} + \beta_2^\tau \log(\text{CulExp})_{ijt} + \varphi_1^\tau \log(\text{dist})_{ijt} + \varphi_2^\tau \text{contig}_{ijt} + \delta_1^\tau \text{comlang}_{ijt} \\ & + \delta_2^\tau \text{colony}_{ijt} + \delta_3^\tau \text{comreligion}_{ijt} + \gamma_1^\tau \text{comlegal}_{ijt} + \gamma_2^\tau \text{FTA}_{ijt} + \nu_{it}^\tau + \nu_{jt}^\tau + \varepsilon_{ijt} \end{aligned} \quad (3.13)$$

²²The way fixed effects should be included in quantile estimation still represents a major concern. According to Paniagua et al. (2015), there is no consensus on how to incorporate fixed effects into a longitudinal quantile decomposition. This issue becomes problematic in gravity analysis. In this paper, I do not address this issue, but I simply apply the methodology developed by Canay (2011).

²³This methodology essentially extends to the longitudinal case the strategy proposed by Eaton and Kortum (2001) and reported in Head and Mayer (2014)

The β_k^τ coefficients (with $k = 1, 2$) refer to the two EP terms (proxied by trade in cultural goods). φ_k^τ , δ_k^τ , and γ_k^τ coefficients refer to geographic, cultural, and institutional proximity measures respectively, while ν_{it} and ν_{jt} account for country-time fixed effects for origin and destination respectively. τ refers to the estimated quantile, which takes into account the heterogeneity of bilateral M&A flows. Finally, ε_{ijt} represents the error term.

3.4.1 Data Description

Bilateral Transactional M&A data Investment data come from the *Thomson Reuters M&A dataset*, provided by the Thomson Financial Securities Data Corporation²⁴. The dataset contains information at origin-destination-year level of aggregation and covers all the *transactions* occurred in the period between 1995 and 2011. Interestingly, the assumption of left-censoring implied by Equation (3.12) is justified by the fact that Thomson Reuters data only records M&A transactions which involve the transfer of at least the 50% of the equity assets. The omission of all those investments that do not reach such threshold can be dealt with as a form of censoring.²⁵ Thus, it is reasonable to think of a null flow as either signaling the absence of any flow from a given country toward another (a “real” zero), or the censoring of all those deals implying less than the 50% equity threshold.²⁶

Geographic, Institutional, and Cultural Variables Distance and contiguity (accounting for the facility of information to circulate as well as for the cost of monitoring), and the two dummies for common legal system and co-participation in a free trade agreement (capturing the ability of an investor to cope with the legal and bureaucratic environment at destination) come from CEPII’s *gravdata*. To capture the effect of the Explicit Cultural References (EP) I use trade in culture-intensive goods (as defined by UNCTAD, 2010) allows to capture the preferential component related the idea of perceived affinity, beyond the effect of actual similarity. Bilateral trade data come from the CEPII’s BACI dataset.²⁷ Finally, the remaining cultural variables to control for cultural similarity (the objective and observable dimensions of CP) come from CEPII’s *gravdata* dataset, and from Melitz and Toubal (2014), which is also available from CEPII.

Table 3.1 reports some selected descriptive statistics for the included variables.

²⁴Data have been accessed mid 2012

²⁵This point constitutes the main limitation of this kind of data, as for obtaining the control into the BoP it is not necessary to possess the 50% of the equity value. For instance, the threshold set by the World Bank to capital movements in order to be considered FDI amounts to a 10% ownership transfer of the total stock of assets.

²⁶Both the advantages and the limitations of using transactional FDI data are further discussed in Appendix 3.A.

²⁷In chapter 1 I discuss the pros and cons of using trade in cultural goods as a proxy of EP.

Table 3.1: Data description: Selected summary statistics

Variable	N	Mean	Std.Dev	25th pct	Median	75th pct	min	MAX
$Inv_{ni,t}$	72278	140.35	1734.65	0	0	0	0	2.07E+05
$\ln Inv_{ni,t}^C$	72278	1.84	1.88	0	1.79	2.83	0	12.24
$\ln CultIMP_{ni,t}$	72278	0.18	3.23	-2.2	0.37	2.56	-6.91	10.48
$\ln CultEXP_{ni,t}$	72278	0.4	3.06	-1.72	0.56	2.58	-6.91	10.48
$\ln dist_{ni}$	72278	8.53	0.93	7.87	8.88	9.19	4.11	9.89
$colony_{ni}$	72278	0.03	0.18	0	0	0	0	1
$lang_{ni}$	72278	0.15	0.35	0	0	0	0	1
$comrelig_{ni}$	72278	0.17	0.26	0.01	0.03	0.25	0	0.99
$contig_{ni}$	72278	0.04	0.19	0	0	0	0	1
$comleg_{ni}$	72278	0.27	0.44	0	0	1	0	1
$FTA_{ni,t}$	72278	0.24	0.43	0	0	0	0	1

The table contains a selection of descriptive statistics for the variables used in the empirical section computed on the estimation sample. The inclusion of both directions of cultural trade and of a large set of controls for CP might raise collinearity issues, threatening estimates' consistency. The issue is addressed in Appendix 3.A, which rules out the possibility that collinearity poses a relevant threat to the consistency of the results.

3.5 Results

Table 3.1 reports the main estimates of the econometric exercise. I have two main objectives. First, I want to identify the contribution of the various components of CP, paying particular attention to the relative contribution of the EP terms. The aim is to find evidence (if any) of the existence of the asymmetric relationship identified in the greenfield case (see Chapter 1). Second, I explore the extent of the potential heterogeneity of the CP-M&A relationship across different levels of bilateral M&A flows.

The main results are reported in Table 3.1 below. For reasons of space, only the estimates for the EP coefficients (namely Cultural Imports and Cultural Exports) are reported in the main text.²⁸ All flows are labelled and commented with respect to the investing country i . Thus, $\ln CultIMP_{ni,t}$ represents the log of the imports in country i of culturally-intensive goods from country n ; analogously, $\ln CultEXP_{ni,t}$ represents the log of the cultural exports from country i to country n . Their coefficients (reported in **bold** when statistically significant) capture the impact of country i and of country n perceived reciprocal affinity on the bilateral M&A flow respectively.

In Table 3.1 emerges that the impact of both EP terms can be identified beyond the role of the objective and symmetric components of CP, the geographical frictions, and the inclusion

²⁸The remaining coefficients are reported in Appendix 3.C.

Table 3.1: Baseline Results: Heterogeneous Impacts of Revealed Preferences on M&A.

Quantile	Dep. Var. : M&A $Inv_{ni,t}$		Tests	
	$\ln CultIMP_{ni,t}$	$\ln CultEXP_{ni,t}$	$\beta_{imp} = \beta_{exp} = 0$	$-\beta_{imp} + \beta_{exp} = 0$
ppml	0.14*** (0.04)	0.21*** (0.04)	$\chi_2^2 = 47.05$	$\chi_1^2 = 1$
$\tau = 10$	0.29* (0.14)	0.06 (0.04)	$\chi_2^2 = 3.01$	$\chi_1^2 = 2.54$
$\tau = 20$	0.38** (0.12)	0.18*** (0.04)	$\chi_2^2 = 12.81$	$\chi_1^2 = 3.61$
$\tau = 30$	0.42*** (0.12)	0.28*** (0.04)	$\chi_2^2 = 22.43$	$\chi_1^2 = 2.61$
$\tau = 40$	0.47*** (0.13)	0.31*** (0.05)	$\chi_2^2 = 21.71$	$\chi_1^2 = 2.94$
$\tau = 50$	0.42*** (0.11)	0.35*** (0.04)	$\chi_2^2 = 33.50$	$\chi_1^2 = 0.42$
$\tau = 60$	0.38** (0.12)	0.35*** (0.07)	$\chi_2^2 = 14.84$	$\chi_1^2 = 0.17$
$\tau = 70$	0.39*** (0.09)	0.39*** (0.08)	$\chi_2^2 = 12.82$	$\chi_1^2 = 0.00$
$\tau = 80$	0.31*** (0.07)	0.41*** (0.03)	$\chi_2^2 = 75.69$	$\chi_1^2 = 1.66$
$\tau = 90$	0.27*** (0.03)	0.41*** (0.06)	$\chi_2^2 = 33.57$	$\chi_1^2 = 8.41$

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard Errors in parentheses. Standard errors for PPML are clustered by country-pair.

The dependent variable $V_{ni,t}$ refers to the aggregate bilateral value of M&A from country i to country n . PPML includes null flows. According to equation (3.12), CQReg treats zeroes as censored. All equations include country-time fixed effect for both i and n . The sample size refers to the specification featuring both cultural imports and cultural exports.

The the fourth and fifth columns report the χ^2 statistics for the wald tests on the coefficients of cultural imports and exports. Test scores reporting opposite-than-expected results are flagged in red.

This table only reports the coefficients related to the EP coefficients. The remaining parameters included in each regression (see Equation (3.13)) are reported in Table c-1 in the appendices.

of the multilateral resistance terms.²⁹ The first row reports the estimates for the benchmark PPML. Quantitatively, EP terms show an asymmetric pattern, which seems to confirm the findings of Chapter 1. However, unlike Greenfields, the difference in magnitude between the two coefficients is not statistically significant. The small χ^2 statistic in the first row of column (4) suggests that, despite the reciprocal affinity of both trading partners remains relevant, the asymmetric patterns detected in the case of Greenfields do not concern M&A. The estimates of the censored quantile regression (for $\tau = [10, 20, 30, 40, 50, 60, 70, 80, 90]$) are listed in rows 2 to 10. In all specifications but $\tau = 10$, both directions of EP are statistically different from zero at the individual level, and never fail to be jointly significant (as shown by the Wald test for joint

²⁹All equations include country×time fixed effects for both the investing and the recipient economy, as suggested by Baldwin and Taglioni (2006) and Baier and Bergstrand (2007) in order to control for the multilateral resistance term (see section 3.3.2).

significance, reported in column (3)). Results can be summarized in two main points. First, the dominant EP channel is not consistent across quantiles. The relative importance of the investing country's relative preference for a given economic partner is quantitatively larger in small flows, but loses relevance in larger and well established channels. The coefficient for Cultural Imports remains larger than the Cultural Exports one until the 7th decile, but its magnitude changes non-monotonically. Its magnitude keeps an upward trend in the early quantiles, to reverse around the 4th decile. In addition, no direction of EP prevails as a driver of bilateral M&A when estimated at mean level. This suggests that, compared to the Greenfield FDI case (investigated in Chapter 1), different mechanisms might be at play. The quantile-wise analysis of the CP-M&A relationship offers a much less straightforward pattern for the relationship between CP and M&A. Second, PPML estimates appear to be systematically underestimated. Figure 3.5 graphically confirms the importance of addressing correctly the quantitative heterogeneity that characterized bilateral FDI exchanges data. The quantile plot (Panel A in) sheds further light on the heterogeneity of the dependent variable, and shows that the importance of EP may be larger than previously stated (Panel B reports the trends for both Cultural Imports and Cultural Exports separately).³⁰ All in all, the quantile decomposition confirms the presence of the "quantitative" heterogeneity discussed in Section 3.3.2. Both the EP coefficient change substantially across deciles, both in quantitative and in statistical terms.

Figure 3.5 is interesting as it shows graphically how the asymmetric patterns detected in Chapter 1 does not hold for M&A, at mean levels as well as across quantiles.³¹

The comparison of the estimates across quantiles (explored both numerically in table 3.1 and graphically in figure 3.5) helps at disentangling the heterogeneous relationship between EP and bilateral M&A flows. Such trends would have passed unnoticed under the usual mean-level estimators, for instance by mean of PPML.³² The exercise reported above also highlights a second major drawback related to PPML when applied to overdispersed data. By construction, PPML tends to overweight large observations, as noticed by Martínez-Zarzoso (2013) and Burger et al. (2009). This feature introduces a non-negligible bias when the dependent variable is highly heterogeneous, as it is the case for aggregate bilateral investment flows. Quantitatively, PPML estimates suggest a stronger role of the destination side preference mechanisms, even if the quantitative asymmetry remains statistically non significant.³³ Conversely, CQReg shows that

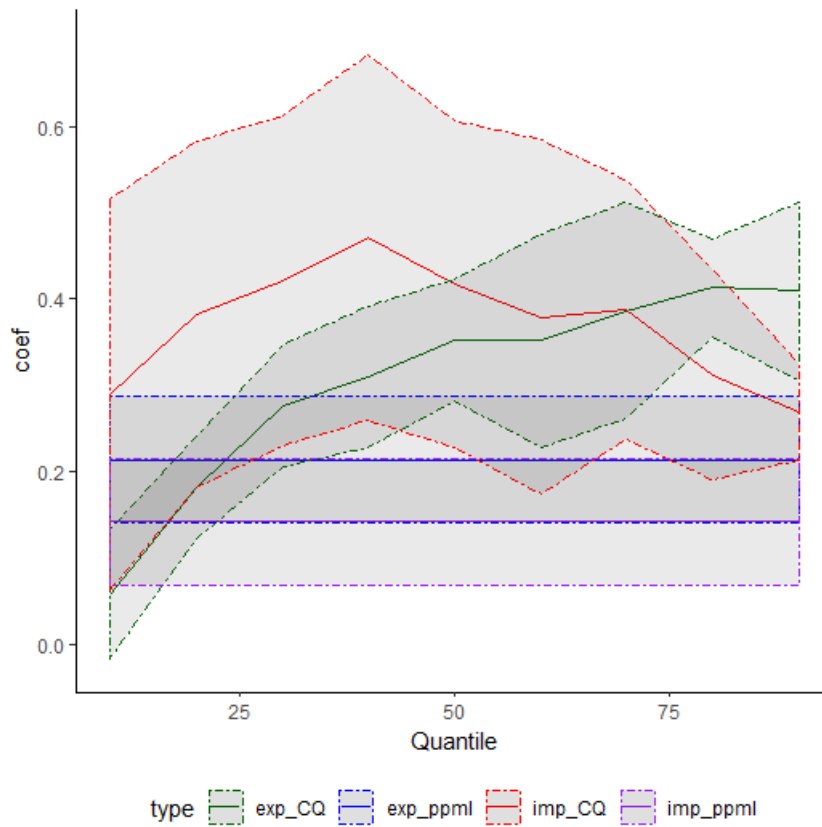
³⁰The issue of the underestimation of the coefficient is not limited to the coefficients of interest but applies to all coefficient included into the regression with the sole exception of the common religion dummy.

³¹Just in three cases, for $\tau = [20, 40, 90]$, the EP coefficients are statistically different from each other, with the relative importance of the EP terms reversing in the higher quantiles. As a matter of facts, the relative importance of the preference awarded by the investing country to a potential economic partner remains stronger than the other way around until the 70th quantile, a pattern which reverses in the higher quantiles. If robust, this finding would suggest that a completely different mechanism is at play for M&A compared to Greenfield FDI (see Chapter 1)

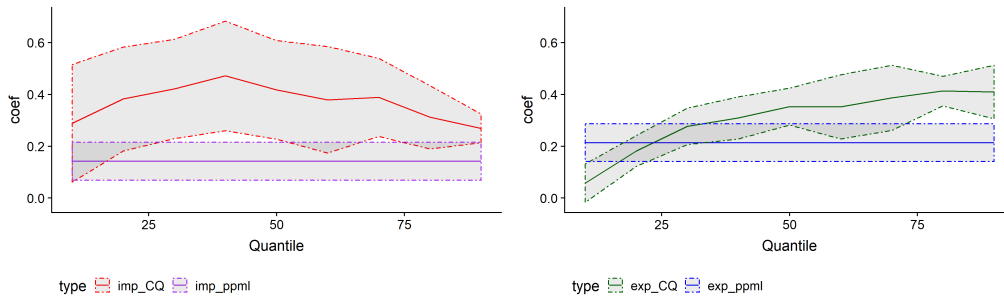
³²The same reasoning would apply to more sophisticated control functions and two stage procedures: despite providing a robust alternative when dealing with large shares of null flows, they would not allow to spot the heterogeneous patterns highlighted by quantile decomposition.

³³This result is in line with the expectations. While I might expect (as confirmed by the empirical evidence) a statistically significant asymmetric pattern for Greenfield FDI, I expected the issue of non-reciprocity in the CP-M&A relationship to remain quantitatively present but not statistically significant. See section 3.3.

Figure 3.1: Censored Quantile Plot of Explicated Cultural Preferences



(a) Shaded area indicate 95% C.I. ppml coefficients reported as solid lines.



(b) Below, imports and exports are reported separately.

this equilibrium is driven by large investments channels, with the origin-side preference channel dominating in smaller bilateral M&A flow.

3.6 Robustness Checks

The results reported in Section 3.5 offer an interesting insight on the relationship between EP and M&A flows, as the asymmetric CP pattern detected in Chapter 1 do not apply to bilateral

M&A.³⁴ Nonetheless, Table 3.1 leaves at least three open questions.

To begin with, the censoring of the dependent variable retains a large share of null bilateral flows, which are likely to severely affect the results (See for instance Head and Mayer, 2014; Figueiredo et al., 2014; Larch et al., 2017).³⁵ The scant related literature generally ignored the issue, limiting to strictly positive channels. In addition, the large number of cultural and institutional variables might generate collinearity in the vector D_{ni} .³⁶ Despite perfect multicollinearity can be excluded, the high correlation among covariates may still bias the coefficients. Table 3.2 estimates equation (3.13) without including the EP terms, to check the robustness of the remaining proxies of cultural similarity. Stable coefficients would support the relevance of the empirical strategy provided, as much as provide additional insight on the CP-M&A relationship. These tests are reported in Section 3.6.2. Finally, the relationship between the non-reciprocal and time varying components of CP and M&A might be non-linear within quantiles. On the one hand, it might be increasing in the geographical distance between two countries. Such “cultural bridge effect” hypothesis (Cuadros et al., 2016) implies that two countries with high cultural affinity may be characterized by larger than expected flows of bilateral investments (compared to other dyads located at the same distance from each other). In terms of heterogeneous impacts, such a bridge effect would be particularly relevant in case it is present (and stronger) at lower quantiles, as it would denote the ability of EP to reduce the detrimental impacts of geographical frictions, by narrowing the perceived distance between countries. On the other hand, the marginal effect of EP on investments could be non constant. The impact of EP might decrease the more two countries appreciate each other: in other words, the more two countries perceive themselves as close, the lower the marginal role of further displays of reciprocal appreciation. Non-linearity tests are reported in Appendix 3.D.

3.6.1 Sensitivity to the Exclusion of Null Flows

Table 3.1 restricts the sample to those channels for which the censored dependent variable is strictly positive, and estimates equation (3.13) by mean of traditional (uncensored) quantile regression.

As expected, accounting for the presence of null flows, and not just for the censoring, dramatically affects the estimates. To begin with, the numerical value of the coefficients decreases sharply when null censored flows are excluded. This is true for all coefficients across all quantiles (see Table c-2 for the remaining coefficients). This is no threat to the consistency of the baseline results, since a direct comparison between Table 3.1 and Table 3.1 would not be appropriate. As a matter of facts, the exclusion of null flows changes the interpretation of the results, not just their magnitude: from the study of the determinants of M&A, to the analysis of those

³⁴This might not be the case for all countries in the sample. Boschma et al. (2016) explore the role of different concepts of distance/proximity on Italian M&A, concluding that a symmetric effect of proximity cannot be taken for granted either.

³⁵As a matter of facts, this is the first study to estimate a structural quantile gravity model of FDI retaining all the null investment flows.

³⁶Multicollinearity tests are provided in the data Appendix 3.A

Table 3.1: Sensitivity to the Exclusion of Null Flows: impact on QReg and ppml estimates.

Quantile	Dep. Var. : M&A Inv _{ni,t}		Tests	
	ln CultIMP _{ni,t}	ln CultEXP _{ni,t}	$\beta_{imp} = \beta_{exp} = 0$	$-\beta_{imp} + \beta_{exp} = 0$
ppml	0.11*** 2.55	0.18*** 3.86	$\chi_2^2 = 28.55$	$\chi_1^2 = 0.70$
$\tau = 10$	-0.032 (-1.30)	-0.00525 (0.20)	$\chi_2^2 = 1.06$	$\chi_1^2 = 0.41$
$\tau = 20$	-0.0256 (-1.27)	0.0173 (0.72)	$\chi_2^2 = 0.85$	$\chi_1^2 = 1.39$
$\tau = 30$	0.02 (0.83)	0.03[†] (1.85)	$\chi_2^2 = 3.80$	$\chi_1^2 = 0.20$
$\tau = 40$	0.1 (0.78)	0.05** (2.6)	$\chi_2^2 = 5.08$	$\chi_1^2 = 1.45$
$\tau = 50$	0.01 (1.13)	0.08*** (4.8)	$\chi_2^2 = 17.55$	$\chi_1^2 = 6.62$
$\tau = 60$	0.06*** (3.49)	0.09** (3.1)	$\chi_2^2 = 48.40$	$\chi_1^2 = 0.55$
$\tau = 70$	0.08*** (5.09)	0.12*** (5.36)	$\chi_2^2 = 59.63$	$\chi_1^2 = 1.77$
$\tau = 80$	0.11*** (7.24)	0.16*** (8.07)	$\chi_2^2 = 129.2$	$\chi_1^2 = 2.3$
$\tau = 90$	0.13*** (5.63)	0.19*** (6.88)	$\chi_2^2 = 32.62$	$\chi_1^2 = 3.76$

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard Errors in parentheses. Standard errors for PPML are clustered by country-pair.

The dependent variable $V_{ni,t}$ refers to the aggregate bilateral value of M&A from country i to country n . Null (censored) flows are excluded. All equations include country-time fixed effect for both i and n . The sample size refers to the specification featuring both cultural imports and cultural exports.

The the fourth and fifth columns report the χ^2 statistics for the wald tests on the coefficients of cultural imports and exports. Test scores reporting opposite-than-expected results are flagged in red.

This table only reports the coefficients related to the EP coefficients. The remaining parameters included in each regression (see Equation (3.13)) are reported in Table c-2 in the appendices.

factors that facilitates them *provided an investment flow already exists*. Thus, Table 3.1 suggests that EP does not play any role in small size channels. The destination side EP term is only significant from the 30th percentile onward, while the investor side appreciation turns significant above the 6th decile. In both cases, the impact is quantitatively modest, if compared to the full sample specification. In addition, the investing side preference coefficient is systematically smaller than the destination side one at all levels. Similarly to Table 3.1, the coefficients for the two directions of EP remain statistically not different from each other in all but the 50th and 90th percentile (test reported in column (4)). Once more, results suggest that the two channels are simultaneously important in driving M&A flows, but that, differently from greenfield FDI, no direction dominates the other. Moreover, the fact that excluding null flows sharply reduces the size of the coefficients of interest signals that EP do play a role in driving the decision to invest in a country: however, once the decision to invest in a country is taken, EP become less

relevant. The perception of an investor for a country, as well as her awareness of how she is perceived abroad, are reasonably taken into higher consideration at the decisional stages. This fact also suggests that non-reciprocity in cultural preferences might still be relevant for bilateral M&A, but that the value data in use are not the most suitable to detect the impact of culture on FDI.³⁷

3.6.2 Stability/Sensibility of the Symmetric Measures of Cultural Proximity

Table 3.2: Sensitivity of the impacts of the symmetric components of D_{ni} on M&A flow

	Dep. Var. : M&A $Inv_{ni,t}$					
	(ppml)	(10 th)	(30 th)	(50 th)	(70 th)	(90 th)
$ln\ dist_{ni}$	-0.602*** (-8.89)	-0.448*** (-10.11)	-0.348*** (-21.06)	-0.279*** (-18.21)	-0.179*** (-13.33)	-0.0782** (-3.23)
$colony_{ni}$	0.499*** (4.38)	0.495*** (5.51)	0.272*** (6.42)	0.270** (3.26)	0.367*** (5.14)	0.922*** (5.55)
$lang_{ni}$	-0.136 (-1.07)	-0.11 (-0.95)	-0.0142 (-0.21)	0.309*** (3.84)	0.387*** (5.2)	0.441* (2.08)
$comrelig_{ni}$	2.280*** (7.58)	-0.214* (-2.10)	0.484*** (4.57)	0.226* (2.15)	0.202*** (5.47)	0.348** (2.72)
$contig_{ni}$	-0.0366 (-0.28)	0.886*** (8.93)	0.917*** (10.7)	0.958*** (9.62)	0.830*** (10.33)	0.967*** (5.75)
$comleg_{ni}$	0.220* (2.15)	-0.0135 (-0.18)	0.205*** (3.73)	0.217*** (3.56)	0.158*** (4.96)	0.241** (3.09)
$FTA_{ni,t}$	0.208 (1.56)	-0.00303 (-0.03)	0.372*** (5.28)	0.465*** (4.03)	0.427*** (3.88)	0.894** (2.76)
Obs	116873	116873	116873	116873	116873	116873
Censoring	-	✓	✓	✓	✓	✓

Notes: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard Errors in parentheses. Standard errors for PPML are clustered by country-pair.

The dependent variable $V_{ni,t}$ refers to the aggregate bilateral value of M&A from country i to country n . PPML includes null flows. According to equation (3.12), CQReg treats zeroes as censored. All equations include country-time fixed effect for both i and n . The sample size refers to the specification featuring both cultural imports and cultural exports.

Table 3.2 reports the results of Equation (3.13), but excluding the EP terms. The comparison with Table c-1 in the appendix (which reports the coefficients for the measures of cultural similar-

³⁷Indeed, as argued in Chapter 1, cultural processes are more likely to affect the extensive margin rather than the intensive one. This could be explained in terms of aggregate flows: the existence of an economic partnership (no matter the number/size of the investments, the quantity and the value of the goods traded, etc.) may offer itself an encouraging signal to prospective investors. This possibility seems to be supported by the trend of the cultural similarity terms too (See for instance the coefficient for common religion in Table c-2). Also in this case, excluding null flows reduces the coefficients in all specifications (including PPML). I consider this fact as an additional evidence of the relevance of the information embedded in null/censored flows, which embed the effects of frictions and attracting factors on the decision to make investments. In this respect, the results by Figueiredo et al. (2015); Cuadros et al. (2016) and Paniagua et al. (2015) might actually be biased downward.

ity in presence of the EP terms) offers interesting insights on the potential impact of collinearity between cultural affinity (captured in terms of explicit preferences) and cultural similarity. The inclusion/exclusion of bilateral cultural trade appears to rescale the usual symmetric dimensions of distance/proximity. Excluding the EP terms biases downward the remaining cultural dimensions. While distance results quantitatively less important (even remaining highly significant in all specifications), contiguity absorbs much of this change. This is not surprising, given that shared borders might imply the historical coevolution of two contiguous countries: cultural affinity, as captured by trade in cultural goods, might partially reflect this issue. Finally, the coefficients related to institutional similarity, such as colonial ties and language, shifts upward. This result may be partially driven by the high correlation across FTA, geographical distance, and cultural trade. Nonetheless, once perfect multicollinearity is excluded (see table a-2), the inclusion of a measure able to capture the time varying and directed dimension of CP results in a redefinition of the impacts of the traditional, symmetric measures of proximity/distance.

3.7 Conclusions

Foreign Direct Investments constitute a potentially important tool for promoting economic growth: they can favor capital accumulation, foster technological circulation, and create occupation. For all these reasons, to understand the factors that increase the attractiveness of a country is a fundamental issue for many governments. However, FDI are not all equal (in terms of economic effects as much as in terms of their determinants). The size of an investment, its degree of reversibility, and the degree of uncertainty are often very different across types of FDI.³⁸ Building upon the results of Chapter 1, I focus on the effects of the potentially asymmetric component of cultural proximity (proxied by trade in cultural goods) on bilateral M&A to investigate the non linear effects of Explicit Cultural Preferences on bilateral M&A. I fit a gravity equation by mean an innovative estimation technique (Censored Quantile Regression with High Dimensional Fixed Effects) to deal with the marked heterogeneity in the data, taking into account the panel structure, the large share of null flows, and the high dispersion of the distribution of bilateral M&A at world level. The quantile decomposition highlights a substantial instability of the coefficients. When focusing specifically on the directed components of cultural preferences, the results are coherent with the idea of CP as a complex construct, characterized by potential non-reciprocity and time variability. The fact that both directions of cultural preferences remain simultaneously significant indicates that there may be some event influencing bilateral economic exchanges beyond the role of traditional measures of CP. Contrarily to greenfield FDI, the analysis shows no evidence of a statistically significant asymmetric pattern between the two direction of EP. This fact suggests that unilateral cultural promotion (attraction) initiatives, though effective means to promote bilateral greenfield FDI, could have no impact on M&A.

Some aspects remain unexplored and would deserve further analyses. Under a methodological perspective, it may be interesting to investigate how CP and its intangible components (which

³⁸Even though the size of certain types of M&A increased substantially in the recent period, making reversibility extremely costly: in this respect, the differences between the two types of investment is fading away

capture the reciprocal affinity) affect cross-border financial flows across the different sectors. Unfortunately, the larger heterogeneity of investment data, joint to the even larger share of null flows at sectoral level, constitutes a severe impediment to this type of analysis. A possible solution would be to focus only on positive FDI flows, a commonly adopted strategy in the related literature. However, the high sensitivity to the inclusion of null flows discussed here suggests that this strategy may lead to biased results. This issue could be solved by applying a longitudinal multilevel quantile regression methodology (Giovannetti et al., 2018), able to isolate the impact of the factors that are specific to a particular bilateral channel from those operating at sectoral and global level. This point is left for future research.

Bibliography

- Adsera, Alícia and Mariola Pytlikova**, “The Role of Language in Shaping International Migration,” *The Economic Journal*, 2015, 125 (586), F49–F81.
- Aggarwal, Raj, Colm Kearney, and Brian Lucey**, “Gravity and culture in foreign portfolio investment,” *Journal of Banking and Finance*, 2012, 36 (2), 525 – 538.
- Aleksynska, Mariya and Olena Havrylchyk**, “FDI from the south: The role of institutional distance and natural resources,” *European Journal of Political Economy*, 2013, 29, 38–53.
- Anderson, James E. and Eric Van Wincoop**, “Gravity with gravitas: a solution to the border puzzle,” *American Economic Review*, 2003, 93 (1), 170–192.
- Baier, Scott L. and Jeffrey H. Bergstrand**, “Do free trade agreements actually increase members’ international trade?,” *Journal of International Economics*, 2007, 71 (1), 72–95.
- Baldwin, Richard and Daria Taglioni**, “Gravity for Dummies and Dummies for Gravity Equations,” Working Paper 12516, National Bureau of Economic Research September 2006.
- Baltagi, Badi H. and Peter Egger**, “Estimation of structural gravity quantile regression models,” *Empirical Economics*, Feb 2016, 50 (1), 5–15.
- Bergstrand, Jeffrey H. and Peter Egger**, “Gravity Equations and Economic Frictions in the World Economy,” in Daniel Bernhofen, Rod Falvey, David Greenaway, and Udo Kreickemeier, eds., *Palgrave Handbook of International Trade*, London: Palgrave Macmillan UK, 2013, pp. 532–570.
- Boschma, Ron, Emanuela Marrocu, and Raffaele Paci**, “Symmetric and asymmetric effects of proximities. The case of M&A deals in Italy,” *Journal of Economic Geography*, 2016, 16 (2), 505–535.
- Brainard, S. Lael**, “An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade,” *The American Economic Review*, 1997, 87 (4), 520–544.
- Burger, Martijn, Frank van Oort, and Gert-Jan Linders**, “On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation,” *Spatial Economic Analysis*, 2009, 4 (2), 167–190.

- Cairns, Alexander P. and Alan Ker**, “Unconditional Quantile Estimation: An Application to the Gravity Framework,” Working Paper 2013.
- Canay, Ivan A.**, “A simple approach to quantile regression for panel data,” *The Econometrics Journal*, 2011, 14 (3), 368–386.
- Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi, and Joanna Mountain**, “Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data,” *Proceedings of the National Academy of Sciences*, 1988, 85 (16), 6002–6006.
- Chakraborty, Suparna, Miao Wang, and M.C. Sunny Wong**, “FDI in Cultural Space: The Role of Bilateral Cultural Distance and Culture Clusters in FDI Expansion,” Technical Report 2017.
- Chan, Jackie and Huanhuan Zheng**, “Migrant Networks, Greenfield FDI, and Cross-border Mergers and Acquisitions,” Technical Report 2017.
- Chernozhukov, Victor and Han Hong**, “An MCMC approach to classical estimation,” *Journal of Econometrics*, 2003, 115 (2), 293 – 346.
- Cuadros, Ana, Joan Martín-Montaner, and Jordi Paniagua**, “Homeward bound FDI: Are migrants a bridge over troubled finance?,” *Economic Modelling*, 2016, 58, 454 – 465.
- , – , and – , “Migration and FDI: The role of job skills,” *International Review of Economics & Finance*, 2019, 59, 318–332.
- Desbordes, Rodolphe and Shang-Jin Wei**, “The effects of financial development on foreign direct investment,” *Journal of Development Economics*, 2017, 127, 153–168.
- Disdier, Anne-Célia, Silvio H. T. Tai, Lionel Fontagné, and Thierry Mayer**, “Bilateral trade of cultural goods,” *Review of World Economics*, 2010, 145 (4), 575–595.
- Dow, Douglas and Sonia Ferencikova**, “More than just national cultural distance: Testing new distance scales on FDI in Slovakia,” *International Business Review*, 2010, 19 (1), 46 – 58.
- Eaton, Jonathan and Samuel Kortum**, “Trade in Capital Goods,” *European Economic Review*, 2001, 45 (7), 1195–1235.
- and – , “Technology, geography, and trade,” *Econometrica*, 2002, 70 (5), 1741–1779.
- Felbermayr, Gabriel J. and Farid Toubal**, “Cultural proximity and trade,” *European Economic Review*, 2010, 54 (2), 279 – 293.
- Figueiredo, Erik, Luiz Renato Lima, and Georg Schaur**, “When elephants mate do ants get crushed? The WTO impact on trade dispersion,” Working Paper, University of Tennessee at Knoxville 2014.
- , – , and **Gianluca Orefice**, “Migration and Regional Trade Agreements: A (New) Gravity Estimation,” *Review of International Economics*, 2015, 24 (1), 99–125.

- Giovannetti, Giorgia, Marco Sanfilippo, and Margherita Velucchi**, “Diverse twins: analysing China’s impact on Italian and German exports using a multilevel quantile regressions approach,” *Applied Economics*, 2018, 50 (28), 3051–3065.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, “Cultural Biases in Economic Exchange?,” *The Quarterly Journal of Economics*, 2009, 124 (3), 1095–1131.
- Head, Keith and John Ries**, “FDI as an Outcome of the Market for Corporate Control: Theory and Evidence,” *Journal of International Economics*, 2008, 74 (1), 2–20.
- **and Thierry Mayer**, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.
- Helble, Matthias**, “Is God good for trade?,” *Kyklos*, 2007, 60 (3), 385–413.
- Hellmanzik, Christiane and Martin Schmitz**, “Taking gravity online: The role of virtual proximity in international finance,” *Journal of International Money and Finance*, 2017, 77, 164 – 179.
- Hofstede, Geert**, *Cultures and organizations. Intercultural cooperation and its importance for survival. Software of the mind*, Mc Graw-Hill, 1991.
- , *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*, Sage publications, 2003.
- Kleinert, Jörn and Farid Toubal**, “Gravity for FDI,” *Review of International Economics*, 2010, 18 (1), 1–13.
- Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin**, “The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects,” Working Paper 2017-7, LeBow College of Business, Drexel University 2017.
- Lee, Chong Wha and Soonchan Park**, “Does Religious Similarity Matter in International Trade in Services?,” *The World Economy*, 2015, 39 (3), 409–425.
- Lewer, Joshua J. and Hendrik Van den Berg**, “Estimating the institutional and network effects of religious cultures on international trade,” *Kyklos*, 2007, 60 (2), 255–277.
- Martínez-Zarzoso, Inmaculada**, “The log of gravity revisited,” *Applied Economics*, 2013, 45 (3), 311–327.
- Melitz, Jacques and Farid Toubal**, “Native language, spoken language, translation and trade,” *Journal of International Economics*, 2014, 93 (2), 351–363.
- **and –**, “Somatic Distance, Trust and Trade,” Working Papers 2018-11, CEPII September 2018.
- Paniagua, Jordi, Erik Figueiredo, and Juan Sapena**, “Quantile regression for the FDI gravity equation,” *Journal of Business Research*, 2015, 68 (7), 1512 – 1518. Special issue on The Spirit of Strategy.

- Powell, James L.**, “Least absolute deviations estimation for the censored regression model,” *Journal of Econometrics*, 1984, *25* (3), 303–325.
- , “Censored regression quantiles,” *Journal of Econometrics*, 1986, *32* (1), 143–155.
- Santos Silva, João M. C. and Silvana Tenreyro**, “The log of gravity,” *The Review of Economics and statistics*, 2006, *88* (4), 641–658.
- **and** – , “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator,” *Economics Letters*, 2011, *112* (2), 220–222.
- Shenkar, Oded**, “Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences,” *Journal of international business studies*, 2001, *32* (3), 519–535.
- Spring, Eva and Volker Grossmann**, “Does bilateral trust across countries really affect international trade and factor mobility?,” *Empirical Economics*, 2016, *50* (1), 103–136.
- Sula, Ozan and Thomas D. Willett**, “The reversibility of different types of capital flows to emerging markets,” *Emerging Markets Review*, 2009, *10* (4), 296–310.
- Tinbergen, Jan**, “An analysis of world trade flows,” *Shaping the world economy*, 1962, pp. 1–117.
- Tung, Rosalie L. and Alain Verbeke**, “Beyond Hofstede and GLOBE: Improving the quality of cross-cultural research,” *Journal of International Business Studies*, Oct 2010, *41* (8), 1259–1274.
- UNCTAD**, “World Investment Report,” Technical Report, United Nation Conference on Trade and Development 1998.
- , *Creative Economy Report. A Feasible Development Option*, UNDP, 2010.
- , “World Investment Report. Investor Nationality: Policy Challenges,” Technical Report, United Nation Conference on Trade and Development 2016.
- , *World Investment Report 2017 - Investment and the Digital Economy*, UNDP, 2017.
- Woo, Park Han, Barnett George A., and Chung Chung Joo**, “Structural changes in the 2003–2009 global hyperlink network,” *Global Networks*, 2011, *11* (4), 522–542.

Appendix

3.A Data Appendix

On the use of Transactional FDI data

The focus of this chapter is on bilateral M&A investment flows. Since official FDI statistics rarely distinguish between different types of FDI, I use *transactional* FDI data. This particular type of data presents a few technical advantages with respect to the traditional BoP figures. To begin with, transactional data are divided in two major investment entry mode, M&A and Greenfield: given the different characteristics of the two types of investments in terms of risk, reversibility, and even motivation, the possibility to divide the two types of flow offers the chance to understand not only the common drivers, but also the specific features that promote them. Second, transactional data keep track of the single investment by collecting information about all the parties involved in each transaction: this allows identifying not only the immediate owner of a subsidiary company, but also the ultimate owner. This is relevant for policy purposes too, as a MNE could follow particular country-specific profit-maximization strategies not only through the actions carried out at the HQ level, but also through its directly controlled partners. Think of those countries that grant special economic treatments to FDI: a locally located MNE may find convenient to invest back home via an immediate partner abroad in order to benefit from such policies, in a sort of round-tripping investment. Finally, transactional data only report the value of the capital invested in origin, and do not update the size of the investment according to further financings that may occur later in the life of a controlled firm. Since I am interested in understanding the role of explicit cultural preferences (EP) on bilateral investment decision, I need to know only the initial value of the capital investment³⁹: later re-investment may be due to factors that have nothing to do with cultural proximity. I am aware that using transactional FDI data also imply a trade-off. Two major limitations apply, relatively to the way data are collected, due to the fact that they are based on direct interviews to MNE's CEO or private sector compilation. First, the bilateral activities of certain countries tend to be systematically underreported: this issue is particularly severe when large MNE are publicly owned and respond to strategic national interest (think of China). Second, these data are usually dominated by large transactions, as they are generally easier to track: thus, large transaction may be overrepresented, to the detriment of the smaller, harder to track, ones. Yet, the advantages of using such form of data more then compensate such issues.

³⁹In the first chapter, I stressed the idea that both cultural proximity and the affinity component captured by EP may be more effective in driving the decision to invest rather than its amount. Unfortunately, because of data availability, I am not able to test the EP-M&A on the number of investments. This issue, though not affecting the quality of the results, make them only partially comparable to the findings discussed in the first chapter.

Further descriptive tables

Table a-1: Correlations across RHS variables

Variable	$\ln \text{CultEXP}_{ni,t}$	$\ln \text{CultIMP}_{ni,t}$	$\text{FTA}_{ni,t}$	$\ln \text{dist}_{ni}$	comrelig_{ni}	comleg_{ni}	colony_{ni}	lang_{ni}
$\ln \text{CultEXP}_{ni,t}$	1							
$\ln \text{CultIMP}_{ni,t}$	0.577	1						
$\text{FTA}_{ni,t}$	0.2543	0.2513	1					
$\ln \text{dist}_{ni}$	-0.2159	-0.2012	-0.5336	1				
comrelig_{ni}	0.0177	0.0117	0.1617	-0.1562	1			
comleg_{ni}	0.0332	0.0138	0.0565	-0.1512	0.2636	1		
colony_{ni}	0.143	0.12	0.0235	-0.036	0.0663	0.1839	1	
lang_{ni}	0.0387	0.028	0.072	-0.1059	0.2597	0.353	0.2059	1
contig_{ni}	0.1652	0.156	0.2091	-0.3812	0.1406	0.1658	0.0993	0.1365

Matrix of Correlation Coefficients for the RHS covariates as in equation (3.13). Coefficients are computed over the estimation sample.

Table a-2: Correlations across RHS variables

Variable	VIF	Sqrt. VIF	Tolerance	R^2
$\ln \text{CultEXP}_{ni,t}$	1.55	1.25	0.6445	0.3555
$\ln \text{CultIMP}_{ni,t}$	1.54	1.24	0.651	0.349
$\text{FTA}_{ni,t}$	1.47	1.21	0.679	0.321
$\ln \text{dist}_{ni}$	1.59	1.26	0.629	0.371
comrelig_{ni}	1.14	1.07	0.8739	0.1261
comleg_{ni}	1.23	1.11	0.816	0.184
colony_{ni}	1.09	1.04	0.9199	0.0801
lang_{ni}	1.22	1.1	0.8228	0.1772
contig_{ni}	1.21	1.1	0.8253	0.1747

Mean VIF = 1.34

Matrix of Collinearity Diagnostic. The table provide the estimated Variance Inflation Factors (VIF), its squared root, the tolerance coefficient, and the R^2 . Additional statistics (eigenvalues and condition index) are available upon request. The text in green flags those coefficients respecting the tolerance criterion: in general, tolerance above 0.6 is to be considered acceptable to avoid perfect collinearity. As for Table (a-1), collinearity is tested over the estimation sample.

Table a-1 shows the matrix of the correlations among the variables included in the empirical model. The major concern is about the high correlation coefficient between the two cultural preference channels which may lead to severe measurement errors in the estimates (since it would make difficult to identify the impact of the single variables in the LHS on the dependent variable).

Table a-2 reports several tests for collinearity. Considering collinearity not to be particularly worrying above a tolerance of 0.6, we can safely assume that we are able estimates the coefficients of interest as well as for any other potential controls included in a relatively precise way.

3.B Strength and Weaknesses of Mean-Value Gravity Estimators

Bilateral data (and particularly, investment flows data) are generally affected *zero-inflation* and *over-dispersion* (See for instance Santos Silva and Tenreyro, 2006, 2011; Head and Mayer, 2014; Yotov et al., 2016), two issues that might heavily bias the estimates when not properly dealt with. Zero-inflation becomes problematic when it concerns the dependent variable, as it reduces data variability: this makes it difficult for standard estimators to obtain consistent estimates. To get rid of the issue, the dependent variable is usually transformed in logarithmic terms. As it drops null observations, it also causes a substantial loss of information (not last the potential existence of some types of friction, either geographical, political, institutional or cultural, that prevents bilateral exchanges). Early studies addressed this issue by adding a unit value to the variable of interest and then taking the log, not to drop any null flow

Early studies solved this issue by taking the log of the dependent variables plus one.⁴⁰ Under a technical perspective, this strategy yields non-robust estimates and is neither robust to the Jensen’s inequality, nor to heteroscedasticity (Santos Silva and Tenreyro, 2006). In order to retain the simplicity of the log-linear form while retaining the null flows into consideration, some alternative procedure has been developed. Eaton and Tamura (1994) proposed a modified Tobit estimator, where the dependent variable Y is replaced with $\ln(Y + a)$, where a (which accounts for the amount of traded value that gets lost during the exchange) is itself a parameter to be estimated in the model. A different but very similar approach is adopted by Eaton and Kortum (2001). Differently, Helpman et al. (2008) developed a revised Heckman correction procedure⁴¹.

Santos Silva and Tenreyro (2006) depart from the idea of linearizing the empirical gravity equation. They proposed a family of modified poisson estimator with constant elasticity, in order to maintain the dependent variable in levels, retaining the multiplicative form of the gravity equation. In this way, the empirical specification remains closer to the theoretical model and circumvents the Jensen’s Inequality problem, with a minimal loss of information. The fact that the coefficients can still be interpreted in terms of elasticity, joint to the robustness to heteroscedasticity⁴² makes the Poisson Pseudo Maximum Likelihood (*ppml*) estimator the benchmark for gravity models in presence of null flows. The main critiques to *ppml* concern its potential inconsistency. On the one hand, Martínez-Zarzoso (2013) warns against the fact that despite *ppml* performs better than many of the proposed alternative estimators in presence of heteroscedasticity, it is outperformed when the data follows different generating processes. For instance, the standard errors of GPML (Gamma Pseudo Maximum Likelihood estimator) are lower when the dependent variable has small shares of null values. Analogously, Feasible Generalized Least Squares (FGLS) perform as well as PPML when the sample is small. She concludes that the choice of the estimator should depends on the structure of the data, rather than being considered in axiomatic terms. A similar conclusion is reached by Head and Mayer (2014) and D’Ambrosio and Montresor (2017). On the other hand, Burger et al. (2009) criticize the effective capability of PPML to account for data over-dispersion. They advocate in favor of a different class of estimators, such as the negative binomial and the zero-inflated poisson. Despite their concern

⁴⁰Indeed, there may be some country pairs that do not trade with/invest in each other for very specific reasons (even excluding the issue related to data misreporting). The intentional drop of those observations would underestimate the relevance of certain types of friction. Adding a unit value to the bilateral flow allows to get $\log(Y + 1) = 0$ when $Y = 0$. Nonetheless, as Head and Mayer (2014) points out, adding an arbitrary unit value to the dependent variable lacks of a “compelling structural interpretation”, and reduces the reliability of the resulting estimates.

⁴¹Nonetheless, according to Santos Silva and Tenreyro (2014), their assumptions on the structure of the error term threatens the consistency of their results

⁴²This point has been recently challenged by some recent works such as the one by Figueiredo et al. (2014, 2015). See section 3.3.2.

remains relevant, the adoption of a negative binomial or a zero-inflated estimator is sensitive to the scale of the dependent variable, so that they only remain optimal exclusively for count data⁴³.

3.C Extended Tables of Results

Table c-1: Impact of CP on M&A flow (aggregate value) at different quantiles - Including Null Flows

	(ppml)	(10 th)	(20 th)	(30 th)	(40 th)	(50 th)	(60 th)	(70 th)	(80 th)	(90 th)
	Dep. Var. : M&A Inv _{ni,t}									
ln CultIMP _{ni,t}	0.142** (3.19)	0.289* (2.1)	0.382** (3.14)	0.421*** (3.61)	0.472*** (3.67)	0.418*** (3.62)	0.379** (3.03)	0.389*** (4.24)	0.312*** (4.2)	0.269*** (8.01)
ln CultEXP _{ni,t}	0.214*** (4.83)	0.0578 (1.28)	0.182*** (5.05)	0.277*** (6.46)	0.310*** (6.26)	0.353*** (8.14)	0.352*** (4.68)	0.387*** (5.04)	0.413*** (11.98)	0.410*** (6.51)
ln dist _{ni}	-0.336*** (-4.57)	-0.742*** (-9.27)	-0.803*** (-10.01)	-0.818*** (-10.79)	-0.798*** (-9.61)	-0.688*** (-11.46)	-0.618*** (-7.57)	-0.612*** (-8.60)	-0.445*** (-13.06)	-0.295*** (-13.79)
colony _{ni}	0.276* (2.3)	0.546*** (8.21)	0.270*** (5.32)	0.304*** (4.45)	0.595*** (9.63)	0.425*** (5.11)	0.334*** (7.82)	0.597*** (5.88)	0.275** (2.78)	0.511*** (4.41)
lang _{ni}	-0.163 (-1.32)	-0.0514 (-0.65)	0.204** (3.06)	0.324* (2.57)	0.554** (2.87)	0.442*** (3.6)	0.317*** (4.05)	0.388* (1.99)	0.331* (2.41)	0.416*** (4.92)
comrelig _{ni}	2.038*** (6.76)	0.307*** (6.76)	0.230*** (6.92)	-0.0445 (-0.51)	0.510*** (11.81)	0.463*** (6.11)	0.321*** (11.14)	0.596*** (8.86)	0.616*** (4.67)	0.647*** (3.77)
contig _{ni}	-0.275 (-1.94)	0.290*** (3.84)	0.231 (1.82)	0.522*** (3.77)	0.295 (1.4)	0.540*** (4.22)	0.631*** (3.52)	0.207 (1.14)	0.828*** (10.62)	0.547*** (6.71)
comleg _{ni}	0.102 (0.96)	0.163 (1.49)	0.248** (2.76)	0.369** (3.13)	0.465*** (4.36)	0.184 (1.81)	0.244* (2.52)	0.107* (2)	0.154 (1.71)	0.335*** (4.09)
FTA _{ni,t}	0.0578 (0.4)	-0.183 (-1.33)	-0.0764 (-0.46)	-0.337 (-1.61)	-0.398 (-1.65)	-0.477** (-2.84)	-0.143 (-0.67)	-0.189 (-1.26)	-0.0574 (-0.37)	0.00247 (-0.08)
Imp×Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Exp×Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs	72278	72278	72278	72278	72278	72278	72278	44680	72278	72278
Censoring	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
F-test	1	2.54	3.61	2.61	2.94	0.42	0.17	0.00	1.66	8.41
Estimator	PPML	CQ	CQ	CQ	CQ	CQ	CQ	CQ	CQ	CQ

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors of PPML are clustered by country-pair.

This table reports all coefficients of the specification summarized in table 3.1. The dependent variable $V_{ni,t}$ represents the aggregate bilateral value of M&A from country i to country n . It includes the zero flows in the PPML. The same flows are treated as censored in the CQ regressions.

The sample size in this table is invariant to the number of covariates included and refers to the regression which features both imports and exports of cultural goods. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed.

F-test over revealed preference channels' equality. $H_0 : \beta(culimp_{ni,t}) = \beta(culexp_{ni,t})$

⁴³Santos Silva and Tenreyro (2011) respond to such critique proving that PPML is reliable even in presence of a very large fraction of null values.

Table c-2: Impact of CP on M&A flow (aggregate value) at different quantiles - Excluding Null Flows

	Dep. Var. : M&A Inv _{ni,t}									
	(ppml)	(10 th)	(20 th)	(30 th)	(40 th)	(50 th)	(60 th)	(70 th)	(80 th)	(90 th)
lnCultIMP _{ni,t}	0.115* (2.55)	-0.032 (-1.30)	-0.0256 (-1.27)	0.0155 (-0.83)	0.013 (-0.78)	0.0153 (-1.13)	0.0563*** (-3.49)	0.0786*** (-5.09)	0.109*** (-7.24)	0.132*** (-5.63)
lnCultEXP _{ni,t}	0.176*** (3.86)	-0.00525 (-0.20)	0.0173 (-0.72)	0.0288 (-1.85)	0.0481** (-2.6)	0.0814*** (-4.8)	0.0874** (-3.1)	0.139*** (-5.36)	0.157*** (-8.07)	0.195*** (-6.88)
ln dist _{ni}	-0.307*** (-4.14)	-0.503*** (-22.11)	-0.441*** (-30.77)	-0.416*** (-38.78)	-0.359*** (-32.23)	-0.326*** (-34.80)	-0.307*** (-25.78)	-0.297*** (-27.53)	-0.243*** (-15.45)	-0.157*** (-10.59)
colony _{ni}	0.289* (2.52)	0.384*** (-3.68)	0.419*** (-7.24)	0.406*** (-5.17)	0.394*** (-7.76)	0.444*** (-9.98)	0.380*** (-7.28)	0.376*** (-4.08)	0.464*** (-17.73)	0.506*** (-7.65)
lang _{ni}	-0.172 (-1.43)	0.278*** (-3.38)	0.244*** (-4.89)	0.201*** (-5.45)	0.207*** (-5.78)	0.132*** (-5.76)	0.224*** (-4.84)	0.200*** (-4.15)	0.16 (-1.41)	0.0775* (-1.99)
comrelig _{ni}	2.012*** (6.33)	0.178 (-1.79)	0.419*** (-5.68)	0.419*** (-6.14)	0.285*** (-4.72)	0.364*** (-8.62)	0.212*** (-5.81)	0.538*** (-10.94)	0.255*** (-4.77)	0.663*** (-5.81)
contig _{ni}	-0.235 (-1.68)	0.501*** (-8.27)	0.460*** (-7.43)	0.404*** (-11.45)	0.645*** (-18.57)	0.440*** (-15.58)	0.400*** (-18.82)	0.327*** (-9.03)	0.482*** (-13.83)	0.550*** (-11.57)
comleg _{ni}	0.13 (1.26)	0.263* (-2.35)	0.135* (-2.52)	0.244*** (-12.03)	0.227*** (-3.67)	0.246*** (-5.09)	0.285*** (-6.29)	0.271*** (-5.73)	0.175*** (-4.03)	0.177* (-2.51)
FTA _{ni,t}	0.128 (0.9)	-0.0611 (-1.03)	0.0157 (-0.43)	-0.0337 (-0.93)	-0.00417 (-0.14)	0.063 (-1.56)	0.0655 (-1.51)	-0.0525 (-1.01)	0.0741* (-1.99)	0.174 (-1.13)
Imp×Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Exp×Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs	48608	48608	48608	48608	48608	48608	48608	27177	48608	48608
Censoring	-	no	no	no	no	no	no	no	no	no
F-test	0.70	0.41	1.39	0.20	1.45	6.62	0.55	1.77	2.3	3.76
Estimator	PPML	CQ	CQ	CQ	CQ	CQ	CQ	CQ	CQ	CQ

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors of PPML are clustered by country-pair.

This table reports all coefficients of the specification summarized in table 3.1. The dependent variable $V_{ni,t}$ represents the aggregate bilateral value of M&A from country i to country n . It does not include null flows in the PPML.

The sample size in this table is invariant to the number of covariates included and refers to the regression which features both imports and exports of cultural goods. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed.

F-test over revealed preference channels' equality. $H_0 : \beta(culimp_{ni,t}) = \beta(culexp_{ni,t})$

3.D Potential non-linearities

The EP-M&A relationship may also be non-linear. There are two potential sources of non-linearity worth exploring. First, the marginal effect of EP on bilateral investment may be characterized by diminishing returns: this means that its marginal effect may be decreasing in the degree of proximity between two given countries. The second and more interesting source of non-linearity is related to the possibility that EP becomes more effective when two countries are geographically more distant. Geographic distance reduces the opportunities to establish contacts between people, a fact that may easily translate into lesser familiarity (and therefore, in lesser exchanges): thus, EP may generate a “bridge effect” that narrows the detrimental effect of distance. Table d-1 tests these two potential sources of non-linearity, reporting the estimates for ppml and for the 25th, 50th, and 75th quantiles respectively.

Panel *a* in table d-1 reports the coefficients for both directions of cultural trade and their quadratic term. The coefficients of the ppml are not stable; moreover, the coefficients of the higher quantiles suggest a self-reinforcing impact of both directions of cultural trade (positive sign in both the level and the squared terms). Thus, the presence of a non-linear quadratic relationship does not seem to be supported by the data. Panel *b* tests the potential bridge effect by displaying the results related to the interaction between the EP terms and geographic distance. The auspicated bridge effect seems to be only partially at work: despite both EP terms maintain the same trend as in figure 3.5, the virtuous relationship with distance turns significant only in the upper quantiles, beyond the median. Notwithstanding this partial evidence in favour, results are not very interesting. For a bridge effect to be effective and economically meaningful, it should be operational at lower quantiles, where geographic remoteness may be the crucial factor preventing the existence of a flow, not just a determinant of its numerical magnitude. Even though

Table d-1: Impact of CP on M&A flow (aggregate value) - Potential non-linearities

Dep. Var. : M&A Inv _{ni,t}				
	(ppml)	(25 th)	(50 th)	(75 th)
Panel (a) - Squared Cultural Trade				
ln CultIMP _{ni,t}	-0.348*** (-4.73)	0.0519 (1.14)	0.116*** (4.04)	0.179*** (4.79)
ln CultEXP _{ni,t}	0.299* (2.55)	0.0213 (1.16)	0.0643*** (4.35)	0.183*** (4.66)
ln CultIMP _{ni,t} ²	-0.175 (-1.42)	0.0428** (2.88)	0.0447*** (6.1)	0.0382*** (6.15)
ln CultEXP _{ni,t} ²	2.032*** (6.7)	0.0288* (2.45)	0.0441*** (13.32)	0.0398*** (7.57)
Panel (b) - Interaction term with distance				
ln CultIMP _{ni,t}	-0.407 (-1.22)	0.259*** (10.82)	0.200*** (26.08)	0.116*** (6.64)
ln CultEXP _{ni,t}	0.622 [†] (1.82)	0.262*** (10.88)	0.312*** (33.97)	0.341*** (25.1)
ln CultIMP × dist _{ni,t}	0.0638 (1.66)	0.00413 (0.22)	0.0168 (1.24)	0.0361*** (5.63)
ln CultEXP × dist _{ni,t}	-0.0484 (-1.22)	-0.0171 (-1.59)	-0.00164 (-0.16)	0.0183*** -4.1
Obs	ppml	72278	72278	72278
Censoring	-	✓	✓	✓
F-test				
Estimator	ppml	CQ	CQ	CQ

Notes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. z-statistics in parentheses. Standard errors of ppml are clustered by country-pair.

The dependent variable $V_{ni,t}$ represents the aggregate bilateral value of M&A from country i to country n . It includes the zero flows in the ppml. The same flows are treated as censored in the CQ regressions.

The sample size in this table is invariant to the number of covariates included and refers to the regression which features both imports and exports of cultural goods. The information which belong to groups with all zeros or missing values are automatically dropped by the estimator as FEs cannot be computed.

Panel (a) reports coefficient estimates for the log and the squared log coefficients of cultural trade. Panel (b) reports the coefficients for the two cultural trade variables and their interaction with distance (in log).

F-test over revealed preference channels' equality. $H_0 : \beta(culimp_{ni,t}) = \beta(culexp_{ni,t})$

the relationship between EP and geographic distance cannot be considered trivial (at least at higher quantiles), its role proved not to be crucial in the definition of new investment flows⁴⁴.

Bibliography

Burger, Martijn, Frank van Oort, and Gert-Jan Linders, "On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation," *Spatial Economic*

⁴⁴This mechanism could be better explored using data on the *number* of bilateral investment flows rather than their value. Unfortunately, the dataset does not have such information. Alternatively, a discrete choice model could be used.

Analysis, 2009, 4 (2), 167–190.

D’Ambrosio, Anna and Sandro Montresor, “Migration and trade flows: new evidence from Spanish regions,” Working Paper 2017.

Eaton, Jonathan and Akiko Tamura, “Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns,” *Journal of the Japanese and International Economies*, 1994, 8 (4), 478 – 510.

– **and Samuel Kortum**, “Trade in Capital Goods,” *European Economic Review*, 2001, 45 (7), 1195–1235.

Figueiredo, Erik, Luiz Renato Lima, and Georg Schaur, “When elephants mate do ants get crushed? The WTO impact on trade dispersion,” Working Paper, University of Tennessee at Knoxville 2014.

– , – , **and Gianluca Orefice**, “Migration and Regional Trade Agreements: A (New) Gravity Estimation,” *Review of International Economics*, 2015, 24 (1), 99–125.

Head, Keith and Thierry Mayer, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, eds., *Handbook of International Economics*, Vol. 4, Elsevier, 2014, chapter 3.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, “Estimating Trade Flows: Trading Partners and Trading Volumes*,” *The Quarterly Journal of Economics*, 2008, 123 (2), 441–487.

Martínez-Zarzoso, Inmaculada, “The log of gravity revisited,” *Applied Economics*, 2013, 45 (3), 311–327.

Santos Silva, João M. C. and Silvana Tenreyro, “The log of gravity,” *The Review of Economics and statistics*, 2006, 88 (4), 641–658.

– **and –**, “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator,” *Economics Letters*, 2011, 112 (2), 220–222.

Silva, João M. C. Santos and Silvana Tenreyro, “Trading Partners and Trading Volumes: Implementing the Helpman–Melitz–Rubinstein Model Empirically,” *Oxford Bulletin of Economics and Statistics*, 2014, 77 (1), 93–105.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch, “An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model,” *World Trade Organization, Geneva*, 2016.