Essays in Applied Economics

Oriol Anguera Torrell

TESI DOCTORAL UPF / ANY 2016

DIRECTOR DE LA TESI
Giacomo A.M. Ponzetto i Nicola Gennaioli
Departament d'Economia i Empresa

Als meus pares

Acknowledgements

I am extremely thankful to Giacomo A.M. Ponzetto and Nicola Gennaioli for their support, guidance and advice. This thesis would not have been possible without all their endless help and patience. Their support has been determinant to finish it.

I also greatly appreciate the different feedback and contributions that I received at different stages of my PhD from Alessandra Bonfiglioli, Christian Brownlees, Rosa Ferrer, Christian Fons-Rosen, Patricia Funk, Gino Gancia, Albrecht Glitz, Stephan Litschig, Joan de Martí and Jaume Ventura. I also owe many thanks to all participants at the International Lunch Seminar at CREI and at the Labor Breakfast Seminar at UPF for their comments and suggestions during these last years.

Special thanks to Marta Araque and Laura Agustí who have been extremely helpful during my PhD.

I would also like to thank my fellow PhD students at UPF for their support and friendship. I am very thankful to Tom, Ciccio, Chris, Kiz, Miguel Karlo, Dmitry, Mapis, Gene, Shangyu, Shengliang, Pau, Flo, Johannes, Burak, Jagdish, Nuno, Andrei, Stephanie, Christopher, Bruno, Miguel Ángel, Fernando, Vicky, Alain, María, Cristina and Ana. I also have a special memory of Michael. I also owe many thanks to my colleagues at CETT-UB.

I am also thankful to many other people who supported me during this process. I especially thank Peng, Davinia, Ángel, François, Sílvia, Maries, Miquel, Marta, Mireia and Sergi.

Last but not least, I am immensely grateful to Xavi who encouraged me in the last stage of this PhD and to my family who has been always there to motivate me.

Abstract

This thesis is composed of three self-contained chapters. In the first chapter, I propose a theoretical model that shows that trusting strangers is one of the channels through which institutions determine positive economic outcomes, such as entrepreneurship, but also negative ones, like corruption. The model predicts that the individual-level relationship between honesty and trust changes depending on institutional quality. I present empirical evidence in support of this prediction. In the second chapter, I study which share of ethnic enclaves in the U.S. can be attributed to spillover forces operating at the industry-location level among compatriots. I propose a micro-founded index that allows the estimation of the strength of such spillovers. I bring this index to the data and I find that immigrants who might have more difficulties in interacting with non-compatriots benefit more from this type of spillovers. In the third chapter, I empirically document that the main patterns on firms' globalization decisions found in the international trade literature on multinational firms extend to the Spanish hotel industry.

Resum

Aquesta tesi està formada per tres capítols independents. En el primer, proposo un model teòric que mostra que la confiança en desconeguts és un dels canals a través dels quals les institucions determinen resultats econòmics positius com l'emprenedoria, però també d'altres negatius com la corrupció. El model prediu que la relació a nivell individual entre honestedat i confiança varia en funció de la qualitat institucional. Presento evidència empírica que dóna suport a aquesta predicció. En el segon capítol, estudio fins a quin punt l'existència d'*spillovers* que operen a nivell d'indústria-localització entre compatriotes determina els enclavaments ètnics als EUA. Amb aquest objectiu, proposo un model que permet estimar la intensitat d'aquests *spillovers*. Estimo aquest índex utilitzant dades del cens del EUA i concloc que els immigrants que poden tenir més dificultats en interaccionar amb no-compatriotes es beneficien més d'aquest tipus d'*spillovers*. En el tercer capítol, documento empíricament que les principals relacions sobre les decisions d'internacionalització que s'han trobat a la literatura de comerç internacional de les empreses multinacionals també s'estenen a la indústria hotelera espanyola.

Preface

This thesis is composed of three self-contained chapters on three different areas of research: (1) trust and corruption, (2) ethnic enclaves in the U.S. and (3) multinational firms' globalization decision.

In the first chapter, I propose a theoretical model where I show that trust towards strangers is one of the channels through which institutions determine economic outcomes, in particular, entrepreneurship and corruption. Additionally I show that the role of trust towards strangers has been overlooked in the literature since high levels of trust do not always enhance desirable economic outcomes. On the one hand, trust helps individuals to take part in economic exchanges aligned with social welfare. On the other, it also facilitates individuals to cooperate for the achievement of corrupt deals. Under this more general view of trust than the unambiguously positive one which dominates in the literature, the model does not go against the well-known negative cross-country correlation between trust towards strangers and corruption. Nevertheless, the model generates a non-trivial new prediction at the individual level. In particular, the individual-level relationship between honesty and trust towards strangers changes depending on the institutional quality of a country. In countries with good institutional quality, more honest individuals are also the more trusting ones. However, in countries with poor institutional quality, more dishonest individuals are the more trusting ones. Using data from individuals of 64 countries from the World Value Survey, I present empirical evidence in support of this prediction.

In the second chapter, I note that immigrants coming from the same country of origin often cluster together in the same region, city or neighborhood in their destination. Such agglomeration suggests positive spillovers, i.e., direct benefits of being close to fellow immigrants. Yet, identifying these spillovers is challenging because they are perfectly confounded by location-specific comparative advantages that could attract immigrants coming from the same country of origin such as proximity to their home country. However, I present suggesting evidence that immigrants also cluster with their compatriots at the industry-location level. This pattern is indicative of spillovers realized at the industry-location level among compatriots. Moreover, it is unlikely that there are industry-location advantages that are specific for the immigrants coming from the same country of origin which would perfectly confound with the described spillovers.

This opens an avenue to credibly estimate a specific type of spillovers while taking into account other agglomeration forces that might bring about ethnic enclaves at the location, as well as at the industry level. Accordingly, I propose a micro-founded index that allows the estimation of such spillovers. I bring this index to the data in the U.S. and I find that it is higher for countries of origin that are more culturally distant from the U.S. and have lower levels of trust towards individuals of other nationalities. These correlations suggest that immigrants who might have more difficulties in interacting with non-compatriots might benefit more from this type of spillovers.

Finally, the third chapter of my thesis is a descriptive project on the firms' globalization decision in a services industry. The international trade literature on multinational firms has shown that firms operating across borders are more productive than those that remain domestic. Additionally, firms engaging in foreign direct investment (FDI) are more productive than those who contract with independent producers or distributors. Much of this evidence has been shown for the manufacturing industry. In this chapter, I empirically analyze whether these same patterns extend to one specific service industry: the Spanish hotel industry. Specifically, I document how productivity correlates with the decision to operate across borders, as well as with the decision to engage in FDI versus contract with third parties. Moreover, I also study their extensive and intensive margins of globalization. The empirical results suggest that the same patterns found for the manufacturing industry extend to the Spanish hotel industry.

Contents

Chapter 1

ENTREPRENEURSHIP, TRUST AND CORRUPTION

1.1 Introduction

Trust towards strangers is associated with a number of socially desirable economic outcomes.[1] At the aggregate level, it is positively correlated with income, growth, financial development, a good performance of larger firms, the quality of institutions, and low corruption (Putnam, 1993; La Porta et al., 1997; Knack and Keefer, 1997; Zak and Knack, 2001; Guiso et al., 2004, 2010; Tabellini, 2008, 2010; Algan and Cahuc, 2010). At the individual level, it is correlated with income, education, and entrepreneurship (Alesina and La Ferrara, 2002; Guiso et al., 2006; Batsaikhan, 2013).

One possible interpretation of these links is that there is a causal effect of trust towards strangers on economic outcomes (Algan and Cahuc, 2010; Tabellini, 2010). Trust might assist individuals in achieving socially desirable economic exchanges with strangers in the presence of asymmetric information. At the same time, there is also the view that trust is shaped by the current environment (Alesina and La Ferrara, 2002). For example, individuals may trust strangers if current institutions successfully enforce the law. In this chapter I show that trust towards strangers may indeed be a channel through which institutions determine economic outcomes. However, this chapter highlights an

[1] In this chapter, I only consider the trust held towards strangers, and I avoid the term "generalized trust". I do this in order to be clear that I only account for the trust held towards not personally known individuals.

ambivalent role for trust that has often been overlooked in the existing literature: trusting strangers might not only facilitate cooperation for the pursuit of socially desirable economic exchanges, but it might also promote cooperation for the achievement of illegal transactions. In this chapter, I study this "dark side" of trusting strangers both theoretically and empirically.

I propose a theoretical model where institutional quality determines the individual levels of trust towards strangers, and those levels of trust affect the following economic outcomes: entrepreneurship and corruption. In the model, trusting strangers is used for cooperation in economic exchanges in which individuals do not completely know if the other part is going to provide what they expect. Whereas trust always assists individuals to bring about transactions that are privately beneficial for them, institutional quality determines whether these exchanges have desirable or undesirable consequences for the rest of society. Individuals willing to engage in lawful transactions trust strangers only when institutions successfully enforce the law. However, individuals who are ready to commit illegal exchanges trust strangers to reciprocate their deals only when institutions cannot enforce the law. Under this more complex view of trust than the unambiguously positive one which dominates in the literature, the model does not contradict the well-known negative correlation between trust towards strangers and corruption across countries. Nevertheless, the model generates a non-trivial new prediction at the individual level. In particular, the individual-level relationship between honesty and trust towards strangers changes depending on the institutional quality of a country. In countries with good enough institutional quality, more honest individuals are also the more trusting ones. However, in countries with poor enough institutional quality, more dishonest individuals are the more trusting ones.

There are many dimensions of trust towards strangers that one can think of. However, I focus on the one that concerns to economic exchanges in which we do not completely know if the other part is going to provide what we expect. For example, in the legal market of second-hand cars, a buyer trusts a seller when the buyer beliefs that the seller is not providing lemons. On the other hand, in the illegal market of cocaine, a buyer trusts a seller when the buyer beliefs that the seller is supplying actual cocaine and not chalk. Hence, trusting strangers might bring about economic exchanges which are socially desirable but also some others which are not. Given the structure of these situations, I interpret trust towards strangers as the individual's belief that a stranger will

be of the type that the individual needs. Certainly, this is not the only interpretation of trust that one can think of, however, it is a natural one in this specific context. Thus, in order to study these two aspects of trusting strangers, I need to introduce a model with two markets, a legal and an illegal one, in a context of asymmetric information.

Accordingly, I propose a model of asymmetric information with a final goods market and a market for bribes. The population is divided into desirable producers, undesirable producers, bureaucrats and consumers. Desirable and undesirable producers need to decide whether to engage in a home activity where they produce for personal consumption or to become entrepreneurs and sell goods to consumers. However, undesirable producers make useless goods when becoming entrepreneurs, and there is a problem of adverse selection since consumers cannot observe if an entrepreneur is a desirable or an undesirable producer. In order to prevent consumers from buying useless goods, there is a regulatory agency composed of bureaucrats. They can distinguish between the two types of producers and are supposed to issue required licenses for becoming an entrepreneur, the purpose of which is to screen out undesirable producers. However, bureaucrats endogenously choose whether to be honest and act in the public interest, or whether to be corruptible and issue licenses to undesirable producers in exchange for a bribe. For simplicity, I assume that desirable producers can always obtain a license without needing to offer a bribe, and that only undesirable producers have to bribe to obtain the required license. I also assume that bureaucrats obtain an institutional payoff for issuing licenses to desirable producers or for denying the entrance of undesirable producers in the entrepreneurial activity. Moreover, there is also asymmetric information between producers and bureaucrats, since the former do not observe if the latter have endogenously chosen to be corruptible or not. The effects of bureaucrats allowing the entry of undesirable producers into the entrepreneurial activity are a reduction in the average quality of goods sold in the market and, as a result, a decline in the equilibrium price. Consequently, desirable producers and consumers are not keen on corruption, i.e., they do not tolerate corruption; whereas undesirable producers always tolerate it. In the model, institutional quality and the proportion of desirable producers are exogenous, and constitute the current environment; while trust towards strangers, the decision of becoming an entrepreneur, and corruption are all endogenous.

As discussed above, in the context of economic exchanges with asymmetric information, a natural interpretation of trust towards strangers is the individual's belief that

a stranger will be of the type that the individual needs. Hence, in the model, trusting strangers is the Bayesian belief that a stranger is of a given type.[2] Correspondingly, consumers' trust towards strangers is the belief that a stranger (an entrepreneur) is a desirable producer and sells high-quality goods. Likewise, undesirable producers' trust towards strangers is the belief that a stranger (a bureaucrat) is corruptible and accepts bribes allowing them to obtain a license. Finally, desirable producers' trust towards strangers is the belief that a stranger (a bureaucrat) is honest and does not allow the entrance of undesirable producers into the entrepreneurial activity through corruption, since this lowers the equilibrium price and therefore the profits of desirable producers. Hence, the trust of consumers and desirable producers brings about cooperation for the achievement of legal economic exchanges. On the contrary, the trust of undesirable producers helps them to cooperate with bureaucrats to circumvent a regulation.

The model has a unique equilibrium that depends on institutional quality and the share of undesirable producers. When institutional quality is good enough, bureaucrats do not accept bribes. Consequently, desirable producers completely trust strangers and become entrepreneurs. Undesirable producers, in their turn, distrust strangers and engage in home production. As a result, consumers fully trust strangers. Altogether, the aggregate level of trust towards strangers is high and corruption is nonexistent. Conversely, when institutional quality is poor enough, bureaucrats are ready to accept bribes. Undesirable producers trust strangers, and as a consequence, some of them become entrepreneurs through corruption. Desirable producers, in turn, distrust strangers and some of them no longer engage in the entrepreneurial activity, since the entrance of undesirable producers in the entrepreneurial activity crowds them out. In consequence, consumers trust strangers less than in the first case. As a whole, if institutional quality is poor enough, the aggregate level of trust is low and corruption is high. Thus, if countries have different levels of institutional quality, the model reproduces the negative cross-country correlation between trust towards strangers and corruption. Not surprisingly, good institutions generate high aggregate levels of trust towards strangers and low corruption.

[2]The literature on trust has taken different approaches to incorporate trust in a theoretical setting. For instance, Zak and Knack (2001) introduce trust as the time not spent in verifying others, Gennaioli et al. (2012) think of trust as a factor that decreases the losses produced by taking risk. And, similar to this chapter, trust has also been modeled as the belief that someone is of a given type or that someone will do a particular action (Butler et al., 2016; Rohner et al., 2013).

Nonetheless, at the individual level, the model generates a not obvious new prediction. Individuals who tolerate corruption (undesirable producers) have the highest levels of trust when institutions are bad. Conversely, individuals who do not tolerate corruption (desirable producers and consumers) hold the highest levels of trust when institutions are good. Therefore, the model predicts that individuals who tolerate corruption more are the more trusting individuals in countries with poor institutions, and the less trusting in countries with good institutions. Using data from individuals of 64 countries from the World Value Survey, I present empirical evidence in support of this prediction. Moreover, I provide evidence that this association is particularly important for white-collar workers, who are the ones interacting more with bureaucrats and who are more directly affected by corruption in order to pursue their businesses. For instance, managers, employers or lawyers might need to ask for a license or to participate in processes intended to gain a procurement contract from the government, and all these procedures are prone to corruption. Using the same data set, I show that in countries with poor institutions, white-collar workers who tolerate corruption more have the higher levels of trust towards strangers on average. And, in countries with good institutions, white-collar workers who tolerate corruption more are the less trusting ones on average. Accordingly, this pattern does not arise for blue-collar workers.

Challenging the established view, this prediction highlights that trust towards strangers is a doubled-edged sword. For instance, in the specific context of entrepreneurship, trusting strangers helps entrepreneurs to follow their honest activities since many deals are done by just shaking hands in the business environment. On the other, entrepreneurs willing to bribe an unknown bureaucrat need to trust the bureaucrat to not report them to the police and accept the bribe, and more importantly, to act as expected since corrupt deals cannot be enforced by any type of legal contract by their nature. Thus, trusting strangers also assists entrepreneurs to accomplish corrupt transactions. Summing up, I show that trusting strangers can promote cooperation for both socially desirable and harmful activities. Consequently, this chapter is also connected to the growing literature which identifies the dark side of social capital (Lampe and Johansen, 2003; Satyanath et al., 2013).[3]

[3]Lampe and Johansen (2003) point out that trusting strangers allows mutual support and cooperation for the pursuit of transactions of illegal goods between strangers. Satyanath et al. (2013) empirically show that networks of civic associations facilitated the rise of the Nazi Party.

Finally, the model also predicts that the higher the share of desirable producers in an economy, the higher the incentives for bureaucrats and undesirable producers to bring about corrupt transactions since they can get higher economic payoffs, and, consequently, the higher institutional quality needed to fight corruption. This suggests that if the number of desirable producers increases, a country should implement policies that improve the quality of institutions.

The rest of this chapter is organized as follows. Section 1.2 presents and discusses the model. Section 1.3 empirically tests the main prediction of the model and, finally, section 1.4 concludes.

1.2 The model

1.2.1 Setup

The economy is inhabited by three groups: producers, consumers and bureaucrats. Each group of the population consists of a continuum of measure one of risk-neutral individuals. The first of them is composed of producers who can be of two types: a fraction $\omega \leq \frac{1}{2}$ of them are undesirable producers (u) and the remaining ones are desirable producers (d). Both types can costlessly produce one unit of a good if they engage in an entrepreneurial activity. However, desirable producers have the ability to produce such a good of high quality (HQ) and undesirable producers, on the other hand, can only make it of low quality (LQ). These goods are sold on the market. Alternatively, these individuals can engage in home production where they produce for personal consumption. The second group of the population is composed of consumers. They, in their turn, either consume one unit or zero units of the goods produced by the entrepreneurs. They positively value high-quality goods and do not derive positive utility from low-quality ones. If there were perfect information, consumers would buy goods only from desirable producers, i.e., desirable producers would become entrepreneurs and undesirable producers would engage in home production. However, I assume the presence of asymmetric information in the sense that consumers only identify the nature of a good after its purchase. In order to avoid that consumers purchase low quality goods, there is a reg-

ulatory agency composed of the last group of the population, the bureaucrats. They can distinguish producers' types and issue licenses in order to prevent undesirable producers from becoming entrepreneurs. These bureaucrats endogenously choose whether to be honest and act in the public interest or whether to be corruptible and issue licenses to undesirable producers in exchange for a bribe. Once again, I assume that there is asymmetric information in the sense that producers do not observe if bureaucrats chose to be honest or corruptible. Bureaucrats and producers, and, entrepreneurs and consumers, are randomly matched. The timing of the model is as follows:

1. I assume that if bureaucrats want to be corruptible, they need to learn how to accept bribes and how the corruption business works in general. In a similar vein, if they want to be honest they need to learn how an honest person behaves. As a result, I impose that bureaucrats choose whether to become honest or corruptible at the beginning of their professional career before knowing if in the future they will be matched with a desirable or an undesirable producer. Hence, if they decide to be honest, they will never accept a bribe. On the other hand, if they decide to be corruptible, they will accept a bribe when offered one. I normalize their wage to zero without loss of generality. If they accept a bribe, their payoff is increased by the bribe. On the contrary, I assume that they receive an extra payoff equal to $\epsilon \in [0, 1]$ if they either issue a license to a desirable producer or refuse a license to an undesirable producer. That is, ϵ is an extra payment that they receive for acting according to the law. I interpret it as a proxy for enforcement or, in other words, for institutional quality of the country.

2. Producers, both desirable and undesirable, need to decide on becoming an entrepreneur or engaging in home production. As previously discussed, they need a license if they want to become entrepreneurs. For the sake of simplicity, I assume that desirable producers can always get one and that the license is free. The payoff of a desirable producer i, in this case, is $U_{d,i}(license) = p$, where p is the endogenous price of the good sold on the market.[4] In contrast, undesirable producers need to bribe a corrupt bureaucrat to obtain a license. If an undesirable producer offers a bribe to a corruptible bureaucrat, they obtain a license and are

[4]The cost of production and the cost of the license is represented by the opportunity cost of foregone home production.

able to sell a good on the market. In such a case, the payoff of an undesirable producer i equals $U_{u,i}(bribe|corrupt) = p - b$, where b is the bribe. If they offer a bribe to an honest bureaucrat they are denied the entrance to the entrepreneurial activity, obtaining a payoff equal to $U_{u,i}(bribe|honest) = 0$.[5] However, they do not know if a bureaucrat is honest or corruptible when they decide whether to offer a bribe or not. Both types of producers, on the other hand, can engage in home production where they have heterogeneous payoffs. In particular, each producer receives a payoff $U_{j,i}(home\ production) = \psi_i$, where $j \in \{d, u\}$ and ψ_i is distributed uniformly on the unit interval. Note that the payoff they obtain in home production does not depend on whether they are desirable or undesirable producers.

3. Finally, if consumers buy a good, their utility equals $U_c(buy|HQ) = 1 - p$ and $U_c(buy|LQ) = -p$ if the good is from high and low quality, respectively. However, when they are matched with an entrepreneur, they do not know if such an entrepreneur has obtained a license through corruption or not, and thus, whether the good is of high or low quality. In the case that they opt not to buy it, they obtain a utility equal to $U_c(don't\ buy) = 0$. I assume that the price equals the highest level that the market can bear.[6] Hence, if only desirable producers become entrepreneurs, consumers are willing to pay one for the good and thus, the equilibrium price equals one. However, as more undesirable producers engage in the entrepreneurial activity, more low quality goods will be sold on the market, and consumers will be willing to pay less than one for the good. As a consequence, as more undesirable producers become entrepreneurs, the price will decrease and the entrepreneurial activity will be less appealing for desirable producers.

[5] I assume that if undesirable producers ask for a license without offering a bribe, the license is denied and they also obtain a payoff equal to zero.

[6] Note that any price $p^* \in [0, q]$, where q is the proportion of entrepreneurs who sell high-quality goods, could be supported as a Perfect Bayesian Equilibrium by some arbitrary beliefs. For instance, assume that consumers believe that a good is from average quality if an entrepreneur posts a price $p^* \in [0, q]$, and they believe that it is certainly from low quality if they see any price different from p^*. Then, consumers would only buy the good if they receive an offer of p^*. Finally, in equilibrium, all entrepreneurs would post a price equal to p^*. However, I select the maximum price which the market can bear (i.e., q) in order to focus on the negative feedback from the average quality of the pool of entrepreneurs to the price. The underlying beliefs of this equilibrium price are intuitively appealing and convenient for my model.

1.2.2 Equilibrium

Definition: *The equilibrium concept is that of a Perfect Bayesian Equilibrium, and consequently it is characterized as:*

1. *the fraction of bureaucrats who choose to be honest (σ)*
2. *the share of desirable producers who become entrepreneurs (γ)*
3. *the proportion of undesirable producers who offer a bribe (δ)*
4. *the consumers' belief that goods are from high quality (t_c)*
5. *the desirable producers' belief that bureaucrats are honest (t_d)*
6. *the undesirable producers' belief that bureaucrats are corruptible (t_u)*
7. *and the equilibrium price (p)*

such that all individuals' strategies are sequentially rational and the beliefs are derived from strategy profiles through Bayes' rule.

The model is solved by backwards induction. In the last step, consumers are randomly matched with entrepreneurs who have obtained a license.[7] Therefore, conditional on the decisions of desirable and undesirable producers and bureaucrats, a consumer buys a good if and only if:

$$t_c(1-p) + (1-t_c)(-p) \geq 0 \tag{1.1}$$

where t_c is the consumer's belief that a good is from high quality, and p is the price. Since the equilibrium price is the highest level which the market can bear, such a price will equal the consumer's belief, $p = t_c$. Then, by Bayes' rule, the equilibrium price is:

$$p = t_c = \frac{(1-\omega)\gamma}{(1-\omega)\gamma + \omega(1-\sigma)\delta} \tag{1.2}$$

[7] I assume that if there are more consumers than entrepreneurs, only a fraction of consumers are matched with entrepreneurs, the remaining ones, do not play any role.

where γ is the fraction of desirable producers who become entrepreneurs, δ is the fraction of undesirable producers who offer a bribe and σ is the fraction of bureaucrats who become honest. Note that if desirable producers are the only ones who engage in the entrepreneurial activity, the price equals one. And, as more undesirable producers become entrepreneurs, the price falls.

In the previous step, conditional on bureaucrats' decision, a desirable producer i asks for a license and becomes an entrepreneur if and only if:

$$t_d p + (1 - t_d)p \geq \psi_i \tag{1.3}$$

where t_d is the desirable producer's belief that a bureaucrat is honest. Since ψ_i is distributed uniformly on the unit interval, the proportion of desirable producers which become entrepreneurs is equal to:

$$\gamma = p \tag{1.4}$$

Similarly, an undesirable producer i offers a bribe to a bureaucrat if and only if:

$$t_u(p - b) + (1 - t_u)0 \geq \psi_i \tag{1.5}$$

where t_u is the undesirable producer's belief that a bureaucrat is corruptible. It follows that the proportion of undesirable producers offering a bribe is equal to:

$$\delta = t_u(p - b) \tag{1.6}$$

For the sake of simplicity, I assume that the bribe is set through Nash Bargaining between bureaucrats and undesirable producers. I assume that each of them has a bargaining power equal to $\frac{1}{2}$, and that if the bargaining process breaks down, each of them obtains a payoff equal to zero. As a consequence, the bribe equals $\frac{p}{2}$.

Note that when producers take their decisions, they have already formed their conjectures about what bureaucrats have chosen in the previous step. Thus, the desirable producers' belief that bureaucrats are honest will be equal to σ in equilibrium. On the other hand, the undesirable producers' belief that bureaucrats are corruptible will be equal to $1 - \sigma$ in equilibrium.

Plugging these beliefs and the equilibrium price, I can find γ and δ as a function of σ and ω:

$$\gamma = \frac{2(1-\omega)}{2(1-\omega)+\omega(1-\sigma)^2} \tag{1.7}$$

$$\delta = \frac{(1-\sigma)(1-\omega)}{2(1-\omega)+\omega(1-\sigma)^2} \tag{1.8}$$

Finally, in the first step, if bureaucrats decide on becoming honest, their expected utility equals:

$$EU_b(honest) = (1-\omega)\gamma\epsilon + \omega\delta\epsilon \tag{1.9}$$

where $(1-\omega)\gamma$ is the probability of being matched with a desirable producer who asks for a license and $\omega\delta$ is the probability of being matched with an undesirable producer who offers a bribe. In contrast, if they become corruptible, their expected utility equals:

$$EU_b(corruptible) = (1-\omega)\gamma\epsilon + \omega\delta b \tag{1.10}$$

Hence, bureaucrats choose to become honest if and only if $\Delta^b = EU_b(honest) - EU_b(corruptible) > 0$. The model has a unique equilibrium which depends on institutional quality (ϵ) and on the proportion of undesirable producers (ω). The following proposition characterizes the equilibrium of the model.

Proposition 1. *There exists a unique equilibrium that depends on ϵ and ω:*

1. ***Honest case:*** *If $\epsilon \geq \frac{1}{2}$, then, all bureaucrats are honest ($\sigma^* = 1$). All desirable producers become entrepreneurs ($\gamma^* = 1$). On the other hand, all undesirable producers engage in home production ($\delta^* = 0$). Desirable and undesirable producers' beliefs are equal to $t_d = 1$ and $t_u = 0$, respectively. Finally, the equilibrium price and consumers' belief are equal to $p^* = t_c = 1$.*

2. ***Intermediate case:*** *If $\frac{1-\omega}{2-\omega} < \epsilon < \frac{1}{2}$, then, a fraction $\sigma^* = \frac{\epsilon\omega - \sqrt{(\epsilon-\epsilon^2)(\omega-\omega^2)}}{\epsilon\omega}$ of bureaucrats become honest and the remaining ones become corruptible. A fraction $\gamma^* = 2\epsilon$ of desirable producers become entrepreneurs. A fraction $\delta^* = \frac{\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}}{\omega}$ of undesirable producers offer a bribe, and hence, a share $\delta^*(1-\sigma^*)$ of undesirable producers become entrepreneurs. Desirable and undesirable*

producers' beliefs are equal to $t_d = \sigma^*$ and $t_u = 1 - \sigma^*$, respectively. Finally, the equilibrium price and consumers' belief are equal to $p^* = t_c = 2\epsilon$.

3. ***Corrupt case:*** *If $\epsilon \leq \frac{1-\omega}{2-\omega}$, then, all bureaucrats are corruptible ($\sigma^* = 0$). A fraction $\gamma^* = \frac{2(1-\omega)}{2-\omega}$ of desirable producers and a fraction $\delta^* = \frac{1-\omega}{2-\omega}$ of undesirable producers become entrepreneurs. Desirable and undesirable producers' beliefs are equal to $t_d = 0$ and $t_u = 1$, respectively. Finally, the equilibrium price and consumers' belief are equal to $p^* = t_c = \frac{2(1-\omega)}{2-\omega}$.*

Proof in Appendix A.

Before describing the results of proposition 1, let me interpret what is trust towards strangers. I define trust towards strangers as *the individual's belief that a stranger will be of the type that the individual needs*. Of course, this is not the only definition of trust that one can think of, however, it is a natural one in the specific context of asymmetric information, and it fits to the standard view on trust.[8] Therefore, let me illustrate what is trust towards strangers in the context of the model:

1. Consumers are randomly matched with entrepreneurs. However, they do not know if an entrepreneur has obtained a license through bribery or not. Then, consumers trust strangers when they believe that a random entrepreneur is a desirable producer. That is to say, consumers' trust towards strangers is the belief that a stranger sells high-quality goods (t_c).

2. In turn, desirable producers can always obtain a license. However, if a bureaucrat is corruptible and allows the entrance of undesirable producers to the entrepreneurial activity, then, the equilibrium price decreases and they are worse

[8] Gambetta (2000) defines trust in the following way: *"Trust or, symmetrically, distrust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such an action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action. When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him. Correspondingly, when we say that someone is untrustworthy, we imply that that probability is low enough for us to refrain from doing so."*. However, previous literature has also made, explicitly or implicitly, the extra assumption that the consequences that derive from trusting strangers do not only need to be beneficial for the private individuals that participate in an exchange, but also for the rest of society. This chapter, conversely, does not impose any effect to the rest of society

off. Hence, desirable producers trust strangers when they believe that a random bureaucrat does not let undesirable producers enter the market. In other words, desirable producers' trust towards strangers is the belief that a stranger does not accept bribes (t_d). Note that in this case the trust of desirable producers does not directly affect them in order to obtain a license since I assumed that they can always obtain it. However, for general equilibrium effects, their levels of trust determine whether they become entrepreneurs or not.

3. Finally, undesirable producers are randomly matched with bureaucrats. Recall that they do not know if a bureaucrat is corrupt or honest when they need to decide whether to offer a bribe or instead engage in home production. Thus, undesirable producers trust strangers if they believe that a random bureaucrat will accept a bribe. Note that when a bureaucrat accepts a bribe, this bureaucrat is performing a beneficial action for an undesirable producer but not for the rest of the society. In brief, undesirable producers' trust towards strangers (t_u) is the belief that a stranger accepts a bribe.

I differentiate between the trust that allows cooperation for corruption (bad trust) and the one that allows cooperation for lawful activities (good trust). Accordingly, the aggregate level of bad trust is the trust of all undesirable producers, $ALBT = \omega t_u$ and, the aggregate level of good trust is the trust of all desirable producers and consumers, $ALGT = (1-\omega)t_d + t_c$.

Having defined trust towards strangers, I now discuss the results of proposition 1. If institutional quality is good enough, that is, $\epsilon \geq \frac{1}{2}$, an economy is in the honest case. In this situation, bureaucrats never accept bribes. Desirable producers, consequently, completely trust strangers and engage in the entrepreneurial activity. In contrast, undesirable producers do not trust strangers since they cannot expect strangers to accept bribes and, for this reason, undesirable producers engage in home production. Consumers, in turn, completely trust strangers since they are all selling high-quality goods. Note that in this case, only the good trust towards strangers exists, that is, trusting strangers does not help to engage in corruption. Yet, trusting strangers allows desirable producers to become entrepreneurs and it also helps consumers to purchase goods. Moreover, as I show below, the aggregate level of trust towards strangers attains its maximum level and

corruption is inexistent.

However, if institutions are poor enough, that is, $\epsilon \leq \frac{1-\omega}{2-\omega}$, the equilibrium is such that all bureaucrats always accept bribes. Accordingly, desirable producers cannot trust strangers not to accept bribes and, it is less appealing for them to engage in the entrepreneurial activity, since the entrance of undesirable producers in the market reduces the equilibrium price. Hence, only a fraction of desirable producers become entrepreneurs. Undesirable producers, on the other hand, have the highest level of trust towards strangers in the economy since they believe that their bribes will be accepted. Since undesirable producers trust strangers, some of them engage in the entrepreneurial activity through corruption. Consumers, for their part, cannot completely trust strangers since some of the entrepreneurs have obtained a license through bribery. In this situation, the aggregate level of trust towards strangers is minimal and the aggregate level of corruption is maximal as I show below. Note that in this case, some of the most trusting individuals are those ones who hold high levels of bad trust. These high levels of bad trust allow undesirable producers to cooperate with bureaucrats to dodge a regulation.

Finally, if institutions have an intermediate quality, that is, $\frac{1-\omega}{2-\omega} < \epsilon < \frac{1}{2}$, then, some bureaucrats are corrupt and the rest are honest. In this situation, the aggregate level of trust towards strangers and the aggregate level of corruption have intermediate levels as I show below.

The above proposition leads to the following corollary which constitutes the main prediction of the model.

Corollary 1. *The individual-level correlation between between tolerance towards corruption and trust towards strangers changes depending on the institutional quality of a country. In particular, individuals who tolerate corruption more are also, on average, the more trusting in countries with poor enough institutional quality, and the less trusting in countries with good enough institutional quality.*

Proof of corollary 1: *By proposition 1.*

In order to clarify the above result, let me ask to the individuals of the model the following two questions: *"Do you tolerate corruption?"* and *"Do you trust strangers?"*.

Desirable producers and consumers never tolerate corruption since they are worse off with it. Conversely, undesirable producers always tolerate corruption since is what allows them to become entrepreneurs. The answers to the second question depend on institutional quality and on the proportion of undesirable producers. That is, the responses depend on whether individuals live in a country with good or bad enough institutional quality. Desirable producers and consumers completely trust strangers in a country with good enough institutional quality, and do not trust in a country with poor enough institutional quality.[9] On the other hand, undesirable producers do not trust strangers in a country with good enough institutional quality, and fully trust strangers in a country with poor enough institutional quality. Consequently, the association between trusting strangers and corruption toleration at the individual level varies across countries. Figure 1.1 illustrates this result. That is, the individual-level correlation between tolerance of corruption and the level of trust towards strangers is negative for countries with good enough institutional quality and positive for countries with poor enough institutional quality. I find empirical support for this prediction in the next section.

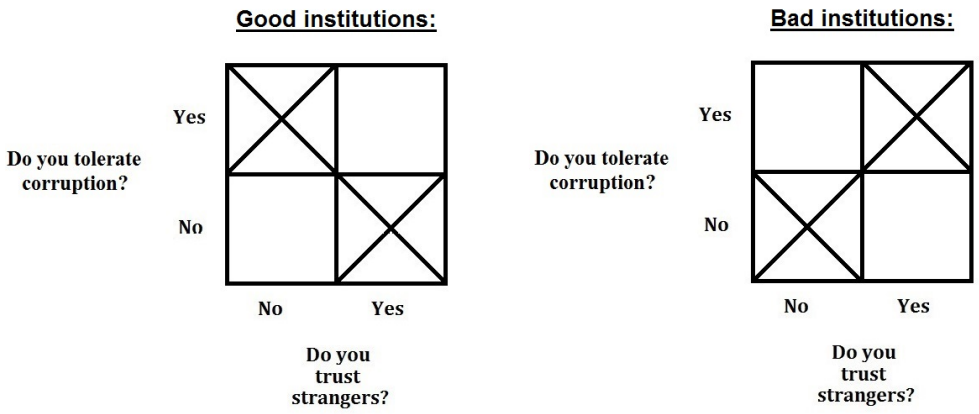

Figure 1.1: Prediction: tolerance of corruption and trust

[9]Consumers trust strangers to some extent in a country with poor enough institutional quality, but less than in a country with good enough institutional quality.

1.2.3 Comparative statics

Figure 1.2 displays for which parameter combinations, each of the equilibrium cases shows up. If institutions are poor enough, that is, ϵ and ω are in region I, a country ends up in the corrupt case. On the other hand, if ϵ and ω are in region II, a country is in the intermediate case. Finally, if ϵ and ω are in region III, a country is in the honest situation. Note that that the lower the share of undesirable producers (ω) in an economy, the higher institutional quality needed to escape from the corrupt case. Why? Bear in mind that if a country is in the corrupt case, the equilibrium price equals $p^* = \frac{2-2\omega}{2-\omega}$, which is a decreasing function of ω. Consequently, the bribe is also a decreasing function of ω. Hence, as more desirable producers there are, it is more appealing for bureaucrats to become corruptible and, at the same time, it is also more tempting for undesirable producers to offer a bribe. Assume that a country is in region II and that it succeeds in reducing the proportion of undesirable producers. Then, if institutional quality remains constant, the country might end up in region I, where corruption increases. This observation suggests that countries with intermediate levels of corruption should improve their institutional quality if the share of undesirable producers decreases and if they do not want corruption to go up.

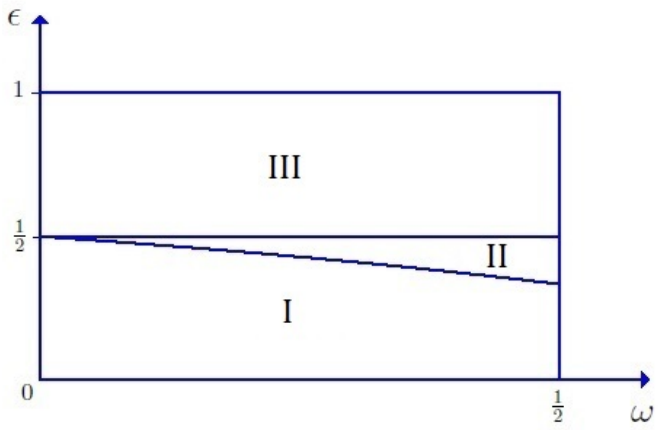

Figure 1.2: Parameters and equilibrium

If different countries are to be found in different regions of the parameter space, the model predicts the negative cross-country correlation between trust towards strangers

and corruption. Figure 1.3 shows that when the aggregate level of trust towards strangers ($ALT = ALGT + ALBT$) increases, the aggregate level of corruption ($ALC = \omega(1 - \sigma^*)\delta^*$) decreases. The upper panel illustrates that the aggregate level of trust achieves its lowest level when institutions are poor enough and hence, a country is in the corrupt case. Then, when institutional quality attains in-between values, a country moves to the intermediate case. In this middle region, the aggregate level of trust increases as institutions improve. Finally, when institutional quality is high enough, a country is in the honest case and the aggregate level of trust attains its highest level. The lower panel shows that the aggregate level of corruption follows the opposite pattern. That is, corruption achieves its highest level when institutions are poor enough. Then corruption starts to fall as institutional quality increases until institutions reach a sufficiently high level in which corruption is nonexistent. It is also interesting to see how the aggregate levels of good and bad trust vary. Figure 1.4 shows that as institutional quality increases the aggregate level of good trust increases and the aggregate level of bad trust decreases.

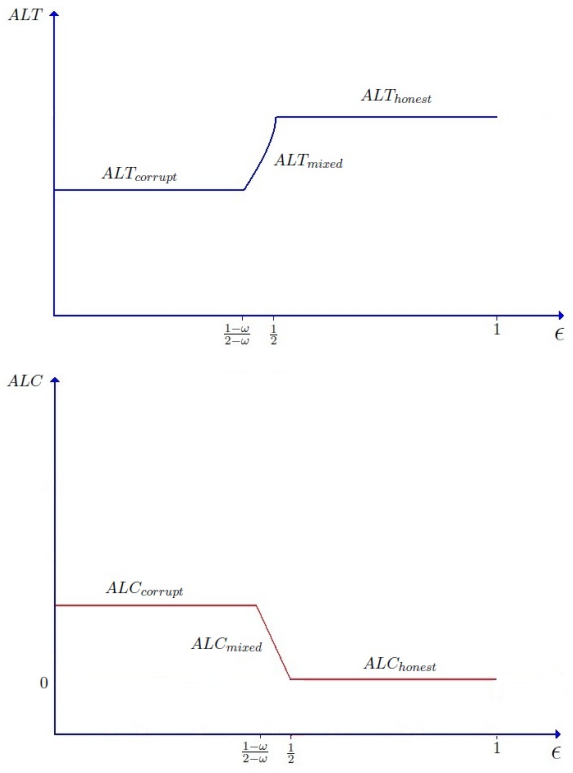

Figure 1.3: Aggregate level of trust and corruption

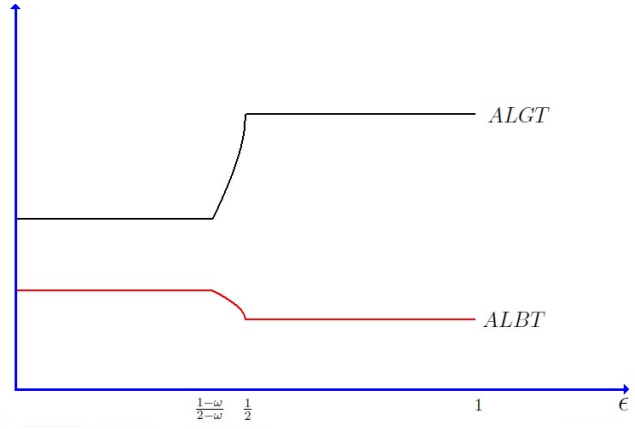

Figure 1.4: Aggregate levels of good and bad trust

The comparative statics on the individual levels of trust are illustrated in Figure 1.5. When institutions are bad, the most trusting individuals are the undesirable producers since they believe that strangers are of the type that they need (corruptible). Conversely, when institutions are good enough, undesirable producers cannot longer trust strangers since all bureaucrats are honest. Analogously, the opposite pattern arises for desirable producers.

Finally, Figure 1.6 shows how the proportion of desirable producers who become entrepreneurs (γ^*) increases as a country achieves better institutions and moves away from the corrupt scenario. Conversely, the proportion of undesirable producers who offer a bribe to a bureaucrat (δ^*) decreases. The reason behind is that the equilibrium price decreases as a country becomes more corrupt (Figure 1.7), and this crowds out desirable producers from the entrepreneurial activity. In the following proposition, I state the above results in a more formal way.

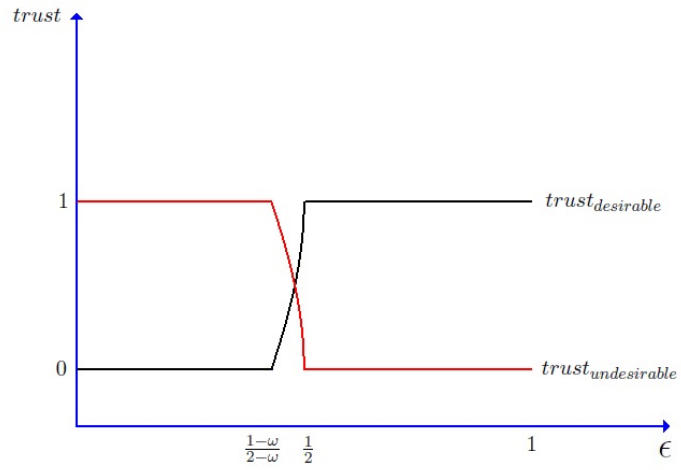

Figure 1.5: Individual level of trust of desirable and undesirable individuals

Proposition 2. *The higher the quality of instititutions,*

1. *the higher the aggregate level of good trust towards strangers:* $\frac{\partial ALGT}{\partial \epsilon} \geq 0$
2. *the lower the aggregate level of bad trust towards strangers:* $\frac{\partial ALBT}{\partial \epsilon} \leq 0$
3. *the higher the aggregate level of trust towards strangers:* $\frac{\partial ALT}{\partial \epsilon} \geq 0$
4. *the lower the aggregate level of corruption:* $\frac{\partial ALC}{\partial \epsilon} \leq 0$
5. *the higher the trust towards strangers of desirable producers:* $\frac{\partial t_d}{\partial \epsilon} \geq 0$
6. *the lower the trust towards strangers of undesirable producers:* $\frac{\partial t_u}{\partial \epsilon} \leq 0$
7. *the higher the proportion of desirable producers who become entrepreneurs:* $\frac{\partial \gamma^*}{\partial \epsilon} \geq 0$
8. *the lower the proportion of undesirable producers who become entrepreneurs:* $\frac{\partial \delta^*}{\partial \epsilon} \leq 0$
9. *the higher the equilibrium price and consumers' trust towards strangers:* $\frac{\partial p^*}{\partial \epsilon} \geq 0$

Proof in Appendix A.

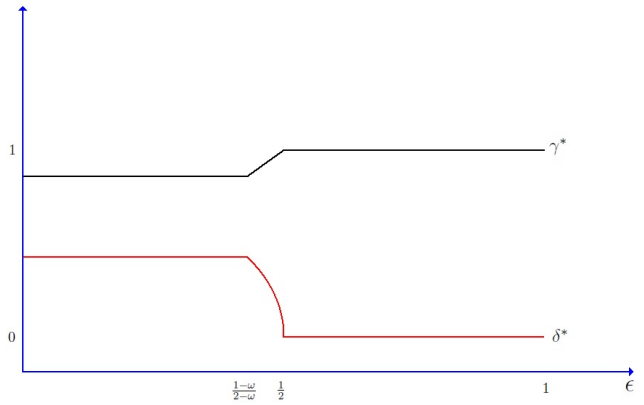

Figure 1.6: Proportion of desirable producers who become entrepreneurs and proportion of undesirable producers who offer a bribe to a bureaucrat

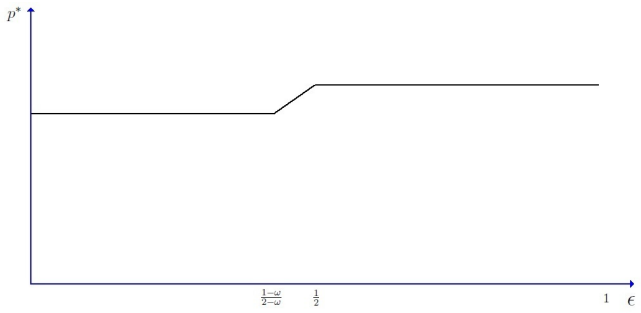

Figure 1.7: Equilibrium price

Finally, the model also predicts that in a country with good institutions, only desirable producers obtain a license to become entrepreneurs. In contrast, in a country with poor institutions, also undesirable but high-trusting individuals engage in the entrepreneurial activity, and, as a consequence, some desirable producers no longer find it appealing to engage in it. Hence, the entrepreneurial licenses are not allocated to the most desirable producers. There exists supporting evidence for this prediction. In corrupt countries, resources are allocated depending on the connections of firms and not because of their productivity (Khwaja and Mian, 2005). However, how are these connections formed? Contacts, or connections, can be thought of as being a function of trust towards strangers. That is, someone who trusts strangers has a comparative ad-

vantage in creating new acquaintances and contacts. In contrast, it is more difficult for someone who cannot trust strangers to make new contacts since they are afraid of being betrayed. All things considered, it seems that trust towards strangers assists undesirable producers in obtaining licenses in countries with poor enough institutional quality.

Summing up, the model illustrated that the current environment determines how much individuals trust strangers, and that those levels of trust affect economic outcomes. That is, depending on institutional quality and on the characteristics of the population, individuals rationally form their levels of trust. Then, those individuals' levels of trust assist them to cooperate with a stranger in situations of asymmetric information. However, while this cooperation is privately beneficial for these individuals, it may have harmful consequences for the rest of society. Furthermore, the model shows that this more general view of trust (good and bad trust) does not go against the negative cross-country correlation between trust towards strangers and corruption. Concretely, it has shown that in corrupt countries, the existence of high levels of bad trust, which is the one that helps individuals to engage in corruption, goes hand-in-hand with low levels of good trust. And, altogether, the aggregate level of trust (good and bad) is low. Moreover, it predicted that the individual-level correlation between tolerance towards corruption and trust towards strangers changes depending on the institutional quality of a country. In those countries with poor enough institutional quality, individuals who tolerate corruption more are also the more trusting ones. Conversely, in countries with good enough institutional quality, individuals who do not tolerate corruption are the more trusting ones. In the next section I test this prediction.

1.3 Test of the main prediction

1.3.1 Data

My main data source is the World Value Survey. This survey interviews a representative national sample of at least 1000 individuals in each country in order to gather comparable data on people's values, beliefs and attitudes. To date, six different waves have

been published: 1981, 1990, 1995, 2000, 2005 and 2010. In order to measure trust towards strangers I use the two most recent releases since they are the only ones which include the following question: "I'd like to ask you how much you trust people from various groups. Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all?". Among the groups included, could be found "People you meet for the first time". This question clearly identifies the trust held towards strangers. I create a variable called *trust_strangers* that equals the answers to this question and which are set to 0, 1, 2 and 3 if people respond "not at all", "not very much", "somewhat" and "completely trust", respectively.

Nonetheless, the literature has mainly been using the answers to the following question as a proxy for trust towards strangers: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". This question is included in all the waves of the World Value Survey. There exist two possible responses: "Need to be very careful" and "Most people can be trusted". I create a variable called *generalized_trust* which equals 0 if individuals answered "Need to be very careful" and 1 if they answered "Most people can be trusted". However, the crux of the matter is to know what people think when they are asked for "most people". Are they thinking just of strangers or also of some of their acquaintances? For this reason, the variable *trust_strangers* identifies trust towards strangers in a more accurate way than *generalized_trust*. However, I also use *generalized_trust* to proxy trust towards strangers as a robustness check.

This same survey includes a question that allows me to proxy for how much individuals tolerate corruption. Specifically, it asks how justifiable is to accept a bribe. In particular, the question is worded as follows: "Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between". Among the included actions it can be found "Someone accepting a bribe in the course of their duties". I create a variable called *tolerance* that equals the answers, which range from 1 ("never justifiable") to 10 ("always justifiable").

Among the controls, I also include several individual characteristics contained in the World Value Survey, such as age, gender, education and size of town. These controls are included to take into account that trust towards strangers and the tolerance of corruption of each individual might be correlated with them. For instance, it could be the case that the individuals' tolerance of corruption and their levels of trust towards strangers are

correlated with the size of the town in which they live. In smaller towns, individuals might trust more strangers since they might believe that their fellow citizens are similar to them. At the same time, they might also be more tolerant towards corruption since in smaller towns, there might be less job opportunities and corruption could permit them to obtain a certain contract or permit.

Finally, the country-level data of institutional quality is the Quality of Government index which takes into account the level of perceived corruption, law and order and bureaucracy quality of a country, and which is published in the International Country Risk Guide (ICRG)[10]. The Quality of Goverment variable ranges from 0 to 1. I create a variable called *institutions* that equals the original one but rescaled from 0 to 10, where the higher the values, the higher institutional quality. I take this index for the years 2005 and 2010 which correspond to the years of the waves of my individual-level data. In 2005, this index was available for 150 countries. Finland was the country with best institutional quality with a score equal to 10. In contrast, Somalia with a score of 0.83 was the country with the lowest institutional quality. As it is generally the case with subjective measures of corruption and of institutional quality, some criticisms about their validity exist. As a consequence, there is a growing empirical literature which attempts to study corruption and institutional quality more objectively.[11] I will, however, use a subjective measure since the more objective ones are not available as a cross-country database. Moreover, Fisman and Miguel (2007) show the legitimacy of using perceptions through a natural experiment: they notice a high correlation between an objective measure of the cheating behavior of UN diplomats and the perceived corruption of their country of origin. Additionally, Olken (2009) finds a correlation, albeit a weak one, between perceptions and a more objective measure of corruption. Appendix B reports summary statistics for all the variables I use and a detailed description of them. It also shows the 64 countries which are included in the analysis with the mean of their institutional quality index.

[10]The ICRG is produced by Political Risk Services which is a private firm providing risk assessments across countries. I downloaded this data from The Quality of Government Institute, http://www.qog.pol.gu.se.

[11]See Olken (2012) for a review.

1.3.2 Main pattern

The main pattern in the empirical analysis is summarized in Figure 1.8, where I group individuals depending on the value that the variable *trust_strangers* takes, and then, per group, I compute the mean of the logarithm of the variable *tolerance*. In high-institutional quality countries, like Norway or Switzerland, individuals with higher levels of trust towards strangers tolerate less corruption on average. Conversely, in low-institutional quality countries, like Mali or Belarus, the more trusting individuals are the more tolerant towards corruption on average.

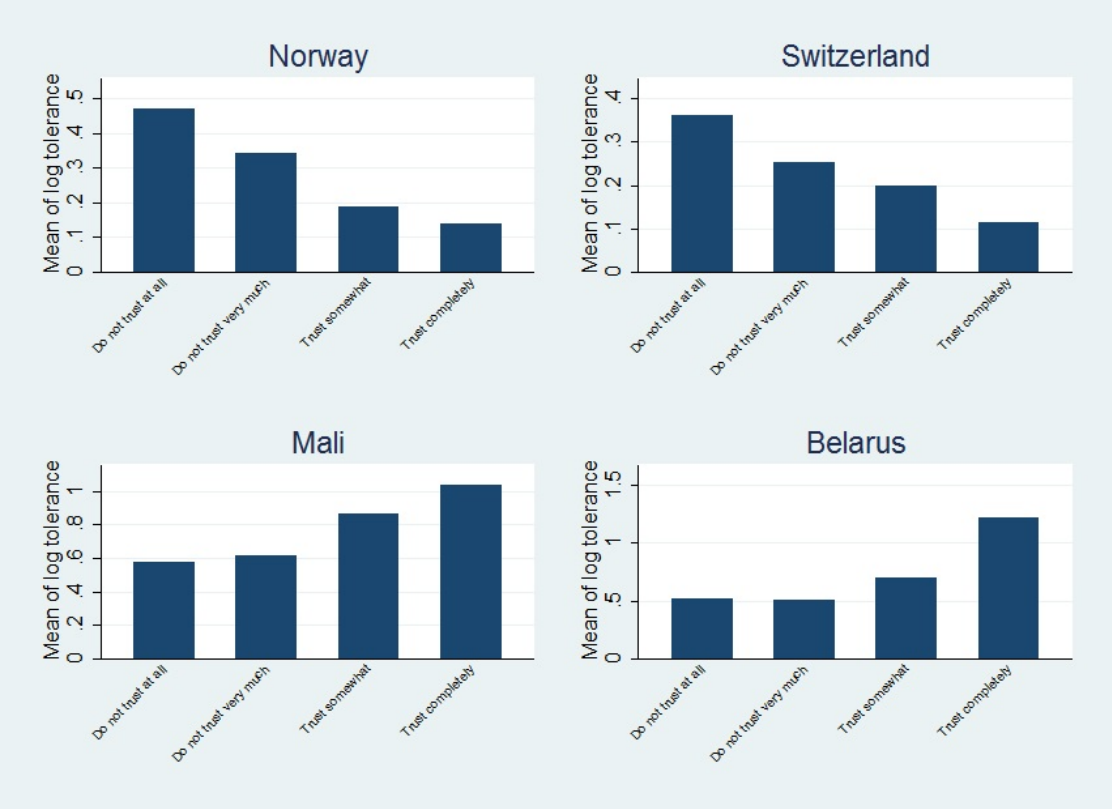

Figure 1.8: Trust towards strangers and tolerance towards corruption

1.3.3 Baseline results

In order to test the main prediction of the model more precisely, I estimate the following interaction model:

$$\begin{aligned} trust\_strangers_{ict} &= \alpha + \lambda\, log(tolerance_{ict}) + \\ &\quad + \eta\, log(tolerance_{ict}) * institutions_{ct} + \beta_{ct} + X^{'}\gamma + \epsilon_{ict} \end{aligned} \tag{1.11}$$

where i, c and t stand for an individual i in a country c in wave t; $trust\_strangers_{ict}$ is the individual level of trust held towards strangers; $tolerance_{ict}$ stands for how tolerant towards corruption an individual is; $tolerance_{ict} * institutions_{ct}$ is the interaction term between the individual level of tolerance towards corruption and the country institutional quality index; β_{ct} are country-wave fixed effects to control for country-wave specific characteristics; and $X^{'}$ are individual controls. Equation 1.11 is estimated using ordinary least squares where standard errors are clustered at the country level in order to account for intra-class correlation between individuals. Note that I take the natural logarithm of the variable *tolerance* since its distribution is positively skewed.[12] In the whole chapter I restrict my sample to all individuals who are aged more than 16 years old. The results of estimating this specification are reported in Table 1.1. Looking first at column 1, it can be seen that individuals who tolerate corruption more are also the more trusting in low-institutional quality countries, i.e., the coefficient on tolerance of corruption is positive and statistically significant. However, this association decreases as the level of institutional quality in a country increases, i.e., the coefficient on the interaction term is negative and statistically significant. Columns 2 shows that this pattern is robust to the inclusion of several individual characteristics in order to control for omitted variable bias. As a robustness check, in Appendix C.1, I also show that these results are also robust to the inclusion of the town's size in which individuals live.

While the estimated coefficients are statistically significant and are informative, it remains to be shown how low the level of institutional quality in a country needs to be in order that individuals who tolerate corruption more are also the more trusting ones.

[12] Nonetheless, the same results for the whole analysis arise if I do not take the natural logarithm.

Table 1.1: Baseline results

	(1) trust_strangers	(2) trust_strangers
log_tolerance	0.141***	0.138***
	(0.0289)	(0.0278)
log_tolerance*institutions	-0.0218***	-0.0200***
	(0.0046)	(0.0043)
age		0.00344***
		(0.0013)
age2		-0.00000378
		(0.0000)
male		0.0317***
		(0.0077)
education		0.0156***
		(0.0039)
Country-Wave FE	yes	yes
N	116657	116657
R^2	0.103	0.107

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses.

Following Brambor and Clark (2006), I graphically show the predicted individual-level correlations between the logarithm of tolerance of corruption and trust towards strangers across the possible range that the institutional quality index in a country can take. In Figure 1.9, I show such a graph when using the estimation results from column 2 in Table 1.1.[13] The solid blue line indicates how the correlation between the variables of interest changes as the institutional quality level in a country varies. The red dashed lines indicate the 95% confidence interval that allows seeing whether the correlation is statistically significant, i.e., whenever the upper and lower bounds are both below or above the zero line.

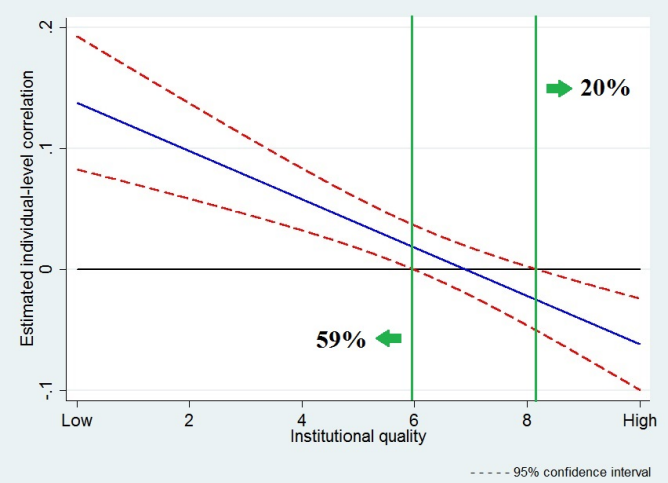

Figure 1.9: Institutional quality and individual-level correlation

Figure 1.9 shows that there exist two thresholds. First, for countries with institutional quality indexes higher than 8.171, the correlation between the logarithm of tolerance of corruption and trust towards strangers is negative and statistically significant. In my sample, 20% of the countries have, on average, institutional quality indexes above this upper threshold. Second, for countries with institutional quality indexes lower than 5.942, the correlation is reversed, and actually becomes positive and statistically significant. In my sample, 59% of the countries have, on average, institutional quality indexes below this lower threshold. This provides empirical support to the prediction of the

[13]The same results hold if I instead use the estimations from column 1.

model. The individual-level correlation between tolerance towards corruption and trust towards strangers varies with institutional quality. In countries with good institutional quality, individuals who tolerate corruption less are more trusting on average. However, in countries with poor institutional quality, individuals who trust strangers more are the ones who also tolerate corruption more. In high-institutional quality countries, non-corrupt individuals might engage in honest activities because they trust strangers to cooperate with them and they do not fear to be cheated. Conversely, in low-institutional quality countries, individuals might engage in corruption by the fact that they can trust strangers to reciprocate their corrupt deals.

In order to see which is the individual-level correlation between the two variables of interest per each country, in Figure 1.10 I also add to the previous graph the estimates that I obtain of regressing trust towards strangers on tolerance of corruption, the individual controls and wave fixed effects for each country separately. The predicted correlation is negative for countries with good institutional quality as Sweden, Norway or the United States. However, it is positive for countries with poor institutional quality like Azerbaijan, Mali or Iraq.

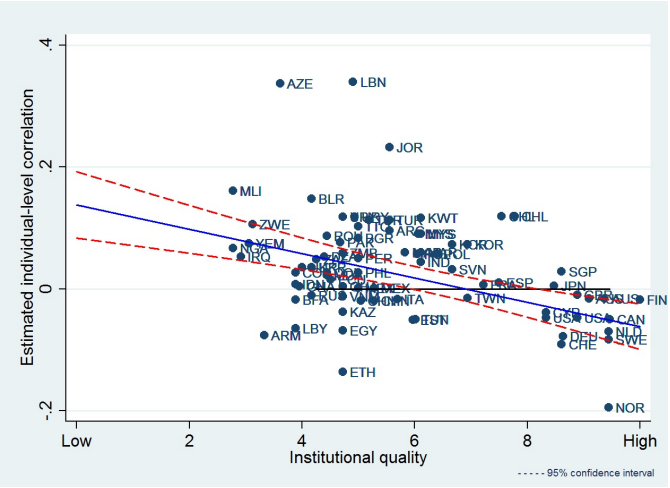

Figure 1.10: Institutional quality and individual-level correlation with countries included

1.3.4 Who are these individuals?

The above analysis has illustrated that individuals who tolerate corruption more are also the more trusting on average, in countries with poor institutions. However, in order to reinforce the idea that trusting strangers assists in the initiation of corrupt deals in this type of countries, it remains to be shown that the results are true for those individuals who more likely engage in corruption. Individuals who are employed in professions which entail interacting with bureaucrats are more likely to engage in corruption. For instance, businesspeople and managers may need to obtain licenses from the government, and these procedures are prone to corruption. As a result, I divide my sample in two groups. The first group includes those who work as managers, employers, lawyers, accountants, teachers, and office workers; i.e., those professions which are more exposed to corruption.[14] The second group contains those individuals who are employed as manual or agricultural workers. In other words, the first group is composed of white-collar workers, while the second of blue-collar workers. However, the information on the professions of individuals is not available for the 2010 wave of the World Value Survey, thus, I restrict my sample to the 2005 wave.

Table 1.2: White-collar vs blue-collar workers

	(1)	(2)
	White collar:	Blue collar:
	trust_strangers	trust_strangers
log_tolerance	0.163***	0.107*
	(0.0431)	(0.0596)
log_tolerance*institutions	-0.0264***	-0.0143
	(0.0056)	(0.0092)
Country FE	yes	yes
N	14911	18037
R^2	0.164	0.092

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. All specifications include controls for age, age2, gender and education.

[14] The World Value Survey classifies lawyers, accountants and teachers in the same category.

Table 1.2 reports the results of estimating the specification of Equation 1.11 separately for white- and blue-collar workers. Since I only have data on the first wave, I include country fixed effects instead of country-wave fixed effects. The coefficient on tolerance of corruption and on the interaction term are only statistically significant for white-collar workers. In Figure 1.11, I graphically illustrate, for each group, how the individual-level correlation between the logarithm of tolerance and trust varies across the possible range that the institutional quality index in a country can take. The graph on the left shows that, on average, the white-collar workers who tolerate corruption more are also the more trusting in countries with poor institutions, and the less trusting in countries with good institutions. The graph on the right, in turn, shows that this is not statistically significant for blue-collar workers. These results provide support for the idea that trusting strangers assists in the initiation of corrupt deals to those individuals who more likely engage in corruption in low-institutional quality countries. In Appendix C.2, I show the results of estimating Equation 1.11 for each type of occupation without grouping individuals.[15] It can be seen that the coefficients are as expected for managers/employers and for supervisory/non-supervisory office workers.

I likewise perform three robustness checks. First, in Appendix C.3, I show that what allows individuals to engage in corrupt deals in low-institutional quality countries is the trust towards strangers and not the trust directed to known people. If I use measures on how individuals trust their family, their neighborhood or their acquaintances, I do not find that individuals who tolerate corruption more are also the more trusting in low-institutional quality countries, and the less trusting in high-institutional quality ones. Second, in Appendix C.4, I show that my results are robust to the use of the commonly used variable *generalized_trust* instead of using the variable *trust_strangers*. And third, in Appendix C.5, I also exhibit that my results are robust to the use of other measures which proxy for the tolerance of corruption.

[15]Individuals occupations are the following: employer/manager of establishment with 10 or more employees; employer/manager of establishment with less than 10 employees; professional worker lawyer, accountant, teacher, etc.; supervisory office worker who supervises others; non-manual office worker who not supervises others; foreman and supervisor; skilled manual worker; semi-skilled manual worker; unskilled manual worker; farmer who has own farm; agricultural worker; and member of armed forces or security personnel.

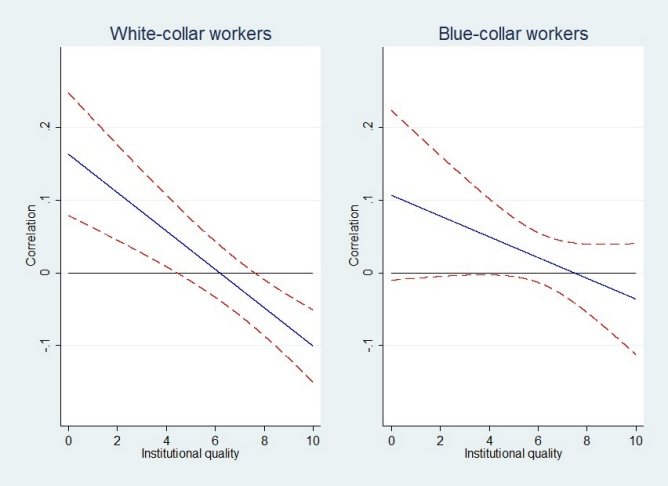

Figure 1.11: White- and blue-collar workers

1.4 Conclusions

In this chapter I have theoretically illustrated that trusting strangers is one of the channels through which institutions affect economic outcomes, particularly, corruption and entrepreneurship. Furthermore, I proposed that trusting strangers is a doubled-edged sword. On the one hand, trust is a social lubricant that helps individuals for the pursuit of outcomes aligned with social welfare. On the other, it also assists individuals in the achievement of corrupt deals.

Good institutions cause that individuals who do not tolerate corruption hold high levels of trust, and that individuals that tolerate it end up not trusting strangers. As a consequence, honest individuals take part in socially desirable economic exchanges, and corrupt individuals do not engage in corruption. However, when bad institutions are in place, the opposite is true. In particular, individuals who tolerate corruption hold high levels of trust, and this allows them to bribe strangers in order to circumvent a regulation. Using data from individuals of 64 countries, I found empirical support for this prediction.

1.5 Appendices

1.5.1 Appendix A

Proof of proposition 1:

1. *If bureaucrats choose to become honest, then, by Bayes' rule,* $t_d = 1$ *and* $t_u = 0$*. Then, a fraction* $\gamma = p$ *of desirable producers engage in the entrepreneurial activity. In contrast, undesirable producers engage in home production (*$\delta^* = 0$*) since* $\psi_i \geq 0 \ \forall i$*. By Bayes' rule,* $t_c = p^* = 1$*, and hence* $\gamma^* = 1$*. Given producers' strategies and beliefs, bureaucrats are also optimizing since they are indifferent between becoming corrupt or honest and their expected utility equals* $(1-\omega)\epsilon$*. However, these strategies are only stable if* $\epsilon \geq \frac{1}{2}$*. In order to see this, assume that* $\sigma = 1-\rho$*, where* ρ *is an infinitesimal which tends to 0. Then,* $\Delta^b < 0$ *if* $\epsilon < \frac{1}{2}$*.* □

2. *Bureaucrats are indifferent between becoming honest or corrupt if and only if* $\Delta^b = 0$*. Plugging* γ *and* δ *from Equations 1.7 and 1.8, I find that the fraction* σ^* *equals* $\frac{\epsilon\omega - \sqrt{(\epsilon-\epsilon^2)(\omega-\omega^2)}}{\epsilon\omega}$*. Then, it is easy to show that* $0 < \sigma^* < 1$ *if and only if* $\frac{1-\omega}{2-\omega} < \epsilon < \frac{1}{2}$*. Next, by Bayes' rule,* $t_d = \sigma^*$ *and* $t_u = 1 - \sigma^*$*. Plugging* σ^**, I find that* $\gamma^* = 2\epsilon$ *and* $\delta^* = \frac{\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}}{\omega}$*. Finally, by Bayes' rule, I find that* $p^* = t_c = 2\epsilon$*.* □

3. *If bureaucrats choose to become corrupt, then, by Bayes' rule,* $t_d = 0$ *and* $t_u = 1$*. Consequently, a fraction* $\gamma = p$ *of desirable producers and a fraction* $\delta = \frac{p}{2}$ *of undesirable producers become entrepreneurs. By Bayes' rule,* $t_c = p_c^* = \frac{2(1-\omega)}{2-\omega}$*,* $\gamma^* = \frac{2(1-\omega)}{2-\omega}$*,* $\delta^* = \frac{1-\omega}{2-\omega}$*. Given these strategies and beliefs, bureaucrats are also optimizing since in the case of becoming honest their expected utility would be lower than the one they obtain being corrupt under the assumption that* $\epsilon \leq \frac{1-\omega}{2-\omega}$*.* □

Proof of proposition 2:

1. *Recall that* $ALGT = (1-\omega)t_d + p^*$. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $ALGT = \frac{2(1-\omega)}{2-\omega}$, *hence,* $\frac{\partial ALGT}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial ALGT}{\partial \epsilon} = \frac{1-2\omega+\omega^2+4\epsilon\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}}{\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}} > 0$ $\forall \epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $ALGT = 2-\omega$, *hence,* $\frac{\partial ALGT}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial ALGT}{\partial \epsilon} \geq 0$ $\square$

2. *Recall that* $ALBT = \omega t_u$. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $ALBT = \omega$, *hence,* $\frac{\partial ALBT}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega)}, \frac{1}{2}]$, *then* $\frac{\partial ALBT}{\partial \epsilon} = \frac{-\omega(1-\omega)}{2\epsilon\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2}} < 0$ $\forall \epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $ALBT = 0$, *hence,* $\frac{\partial ALBT}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial ALBT}{\partial \epsilon} \leq 0$ $\square$

3. *Recall that* $ALT = (1-\omega)t_d + \omega t_u + p^*$. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $ALT = 2-\omega$, *hence,* $\frac{\partial ALT}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial ALT}{\partial \epsilon} = \frac{1-3\omega+2\omega^2+4\epsilon\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}}{2\epsilon\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}} > 0$ $\forall \epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $ALT = \omega + \frac{2-2\omega}{2-\omega}$, *hence,* $\frac{\partial ALT}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial ALT}{\partial \epsilon} \geq 0$ $\square$

4. *Recall that* $ALC = \omega(1-\sigma^*)\delta^*$. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $ALC = \omega\frac{1-\omega}{2-\omega}$, *hence,* $\frac{\partial ALC}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial ALC}{\partial \epsilon} = -2(1-\omega) < 0$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $ALC = 0$, *hence,* $\frac{\partial ALC}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial ALC}{\partial \epsilon} \leq 0$ $\square$

5. *Note that* $t_d = \sigma^*$. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $t_d = 0$, *hence,* $\frac{\partial t_d}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial t_d}{\partial \epsilon} = \frac{1-\omega}{2\epsilon\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}} > 0$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $t_d = 1$, *hence,* $\frac{\partial t_d}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial t_d}{\partial \epsilon} \geq 0$ $\square$

6. *Note that* $t_u = 1-\sigma^*$. *Then, by the previous point,* $\frac{\partial t_u}{\partial \epsilon} \leq 0$ $\square$

7. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $\gamma^* = \frac{2-2\omega}{2-\omega}$, *hence,* $\frac{\partial \gamma^*}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial \gamma^*}{\partial \epsilon} = 2 > 0$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $\gamma^* = 1$, *hence,* $\frac{\partial \gamma^*}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial \gamma^*}{\partial \epsilon} \geq 0$ $\square$

8. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $\delta^* = \frac{1-\omega}{2-\omega}$, *hence,* $\frac{\partial \delta^*}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial \delta^*}{\partial \epsilon} = \frac{(1-\omega)(1-4\epsilon)}{2\sqrt{\epsilon(1-2\epsilon)(\omega-\omega^2)}} < 0$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $\delta^* = 0$, *hence,* $\frac{\partial \delta^*}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial \delta^*}{\partial \epsilon} \geq 0$ $\square$

9. *If* $\epsilon \in [0, \frac{1-\omega}{2-\omega}]$, *then* $p^* = \frac{2-2\omega}{2-\omega}$, *hence,* $\frac{\partial p^*}{\partial \epsilon} = 0$. *If* $\epsilon \in [\frac{1-\omega}{2-\omega}, \frac{1}{2}]$, *then* $\frac{\partial p^*}{\partial \epsilon} = 2 > 0$. *Finally, if If* $\epsilon \in [\frac{1}{2}, 1]$, *then* $p^* = 1$, *hence,* $\frac{\partial p^*}{\partial \epsilon} = 0$. *Therefore,* $\frac{\partial p^*}{\partial \epsilon} \geq 0$ $\square$

1.5.2 Appendix B

Variable	Description and source
trust_strangers	This variable equals the individual answers to: "Could you tell me whether you trust people you meet for the first time completely, somewhat, not very much or not at all?" The responses are set to 0, 1, 2 and 3 if an individual responded not at all, not very much, somewhat and completely trust respectively. Source: World Value Survey 2005 and 2010
tolerance	This variable equals the individual answers to: "Please tell me whether you think it can always be justified, never be justified, or something in between for the following action: Someone accepting a bribe in the course of their duties" The responses are set to 1, 2 ,..., 10, where the higher the numbers, the higher the justification. Source: World Value Survey 2005 and 2010
education	This variable equals the individual's reported level of education. It equals 1 if no formal education, 2 if incomplete primary school, 3 if complete primary school, 4 if incomplete secondary school (technical/vocational type), 5 if complete secondary school (technical/vocational type), 6 if incomplete secondary (university-preparatory type, 7 if complete secondary (university-preparatory type), 8 if some university-level education (without degree) and 9 if university-level education (with degree). Source: World Value Survey 2005 and 2010

Variable	Description and source
age	This variable equals the individual's reported age. Source: World Value Survey 2005 and 2010
male	This variable equals 1 if an individual reported to be a male and 0 if they reported to be a female. Source: World Value Survey 2005 and 2010
generalized_trust	This variable equals the individual answers to: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" The responses are set to 0 and 1 if an individual responded need to be very careful and most people can be trusted respectively. Source: World Value Survey 2005 and 2010
trust_family	This variable equals the individual answers to: "Could you tell me whether you trust your family completely, somewhat, not very much or not at all?" The responses are set to 0, 1, 2 and 3 if an individual responded not at all, not very much, somewhat and completely trust respectively. Source: World Value Survey 2005 and 2010
trust_neighborhood	This variable equals the individual answers to: "Could you tell me whether you trust your neighborhood completely, somewhat, not very much or not at all?" The responses are set to 0, 1, 2 and 3 if an individual responded not at all, not very much, somewhat and completely trust respectively. Source: World Value Survey 2005 and 2010

Variable	Description and source
trust_acquaintances	This variable equals the individual answers to: "Could you tell me whether you trust people you know personally completely, somewhat, not very much or not at all?" The responses are set to 0, 1, 2 and 3 if an individual responded not at all, not very much, somewhat and completely trust respectively. Source: World Value Survey 2005 and 2010
cheat_taxes	This variable equals the individual answers to: "Please tell me whether you think it can always be justified, never be justified, or something in between for the following action: Cheating on taxes if you have a chance" The responses are set to 1, 2,..., 10, where the higher the numbers, the higher the justification. Source: World Value Survey 2005 and 2010
claim_gov_benefits	This variable equals the individual answers to: "Please tell me whether you think it can always be justified, never be justified, or something in between for the following action: Claiming government benefits to which you are not entitled" The responses are set to 1, 2,...,10 where the higher the numbers, the higher the justification. Source: World Value Survey 2005 and 2010
avoid_fare_transport	This variable equals the individual answers to: "Please tell me whether you think it can always be justified, never be justified, or something in between for the following action: Avoiding a fare on public transport" The responses are set to 1, 2 ,..., 10, where the higher the numbers, the higher the justification. Source: World Value Survey 2005 and 2010

Variable	Description and source
manager≥10	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "employer/manager of establishment with 10 or more employees" Source: World Value Survey 2005 and 2010
manager<10	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "employer/manager of establishment with less than 10 employees" Source: World Value Survey 2005 and 2010
professional_worker	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "professional worker lawyer, accountant, teacher, etc." Source: World Value Survey 2005 and 2010
supervisory_office	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "supervisory office worker: supervises others" Source: World Value Survey 2005 and 2010
non-supervisory_office	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "non-manual: office worker: non-supervisory" Source: World Value Survey 2005 and 2010

Variable	Description and source
supervisor_manual	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "foreman and supervisor" Source: World Value Survey 2005 and 2010
skilled_manual	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "skilled manual worker" Source: World Value Survey 2005 and 2010
semi-skilled_manual	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "semi-skilled manual worker" Source: World Value Survey 2005 and 2010
unskilled_manual	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "unskilled manual worker" Source: World Value Survey 2005 and 2010
farmer	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "farmer: has own farm" Source: World Value Survey 2005 and 2010
agricultural	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "agricultural worker" Source: World Value Survey 2005 and 2010

Variable	Description and source
army	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "member of armed forces, security personnel" Source: World Value Survey 2005 and 2010
unemployed	This variable equals 1 if an individual reported that their profession/occupation in which they are doing most of their job is "never had a job" Source: World Value Survey 2005 and 2010
town_size	This variable equals 1 if the size of the town in which each individual lives is under 2000 people, 2 if the size is between 2000 and 5000, 3 if it is between 5000 and 10000, 4 if it is between 10000 and 20000, 5 if it is between 20000 and 50000, 6 if it is between 50000 and 100000, 7 if it is between 100000 and 500000, 8 if it is 500000 and more. Source: World Value Survey 2005 and 2010
institutions	This variable equals the mean value of the three components of the political risk rating of the International Country Risk Guide (ICRG) which are: "Corruption", "Law and Order" and "Bureaucracy Quality". It ranges from 0 to 10, where higher values indicate higher quality of institutions. Source: The Quality of Goverment Institute 2005 and 2010

Table 1.3: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
trust_strangers	0.957	0.781	0	3	116657
log_tolerance	0.352	0.625	0	2.303	116657
education	5.545	2.445	1	9	116657
age	41.853	16.631	15	98	116657
male	0.480	0.500	0	1	116657
generalized_trust	0.257	0.437	0	1	111893
trust_family	2.815	0.476	0	3	115236
trust_neighborhood	1.880	0.809	0	3	115236
trust_acquaintances	2.002	0.772	0	3	115236
log_cheat_taxes	0.475	0.696	0	2.303	112188
log_claim_gov_benefits	0.606	0.770	0	2.303	112188
manager$\geq$10	0.293	0.169	0	1	33539
manager$<$10	0.058	0.234	0	1	33539
professional_worker	0.145	0.352	0	1	33539
supervisory_office:	0.075	0.264	0	1	33539
non-supervisory_office	0.137	0.344	0	1	33539
supervisor_manual	0.023	0.149	0	1	33539
skilled_manual	0.174	0.379	0	1	33539
semi-skilled_manual	0.101	0.302	0	1	33539
unskilled_manual	0.120	0.325	0	1	33539
farmer	0.049	0.517	0	1	33539
agricultural	0.071	0.256	0	1	33539
army	0.018	0.132	0	1	33539
town_size	4.943	2.444	1	8	80567
institutions	5.770	1.984	2.778	10	64

Table 1.4: Included countries I

Country	Institutional Quality Mean
Nigeria	2.777
Mali	2.777
Iraq	2.916
Yemen	3.055
Zimbabwe	3.125
Armenia	3.333
Azerbaijan	3.611
Indonesia	3.888
Libya	3.888
Burkina Faso	3.888
Ukraine	4.085
Belarus	4.166
Colombia	4.166
Russia	4.166
South Africa	4.259
Ghana	4.340
Algeria	4.398
Romania	4.444
Moldova	4.444
Ecuador	4.513
Pakistan	4.675
Egypt	4.722
Kazakhstan	4.722
Vietnam	4.722
Ethiopia	4.722
Zambia	4.722
Uruguay	4.826
Lebanon	4.907
Bulgaria	5
Peru	5
Trinidad and Tobago	5
Philippines	5

Table 1.5: Included countries II

Country	Institutional Quality Mean
Mexico	5.138
China	5.162
Turkey	5.370
Jordan	5.555
Argentina	5.555
Italy	5.694
Morocco	5.972
Estonia	5.972
Tunisia	6.018
Malaysia	6.087
India	6.111
Kuwait	6.111
Poland	6.215
Slovenia	6.666
Korea, South	6.805
Taiwan	6.944
France	7.222
Spain	7.5
Chile	7.662
Cyprus	8.333
Japan	8.472
United States	8.611
Switzerland	8.611
Singapore	8.611
Germany	8.634
United Kingdom	8.888
Australia	9.236
Sweden	9.444
Norway	9.444
Netherlands	9.444
Canada	9.467
Finland	10

1.5.3 Appendix C

Appendix C.1: Including town size as a control

Table 1.6: Town size included

	(1) trust_strangers	(2) trust_strangers	(3) trust_strangers
log_tolerance	0.146***	0.143***	0.141***
	(0.0293)	(0.0287)	(0.0285)
log_tolerance*institutions	-0.0228***	-0.0209***	-0.0208***
	(0.0048)	(0.0045)	(0.0045)
age		0.00371**	0.00370**
		(0.0014)	(0.0014)
age2		-0.00000740	-0.00000699
		(0.0000)	(0.0000)
male		0.0342***	0.0334***
		(0.0087)	(0.0087)
education		0.0153***	0.0164***
		(0.0049)	(0.0049)
town_size			-0.00659**
			(0.0033)
Country-Wave FE	yes	yes	yes
N	80567	80567	80567
R^2	0.115	0.120	0.120

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses.

Appendix C.2: Different occupations

Table 1.7: Different occupations I

	(1) Manager $\geq$10: trust_strangers	(2) Manager $<$10: trust_strangers
log_tolerance	0.266	0.293 **
	(0.1689)	(0.1208)
log_tolerance*institutions	-0.046**	-0.046***
	(0.0224)	(0.0156)
N	982	1949
R^2	0.200	0.185

	(3) Professional worker: trust_strangers	(4) Supervisory office: trust_strangers
log_tolerance	0.118	0.266***
	(0.0792)	(0.0880)
log_tolerance*institutions	-0.026***	-0.034***
	(0.0123)	(0.0120)
N	4851	2520
R^2	0.187	0.217

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. All specifications include controls for age, age2, gender, education and country FE.

Table 1.8: Different occupations II

	(5) Non-supervisory office: trust_strangers	(6) Supervisor manual: trust_strangers
log_tolerance	0.124**	0.018
	(0.0601)	(0.1821)
log_tolerance*institutions	-0.0166*	0.008
	(0.0082)	(0.0239)
N	4609	764
R^2	0.140	0.158

	(7) Skilled manual: trust_strangers	(8) Semi-skilled manual: trust_strangers
log_tolerance	0.132	0.0957
	(0.0810)	(0.1112)
log_tolerance*institutions	-0.0179	-0.0152
	(0.0110)	(0.0176)
N	5822	3393
R^2	0.123	0.096

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. All specifications include controls for age, age2, gender, education and country FE.

Table 1.9: Different occupations III

	(9) Unskilled manual: trust_strangers	(10) Farmer: trust_strangers
log_tolerance	0.129	0.201
	(0.0777)	(0.2091)
log_tolerance*institutions	-0.0131	-0.0360
	(0.0154)	(0.0337)
N	4031	1658
R^2	0.090	0.104

	(11) Agricultural: trust_strangers	(12) Army: trust_strangers
log_tolerance	0.122	0.325
	(0.3470)	(0.2237)
log_tolerance*institutions	-0.0254	-0.0557
	(0.0708)	(0.0352)
N	2369	591
R^2	0.087	0.212

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. All specifications include controls for age, age2, gender, education and country FE.

Appendix C.3: Trust towards whom?

Trust might be directed to different types of people, for instance, trust towards a close circle of family and friends as well as trust towards strangers. Next, I will show that my empirical results only hold for trust towards strangers. In particular, the World Value Survey asks individuals how much they trust their family, their neighborhood or people they know personally. Table 1.10 reports the results of estimating Equation 1.11 when using the different measures of trust. Column 1 shows that the coefficients are only statistically significant and with the expected sign for my original variable of trust towards strangers. Conversely, this same pattern does not arise when I use the measures of trust towards people closer to the individuals. In fact, individuals who have more trust towards family, people they know personally or neighbors tolerate less corruption on average, in both types of countries. One way of interpreting these results is the following. In countries with good institutions, individuals who do not tolerate corruption trust everyone: strangers and non-strangers. However, in countries with poor institutions, those same individuals may end up only trusting those who are closer to them since they may fear being betrayed by corrupt individuals.

Table 1.10: Trust towards different types of people

	(1) trust_strangers	(2) trust_family	(3) trust_acquaintances	(4) trust_neighborhood
log_tolerance	0.138***	-0.0221	0.0128	-0.0183
	(0.0282)	(0.0265)	(0.0272)	(0.0287)
log_tolerance*institutions	-0.0201***	-0.00431	-0.00911**	-0.00793*
	(0.0044)	(0.0039)	(0.0042)	(0.0047)
age	0.00340**	-0.00211**	0.00678***	-0.00235*
	(0.0013)	(0.0009)	(0.0014)	(0.0014)
age2	-0.00000288	0.0000259***	-0.00000473	0.0000409***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
male	0.0327***	0.00673	0.0413***	0.00992
	(0.0077)	(0.0054)	(0.0113)	(0.0094)
education	0.0156***	0.00687***	-0.00422	0.0165***
	(0.0039)	(0.0019)	(0.0038)	(0.0034)
Country-Wave FE	yes	yes	yes	yes
N	115236	115236	115236	115236
R^2	0.107	0.059	0.124	0.121

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses.

Appendix C.4: Generalized trust

As discussed above, the literature has been using *generalized_trust* to measure trust towards strangers. Yet, it is not clear what this variable identifies. Hence, it is fundamental to identify what captures the responses to "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". To this aim, I first show the correlation between all the different measures of trust; i.e., *generalized_trust*, *trust_strangers*, *trust_family*, *trust_acquaintances*, and *trust_neighborhood*. Table 1.11 shows that *generalized_trust* is highly correlated with trust towards strangers, trust towards acquaintances and trust towards neighbors. Second, I run an OLS regression at the individual level with *generalized_trust* on the left hand side and the other measures on the right hand side. Table 1.12 confirms that *generalized_trust* correlates significantly and positively with trust towards strangers, trust towards acquaintances and trust towards neighbors. The second column shows that this is robust to the inclusion of individual controls. Therefore, *generalized_trust* seems to be, at best, a contaminated measure of trust towards strangers.

Table 1.11: Raw correlations

	generalized_trust	trust_stran.	trust_family	trust_acquain.
trust_strangers	0.28			
trust_family	0.05	0.07		
trust_acquaintances	0.19	0.37	0.22	
trust_neighborhood	0.18	0.36	0.28	0.45

Nonetheless, I also use *generalized_trust* to proxy trust towards strangers as a robustness check. Hence, I estimate Equation 1.11 but using the alternative measure of trust. Table 1.13 provides evidence that my results are robust to the use of *generalized_trust* when using the same database as in the rest of the chapter.

However, the variable *generalized_trust* exists for the six waves of the World Value Survey. To make use of all the available data, I run again the same specification with the six waves and Table 1.14 shows that my results are robust. Columns 1 and 2 include data for the six waves. Column 2 only includes data for all the waves but the first one since in that one there is no available data on education.

Table 1.12: Different measures of trust

	(1) generalized_trust	(2) generalized_trust
trust_strangers	0.102***	0.101***
	(0.0098)	(0.0094)
trust_family	0.00758	0.00545
	(0.0067)	(0.0063)
trust_acquaintances	0.0292***	0.0272***
	(0.0039)	(0.0038)
trust_neighborhood	0.0345***	0.0373***
	(0.0035)	(0.0034)
Country-Wave FE	yes	yes
N	111893	111893
R^2	0.195	0.198

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. Column 2 includes controls for age, age2, gender and education.

Table 1.13: Generalized Trust

	(1) generalized_trust	(2) generalized_trust
log_tolerance	0.0420***	0.0406***
	(0.0129)	(0.0124)
log_tolerance*institutions	-0.00697***	-0.00638***
	(0.0024)	(0.0022)
age		0.000220
		(0.0008)
age2		0.00000553
		(0.0000)
male		0.00728**
		(0.0032)
education		0.0130***
		(0.0028)
Country-Wave FE	yes	yes
N	111893	111893
R^2	0.144	0.148

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses.

Table 1.14: Generalized Trust: all the waves

	(1) generalized_trust	(2) generalized_trust	(3) generalized_trust
log_tolerance	0.0397***	0.0398***	0.0431***
	(0.0112)	(0.0112)	(0.0120)
log_tolerance*institutions	-0.00648***	-0.00649***	-0.00655***
	(0.0020)	(0.0020)	(0.0021)
age		0.000803	0.000735
		(0.0006)	(0.0006)
age2		-0.00000764	-0.00000329
		(0.0000)	(0.0000)
male		0.00737**	0.00573*
		(0.0028)	(0.0029)
education			0.0122***
			(0.0027)
Country-Wave FE	yes	yes	yes
N	257657	257657	234142
R^2	0.124	0.124	0.131

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses. Columns 1 and 2 include data for the 6 waves of the World Value Survey. Column 3 includes data for the waves 2, 3, 4, 5 and 6 of the Wolrd Value Survey.

Appendix C.5: Different types of tolerance

The World Value Survey also provides other measures which allow me to proxy for how tolerant towards corruption individuals are. For instance, it asks how much they can tolerate cheating on taxes or claiming government benefits to which they are not entitled. Table 1.15 shows that the same pattern arises when using either "justify accepting a bribe", "justify cheating on taxes" or "justify claiming government benefits" as a proxy for the tolerance of corruption of individuals.

Table 1.15: Different types of tolerance

	(1) bribes:	(2) taxes:	(3) gov_benefits:
	trust_strangers	trust_strangers	trust_strangers
log_tolerance	0.142***	0.0883***	0.0521*
	(0.0286)	(0.0291)	(0.0274)
log_tolerance*institutions	-0.0205***	-0.0143***	-0.00919**
	(0.0043)	(0.0044)	(0.0043)
age	0.00341**	0.00339**	0.00324**
	(0.0014)	(0.0014)	(0.0013)
age2	-0.00000278	-0.00000299	-0.00000176
	(0.0000)	(0.0000)	(0.0000)
male	0.0311***	0.0321***	0.0314***
	(0.0079)	(0.0078)	(0.0079)
education	0.0158***	0.0157***	0.0156***
	(0.0041)	(0.0041)	(0.0041)
Country-Wave FE	yes	yes	yes
N	112188	112188	112188
R^2	0.108	0.107	0.107

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the country level are in parentheses.

Chapter 2

IDENTIFYING LABOR MARKET SPILLOVERS AMONG IMMIGRANTS TO THE U.S.

2.1 Introduction

Immigrants tend to agglomerate geographically with other immigrants from the same country of origin. For instance, in the U.S., 64% of Cuban immigrants were living in the Miami-Hialeah metropolitan statistical area (MSA) in 2000. Similarly, 78% of Armenian immigrants were located in the Los Angeles-Long Beach MSA. Important reasons for why immigrants cluster is that they may benefit from living next to their compatriots. For instance, these can help to find a job, provide information about cheap housing or make less traumatic the process of migrating.[1] I will in the following refer to all these advantages that immigrants obtain when clustering with their compatriots as origin-specific spillover benefits.[2]

This spillover concept has gained a central place in the literature studying agglomeration of immigrants in the U.S. (Bartel, 1989; Dunlevey, 1991; Zavodny, 1997; Jaeger, 2000, 2008; Bauer et al., 2007). Nonetheless, ethnic enclaves might also emerge in the absence of origin-specific spillover benefits. For instance, Cuban immigrants might

[1]For example, Munshi (2003) finds that Mexican immigrants are more likely to obtain a job if they are agglomerated with more immigrants coming from their same community.

[2]The immigration literature usually refers to these advantages as network benefits.

simply decide to settle in Miami-Hialeah because it is close to Cuba and it has a similar weather. I will in the following refer to these location characteristics which benefit the immigrants coming from a country of origin as origin-specific comparative advantage benefits for this location. Accordingly, ethnic enclaves might emerge as a combination of both types of origin-specific forces operating at the location level: spillovers and comparative advantages. The coexistence of both agglomeration forces creates an identification issue since the two operate at the same dimension, and consequently, ethnic enclaves cannot be completely attributed to origin-specific spillovers.

However, immigrants from the same country of origin agglomerate not only at the location level, but also at the industry-location one. Figure 2.1 displays a matrix representing which percentage of immigrants from two countries of origin and who live in a MSA work in a given industry. In particular, this matrix shows these percentages across two MSAs and two industries. Using data from the 2000 U.S. Census, Figure 2.1 shows that immigrants from Germany agglomerate in the *Construction* industry in Dallas - Forth Worth whereas they agglomerate in the *Finance and Insurance* industry in Seattle - Everett. On the other hand, immigrants from Russia follow an opposite pattern. Russian immigrants agglomerate in *Finance and Insurance* in Dallas - Fort Worth and in *Construction* in Seattle - Everett. Specifically, 7% of Germans who live in Dallas - Fort Worth work in *Construction* and none of them in *Finance and Insurance*. Contrarily, only 2% of immigrants from Germany who live in Seattle - Everett work in *Construction* whereas 6% in *Finance and Insurance*. This suggests that immigrants from the same country of origin also cluster in industry-MSA pairs.

	Construction	Finance and insurance
Dallas – Fort Worth, TX	Germany: 7% Russia: 0%	Russia: 10% Germany: 0%
Seattle – Everett, WA	Russia: 6% Germany: 2%	Germany: 6% Russia: 0%

Figure 2.1: Agglomeration at the industry-MSA level

Which forces could make immigrants from the same country settle in the same industry-MSA pair? Compatriots might end up in the same industry-MSA pair because established country fellows might provide valuable information to newcomers about job

vacancies in the industry in which they work. Furthermore, some country fellows might be running businesses and they might hire newcomers or provide helpful information for establishing a business in the same industry. Alternatively, compatriots working and living in the same industry-location pair might learn from each other more than what they would do from other individuals who also live and work in the same industry-location pair. Country fellows might be similar and they might feel sympathy for each other facilitating this learning. This latter reason is the typical knowledge-spillover benefit described by Marshall (1890) that could materialize among compatriots who work and live in the same industry-location pair. I refer to all these advantages as origin-specific spillover benefits at the industry-location level. Therefore, immigrants might cluster with their compatriots in industry-MSA pairs due to origin-specific spillovers at this level.

Moreover, it is unlikely that there are industry-location advantages that are specific for the immigrants coming from the same country of origin which would perfectly confound with the described spillovers. This is not stating that industry-location comparative advantages do not exist, but that they are not origin-specific. For instance, an industry-location could be thriving and become an industry-location advantage, but that would attract immigrants from all countries of origin, or at least all those who are good at that industry. At the same time, I consider that immigrants from a particular country of origin might have a comparative advantage to work in certain industries, but this would bring about a cluster of compatriots at the same industry in the different locations (Mandorff, 2007; Kerr and Mandorff, 2015). Going back to the example in Figure 2.1, an average German immigrant might have a comparative advantage to work in *Finance and Insurance* everywhere, that is, she might be good at this industry. Yet, it is unlikely that she has a comparative advantage to work in *Finance and Insurance* in Seattle - Everett and not in Dallas - Fort Worth. Therefore, clusters of compatriots at the industry-location level seem to be due to origin-specific spillover forces. Hence, the non-existence of origin-specific comparative advantages for industry-location pairs is my key identification assumption, and it opens an avenue to credibly estimate a specific type of spillover forces less likely to be confounded with comparative advantages.[3]

[3]The assumption that there are no origin-specific comparative advantages for industry-location pairs might fail to convince for some cases. For example, Mexican immigrants might have a comparative advantage to work in restaurants in San Jose and not in New York since a lot of restaurants' customers in

Precisely, I introduce in this chapter a measure allowing to estimate the particular type of spillovers that materializes among compatriots at the industry-location pair. To this aim, I propose a model that builds on the seminal work of Ellison and Glaeser (1997) from the urban economics literature. In my model, immigrants sequentially choose in which industry-location pair to work and live when they move to the U.S. This model micro-founds a country-level index which estimates the strength of origin-specific spillover forces operating at the industry-location level. Then, I bring this index to the data and I find that, on average across countries of origin, at least 1.55% of the immigrants in the U.S. benefit from this specific type of spillover. However, this estimation is only a lower bound and it suggests that spillover benefits happening at the industry-MSA level with immigrants' compatriots are economically meaningful. Moreover, I also show that at least 7.11% of the agglomeration of immigrants at the MSA level is driven by origin-specific spillovers at the industry-MSA level.

Furthermore, there is substantial variation of the estimated index across countries. The index is higher for those countries of origin which, on average, are more culturally distant from the U.S. and have lower levels of trust towards individuals of other nationalities. A possible interpretation for these relationships is that immigrants from these countries of origin might find it more difficult to find a job when they are not clustered with their compatriots. For instance, immigrants who are from countries of origin which are more culturally distant from the U.S., as well as immigrants from countries of origin that cannot trust individuals from other nationalities, might have more difficulties interacting with other individuals in the U.S. and they might find it more difficult to find a job. Yet, these immigrants might resort to their compatriots to find a job in the same industry in which their compatriots work.

My chapter spans two strands of related research. First, it is related to a literature which studies the determinants of immigrant locations in the U.S. (Bartel, 1989; Dunlevey, 1991; Zavodny, 1997; Jaeger, 2000, 2008; Bauer et al., 2007). Papers in this line of research establish that immigrants tend to choose locations with high concentrations of immigrants from their same country of origin or region. These concentrations are robust to controlling for distance from the country of origin to the possible final destination, the language in the final destination, etc., to take into account origin-specific

San Jose are also Mexican and they might tend to demand Mexican food, or to order in Spanish. I deal with this concern in Section 2.5.5.

comparative advantages for locations. Therefore, they argue that the fact that immigrants are attracted to destinations with high concentrations of immigrants from their same country of origin shows that immigrants enjoy spillover benefits for being collocated with their compatriots. Yet, there might exist a myriad of possible variables, apart from the ones considered which might represent an origin-specific comparative advantage for a location; for instance, location specific policies beneficial for the immigrants of some specific countries, or simply unobserved origin-specific tastes for locations. These unobserved origin-specific comparative advantages for locations might be common for the different waves of immigrants who choose a destination in the U.S. Thus, one would find that a recently arrived wave of immigrants would be attracted to the same destination where there were a high concentration of their compatriots, even with the non-existence of origin-specific spillover payoffs. My chapter is distinguished from this research in the sense that I consider spillover benefits which are realized among compatriots not only at the location level but at the industry-location one. Going one step further is what allows me to mitigate the issue of other agglomeration forces which perfectly confound with spillover ones. Moreover, I quantify which share of ethnic enclaves can be attributed to origin-specific spillovers at the industry-location level.

My chapter is also related to a strand of the urban economics literature which studies the agglomeration of plants of a same industry (e.g., Ellison and Glaeser, 1997; Maurel and Sedillot, 1999; Duranton and Overman, 2005). This research has the exact same problem as the one on immigrant locations: spillovers at the industry level are perfectly confounded with industry-specific comparative advantages for locations. For example, wine producing plants might be clustered in California to enjoy industry-specific spillover benefits but also because California is close to the vineyards which are necessary for wine producing plants. My theoretical model builds on the seminal work by Ellison and Glaeser (1997, 1999) which proposes a location choice model for firms. I use their framework to study immigrants' choices instead of firms choices and I extend their model by allowing immigrants to choose along two dimensions (industry and location) instead of only one.

The rest of this chapter is organized as follows. Section 2.2 discusses some stylized facts and the different agglomeration forces affecting the immigrants' choice on where to live and work when they move to the U.S. Section 2.3 presents and discusses the model. Section 2.4 explains the empirical implementation. Section 2.5 brings the micro-

founded index to the data and, finally, Section 2.6 concludes.

2.2 Stylized facts

2.2.1 Concentration at the MSA level

Immigrants to the U.S. tend to cluster at the same MSA with their compatriots. There exist some countries of origin which clearly display a high concentration of their migrants. For example, Figure 2.2 shows that 70% of the immigrants coming from the Dominican Republic were living in New York - Northeastern New Jersey MSA in the year 2000 according to the U.S. Census. It is remarkable to notice Boston is the second MSA with the largest concentration of immigrants from the Dominican Republic with only 7% of them. Thus, Dominicans clearly show a tendency to concentrate with their compatriots in New York - Northeastern New Jersey when moving to the U.S. On the other hand, Figure 2.3 shows that immigrants coming from Canada also concentrate with other Canadian immigrants but in a less massive way. For instance, the MSA with a largest concentration of Canadians is Los Angeles - Long Beach with only 7% of them. Therefore, immigrants tend to form ethnic enclaves at the location level. Yet, the degree of clustering varies across countries of origin.

As discussed above, immigrants might cluster at the location level due to the existence of origin-specific spillover forces at the location level, or due to origin-specific comparative advantage forces for locations. Therefore, both type of origin-specific agglomeration forces can create ethnic enclaves at the location level and we cannot tell them apart.

2.2.2 Concentration at the industry level

Immigrants to the U.S. also tend to agglomerate with their compatriots in the same industry (Mandorff, 2007; Kerr and Mandorff, 2015). For instance, Figure 2.4 shows that 29% of immigrants coming from the Philippines were working in the *Health Care and Social Assistance* industry. It is noteworthy that the second industry with the largest concentration of Filipino immigrants is *Manufacturing* with only 13% of them. Similarly,

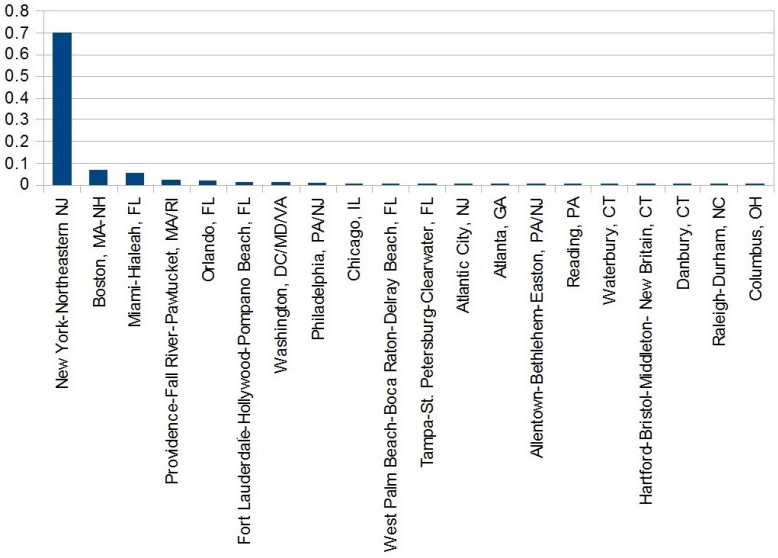

Figure 2.2: Agglomeration of immigrants from the Dominican Republic at the MSA level

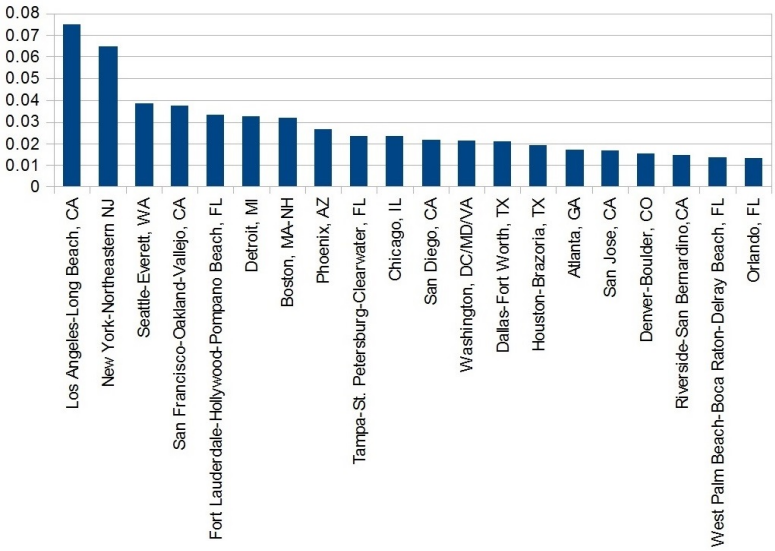

Figure 2.3: Agglomeration of immigrants from Canada at the MSA level

Figure 2.5 shows that 21 and 17% of Chinese immigrants were working in *Manufacturing* and *Accommodation and Food Services* respectively. One explanation for this agglomeration is the existence of origin-specific comparative advantages for industries. For instance, the food-service is a well established industry in China, and as a consequence, Chinese immigrants might have a comparative advantage to work in eating and drinking places. One could also think of the existence of origin-specific spillovers operating at the industry level. For example, a Chinese cook in New York might benefit from the fact that there are other Chinese cooks in Boston since the agglomeration of Chinese immigrants in eating and drinking places might bring about a greater specialization in the imports of Chinese food products to the U.S. This second reason might sound less likely to occur but should not be dismissed. Therefore, once again, there is a coexistence of origin-specific agglomeration forces operating at the same level. Origin-specific comparative advantages for industries and origin-specific spillovers operating at the industry level might create a cluster of compatriots in some industries. My chapter also takes into account these agglomeration forces at the industry level.

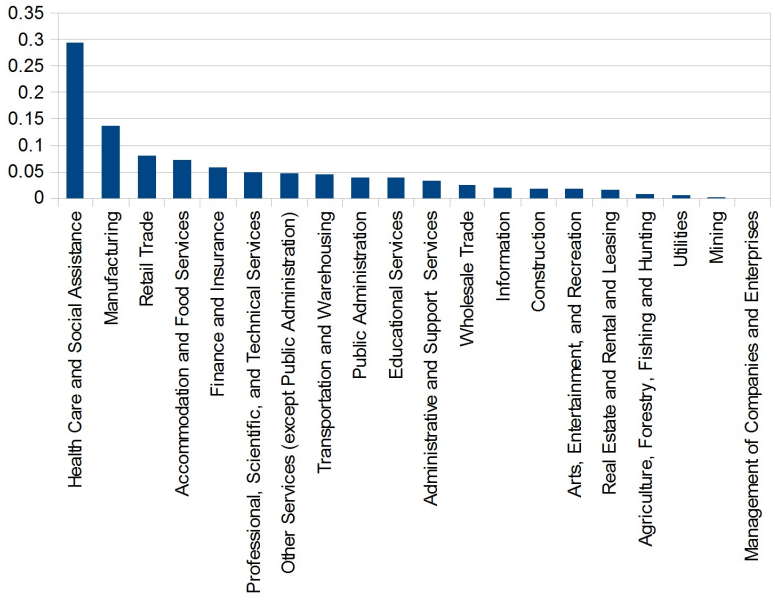

Figure 2.4: Agglomeration of Filipino immigrants at the industry level

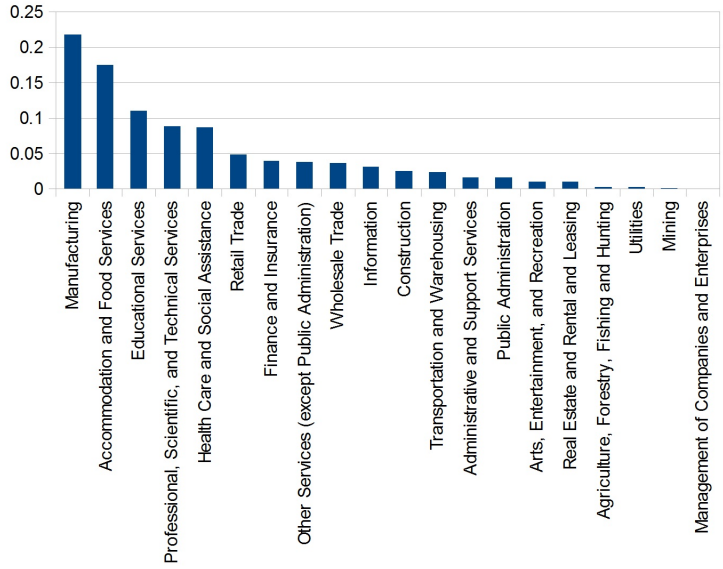

Figure 2.5: Agglomeration of Chinese immigrants at the industry level

2.2.3 Concentration at the industry-MSA level

Immigrants also agglomerate with their compatriots at the industry-MSA level. Figures 2.6 and 2.7 display a matrix representing the percentage of immigrants coming from a given country of origin and living in a particular MSA that work in a given industry. Inspection of these percentages shows that immigrants coming from the same country of origin concentrate in one industry in a MSA and to a completely different one in another MSA. At the same time, they show that immigrants coming from a different country of origin follow an opposite pattern. For instance, immigrants coming from Korea agglomerate in *Professional, scientific and technical services* in San Francisco - Oakland - Vallejo MSA and in Orlando MSA, whereas they agglomerate in *Administrative and waste management services* in Dallas - Fort Worth MSA. On the other hand, immigrants from Jamaica follow an opposite pattern.

This type of pattern is likely to reflect origin-specific spillovers operating at the industry-MSA level. For example, already established compatriots might provide valuable information about job vacancies in the firm in which they are employed to the newcomers, or provide good referrals about the newcomers to their employers. This

might generate a cluster of them at the industry-MSA level. Notice that I do not consider the existence of origin-specific comparative advantages for industry-MSA pairs. Nonetheless, I allow an industry to be better in one MSA than to another one, but this is not origin-specific. For instance, *Professional, scientific and technical services* might be a better industry in Orlando than in Dallas - Fort Worth, and represent an industry-location advantage. Yet, this would represent a comparative advantage for immigrants coming from any country of origin, or at least, for those who are good at this industry. Correspondingly, it is not likely that an average Korean immigrant has a comparative advantage to work in *Professional, scientific and technical services* in Orlando and not to work in the same industry in Dallas - Fort Worth. Yet, I allow that an average Korean immigrant has a comparative advantage to work in *Professional, scientific and technical services* in any location, and a comparative advantage to live in Orlando while working in any industry. The non-existence of origin-specific comparative advantages for industry-MSA pairs is my key identification assumption.

	Professional, scientific and technical services	Administrative and waste management services
San Francisco – Oakland – Vallejo, CA	Korea: 5% Jamaica: 0%	Jamaica: 21% Korea: 2%
Dallas – Fort Worth, TX	Jamaica: 11% Korea: 2%	Korea: 8% Jamaica: 3%
Orlando, FL	Korea: 7% Jamaica: 2%	Jamaica: 7% Korea: 0%

Figure 2.6: Agglomeration at the industry-MSA level: Korea and Jamaica

	Finance and Insurance	Real estate and rental and leasing
Atlanta, GA	Canada: 7% Taiwan: 2%	Taiwan: 5% Canada: 2%
Phoenix, AZ	Taiwan: 13% Canada: 2%	Canada: 6% Taiwan: 0%

Figure 2.7: Agglomeration at the industry-MSA level

2.2.4 Herfindahl index at the different levels

There seems to be agglomeration of immigrants to the U.S. with their compatriots at the MSA level, at the industry level and at the industry-MSA level. Following other studies on the concentration of immigrants (Jaeger, 2000, 2008), I use the Herfindahl index, a common measure of market concentration, to summarize the degree of agglomeration. I compute this index at each of the considered levels for each of the 50 top sending countries of origin. That is, I calculate $H^{level} = \sum_{l \in level} (s_l^c)^2$, where $level$ is either all the possible MSAs, or all the possible industries or all the possible industry-MSA pairs, and s_l^c is the share of immigrants from a country of origin c who are in l. Table 2.1 provides the summary statistics of these Herfindahl indexes at each dimension. On average, the largest agglomerations happen to be at the MSA level and the lowest ones at the industry-MSA level by construction. Guyana is the country with the largest concentration at the MSA with a Herfindahl index equal to 0.619, and also at the industry-MSA level with a Herfindahl index equal to 0.022. Greece is the country with the largest concentration at the industry level with a Herfindahl index equal to 0.070. Although the concentration at the industry-MSA level exhibit low numbers, there is still some agglomeration allowing to study spillovers among compatriots realized at this dimension. Accordingly, in the next section I introduce a model to study agglomeration of compatriots at the industry-MSA level while also considering agglomeration at the other dimensions.

Table 2.1: Herfindahl indeces

Level	Mean	Std. Dev.	Min.	Max.	N
MSA	0.189	0.01549	0.047	0.619	50
Industry	0.042	0.00008	0.028	0.070	50
Industry-MSA	0.009	0.00002	0.070	0.022	50

2.3 Sequential industry-location choice model

In this section, I present and discuss a sequential choice model to study the immigrants' decision on where to work and live when they move to the U.S. I assume that immi-

grants, from different countries of origin, sequentially choose an industry-location pair. I introduce several agglomeration forces affecting the immigrants' decision on where to work and live. I include origin-specific agglomeration forces for industries and for locations motivated by the stylized facts from the previous section, bringing about agglomerations of compatriots at the industry level and at the location level. These consist of spillover and comparative advantage forces. Likewise, I introduce origin-specific agglomeration forces for industry-location pairs consisting only of spillover forces.

Moreover, I also incorporate non-country specific agglomeration forces operating at the industry-location level. These include non-country specific comparative advantages for industry-location pairs and non-country specific spillovers operating at the industry-location level. By non-country specific comparative advantages I refer to the fact that an average immigrant from any country of origin might find profitable to work and live in a particular industry-location pair. For example, assume that being a taxi driver in New York were very profitable for all the individuals in general because there were lots of customers, that would attract individuals from any origin to this industry-location. By non-country specific spillovers I mean that an immigrant might benefit from working in the same industry and being located in the same location as other workers (natives or immigrants). For example, an average immigrant might benefit to be an engineer in Palo Alto if other individuals also work in the engineering in Palo Alto. Both types of non-country specific payoffs operating at the industry-location level can create a cluster of individuals from any country of origin in an industry-location pair.

Following Ellison and Glaeser (1997, 1999), I assume that immigrants have rational expectations and that the payoffs P_{ckim} that an immigrant k from country of origin c receives when choosing an industry-location pair (i, m) are given by:

$$P_{ckim} = \bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM} + g_{cim}(v_{c1}, v_{c2}, ..., v_{ck-1}) + \epsilon_{ckim} \tag{2.1}$$

where:

1. $\bar{P}_{ci}^{CI}$ is a parameter capturing the expected profitability of working in an industry i for the average immigrant from country of origin c. This profitability includes observed and unobserved origin-specific comparative advantages for an industry i and origin-specific spillovers which materialize at the industry i level.

2. $\bar{P}_{cm}^{CM}$ is a parameter capturing the expected profitability of living in a location

m for the average immigrant from country of origin c. It includes observed and unobserved origin-specific comparative advantages for a location m and origin-specific spillovers operating at the location m.

3. $\bar{P}_{im}^{IM}$ is a parameter capturing the expected profitability of working and living in an industry-location pair (i, m) for the average immigrant from any country of origin. It includes observed and unobserved non-country specific comparative advantages for an industry-location pair (i, m) and non-country specific spillovers operating at the industry-location pair (i, m).

4. $g_{cim}(v_1, v_2, ..., v_{k-1})$ is a function capturing the benefits of origin-specific spillovers operating at the industry-location level created by other immigrants from the same country of origin who have previously chosen in which industry to work and where to live. This function depends on the decisions of immigrants from the same country of origin who previously chose an industry-location pair, where v_j stands for the optimal choice of an immigrant j. I describe this function below.

5. ϵ_{ckim} is an additional independent random variable capturing idiosyncratic tastes that an immigrant k has for an industry-location pair (i, m). I assume that ϵ_{ckim} are drawn from a type-I extreme value distribution.

As a starting point, let me analyze how immigrants would take their decision in the absence of origin-specific spillovers operating at the industry-location level. To this aim, I assume that $g_{cim}(v_1, v_2, ..., v_{k-1}) = 0$. Then, the probability that an immigrant k from country of origin c chooses an industry-location (i, m) can be obtained from the random utility maximization framework introduced by McFadden (1974):

$$Prob\left(v_{ck} = (i, m)\right) = \frac{exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right)}{\sum_{i'}\sum_{m'} exp\left(\bar{P}_{ci'}^{CI} + \bar{P}_{cm'}^{CM} + \bar{P}_{i'm'}^{IM}\right)} \tag{2.2}$$

where v_{ck} stands for the optimal choice of an immigrant k from country of origin c. This probability is independent of the immigrant k but not of her country of origin c. For notational purposes, I will write this probability to be p_{cim}.

To clarify the mechanics of this simple model let me introduce a metaphor introduced by Ellison and Glaeser (1997) in which they view the mechanics of the model as

if the decision of each individual would be obtained by throwing a dart on a dartboard. In my case, the dartboard would be composed by all the industry-location pairs, and each immigrant would be seen as a dart. However, these darts would be rigged in such a way that the probability of landing on each industry-location pair (i, m), would be equal to p_{cim}. For instance, consider that there were 4 immigrants from an imaginary country called *Crossland*. Then, as Figure 2.8 shows, each immigrant can be represented by a dart, and each dart ends up on an industry-location cell with the derived probability.

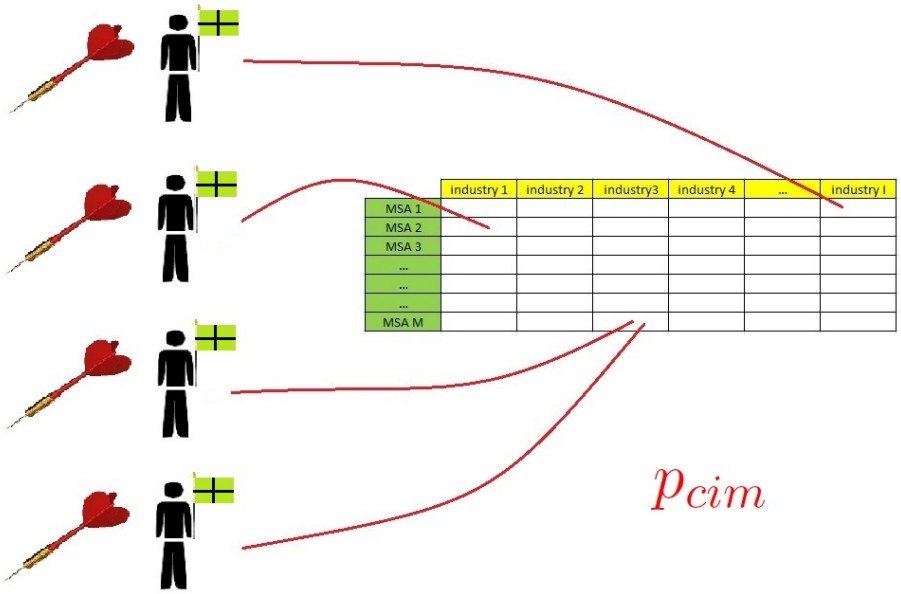

Figure 2.8: Dartboard metaphor

I can compute the expected share of immigrants from country of origin c who will choose to work and live in an industry-location pair (i, m):

$$\hat{S}_{cim} \equiv \mathbb{E}\big[S_{cim}\big] = \frac{N_c\, p_{cim}}{N_c} = p_{cim} \tag{2.3}$$

where S_{cim} stands for the share of immigrants from country of origin c who chose industry i and location m, and N_c for the number of immigrants from country of origin c who immigrated to the U.S.

Now, I introduce the origin-specific spillovers operating at the industry-location level. Specifically, following Ellison and Glaeser (1997) I assume that the function

$g_{cim}(v_{c1}, v_{c2}, ..., v_{ck-1})$ is given by:

$$g_{cim}(v_{c1}, v_{c2}, ..., v_{ck-1}) = (-\infty) \sum_{\substack{j \in c \\ j<k}} e^c_{jk}(1 - u_{jim}) \tag{2.4}$$

where j stands for an immigrant from the same country of origin c who has previously chosen location, e^c_{jk} is a Bernoulli variable equal to 1 with probability δ^c that indicates if an origin-specific spillover operating at the industry-location level exists between any pair of immigrants j and k from the same country of origin c, and u_{jim} is an indicator variable equal to 1 if immigrant j has chosen industry-location (i, m), and 0 otherwise. This is the notation that Ellison and Glaeser (1997) used for lexicographic preferences. To illustrate this, consider the first two individuals from a given country of origin who immigrate to the U.S. Assume that the first one, $j = 1$, has already chosen to work and to live in an industry-location pair (i, m), while the second, $k = 2$, is about to choose it. If there exists an origin-specific spillover between them operating at the industry-location level, then $e^c_{jk} = 1$. If immigrant k considers to go to any industry-location pair $(i', m') \neq (i, m)$, then $u_{ji'm'} = 0$, and consequently, $g_{ci'm'}(v_{cj}) = -\infty$. However, if she picks the industry-location pair (i, m), then $g_{cim}(v_{cj}) = 0$. Thus, the immigrant k will obtain a negative infinity payoff if picking an industry-location pair different than the one which j has chosen. This implies that if there exists an origin-specific spillover between two immigrants j and k, the immigrant k will end up in the exact same industry and location pair as the one that immigrant j has selected. That is, immigrants have lexicographic preferences. They first check whether they have origin-specific spillovers at the industry-location level with one of their compatriots. In that case, they choose to work and live in the same industry-location pair as this compatriot nor taking account of their idiosyncratic tastes neither of the other agglomeration forces. On the contrary, they choose an industry-location pair according to their idiosyncratic tastes and the other agglomeration forces which operate at the three pairwise interactions between country of origin, industry and location. Therefore, what determines the intensity of how frequently immigrants enjoy spillovers at the industry-location level with their compatriots is δ^c and hence, it is the main object of my analysis.

This way of formalizing the origin-specific spillovers is convenient to make the model tractable. It assumes that immigrants benefit from this particular type of spillover

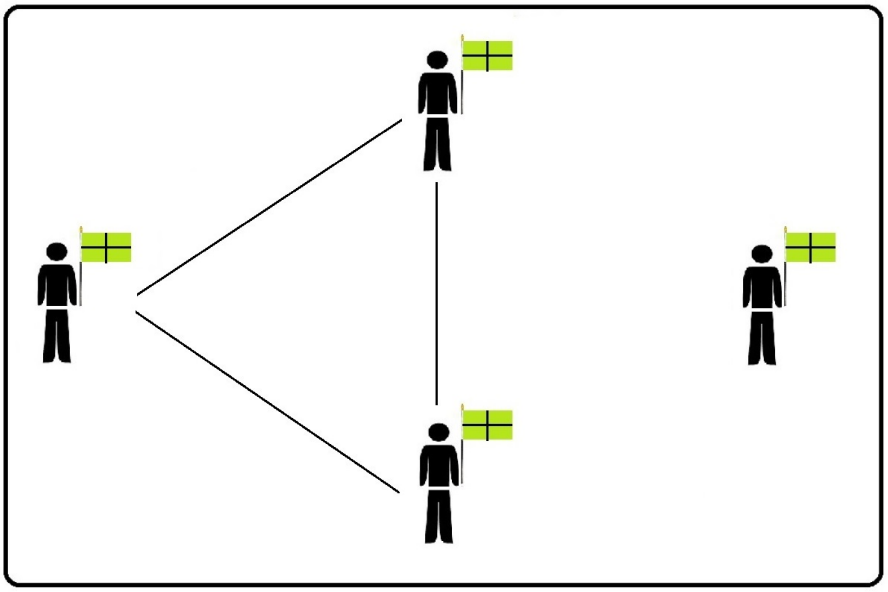

Figure 2.9: First stage

by choosing the exact same industry-location pair as the one chosen by their compatriots they have the spillover with. Besides, I also assume that the existence of spillovers is a symmetric and transitive relationship ($e^c_{jk} = 1 \Rightarrow e^c_{kj} = 1$ and $e^c_{jk} = 1,\ e^c_{kl} = 1 \Rightarrow e^c_{jl} = 1$). Ellison and Glaeser (1997) show that these assumptions, jointly with rational expectations, prevent the situation in which an immigrant would have an origin-specific spillover at the industry-location level with two immigrants who would had previously chosen different industry-location pairs. Moreover, since immigrants have rational expectations, the immigrants' order of choosing an industry-location pair does not matter for the final distribution of immigrants across industry-location pairs.

Similar to Ellison and Glaeser (1997) one can view this model to be equivalent to a two-stage process in which first, there is a stochastic process depending on the probability δ^c such that groups immigrants from the same country of origin in clusters, and then, each of these clusters is assigned to an industry-location pair with probability p_{cim}. For example, going back to the example of the four immigrants from *Crossland*, imagine that in the first stage the stochastic process groups three of them in a cluster as shown in Figure 2.9. Then, in the second stage, the cluster is thrown to the dartboard as a single dart and the fourth one is thrown alone as displayed in Figure 2.10

In the full-fledged model, i.e. including the origin-specific spillovers operating at

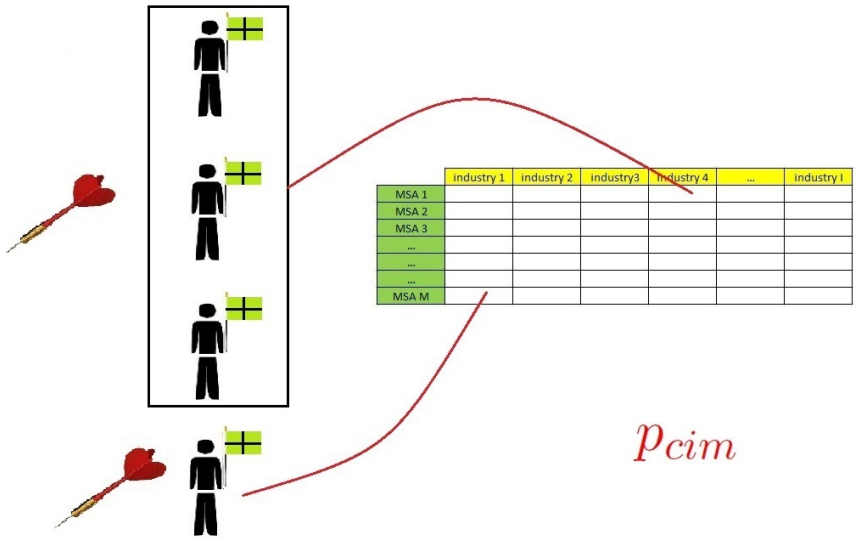

Figure 2.10: Second stage

the industry-location level as well as the other agglomeration forces, the expected share of immigrants from country of origin c who work and live in an industry-location pair (i, m) is:

$$\hat{S}_{cim} \equiv \mathbb{E}\big[S_{cim}\big] = \mathbb{E}\left[\sum_{k \in c} \frac{u_{kim}}{N_c}\right] = \sum_{k \in c} \frac{\mathbb{E}\big[u_{kim}\big]}{N_c} = \sum_{k \in c} \frac{p_{cim}}{N_c} = p_{cim} \tag{2.5}$$

where once again u_{kim} is an indicator variable equal to 1 if immigrant k has chosen industry-location (i, m), and 0 otherwise. To obtain Equation 2.5 I used $\mathbb{E}\left(u_{kim}\right) = Prob\left(u_{kim} = 1\right) = p_{cim}$. The probability that an individual ends up in an industry-location pair (i, m) is independent of δ^c. Intuitively, immigrants who do not belong to a cluster end up in (i, m) with probability p_{cim}. Likewise, if they belong to a cluster, they also end up with same probability p_{cim} in cell (i, m). Therefore, the version of the model without origin-specific spillovers operating at the industry-location level and the full-fledged version deliver the exact same expected share of immigrants in each industry-location cell. Yet, the two versions of the model deliver different variances of the final distribution of S_{cim}. This is illustrated below by means of a simple example.

For instance, I consider a very simple case in which there are only two industries,

two locations and two immigrants from a country of origin c. Moreover, I assume that $\bar{P}_{ci}^{CI} = \bar{P}_{cm}^{CM} = \bar{P}_{im}^{IM} = 1 \; \forall i$ and $\forall m$, and that there do not exist origin-specific spillovers operating at the industry-location level, i.e. $\delta^c = 0$. In this hypothetical situation, the possible allocations of immigrants across industry-location pairs are depicted in Figure 2.11. Each matrix represents a possible allocation of immigrants across the four industry-location pairs, and each element of the matrix stands for the number of immigrants from country c that are in each industry-location pair. For example, in the upper-left matrix, the two immigrants from country of origin c chose industry-location pair (i_1, m_1). The number close to the arrow is the probability that this allocation is the realized one given the assumed parameters. Therefore, the expected share of immigrants from country of origin c who are in any industry-location pair (i, m) is $\hat{S}_{cim} = \frac{1}{4}$.

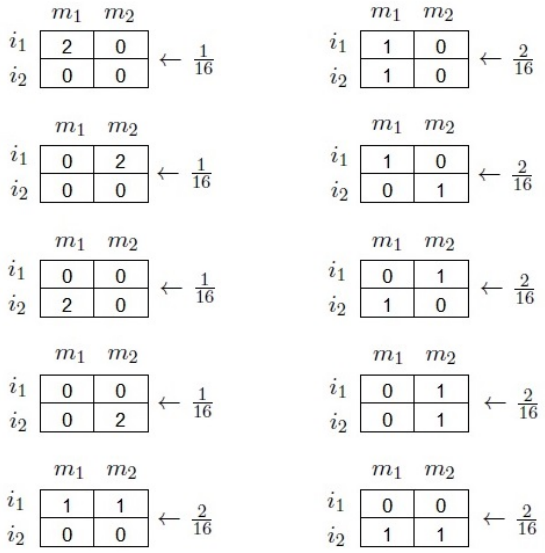

Figure 2.11: Distribution with $\delta^c = 0$

Alternatively, I now assume the other extreme case in which origin-specific spillovers operating at the industry-location level are very strong such that $\delta^c = 1$. In this case, the two immigrants are always agglomerated at the same industry-location pair and the possible allocations across industry-location pairs are depicted in Figure 2.12. Once again, the expected share of immigrants from country of origin c who are in any industry-location pair (i, m) is $\hat{S}_{cim} = \frac{1}{4}$.

In these two extreme considered cases, $\hat{S}_{cim} = \frac{1}{4}$ independently of the strength of

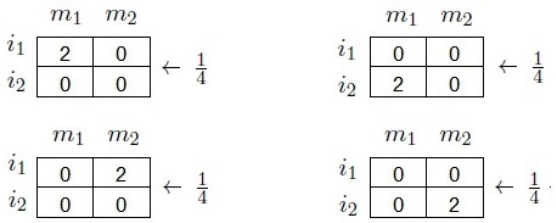

Figure 2.12: Distribution with $\delta^c = 1$

origin-specific spillovers operating at the industry-location level δ^c. However, $Var(S_{cim})$ changes with δ^c and it is higher the higher δ^c. In this example, $Var(S_{cim}) = 0.4375$ when $\delta^c = 0$ and $Var(S_{cim}) = 0.8125$ when $\delta^c = 1$. Intuitively, when origin-specific spillovers at the industry-location level are the norm rather than the exception, immigrants coming from the same country of origin are clustered, and consequently, the variance of the final distribution is big. Therefore, one can learn something about the strength of this particular type of origin-specific spillover by exploring a measure that takes into account the second moment of the distribution of the shares of immigrants from a country of origin in a particular industry-location pair. To this aim and following Ellison and Glaeser (1999), I consider the following concentration measure: $G^{CIM} \equiv \sum_i \sum_m \left(S_{cim} - \hat{S}_{cim}\right)^2$. The expected value of G^{CIM} provides a relationship between the shares of immigrants from a country of origin c who are located in an industry-location (i, m) and the strength of origin-specific spillovers operating at the industry-location level. I characterize this relationship in the following proposition.

Proposition 3. *If immigrants choose location accordingly to the described model,*

$$\mathbb{E}\left[G^{CIM}\right] = \frac{1}{N_c}\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right) + \delta^c \left(1 - \frac{1}{N_c}\right)\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right) \quad (2.6)$$

Proof in Appendix A.

From Equation 2.6, I can isolate δ^c:

$$\delta^c = \frac{\mathbb{E}\left[G^{CIM}\right] - \frac{1}{N_c}\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right)}{\left(1 - \frac{1}{N_c}\right)\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right)} \tag{2.7}$$

Given that G^{CIM} enters linearly in the numerator of Equation 2.7, I propose an unbiased estimator of δ^c in the following corollary.

Corollary 2. *An unbiased estimator of the importance of origin-specific spillovers operating at the industry-location level δ^c is given by:*

$$\hat{\delta}^c = \frac{G^{CIM} - \frac{1}{N_c}\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right)}{\left(1 - \frac{1}{N_c}\right)\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right)} \tag{2.8}$$

and plugging G^{CIM}:

$$\hat{\delta}^c = \frac{\frac{\sum_i \sum_m \left(S_{cim} - \hat{S}_{cim}\right)^2}{\left(1 - \sum_i \sum_m \hat{S}^2_{cim}\right)} - \frac{1}{N_c}}{1 - \frac{1}{N_c}} \tag{2.9}$$

The parameter δ^c can be interpreted as the fraction of pairs of immigrants from the same country of origin who benefit from spillovers which materialize among compatriots at the industry-location level. Notice that $\hat{\delta}^c$ is an unbiased estimator of δ^c. Positive values of the index should be interpreted as displaying a clustering of compatriots at the industry-location level beyond that expected by agglomeration forces operating at the three pairwise interactions between country of origin, industry and location. Similarly, negative values should be interpreted as displaying a distribution of compatriots more diluted than what should be expected by the agglomeration forces operating at the three pairwise interactions.

2.4 Empirical implementation

2.4.1 Empirical strategy

To obtain the unbiased estimator $\hat{\delta^c}$ of Equation 2.9 for every country, I need data on the shares of immigrants from every country of origin c who are in each industry-location pair, S_{cim} $\forall(i,m)$, and the number of immigrants from each country of origin, N_c $\forall c$; I also need to obtain $\hat{S}_{cim}$. Recall that in the described model:

$$\hat{S}_{cim} = p_{cim} = \frac{exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right)}{\sum_{i'}\sum_{m'} exp\left(\bar{P}_{ci'}^{CI} + \bar{P}_{cm'}^{CM} + \bar{P}_{i'm'}^{IM}\right)} \tag{2.10}$$

where p_{cim} is the probability that an immigrant k from country of origin c chooses the industry-location pair (i,m). This probability depends on the agglomeration forces operating at the three pairwise interactions between country of origin, industry and location. One of the advantages of studying agglomeration of immigrants at the triple interaction is that we can proxy any payoff operating at the three pairwise interactions ($\bar{P}_{ci}^{CI}$, $\bar{P}_{cm}^{CM}$ and $\bar{P}_{im}^{IM}$) making use of fixed effects. Therefore, Equation 2.10 can be estimated using a conditional logit model where $\bar{P}_{ci}^{CI} \equiv \beta_{ci}^{CI}\gamma_{ci}^{CI}$ where γ_{ci}^{CI} is a dummy variable equal to 1 if an immigrant is from country of origin c and is working in industry i and β_{ci}^{CI} is a parameter to be estimated. I analogously define $\bar{P}_{cm}^{CM}$ and $\bar{P}_{im}^{IM}$. The corresponding log-likelihood is:

$$logL_{CL} = \sum_k \sum_i \sum_m d_{kcim}\, log(p_{cim}) = \sum_c \sum_i \sum_m N_{cim}\, log(p_{cim}) \tag{2.11}$$

where d_{ckim} is a dummy equal to one if immigrant k is from country of origin c and has picked an industry-location pair (i,m) and N_{cim} is the number of immigrants from country of origin c who have chosen an industry-location pair (i,m). Estimating this conditional logit model is a calculation intensive task and can present some problems given the large number of industry-location pairs. However, following Guimaraes et al. (2003) and Schmidheiny and Brulhart (2011), I show that there is an equivalence between estimating this conditional logit and a Poisson regression in the following lemma.

Lemma 1. *It is equivalent to estimate the conditional logit model*

$$p_{cim} = \frac{exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right)}{\sum_{i'}\sum_{m'} exp\left(\bar{P}_{ci'}^{CI} + \bar{P}_{cm'}^{CM} + \bar{P}_{i'm'}^{IM}\right)} \tag{2.12}$$

and a Poisson model in which N_{cim} is independently Poisson distributed with,

$$\mathbb{E}\left[N_{cim}\right] = exp\left(\bar{\gamma}_{ci}^{CI} + \bar{\gamma}_{cm}^{CM} + \bar{\gamma}_{im}^{IM}\right) \tag{2.13}$$

where $\bar{\gamma}_{ci}^{CI}$, $\bar{\gamma}_{cm}^{CM}$ and $\bar{\gamma}_{im}^{IM}$ are fixed effects at the country-industry, country-location and industry-location levels respectively.

Proof in Appendix A.

Therefore, I can obtain $\hat{S}_{cim}$ using a Poisson regression model. Then, using S_{cim}, N_c and $\hat{S}_{cim}$ I can obtain $\hat{\delta^c}$. Nonetheless, I need to provide a measure of statistical significance to be in position to say something about non-zero values of $\hat{\delta^c}$. To deal with this issue, in the next sub-section I develop a bootstrap methodology.

2.4.2 Confidence interval for my index

Following Cassey and Smith (2014) I propose a way of computing confidence intervals for my index. In particular, I simulate many times the allocation of immigrants across the U.S. assuming that origin-specific spillovers operating at the industry-location level are turned off, $\delta^c = 0$, and that immigrants choose an industry-location pair taking into account only the other agglomeration forces and their idiosyncratic tastes. After each simulation I compute $\hat{\delta^c}$. Since I turned off δ^c to 0, each realization of $\hat{\delta^c}$ is due to the other agglomeration forces and idiosyncratic tastes. This allows me to construct a confidence interval for each $\hat{\delta^c}$. Specifically, I adopt the following kind of bootstrap methodology:

1. For any country of origin, I take the estimations of the Poisson model of Equation 2.13, and I compute the estimated $\hat{p}_{cim}$. I also take the number of immigrants

N_c from each country of origin to match the ones in the data.

2. I randomly allocate each of the immigrants from each country of origin c to one of the possible industry-location pairs with probability $\hat{p}_{cim}$.

3. Then, I compute the actual share of immigrants from each country of origin in each industry-location pair, S_{cim}. Since I also have $\hat{S}_{cim} = \hat{p}_{cim}$ and N_c, I can obtain $\hat{\delta}^c$ according to Equation 2.9.

4. Finally, I repeat steps 2, and 3 for 10,000 times.

Given that I assume that there are no origin-specific spillovers operating at the industry-location level, each realization of $\hat{\delta}^c$ is due to the other agglomeration forces operating at the three pairwise interactions and to the idiosyncratic tastes. Hence, the expected value of $\hat{\delta}^c$ equals zero. Ordering the 10,000 $\hat{\delta}^c$, I can obtain the critical values for the interval containing the 95% of the observations. This allows me to test whether the estimated index for each country of origin is statistically different from zero.

2.5 Estimation

2.5.1 Data

The order in which immigrants choose an industry-location pair is not important in my model. As a consequence, I use the database that provides the most complete picture of how immigrants are distributed across industry-location pairs in the U.S. I use the 2000 U.S. Census, five percent sample, provided by the IPUMS-USA. This is the latest data available from a decennial Census. After this year, several rounds of the American Community Survey (ACS) have been conducted every year. However, the ACS interviews a much more reduced number of individuals than the Census.

The U.S. Census includes information about foreign born individuals independently of their legal status. Therefore, illegal immigrants are also included in the data. I restrict the sample to employed immigrants who were at least 24 years old at their arrival to the U.S. to exclude those immigrants who got their education in the U.S.

To compute the unbiased index of Equation 2.9, I only need data on where immigrants who are in the U.S. were born, in which industry they work and in which location they live. The Census provides such information. Employed immigrants are classified in 89 industries at the 3-digit NAICS 1997 classification level plus *Construction* which is only available at the 2-digit NAICS 1997 level.[4] I estimate my index at the MSA level since these are considered to be good approximations to labor markets (Bartel, 1989). I restrict my sample to the 40 top MSAs in which employed immigrants live and to the 50 top sending countries. This selected sample includes 300,177 employed individuals who make up for the 75% of all the employed immigrants in the Census who arrived with more than 24 years old and for whom there is information at the required levels. Appendix B reports the countries, industries and MSAs included in the analysis.

2.5.2 Origin-specific spillovers at the industry-MSA level

Using the MSA as the location variable, I first estimate the Poisson model of Equation 2.13.[5] With these estimates, I compute $\hat{S}_{cim}$. Finally, using S_{cim}, N_c and $\hat{S}_{cim}$ I obtain $\hat{\delta^c}$. The estimation of this index, $\hat{\delta^c}$, can be interpreted as the estimated fraction of pairs of immigrants from the same country of origin who benefit from spillovers operating at the country of origin-industry-MSA level.

Tables 2.2 and 2.3 provide the estimates of my spillovers index. The first column displays the country of origin, the second one the number of immigrants from each country of origin included in the sample, the third one the estimated $\hat{\delta^c}$, and the fifth one whether it is statistical significant or not at the following levels: * 90%, ** 95%, and *** 99%. The index is statistically significant for 42 countries of origin. The country with a highest statistically significant index is Cambodia with $\hat{\delta}^{KHM} = 0.140\%$. This means that 0.140% of the possible pairs of immigrants from Cambodia benefit from this specific type of spillover. Given that there are 1,113 Cambodian immigrants in the sample, there exist $\binom{1,113}{2} = 618,828$ possible pairs, and the index estimates that 867 pairs of Cambodians benefit from spillovers at the industry-MSA level among

[4] *Construction* which is an industry at the 2-digits NAICS 1997 contains two industries at the 3-digits NAICS 1997 level: *Prime Contracting* and *Trade Contracting*.

[5] To this aim, I use the Poisson estimation technique described in Santos Silva and Tenreyro (2006).

them. Nonetheless, how many immigrants coming from Cambodia are needed to form 867 pairs? On the one hand, all the 1,113 individuals might be needed to form these pairs. On the other, these pairs could be formed with approximately 40 individuals. Accordingly, I propose an alternative way of interpreting the magnitude of the index by computing the minimum number of individuals needed to form the pairs who benefit from spillovers. I illustrate this below.

To show why this is an intuitive measure, recall that δ^c is the probability that any pair of immigrants have a spillover between them, and that the existence of spillovers is a transitive and symmetric relation. Then, consider a case in which there are six immigrants from *Crossland* and three pairs of them have a spillover. A minimum of three immigrants are needed to form three pairs; in this case, only three immigrants benefit from spillovers as in Figure 2.13. Yet, it might also be that all six immigrants have a spillover with another immigrant as in Figure 2.14. I can compute the minimum number of immigrants from that country of origin, x^c, who benefit from these spillovers by solving the following equation:

$$\binom{x^c}{2} = \delta^c \binom{N_c}{2} \tag{2.14}$$

such that $x^c = \frac{1+\sqrt{1+8\delta^c\binom{N_c}{2}}}{2}$. I define $\rho^c \equiv \frac{x^c}{N_c}$ to be the minimum share of immigrants from country of origin c who benefit from these spillovers. As an illustrative example, if $\hat{\delta}^{KHM} = 0.140\%$ and $N_{KHM} = 1,113$ individuals, this means that at least 42 immigrants from Cambodia benefit from these spillovers, namely 3.788% of the total. This minimum share ρ^c of immigrants who benefit from spillovers at the industry-MSA level among compatriots is reported in the fourth column of Tables 2.2 and 2.3. On average across countries of origin, at least 1.553% of the immigrants in the sample benefit from this specific type of spillover.

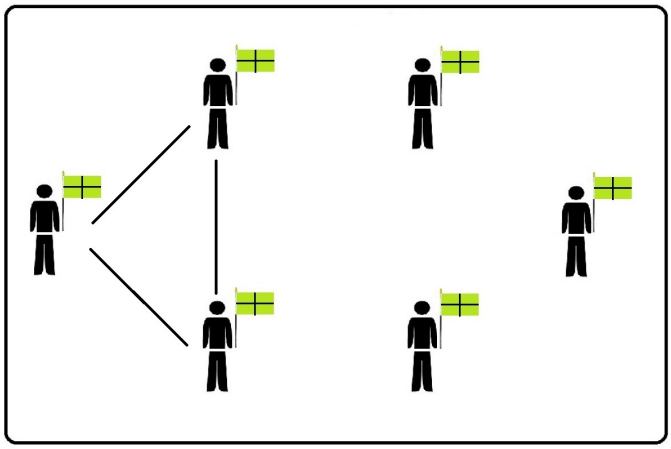

Figure 2.13: Three pairs composed of three people

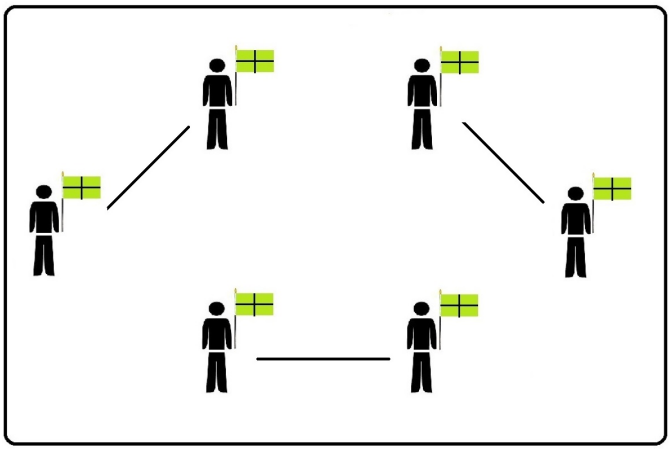

Figure 2.14: Three pairs composed of six people

Table 2.2: Index estimates I

Country of origin	N_c	$\hat{\delta}^c$	ρ^c	Stat. Significance
Cambodia	1113	0.140%	3.788%	***
Brazil	3141	0.116%	3.424%	***
Haiti	6187	0.082%	2.879%	***
Portugal	1744	0.074%	2.755%	***
Poland	6359	0.066%	2.577%	***
Bangladesh	1162	0.059%	2.470%	***
Korea	11317	0.058%	2.404%	***
Nigeria	2177	0.051%	2.280%	***
Yugoslavia	2498	0.046%	2.157%	***
China	16489	0.045%	2.129%	***
Ireland	1594	0.044%	2.124%	***
Pakistan	2720	0.043%	2.102%	***
Honduras	3172	0.040%	2.018%	***
El Salvador	8956	0.035%	1.876%	***
Japan	4361	0.033%	1.833%	***
Thailand	1398	0.029%	1.731%	***
Ghana	1135	0.028%	1.712%	**
Romania	2019	0.028%	1.687%	***
Israel	1370	0.025%	1.603%	**
Lebanon	1140	0.024%	1.595%	***
Russia	4688	0.023%	1.518%	***
Laos	1078	0.022%	1.545%	***
Greece	1419	0.021%	1.476%	*
Vietnam	11203	0.019%	1.385%	***
India	15130	0.019%	1.368%	***

Table 2.3: Index estimates II

Country of origin	N_c	$\hat{\delta^c}$	ρ^c	Stat. Significance
Italy	2991	0.018%	1.344%	***
Taiwan	5143	0.017%	1.306%	***
Egypt	1765	0.017%	1.323%	**
Philippines	22008	0.015%	1.231%	***
Ecuador	3676	0.015%	1.226%	**
Colombia	7515	0.014%	1.196%	***
Iran	3804	0.013%	1.175%	***
Guatemala	4894	0.013%	1.154%	***
Canada	6091	0.013%	1.139%	***
Venezuela	1238	0.012%	1.120%	
Ukraine	3607	0.012%	1.093%	**
Peru	4536	0.011%	1.041%	***
Hong Kong	2073	0.010%	1.046%	*
United Kingdom	6902	0.009%	0.973%	***
France	1665	0.008%	0.950%	
Jamaica	7922	0.008%	0.905%	***
Mexico	58960	0.007%	0.842%	***
Germany	4161	0.007%	0.829%	***
Puerto Rico	6701	0.003%	0.580%	**
Argentina	1918	0.003%	0.593%	
Nicaragua	2897	0.002%	0.485%	
Cuba	12348	0.001%	0.336%	
Dominican Republic	7977	0.000%	0.220%	
Trinidad and Tobago	2693	-0.001%		
Guyana	3122	-0.013%		

2.5.3 How does the index compare to agglomeration forces at the MSA level?

The immigration literature establishes that immigrants tend to cluster with their compatriots at the location level when moving to the U.S. They argue that immigrants might benefit from being collocated with their compatriots at the same location. For instance, compatriots might provide financial aid or valuable information of affordable housing or make less traumatic the process of migrating. Among these, immigrants might also benefit from being clustered with their compatriots since there might exist job related spillovers. For example, they might provide valuable information about job vacancies.

However, it is difficult to attribute the agglomeration of immigrants at the location level to these spillovers since they could be the result of origin-specific comparative advantages for locations. Yet, my index estimates the strength of a subset of these spillover forces which are not only realized at the location level but at the industry-location one. For example, immigrants might provide information about job vacancies to their compatriots in the same industry in which they work, they might hire a compatriot in their own business, or they might assist them in establishing a business in the same industry (since they have information about the needed procedures, about business opportunities, or contacts about good co-ethnics workers, etc.). Therefore, the spillovers which I consider in this chapter are a subset of the ones which the literature has recognized.

Accordingly, in this sub-section I compute which percentage of origin-specific agglomeration forces at the location level are accounted by origin-specific spillovers at the industry-location level. To this aim, I adapt the standard Ellison and Glaeser (1997) index to measure origin-specific agglomeration forces at the location level. That is, for each country of origin I compute:

$$\widehat{EG^c} = \frac{\frac{\sum_m (S_{cm} - x_m)^2}{\left(1 - \sum_m x_m^2\right)} - \frac{1}{N_c}}{1 - \frac{1}{N_c}} \tag{2.15}$$

where S_{cm} stands for the share of immigrants from country of origin c who are located in MSA m, and x_m stands for the share of immigrants coming from any country of origin who are located in MSA m. This index captures the strength of origin-specific agglomeration forces for locations. These forces can either be origin-specific spillovers

or origin-specific comparative advantages. This adapted index can be interpreted as the probability that any pair of immigrants from the same country of origin are agglomerated at the same MSA due to origin-specific agglomeration forces at the MSA level.

Tables 2.4 and 2.5 shows how my index compares to this adapted index. The second column displays the estimates for my index, i.e., the exact same estimates shown in the previous sub-section ($\hat{\delta}^c$). Column 3, in its turn, displays the estimates corresponding to the adapted Ellison and Glaeser (1997) index ($\widehat{EG}^c$). To ease the interpretation of these indeces, column 4 reports the minimum share of immigrants who benefit from origin-specific spillovers at the industry-MSA level (ρ^c). Likewise, column 5 shows the minimum share of immigrants who are agglomerated at the same MSA due to origin-specific agglomeration forces at the MSA level (ρ^c_{EG}). Finally, column 6 displays the share of the minimum number of individuals who are agglomerated at the MSA due to origin-specific agglomeration forces that can be accounted for origin-specific spillover forces at the industry-location level ($\frac{\rho^c}{\rho^c_{EG}}$). These numbers in the sixth column are quite large. For instance, this share equals to 15.062% for Cambodia. On average across countries, this share equals to 7.114%. Hence, I estimate that at least 7.114% of the agglomeration of immigrants from the same country of origin at the MSA level is driven by origin-specific spillovers at the industry-MSA level.

Table 2.4: Index estimates vs. agglomeration forces at the MSA level I

Country of origin	$\hat{\delta}^c$	$\widehat{EG^c}$	ρ^c	ρ^c_{EG}	Share
Cambodia	0.140%	6.308%	3.788%	25.150%	15.062%
Brazil	0.116%	5.642%	3.424%	23.765%	14.408%
Haiti	0.082%	11.528%	2.879%	33.959%	8.477%
Portugal	0.074%	13.193%	2.755%	36.340%	7.582%
Poland	0.066%	16.806%	2.577%	41.000%	6.285%
Bangladesh	0.059%	16.392%	2.470%	40.512%	6.097%
Korea	0.058%	2.973%	2.404%	17.246%	13.942%
Nigeria	0.051%	4.556%	2.280%	21.363%	10.671%
Yugoslavia	0.046%	3.973%	2.157%	19.949%	10.814%
China	0.045%	2.589%	2.129%	16.094%	13.226%
Ireland	0.044%	4.637%	2.124%	21.559%	9.853%
Pakistan	0.043%	3.430%	2.102%	18.535%	11.339%
Honduras	0.040%	2.466%	2.018%	15.717%	12.842%
El Salvador	0.035%	7.797%	1.876%	27.927%	6.718%
Japan	0.033%	1.934%	1.833%	13.916%	13.170%
Thailand	0.029%	4.387%	1.731%	20.974%	8.251%
Ghana	0.028%	10.861%	1.712%	32.986%	5.191%
Romania	0.028%	1.915%	1.687%	13.861%	12.169%
Israel	0.025%	1.328%	1.603%	11.556%	13.876%
Lebanon	0.024%	4.149%	1.595%	20.404%	7.816%
Russia	0.023%	5.364%	1.518%	23.169%	6.554%
Laos	0.022%	9.767%	1.545%	31.285%	4.938%
Greece	0.021%	5.922%	1.476%	24.362%	6.060%
Vietnam	0.019%	7.413%	1.385%	27.230%	5.087%
India	0.019%	2.339%	1.368%	15.296%	8.942%

Table 2.5: Index estimates vs. agglomeration forces at the MSA level II

Country of origin	$\hat{\delta}^c$	$\widehat{EG^c}$	ρ^c	ρ^c_{EG}	Share
Italy	0.018%	7.392%	1.344%	27.200%	4.943%
Taiwan	0.017%	3.749%	1.306%	19.371%	6.742%
Egypt	0.017%	4.174%	1.323%	20.454%	6.467%
Philippines	0.015%	3.535%	1.231%	18.802%	6.546%
Ecuador	0.015%	27.420%	1.226%	52.371%	2.341%
Colombia	0.014%	7.835%	1.196%	27.995%	4.273%
Iran	0.013%	11.334%	1.175%	33.675%	3.488%
Guatemala	0.013%	9.217%	1.154%	30.366%	3.801%
Canada	0.013%	3.018%	1.139%	17.378%	6.556%
Venezuela	0.012%	11.095%	1.120%	33.336%	3.359%
Ukraine	0.012%	6.511%	1.093%	25.526%	4.280%
Peru	0.011%	2.917%	1.041%	17.088%	6.092%
Hong Kong	0.010%	4.486%	1.046%	21.199%	4.933%
United Kingdom	0.009%	1.402%	0.973%	11.847%	8.214%
France	0.008%	1.025%	0.950%	10.152%	9.356%
Jamaica	0.008%	12.630%	0.905%	35.543%	2.547%
Mexico	0.007%	6.667%	0.842%	25.822%	3.262%
Germany	0.007%	1.673%	0.829%	12.943%	6.406%
Puerto Rico	0.003%	6.683%	0.580%	25.857%	2.243%
Argentina	0.003%	1.813%	0.593%	13.487%	4.398%
Nicaragua	0.002%	22.902%	0.485%	47.865%	1.012%
Cuba	0.001%	49.277%	0.336%	70.199%	0.479%
Dominican Republic	0.000%	33.156%	0.220%	57.584%	0.383%
Trinidad and Tobago	-0.001%	20.867%		45.691%	
Guyana	-0.013%	39.588%		62.925%	

2.5.4 Variation of the estimated index acroos countries

There is substantial variation of the estimated index across countries of origin. Consequently, I examine next whether the estimated index is related to some characteristics of the country of origin. First, I start by constructing a variable measuring the cultural distance from each country of origin to the U.S. I characterize the culture of each country of origin with the answers to four questions asked in the World Value Survey, following Tabellini (2008, 2010). The World Value Survey interviews a representative national sample of at least 1,000 individuals in many countries in order to gather comparable data on people's values, beliefs and attitudes. Tabellini (2008, 2010) argues that a culture can be described using the answers of four questions included in this survey which measure (i) generalized trust, (ii) tolerance and respect for other people as a desirable quality that children should be encouraged to learn, (iii) obedience as another quality that children should have, and (iv) a feel of free choice and control over own life. To date, six different waves have been published: 1981, 1990, 1995, 2000, 2005 and 2010. I aggregate all the waves and compute a country average for each of these questions creating a vector of 4 traits characterizing the culture of each country. Then, I measure the Euclidean distance from each country of origin's vector to the one from the U.S. Appendix C reports summary statistics for all the variables I use in this sub-section and a detailed description of them.

Figure 2.15 plots the relationship between cultural distance to the U.S. and the logarithm of the estimated spillovers index $\hat{\delta}^c$. The green dots stand for those countries of origin such that the estimated index is statistically significant, and the maroon crosses for the rest. I use this notation for the rest of this section. There is a positive and statistically significant relationship between the two variables. Immigrants coming from countries of origin that are more culturally distant from the U.S., tend to agglomerate more at the industry-MSA level with their compatriots. A possible interpretation of this relationship is that being more culturally distant might entail more difficulties in interacting with natives once in the U.S. In such a case, immigrants might find it beneficial to be clustered with their compatriots. For instance, this can facilitate the emergence of knowledge-spillovers between them. Moreover, compatriots might help in the search of a job or assist them in creating a business in the same industry.

Figure 2.15: Cultural distance

Along the same lines, I evaluate next if the willingness to be clustered with compatriots at the industry-MSA level is correlated with the immigrants' levels of trust towards foreigners. Immigrants who cannot trust individuals from other countries of origin (including U.S. natives) might find it advantageous to be clustered with their compatriots. Combining my estimates with trust data from the World Value Survey, I test this hypothesis. In particular, I use the last two waves (2005 and 2010) which include the following question: "I'd like to ask you how much you trust people from various groups. Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all?". Among the groups included, could be found "People from another nationality". I create a variable that equals the country average of the answers to this question and which are set to 0, 1, 2 and 3 if people respond "not at all", "not very much", "somewhat" and "completely trust", respectively. Figure 2.16 displays the relationship between this measure of how each nationality can trust foreigners and the spillovers index. This relationship is negative and statistically significant. This result is aligned with the social capital literature which sees the capacity to trust others as a social lubricant at the workplace (La Porta et al., 1997). Hence, immigrants who can trust non-compatriots might not need to be as clustered as the ones who cannot.

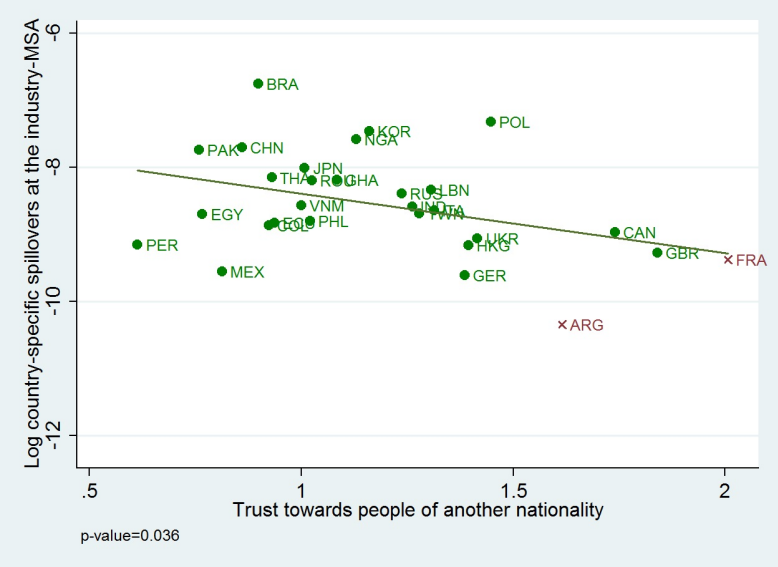

Figure 2.16: Trust towards other nationalities

Other dimensions that could play a role in explaining why the immigrants from some countries of origin find it profitable to cluster with their compatriots at the industry-MSA level are the language distance, the age at arrival and the level of education. Accordingly, I analyze below how each of these dimensions correlates with the estimated index of spillovers at the industry-MSA level among compatriots. First, I show the relationship between a measure of how distant is the language spoken in each country of origin from the English and the spillovers index. The language spoken in each country is obtained from the World Factbook (CIA) and the language distance from the English from Chiswick and Miller (2004). Figure 2.17 displays the relationship between the logarithms of the estimated index and language distance. There is a positive correlation between the two variables, but it is only statistically significant with an 83.3% confidence level. This suggests that immigrants from countries of origin in which the spoken language is more distant from the English find it more beneficial to agglomerate at the industry-MSA level with their compatriots when moving to the U.S. For example, Japanese is very different from the English, and for immigrants coming from Japan might be very important to be with their compatriots at the workplace in order to better communicate and be more productive. This result is in line with Hellerstein et al. (2008) which find that approximately one third of workplace racial segregation in the U.S. is

due to language proficiency. In a similar vein, Glitz (2014) shows that racial workplace segregation in Germany, both from natives and also from other ethnic groups, is higher if workers do not share a common language.

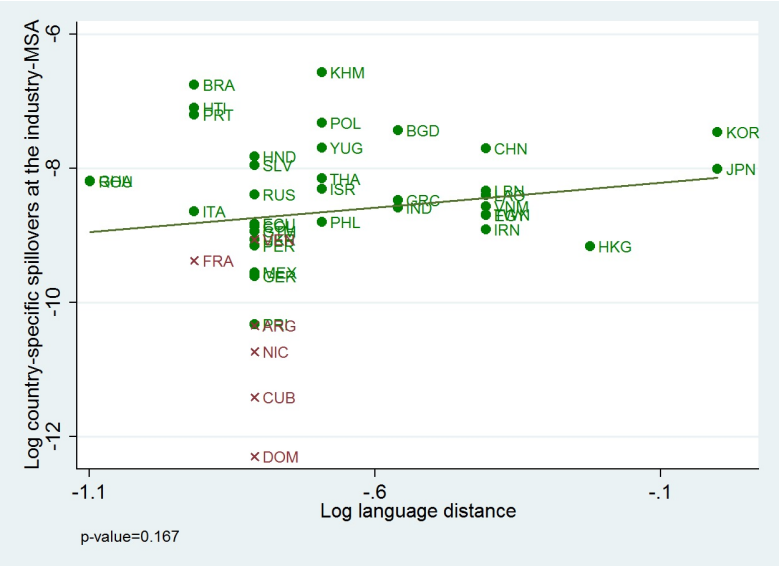

Figure 2.17: Language distance

The next dimension I check is the average age at arrival to the U.S. for each country of origin. In this case, I compute the average age at arrival of the immigrants included in the sample, for each country of origin. Figure 2.18 displays the relationship between the average age at arrival and the logarithm of the estimated index per each country. The relationship is negative but only statistically significant with an 88.8% confidence level. One possible interpretation for this relationship might be that when immigrants arrive to the U.S. at their late thirties, they might have acquired some work experience in a specific industry. This experience might help them in obtaining a job and they might not need to resort to their compatriots. Along the same lines, Ioannides and Datcher Loury (2004) review evidence that shows that the use of informal contacts to obtain a job decreases with age or work experience.

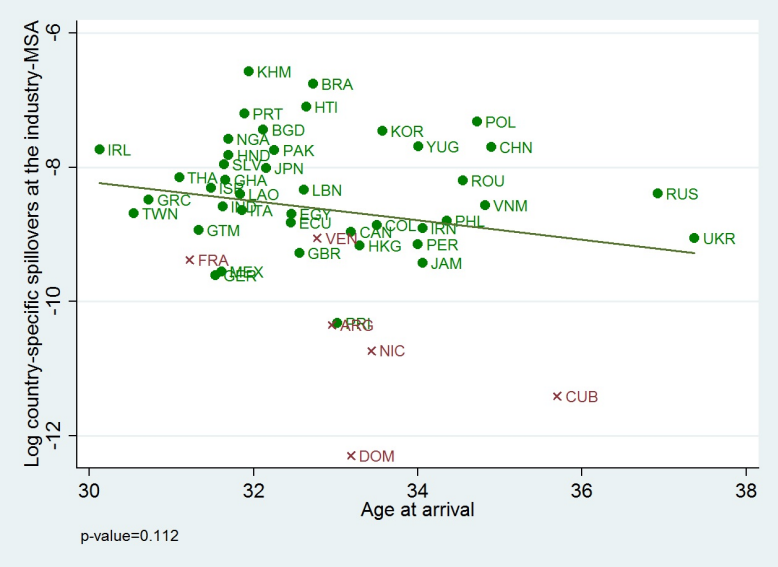

Figure 2.18: Age at arrival

Finally, I examine the relationship between the level of education and the propensity to agglomerate with their compatriots. There are two competing hypothesis regarding this relationship. On the one hand, more educated immigrants might work in some type of industries in which spillovers might be higher and consequently they might benefit more from spillovers materialized at the industry-MSA level among compatriots. On the other, more educated immigrants might interact more than uneducated ones with all the workers at the industry-MSA level, independently of also being from the same country of origin. Therefore, they might also require less to be clustered with their compatriots at the industry-MSA level to benefit from spillovers. Likewise, Ioannides and Datcher Loury (2004) document that the probability of obtaining a job through informal contacts is lower for individuals with high levels of education. Accordingly, per each country of origin of my sample, I compute the percentage of immigrants included in the sample who have some college education. Nonetheless, Figure 2.19 shows that there does not seem to be any relationship between the variable measuring education and the logarithm of the spillovers index.

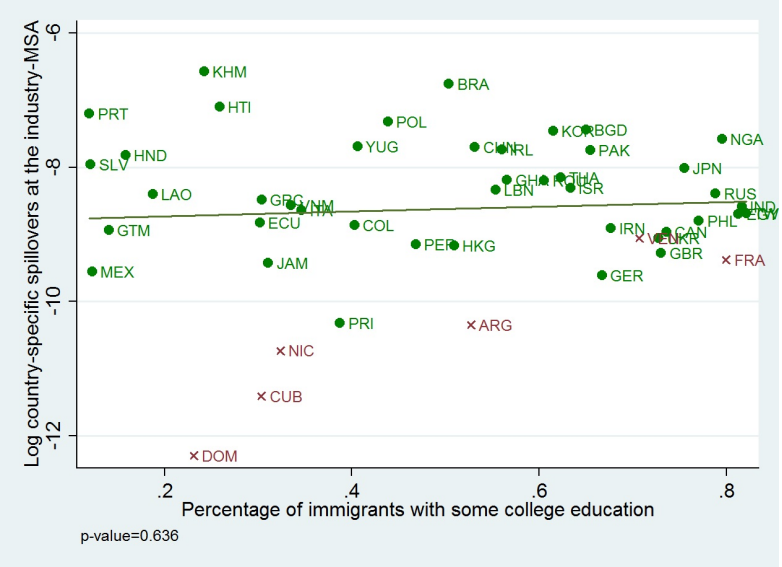

Figure 2.19: Education

2.5.5 Origin-specific comparative advantages for industry-location pairs

In this sub-section, I revisit the assumption of the nonexistence of origin-specific comparative advantages for industry-MSA pairs since it might fail to convince for some cases. For instance, Mexican immigrants might have a comparative advantage to work in restaurants in San Jose and not in New York since a lot of restaurants' customers in San Jose are also Mexican and they tend to demand Mexican food, or to order in Spanish. Thus, the intensity in which workers in each industry need to interact with customers as well as the share of compatriots in a location might constitute an origin-specific comparative advantage for industry-MSA pairs.

To this aim, I add a variable to control for the possibility of this kind of origin-specific comparative advantages for industry-MSA pairs in the estimation of p_{cim} in Equation 2.12. This variable equals the interaction between a variable measuring how important is to deal with customers in each industry and another measuring the percentage of immigrants from same country of origin in each MSA. To construct a variable measuring how important is to interact with customers in each industry, I use the O*Net

database (2002). This database reports information regarding the most important tasks that workers need to do at each possible occupation. In particular, it reports how frequently workers need to deal with customers in a scale of 1 to 5. I create a dummy equal to 1 for those occupations in which the variable takes values higher or equal than 3. Then, I use the occupation-industry matrix from the Bureau of Labor Statistics (2000) which specifies the share of each occupation in each industry. Finally, multiplying the share of each occupation by the dummy and collapsing at the industry level, I obtain a measure on how important is to deal with customers in each industry. Hence, p_{cim} is equal to:

$$p_{cim} = \frac{exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM} + CA_{cim}\right)}{\sum_{i'}\sum_{m'} exp\left(\bar{P}_{ci'}^{CI} + \bar{P}_{cm'}^{CM} + \bar{P}_{i'm'}^{IM} + CA_{cim}\right)} \tag{2.16}$$

where CA_{cim} is defined as the interaction between how important is to deal with customers in each industry and the percentage of immigrants from same country of origin in each MSA. Nonetheless, Equation 2.16 is only estimated for 75 industries instead of 90 as above. This is because the occupation-industry matrix of the year 2000 uses a different industry code (SIC 1987) and I cannot obtain a perfect match with the industry codes used in the Census. As a consequence, I use fewer immigrants than before, specifically, I use 267,375 immigrants.

Tables 2.6 and 2.7 provides the estimates for my index when including this extra variable for this sub-sample.[6] The average of the estimated index across countries equals 0.032% and the index is statistically significant for 40 countries of origin. Moreover, the minimum share of immigrants who benefit from these spillovers equals to 1.684% on average across countries. Thus, even when I control for the possibility of one type of origin-specific comparative advantages for industry-location pairs, my spillovers index is statistically different from zero for many countries.

[6]These estimates are not directly comparable to the ones from Tables 2.2 and 2.3 since the two used samples differ.

Table 2.6: Index estimates revisited I

Country of origin	N_c	$\hat{\delta}^c$	ρ^c	Stat. Significance
Cambodia	1020	0.160%	4.042%	***
Brazil	2638	0.152%	3.914%	***
Haiti	5668	0.098%	3.142%	***
Portugal	1580	0.085%	2.938%	***
Poland	5716	0.076%	2.763%	***
Bangladesh	1083	0.071%	2.703%	**
Korea	10338	0.067%	2.599%	***
Nigeria	1945	0.056%	2.398%	***
Yugoslavia	2330	0.052%	2.304%	***
China	14890	0.051%	2.260%	***
Pakistan	2523	0.051%	2.272%	***
Ireland	1421	0.049%	2.255%	***
Japan	3836	0.040%	2.017%	***
Honduras	2738	0.037%	1.939%	***
Ghana	1040	0.035%	1.922%	**
Israel	1236	0.032%	1.842%	**
Romania	1861	0.032%	1.817%	***
Thailand	1283	0.031%	1.813%	**
Lebanon	1045	0.029%	1.745%	**
Russia	4302	0.028%	1.691%	***
Egypt	1604	0.024%	1.584%	**
Vietnam	10298	0.022%	1.496%	***
Greece	1332	0.022%	1.522%	
India	13691	0.021%	1.465%	***
El Salvador	7691	0.020%	1.407%	***

Table 2.7: Index estimates revisited II

Country of origin	N_c	$\hat{\delta^c}$	ρ^c	Stat. Significance
Taiwan	4355	0.019%	1.393%	***
Colombia	6492	0.019%	1.380%	***
Italy	2792	0.018%	1.356%	***
Philippines	19445	0.018%	1.326%	***
Guatemala	4138	0.016%	1.294%	***
Canada	5476	0.015%	1.238%	***
Iran	3468	0.015%	1.230%	***
Laos	970	0.015%	1.267%	*
Venezuela	1073	0.015%	1.252%	
Ecuador	3314	0.014%	1.216%	*
Ukraine	3323	0.013%	1.163%	***
Peru	4006	0.012%	1.126%	***
France	1470	0.011%	1.091%	
Hong Kong	1799	0.010%	1.031%	
Jamaica	7181	0.010%	1.004%	**
United Kingdom	6251	0.010%	1.004%	***
Mexico	51502	0.009%	0.923%	***
Germany	3697	0.007%	0.825%	***
Nicaragua	2488	0.005%	0.743%	
Puerto Rico	5906	0.004%	0.659%	**
Dominican Republic	7258	0.002%	0.489%	
Cuba	10947	0.001%	0.308%	
Trinidad and Tobago	2391	-0.001%		
Argentina	1720	-0.001%		
Guyana	2804	-0.014%		

2.6 Conclusions

Immigrants from the same country of origin tend to cluster at the location level when moving to the U.S. Such agglomeration pattern suggests that immigrants might benefit from being collocated with their compatriots at the same location. For instance, compatriots might provide financial aid, a shelter at arrival or valuable information about job vacancies. In this chapter I referred to these benefits as origin-specific spillovers at the location level. Nonetheless, it is difficult to attribute such agglomeration of immigrants to these spillovers since they could be the result of other agglomeration forces. For instance, there might exist origin-specific comparative advantages for locations such as proximity to their home country.

In this chapter, I presented suggesting evidence that immigrants also cluster with their compatriots at the industry-location level. This pattern is indicative of spillovers realized at the industry-location level among compatriots. Moreover, I argued that it is unlikely that there are industry-location advantages that are origin-specific which opens an avenue to credibly estimate the strength of a specific type of spillovers which materialize among compatriots at the industry-location level. Accordingly, I proposed an industry-location discrete choice model for immigrants which micro-founds an index able to estimate this particular type of spillovers. Finally, I brought this index to the data and I found that this index is higher for those countries of origin which, on average, are more culturally distant from the U.S. and have lower levels of trust towards individuals of other nationalities. This suggests that immigrants who might have more difficulties in interacting with non-compatriots might benefit more from these specific type of spillovers.

2.7 Appendices

2.7.1 Appendix A

Proof of Proposition 3:

$$
\begin{aligned}
\mathbb{E}\big[G^{CMI}\big] &= \mathbb{E}\Bigg[\sum_i \sum_m \Big(S_{cim} - \hat{S}_{cim}\Big)^2\Bigg] \\
&= \sum_i \sum_m \mathbb{E}\Big[\Big(S_{cim} - \hat{S}_{cim}\Big)^2\Big] \\
&= \sum_i \sum_m \Big\{Var\Big(S_{cim} - \hat{S}_{cim}\Big) + \mathbb{E}\big[S_{cim} - \hat{S}_{cim}\big]^2\Big\} \\
&= \sum_i \sum_m \Big\{Var\left(S_{cim}\right) + \mathbb{E}\big[S_{cim} - \hat{S}_{cim}\big]^2\Big\} \\
&= \sum_i \sum_m \Big\{Var\left(\sum_{k\in c} \frac{u_{kim}}{N_c}\right) + \mathbb{E}\big[S_{cim} - \hat{S}_{cim}\big]^2\Big\} \\
&= \sum_i \sum_m \Big\{\sum_{k\in c} \frac{1}{N_c^2} Var(u_{kim}) + \sum_{\substack{l\in c \\ l\neq k}} \frac{1}{N_c^2} Cov(u_{kim}, u_{lim}) + \\
&+ \mathbb{E}\big[S_{cim} - \hat{S}_{cim}\big]^2\Big\}
\end{aligned}
$$

In the described model, $Var(u_{kim}) = p_{cim} - p_{cim}^2$*,* $Cov(u_{kim}, u_{lim}) = \delta^c(p_{cim} - p_{cim}^2)$ *and* $\mathbb{E}\big[S_{cim} - \hat{S}_{cim}\big] = 0$*. Therefore,*

$$
\mathbb{E}\big[G^{CIM}\big] = \sum_i \sum_m \Big\{\frac{1}{N_c}(p_{cim} - p_{cim}^2) + \delta^c \frac{(N_c - 1)}{N_c}(p_{cim} - p_{cim}^2)\Big\}
$$

and since $\sum_i \sum_m p_{cim} = 1$*,*

$$\mathbb{E}\big[G^{CIM}\big] = \frac{1}{N_c}\left(1 - \sum_i \sum_m p_{cim}^2\right) + \delta^c \left(1 - \frac{1}{N_c}\right)\left(1 - \sum_i \sum_m p_{cim}^2\right)$$

Finally, since $\hat{S}_{cim} = p_{cim}$*:*

$$\mathbb{E}\big[G^{CIM}\big] = \frac{1}{N_c}\left(1 - \sum_i \sum_m \hat{S}_{cim}^2\right) + \delta^c \left(1 - \frac{1}{N_c}\right)\left(1 - \sum_i \sum_m \hat{S}_{cim}^2\right)$$

□

Proof of Lemma 1: *If N_{cim} is independently Poisson distributed with $\mathbb{E}\big[N_{cim}\big] = exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right)$, I can write the log-likelihood function of a Poisson model as:*

$$logL_P = \sum_c \sum_i \sum_m \Big\{ - exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right) + \\ + N_{cim}\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right) - log\ N_{cim}! \Big\}$$

I define $f_{cim}\ \forall(c,i,m)$ as follows:

$$f_{cim} \equiv \left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right) - \left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM}\right)$$

Therefore, I can rewrite the log-likelihood function as:

$$logL_P = \sum_c \sum_i \sum_m \Big\{ - exp\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM} + f_{cim}\right) + \\ + N_{cim}\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM} + f_{cim}\right) - log\ N_{cim}! \Big\}$$

Then, I can take the first order condition with respect to $\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM}\right)$:

$$\frac{\partial\ logL_P}{\partial\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM}\right)} = \sum_i \sum_m \Big\{ - exp\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM} + f_{cim}\right) + N_{cim} \Big\} = 0$$

and this implies that:

$$exp\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM}\right) = \frac{N_c}{\sum_i \sum_m exp\left(f_{cim}\right)}$$

Plugging back $exp\left(\bar{P}_{c1}^{CI} + \bar{P}_{c1}^{CM} + \bar{P}_{11}^{IM}\right)$ into the log-likelihood I obtain the concentrated log-likelihood of the Poisson model:

$$logL_P = \sum_c \sum_i \sum_m \left\{ - \frac{N_c}{\sum_{i'} \sum_{m'} exp\left(f_{ci'm'}\right)} exp\left(f_{cim}\right) + \right.$$
$$\left. + N_{cim}\, log\, \left(\frac{N_c}{\sum_{i'} \sum_{m'} exp\left(f_{ci'm'}\right)} exp\left(f_{cim}\right) \right) - log\, N_{cim}! \right\}$$

Note that $\frac{exp(f_{cim})}{\sum_{i'} \sum_{m'} exp\left(f_{ci'm'}\right)}$ *is equivalent to the probability that an immigrant from country of origin* c *chooses an industry-location* (i, m)*:*

$$\frac{exp\left(f_{cim}\right)}{\sum_{i'} \sum_{m'} exp\left(f_{ci'm'}\right)} = \frac{exp\left(\bar{P}_{ci}^{CI} + \bar{P}_{cm}^{CM} + \bar{P}_{im}^{IM}\right)\ exp\left(-\bar{P}_{c1}^{CI} - \bar{P}_{c1}^{CM} - \bar{P}_{11}^{IM}\right)}{exp\left(-\bar{P}_{c1}^{CI} - \bar{P}_{c1}^{CM} - \bar{P}_{11}^{IM}\right) \sum_{i'} \sum_{m'} exp\left(\bar{P}_{ci'}^{CI} + \bar{P}_{cm'}^{CM} + \bar{P}_{i'm'}^{IM}\right)} =$$
$$= p_{cim}$$

Therefore, I can re-write the log-likelihood of the Poisson model as follows:

$$logL_P = \sum_c \sum_i \sum_m N_{cim}\, log\, p_{cim} - \sum_c N_c + \sum_c \sum_i \sum_m N_{cim}\, log\, N_c -$$
$$- \sum_c \sum_i \sum_m log\, N_{cim}!$$

The first term in the log-likelihood of the Poisson model equals the log-likelihood of the conditional logit model. The other three terms are just constants. Therefore, the estimates for both models are going to be exactly the same.□

2.7.2 Appendix B

Table 2.8: Included countries of origin I

Countries	Number of immigrants
Mexico	58960
Philippines	22008
China	16489
India	15130
Cuba	12348
Korea	11317
Vietnam	11203
El Salvador	8956
Dominican Republic	7977
Jamaica	7922
Colombia	7515
United Kingdom	6902
Puerto Rico	6701
Poland	6359
Haiti	6187
Canada	6091
Taiwan	5143
Guatemala	4894
Russia	4688
Peru	4536
Japan	4361
Germany	4161
Iran	3804
Ecuador	3676
Ukraine	3607

Table 2.9: Included countries of origin II

Countries	Number of immigrants
Honduras	3172
Brazil	3141
Guyana	3122
Italy	2991
Nicaragua	2897
Pakistan	2720
Trinidad and Tobago	2693
Yugoslavia	2498
Nigeria	2177
Hong Kong	2073
Romania	2019
Argentina	1918
Egypt	1765
Portugal	1744
France	1665
Ireland	1594
Greece	1419
Thailand	1398
Israel	1370
Venezuela	1238
Bangladesh	1162
Lebanon	1140
Ghana	1135
Cambodia	1113
Laos	1078

Table 2.10: Included MSAs I

MSA	Number of immigrants
New York-Northeastern NJ	64694
Los Angeles-Long Beach, CA	50745
Miami-Hialeah, FL	18116
San Francisco-Oakland-Vallejo, CA	17094
Chicago, IL	16157
Washington, DC/MD/VA	12183
Houston-Brazoria, TX	9533
San Jose, CA	8708
Dallas-Fort Worth, TX	8695
Boston, MA-NH	8032
San Diego, CA	6799
Fort Lauderdale-Hollywood-Pompano Beach, FL	6161
Riverside-San Bernardino,CA	5434
Atlanta, GA	5236
Phoenix, AZ	5090
Philadelphia, PA/NJ	4326
Seattle-Everett, WA	4220
Orlando, FL	3817
Tampa-St. Petersburg-Clearwater, FL	3367
Las Vegas, NV	3252

Table 2.11: Included MSAs II

MSA	Number of immigrants
Detroit, MI	3000
West Palm Beach-Boca Raton-Delray Beach, FL	2918
Denver-Boulder, CO	2854
Sacramento, CA	2485
Honolulu, HI	2454
Portland, OR-WA	2394
Baltimore, MD	2030
El Paso, TX	1873
Fresno, CA	1871
Austin, TX	1749
San Antonio, TX	1671
Ventura-Oxnard-Simi Valley, CA	1638
McAllen-Edinburg-Pharr-Mission, TX	1592
Minneapolis-St. Paul, MN	1549
Raleigh-Durham, NC	1548
Hartford-Bristol-Middleton- New Britain, CT	1546
Providence-Fall River-Pawtucket, MA/RI	1507
Cleveland, OH	1474
Monmouth-Ocean, NJ	1246
Salt Lake City-Ogden, UT	1119

Table 2.12: Included industries I

Industry	Number of immigrants
Construction	21768
Food Services and Drinking Places	20649
Administrative and Support Services	16565
Professional, Scientific and Technical Services	16434
Educational Services	14836
Hospitals	14404
Ambulatory Health Care Services	10904
Computer and Electronic Product Manufacturing	9188
Accommodation Services	7675
Personal and Laundry Services	7581
Food and Beverage Stores	6990
Clothing Manufacturing	6682
Private Households	6413
Social Assistance	6302
Repair and Maintenance	6177
Wholesale Trade, Durable Goods	5829
Wholesale Trade, Nondurable Goods	5604
Nursing and Residential Care Facilities	5364
Real Estate	5185
Food Manufacturing	4460
Miscellaneous Manufacturing	4228
Monetary Authorities - Central Bank	3750
Fabricated Metal Product Manufacturing	3731
General Merchandise Stores	3687
Transportation and Equipment Manufacturing	3438
Religious, Grant-Making, Civic, and Prof. Organizations	3422
Transit and Ground Passenger Transportation	3420
Clothing and Clothing Accessories Stores	3229
Broadcasting and Telecommunications	2908
Insurance Carriers and Related Activitites	2881

Table 2.13: Included industries II

Industry	Number of immigrants
Amusement, Gambling and Recreation Industries	2865
Chemical Manufacturing	2850
Crop Production	2733
Machinery Manufacturing	2472
Miscellaneous Store Retailers	2310
Truck Transportation	2308
Securities, Commodity Contracts, Fin. Activitites	2152
Plastics and Rubber Products Manufacturing	2136
Motor Vehicle and Part Dealers	2062
Electronics and Appliance Stores	2061
Support Activities for Transportation	1981
Publishing Industries	1980
Executive, Legislative and Other General Gov. Support	1904
furniture amd Related Product Manufacturing	1897
Printing and Related Support Activitites	1871
Health and Personal Care Stores	1722
Postal Service	1576
Air Transportation	1495
Textile Mills	1465
Electrical Equipment, Appliance and Component Manu.	1422
Credit Intermediation and Related Activitites	1401
Justice, Public Order, and Safety Activitites	1274
Information Services and Data Processing Services	1269
Performing Arts, Spectator Sports and Related Industries	1253
Furniture and Home Furnishings Stores	1202
Building Material and Garden Equip. and Supplies Dealers	1134
Textile Product Mills	1107
National Security and International Affairs	1081
Non-Store Retailers	1075
Gasoline Stations	1032

Table 2.14: Included industries III

Industry	$Numberofimmigrants$
Primary Metal Manufacturing	986
Non-Metallic Mineral Product Manufacturing	980
Utilities	971
Paper Manufacturing	882
Administration of Human Resource Programs	874
Rental and Leasing Services	815
Warehousing and Storage	756
Motion Picture and Sound Recording Industries	727
Waste Management and Remediation Services	696
Couriers and Messengers	645
Admisitration of Economic Programs	621
Wood Product Manufacturing	615
Sporting Goods, Hobby, Book and Music Stores	603
Leather and Allied Product Manufacturing	502
Animal Production	423
Beverage and Tobacco Product Manufacturing	341
Heritage Institutions	254
Petroleum and Coal Products Manufacturing	228
Support Activitites for Mining and Oil and Gas Extraction	211
Administration of Environmental Quality Programs	207
Rail Transportation	197
Support Activitites for Agriculture and Forestry	187
Water Transportation	176
Scenic and Sightseeing Transportation	89
Fishing, Hunting and Trapping	86
Mining (except Oil and Gas)	85
Management of Companies and Enterprises	84
Forestry and Logging	69
Oil and Gas Extraction	59
Piepeline Transportation	14

2.7.3 Appendix C

Variable	Description and source
Cultural Distance	This variable equals the cultural distance between U.S. and each country of origin. Source: Own elaboration based on World Value Survey 2005 and 2010
Trust towards people of another nationality	This variable equals the average per country of the individual answers to: "Could you tell me whether you trust people you meet for the first time completely, somewhat, not very much or not at all?" The responses are set to 0, 1, 2 and 3 if an individual responded not at all, not very much, somewhat and completely trust respectively. Source: World Value Survey 2005 and 2010
Language distance	This variable equals the language distance between U.S. and each country of origin. Source: Chiswick (2004)
Age at arrival	This variable equals the country of origin average of the age of immigration to the U.S. Source: Own elaboration based on IPUMS Census (2000)
Education	This variable equals the country of origin percentage of immigrants with some college education. Source: Own elaboration based on IPUMS Census (2000)

Table 2.15: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Cultural Distance	0.311	0.102	0.072	0.568	31
Trust towards people of another nationality	1.178	0.337	0.614	2.007	29
Language distance	0.523	0.151	0.333	1	42
Age at arrival	32.830	1.546	30.122	37.372	48
Education	0.501	0.222	0.120	0.821	48

Chapter 3

SERVICE FIRMS: WHO INTERNATIONALIZES AND HOW? EVIDENCE FROM THE SPANISH HOTEL INDUSTRY

3.1 Introduction

Recently, the international trade literature on multinational firms has been focused on analyzing which firms operate across borders. Firm-level productivity has been proven to be a key factor (Bernard and Jensen, 1995, 1999; Helpman et al., 2004). Additionally, there has also been interest in analyzing how firm-level productivity varies with the choice between foreign direct investment (FDI) and contracting with independent producers or distributors when operating across borders (Antràs and Helpman, 2004, 2008; Antràs and Yeaple, 2014).

There exists empirical evidence showing that firms operating across borders are more productive than domestic firms (Helpman et al., 2004). Moreover, firms engaging in FDI are more productive than those which contract with independent producers or distributors when operating abroad (Tomiura, 2007; Kohler and Smolka, 2009; Corcos et al., 2013; Antràs and Yeaple, 2014). Much of this evidence has been shown for the manufacturing industry. Yet, eating a McDonald's burger in Sri Lanka, drinking an

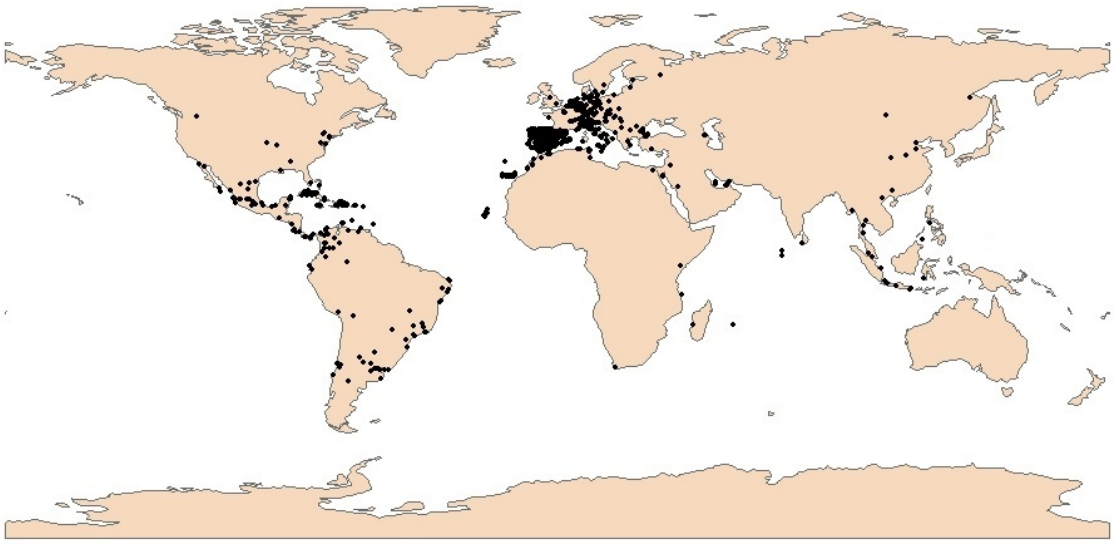

Figure 3.1: Spanish hotels across the world

Starbucks coffee in Moscow or sleeping in a Hilton hotel in Poland are icons of what globalization represents nowadays. Despite the importance of the service sector, it is greatly ignored in the international trade literature on multinational firms.[1]

This chapter tries to fill this gap by studying how firm-level productivity varies with the internationalization decision in a specific service industry in Spain: the hotel industry. The Spanish hotel industry is specially relevant since Spanish hotels are spread out all around the globe as shown in Figure 3.1. There exist 968 Spanish hotels outside of Spain accounting for more than 250,000 rooms. Among the biggest 300 hotel firms in the world, 9.3% of them are Spanish according to the HotelsMag ranking.[2] Along the same lines, the average outward Spanish FDI in the hotel and restaurant industry more than triples that made by the U.S. between 2003 and 2012 according to the OECD database on FDI flows by industry. Moreover, Spain ranks third by international tourists arrivals according to the World Tourism Organization (UNWTO) and ranks second by tourism revenue.[3]

[1] Buch et al. (2009) is an exception for the German banking industry.

[2] This ranking can be consulted in the HotelsMag website: http://www.hotelsmag.com/

[3] According to "la Caixa" Research center, the total contribution of tourism industry to Spanish GDP reached a 10.9% in 2012, and employment in tourism accounted for an 11.9% of the total. This information is available in the following link: http://www.caixabankresearch.com/en/detail-news?lastnewsportal_articleData=363734,10180,1.5

There is one fundamental difference between manufacturing industries and the hotel industry. In manufacturing, only the most productive firms find it optimal to become exporters (Bernard and Jensen, 1995, 1999; Melitz, 2003). In the hotel industry, basically all firms are exporters. According to the World Tourism Organization, the revenues generated from the consumption of foreign tourists should be accounted as an export. Thus, hotel firms become exporters when they host foreign tourists.[4] Hence, the decision for a hotel firm to become an exporter is not completely in their hands. On the other hand, hotel firms can decide whether or not to globalize by operating in other markets.

The purpose of this chapter is to empirically document whether the patterns found for manufacturing firms' globalization decisions extend to the hotel industry. To this aim, I use the Spanish hotel census that does not only include hotel properties located in Spain, but also all hotel properties belonging to Spanish firms that are located abroad. Moreover, this data source also includes the organizational mode of each of these properties. Thus, for each Spanish hotel firm I have information on the extent to which it operates across borders as well as whether engages in FDI or contract with third parties. I combine this data source with balance sheet information of hotel firms. This allows me to study if there exists a positive relationship between hotel firm-level productivity and the decision to engage in global activities. Besides this, I also evaluate to what extent firm-level productivity relates to the extensive and intensive margins of this internationalization decision and with the different internationalization modes (FDI versus contracting with third parties).

The empirical results show that the same stylized facts for multinational firms in manufacturing industries are reproduced in the Spanish hotel industry. Hotel firms that also operate abroad are more productive than those that remain domestic on average. Moreover, firms engaging in FDI are, on average, more productive than those contracting with third parties when operating abroad. Finally, I also document that more productive hotel firms operate in more countries, with more brands and with more properties per brand-country on average.

This chapter spans two strands of related research. First, it is related to the literature on multinational firms that empirically studies how firm-level productivity is correlated

[4]In fact, tourism was the second largest Spanish exporter sector in 2013 according to the Spanish balance of payments. This ranking has been realized by Hosteltur and can be seen in the following link (in Spanish): http://www.hosteltur.com/144680_turismo-segundo-sector-exportador-espana.html

with the internationalization decision (Helpman et al., 2004; Tomiura, 2007; Kohler and Smolka, 2009; Corcos et al., 2013; Antràs and Yeaple, 2014). I extend this literature by showing that those same patterns are reproduced in the hotel industry.[5] Second, this chapter is also related to the hospitality management literature that studies the different modes of globalization of hotel firms (Contractor and Kundu, 1998a,b; Rodríguez, 2002; Chen and Dimou, 2005; Quer et al., 2007; León-Darder et al., 2011; Martorell et al., 2013). This line of research takes a managerial view and does not take into account how firm-level productivity affects the globalization decision.[6] This chapter spans this literature by framing the same research question in the context of the international trade theory on multinational firms.

The rest of this chapter is organized as follows. Section 3.2 presents and discusses the data set. Section 3.3 conducts the empirical analysis and, finally, Section 3.4 concludes.

3.2 Data

The population of interest consists of all hotel firms operating exclusively in Spain, as well as Spanish firms operating across borders. To study the internationalization decision of hotel firms, I need firm-level information on (1) whether they operate only domestically or also across borders; (2) the organizational mode of each of their properties; and (3) their productivity. To this aim, I construct a database using information from two data sources.

On the one hand, I use the 2015 Spanish hotel census provided by *Alimarket*. It is a cross-section that is representative of all hotel properties located in Spain. Moreover, it includes all properties located abroad but belonging to Spanish hotel firms. For each of these properties, there is information on its name, the firm it belongs to, the commercial brand, location, number of rooms, category and organizational mode (ownership & management, lease & management, management contract or franchise contract). In this

[5] Lin and Thomas (2008) also study the FDI versus outsourcing decision for 11 hotel brands in the U.S. Nonetheless, their data does not include productivity measures and they correlate the rooms of each property with the organizational mode.

[6] Quer et al. (2007) is an exception. They consider firms' returns on assets as a variable that might affect the organizational entry mode when globalizing.

census, some firms constitute a hotel chain and operate several properties under one or several brands.

On the other hand, I use the Iberian Balance Sheet Analysis System database (SABI) provided by *Informa D&B* and *Bureau Van Dijk*. SABI is a panel database that gathers financial information for more than two million firms operating in Spain. Both data sources include the fiscal identification code of each firm. These codes allow me to combine both sources of information to create the database for this chapter. Yet, not all firms included in the census also appear in SABI, and consequently, I lose some hotel firms.[7]

The final sample includes 3,729 hotel properties in Spain accounting for 73% of all hotel rooms available in Spain according to the census. Among these properties, 95% of them belong to 1,782 Spanish firms and the remaining to 21 foreign hotel firms. The final sample also includes 805 hotel properties outside Spain but belonging to 49 Spanish hotel firms. These account for 89% of all the internationally supplied rooms by Spanish firms.

I also use some additional sources of information to proxy for some measures varying across countries. For instance, GDP per capita, cultural distance, and an institutional quality measure are taken into account when I cannot include country fixed effects. In the Appendix I report a detailed description of all the variables I use in this chapter as well as their summary statistics.

3.3 Empirical Analysis

3.3.1 TFP estimation

To conduct my empirical analysis, I first need to obtain an estimation of firm-level productivity. Following the empirical literature on multinational firm boundaries, I proxy firm-level productivity with an estimation of firm's revenue total factor productivity (TFP). Accordingly, I assume a Cobb-Douglas production function for each hotel firm.

[7]Nonetheless, these observations tend to correspond to hotel firms operating only one small property in Spain. It is likely that these firms have low levels of productivity. Thus, more productive domestic firms might be overrepresented in the sample with respect to less productive domestic firms. This is not an issue for this chapter, since it can only underestimate the correlations in the empirical analysis.

A direct estimation of the productivity parameter by ordinary least squares might produce biased estimates due to the correlation between inputs and unobservable productivity shocks which generate a simultaneity problem (Olley and Pakes, 1996). Levinsohn and Petrin (2003) propose to use intermediate inputs as a proxy for these unobservable productivity shocks. In this chapter, I follow the Levinsohn and Petrin (2003) approach to estimate each hotel firm TFP using their balance sheet data.

I use revenue, tangible assets and number of workers as proxies for output, capital and labor respectively. I draw on total expenditure in materials to proxy for intermediate inputs. The latest available information on their balance sheets corresponds to the year 2014. The Levinsohn and Petrin (2003) approach requires to use at least two years of data. Accordingly, I use the balance sheets of these firms for several years, specifically between 2008 and 2014.[8] I first compute TFP for each firm and for each year between 2008 and 2014 and then, I obtain an average per each firm. I use these TFP estimates in the remainder of the section.

3.3.2 TFP and Size

The models with heterogeneous firms that study the internationalization decision have a common denominator in their predictions: more productive firms are bigger. This relationship is also empirically documented. Thus, I start by showing such a pattern in the context of the hotel industry. I use the total number of rooms that each firm supplies as a proxy for their size. The number of rooms should be a reasonable measure if hotel firms have rational expectations, and consequently, adapt the number of rooms to their demand. Figure 3.2 plots the relationship between the logarithms of the number of rooms of each firm and the estimated TFP taking into account only those properties that are located in Spain.[9] This figure shows a positive and statistically significant correlation between the two variables.

[8] Revenue, tangible assets and expenditure in materials are deflated using the Spanish GDP deflator from World Bank.

[9] This relationship is robust to considering also those rooms belonging to properties located abroad but belonging to Spanish firms.

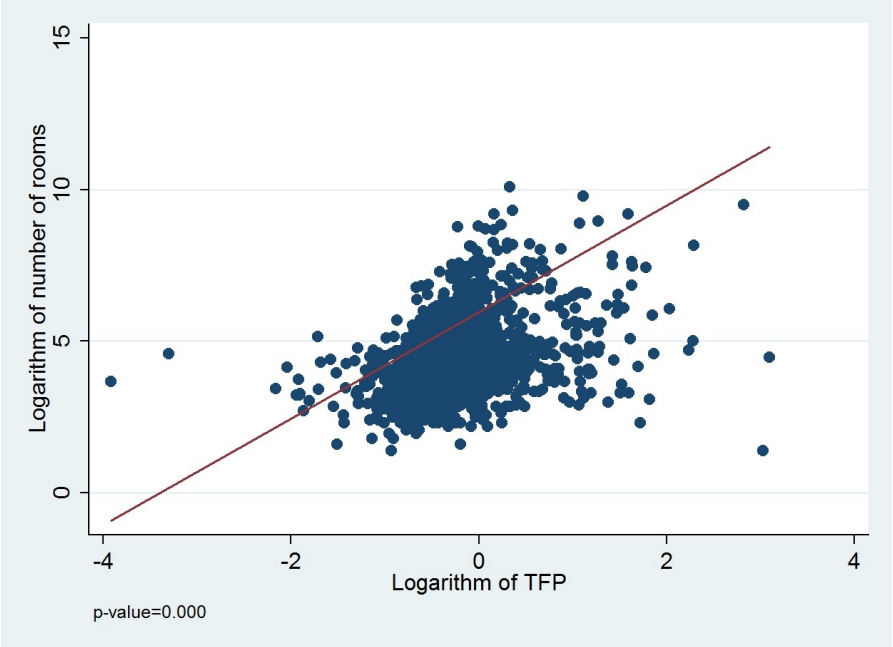

Figure 3.2: TFP and size

3.3.3 TFP and Internationalization

This sub-section evaluates whether hotel firms that internationalize are more productive than the ones that remain domestic. My database includes Spanish and foreign firms operating in Spain, as well as Spanish firms operating across borders. The international trade literature on multinational firms predicts that only the most productive firms find it optimal to operate across borders (Helpman et al., 2004; Antràs and Helpman, 2004, 2008; Antràs and Yeaple, 2014). To test this prediction using my database, I need to show that domestic Spanish firms are less productive than internationalized hotel firms.

I start by estimating the probability density function of firms' TFP for three different groups of firms: domestic Spanish firms (DSF), internationalized Spanish firms (ISF) and internationalized foreign firms which operate in Spain (IFF). Figure 3.3 shows the probability density functions of firms' TFP for these three different groups of firms. On average, DSF are the least productive, whereas ISF, as well as IFF, are the most productive ones. Moreover, this graph suggests that, on average, there are no differences in productivity between ISF and IFF.

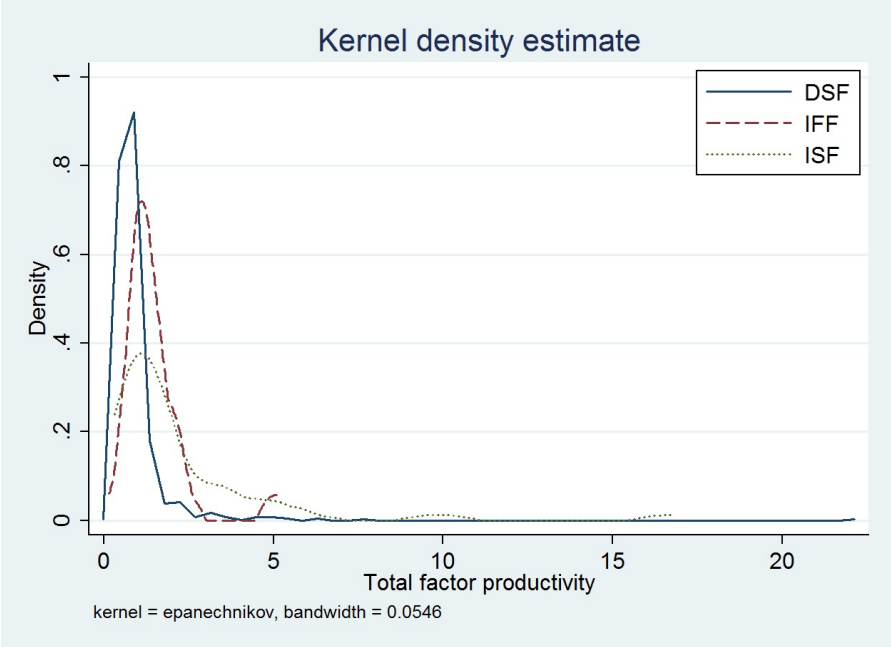

Figure 3.3: Probability density function of firm's TFP

I non-parametrically test this sorting pattern using a Kolmogorov-Smirnov test. This test has been used in other papers studying whether the distribution function of TFP for one group dominates the corresponding distribution function for another group (e.g. Delgado et al., 2002; Benfratello and Razzolini, 2008; Kohler and Smolka, 2009). This test consists in testing the following two sets of hypothesis:

$$H_0: \ F(z) - G(z) = 0 \ \forall z \in \mathbb{R} \ \ vs \ \ H_1: \ F(z) - G(z) \neq 0 \ for \ some \ z \in \mathbb{R}$$
$$H_0: \ F(z) - G(z) \leq 0 \ \forall z \in \mathbb{R} \ \ vs \ \ H_1: \ F(z) - G(z) > 0 \ for \ some \ z \in \mathbb{R}$$

The distribution function $F(z)$ stochastically dominates the distribution function $G(z)$, when the first null hypothesis of the first test can be rejected and when the null hypothesis of the second test cannot be rejected. Accordingly, I test each pairwise comparison of the TFP distribution functions (TFPD) of DSF, ISF and IFF. Table 3.1 displays the results of these tests. The first column describes the tested null hypothesis, and the second, third and fourth columns report the p-value for each pairwise comparison respectively. These results confirm what Figure 3.3 suggests. The underlying

productivity distribution of Spanish domestic firms is different from the ones of internationalized Spanish and foreign firms. Moreover, the productivity distributions of the last two stochastically dominate the one of the former. Therefore, these results provide evidence that, on average, only the more productive firms operate across borders, as in the international literature on multinational firms.

Table 3.1: Kolmogorov-Smirnov Test

K-S Test Null Hypothesis	i =DSF, j=ISF	i =DSF, j=IFF	i =ISF, j=IFF
	p-value	p-value	p-value
$TFPD^i - TFPD^j$=0	0.000	0.000	0.432
$TFPD^i - TFPD^j \leq 0$	0.000	0.000	0.218
$TFPD^i - TFPD^j \leq 0$	1.000	0.917	0.742

As a robustness check, I regress the TFP measure for each firm on dummies for internationalized Spanish firms (ISF) and internationalized foreign firms (IFF). This regression is similar to the *exporter premia* analysis conducted by Bernard et al. (2007). That is, I estimate the following equation:

$$TFP_j = \alpha + \beta ISF_j + \gamma IFF_j + \epsilon_j \tag{3.1}$$

where j stand for a hotel firm j; TFP_j is the firm-level TFP; ISF_j is a dummy variable equal to one for internationalized Spanish firms; and IFF_j is a dummy variable equal to one for internationalized foreign firms. Equation 3.1 is estimated using ordinary least squares with robust standard errors. Column 1 of Table 3.2 shows that internationalized firms, both Spanish and foreign ones, are more productive than domestic firms on average.

Since, I also have information on the number of rooms for each property, and consequently, for each firm, I also analyze whether internationalized firms are bigger than domestic ones. To this aim, I estimate two additional equations. On the one hand, I regress the total number of rooms of each hotel firm on the dummies for ISF and IFF. On the other, I estimate an equation at the property level where I regress the number of rooms of each property on the same dummies and country fixed effects. That is, I estimate the following two equations:

$$Rooms_j = \alpha + \beta ISF_j + \gamma IFF_j + \epsilon_j \tag{3.2}$$

$$Rooms_{hcj} = \alpha + \beta ISF_j + \gamma IFF_j + \delta_c + \epsilon_{hcj} \tag{3.3}$$

where h, c and j stand for a hotel property h in a country c from a hotel firm j; $Rooms_j$ and $Rooms_{hcj}$ equal the number of rooms of a hotel firm and of a hotel property respectively; ISF_j and IFF_j are defined as in Equation 3.1 and δ_c are country fixed effects to control for country specific characteristics. Equation 3.2 and 3.3 are estimated using ordinary least squares with robust and clustered standard errors at the firm level respectively. Columns 2 and 3 of Table 3.2 show that the coefficients on the dummies variables are always positive and statistically significant. This evidence suggests that internationalized hotel firms are bigger in terms of rooms than domestic firms.

Table 3.2: Internationalization premia

	(1) TFP	(2) Rooms	(3) Rooms
SMNF	1.364***	7819.9***	76.14***
	(0.4036)	(2229.8305)	(23.0864)
FMNF	0.541***	1262.1**	52.90**
	(0.2096)	(528.0842)	(20.6588)
Constant	0.904***	156.3***	101.6***
	(0.0238)	(8.2134)	(22.6670)
Country FE	No	No	Yes
N	1803	1803	4528
R^2	0.042	0.200	0.329

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns 1, and 2 report robust standard errors are in parentheses. Column 3 reports clustered standard errors at the firm level are in parentheses.

3.3.4 TFP and Extensive and Intensive Margins

In the previous sub-section I showed that, on average, only the more productive hotel firms operate across borders. Nonetheless, Spanish hotel firms internationalize at different levels. Some of them are present in only one country with only one property, whereas others are present in multiple countries, with multiple brands and with multiple properties. This stylized fact is analogous to the one described by Bernard et al. (2007) and Bernard et al. (2014) for multi-product exporters. Accordingly, in this section, I draw on the methodology recently applied by Bernard et al. (2014) to document how the extensive and intensive margins of the internationalization decision correlate with firm-level productivity. To this aim, I decompose the number of foreign properties a hotel firm j has (H_j) into the number of foreign countries in which it operates (C_j), the number of brands it internationally operates (B_j), a measure of coverage that equals the share of brand-country pairs in which it operates (D_j), and the average foreign properties per brand-country in which it operates ($\bar{H}_j$):

$$H_j = C_j B_j D_j \bar{H}_j \tag{3.4}$$

$$D_j = \frac{o_{bcj}}{C_j B_j} \tag{3.5}$$

$$\bar{H}_j = \frac{1}{o_{bcj}} \sum_b \sum_c H_{bcj} \tag{3.6}$$

where o_{bcf} is the number of brand-country pairs in which a hotel operates. Therefore, the first three components proxy the extensive margin of a hotel firm, whereas the fourth measures its intensive margin. Table 3.3 reports the cross-section regressions for the logarithms of the number of foreign properties and for each of its four components on the logarithm of TFP, and some firm controls which are also obtained from their balance sheets. Column 1 shows that more productive hotel firms have more foreign establishments on average. Considering their extensive margins, columns 2 and 3 show that more productive hotel firms operate in more countries and with more brands. Column 4 shows that the coverage measure is negatively correlated with firm productivity. Therefore, more productive firms operate with more brands and in more countries, but do not operate with each brand in every country. Looking at the intensive margin, column 5

shows that the correlation between firm productivity and the average number of properties per brand-country is positive but it is only statistically significant at the 84.3% confidence level.

These results go in accordance to the international trade literature when studying the intensive and extensive margins of international trade (e.g., Bernard et al., 2014). However, I analyze the decision to internationalize instead of the decision to export. I find that, on average, more productive firms have more foreign units (extensive margin), operate in more countries (extensive margin), operate more brands (extensive margin) and have more units per each brand-country (intensive margin).

3.3.5 TFP and Modes of internationalization

The decision between engaging in FDI or contracting with third parties is the object of study in the international trade literature on multinational firm boundaries. The different models in this literature have a common feature: contracts are not perfect.

Antràs and Helpman (2004, 2008) propose a model with heterogeneous firms to study the decision between vertical FDI and foreign outsourcing [10] In these models, FDI entails a higher fixed cost of entry but allows more productive firms to obtain higher profits. Tomiura (2007), Kohler and Smolka (2009) and Corcos et al. (2013) provide empirical support for this prediction in the manufacturing industry. Along the same lines, Antràs and Yeaple (2014), in a recent survey, build on Melitz (2003) to formalize the decision between horizontal FDI and licensing with heterogeneous firms. In their model, FDI requires a higher fixed cost of entry but allows more productive firms to obtain higher profits.[11]

The organizational choice of hotel firms can be considered as either a horizontal FDI versus licensing decision, or as a vertical FDI versus outsourcing decision. The interpretation of the former is that hotel firms duplicate an activity by opening a new property in a new country. Hence, they can either open a hotel and manage it by themselves or

[10] Their model also allows for vertical integration and outsourcing within the same country of the firm.

[11] Helpman et al. (2004) also adapt the Melitz (2003) model to allow firms to engage in horizontal FDI. In their model, firms decide between exporting or FDI. Exporting requires firms paying a fixed cost. Horizontal FDI involves an even higher fixed cost than exporting but it saves on transport costs bringing about higher profits for more productive firms. Yet, they do not consider the possibility of licensing independent producers

Table 3.3: TFP and the margins of internationalization

	(1) lnforeignproperties	(2) lnforeignmarkets	(3) lnbrands	(4) lndensity	(5) lnaverage
Ln(TFP)	0.658**	0.444**	0.242*	-0.168*	0.140
	(0.3099)	(0.2063)	(0.1293)	(0.0929)	(0.0973)
Ln(workers)	0.417***	0.297***	0.142***	-0.0962**	0.0743**
	(0.1247)	(0.0916)	(0.0525)	(0.0358)	(0.0348)
Ln(capital per worker)	-0.0471	-0.0357	-0.0132	0.00711	-0.00527
	(0.1123)	(0.0777)	(0.0568)	(0.0414)	(0.0366)
Age	0.00950	-0.00587	0.00861	-0.00491	0.0117
	(0.0190)	(0.0109)	(0.0074)	(0.0056)	(0.0088)
Constant	-0.958*	-0.606	-0.571**	0.368**	-0.149
	(0.5207)	(0.3860)	(0.2201)	(0.1527)	(0.1718)
N	49	49	49	49	49
R^2	0.385	0.362	0.312	0.286	0.194

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses.

license a third party to operate with their brand in exchange of some rents. The interpretation of the latter is that the hotel production process can be sliced in two stages. Hotel firms' headquarters provide support and assistance to downstream producers, including hotel operations assistance, sales and marketing support (e.g., the reservation system), and purchasing power. Downstream producers, in their turn, are the final producers of the service. That is, vertical integration or outsourcing occur with a downstream producer rather than with a supplier of intermediate inputs as is generally the case in the manufacturing industry. Nonetheless, in either case, the discussed theoretical models predict that firms engaging in FDI are more productive than firms contracting with third parties.

In this sub-section, I assess whether this prediction holds in the context of the hotel industry. Nonetheless, my database discriminates among four different organizational modes. Specifically, each property can be operated according to one of the following modes: ownership & management, lease & management, management contracts and franchise contracts. The first of them, ownership & management, means that a property is owned and managed by the firm. In the second of them, lease & management, the firm rents the property from a third party and manages the property as it owned it. Whereas, in the fourth case, a franchise contract, the firm contracts with a third party who supplies and manages the property. This property is operated under the brand of the hotel firm which receives some fees and royalties in exchange. Finally, a management contract is similar to a franchise contract but with one main difference: the hotel firm supplies the general manager to the third party. That is, the property belongs to the third party and also all the employees, except the general manager who belongs to the hotel firm. In this way, the hotel firm exerts some control in the management of their branded hotels. The main characteristics of each business model are displayed in Figure 3.4.

Therefore, hotel firm globalization decisions is not a dichotomous choice (FDI versus contracting with third parties). They can choose among four different options. The fixed costs associated with each mode as well as the control that the hotel firm can exert in each mode increase in the following order: franchise contracts, management contracts, lease & management and ownership & management. Thus, in Figure 3.5, I order the different organizational modes from pure contracting with third parties to complete FDI. However, possible contractual frictions as rent dissipation, hold-up problems involved or quality dilution issues are likely to decrease in the aforementioned order.

	FRANCHISE CONTRACTS	MANAGEMENT CONTRACTS	LEASE & MANAGE	OWNERSHIP & MANAGEMENT
Does the chain own the property?	No	No	No	Yes
Does the chain manage the property?	No	Yes	Yes	Yes
Does the chain own the operating company?	No, it belongs to the 3rd party.	No, it belongs to the 3rd party.	Yes	Yes
Do the employees belong to the chain?	No, they belong to the 3rd party.	No, they belong to the 3rd party.	Yes	Yes
Revenue for the chain	Fees from the 3rd party	Fees from the 3rd party	All revenues of hotel	All revenues of hotel

Figure 3.4: Business models

Hence, considering these observations jointly with the discussed predictions on multinational firm boundaries, I study whether hotel firms rank to this same order depending on their firm-level productivity. Nonetheless, I start my analysis by grouping ownership & management and lease & management as FDI, while the others as contracting with third parties. Later on, I discriminate among the four modes.

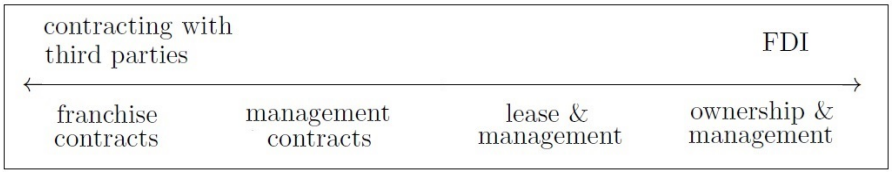

Figure 3.5: Organizational modes

My database includes 805 properties that are located outside Spain but belong to 49 hotel firms. These hotel properties account for 85% of all Spanish hotels that are located abroad. Columns 2 and 3 of Table 3.4 report the absolute numbers and the percentages of properties that correspond to each business model respectively. Franchise contracts are the least chosen option among Spanish hotel chains when operating internationally, whereas the other three are chosen approximately with the same frequency. Following Kohler and Smolka (2009), column 4 of Table 3.4 shows the number of hotel firms which use each mode in a mutually inclusive way. That is, a firm appears in all business models that this firm uses (e.g., when a firm does franchising and ownership &

management, this firm is counted once for every category). This column shows that few firms decide upon franchise contracts.

Table 3.4: Business model strategies

Mode	Number of properties	Percentage	Number of firms
Franchise contract	15	1.86%	4
Management contract	288	35.78%	22
Lease & management	218	27.08%	15
Ownership & management	284	35.28%	38

Moreover, only 57% of Spanish hotel firms engage in a single business model for all their properties. The remaining 43% combine the different strategies among their properties. It might be that depending on the category of the property or in the institutional environment, firms find it optimal to choose one or another organizational mode. The hospitality management literature found that some country-specific characteristics might be determining the organizational mode decision (Contractor and Kundu, 1998a,b; Rodríguez, 2002; Chen and Dimou, 2005; Quer et al., 2007; León-Darder et al., 2011; Martorell et al., 2013). For instance, institutional quality or cultural distance are proven to play an important role. As a consequence, firms might decide to engage in FDI in countries with bad institutional quality and to contract with third parties in countries with good institutional quality. To this aim, I propose to study the relationship between firms productivity and the likelihood of engaging in FDI in each of their properties but controlling for country fixed effects. Accordingly, I start by estimating the following linear probability model:

$$FDI_{hcj} = \alpha TFP_j + \beta stars_{hcj} + X_j^{'}\gamma + \delta_c + \epsilon_{hcj} \tag{3.7}$$

where h, c and j stand for a hotel property h in a country c from a hotel firm j; FDI_{hcj} is a dummy variable equal to one when a hotel property is either owned & managed or leased & managed by the firm; TFP_j is the firm-level TFP; $stars_{hcj}$ is equal to the number of stars of each property; $X_j^{'}$ are some firm controls; and δ_c are country fixed effects to control for country specific characteristics. Equation 3.7 is estimated using ordinary least squares with clustered standard errors at the firm level. Columns 1 and 2 of Table 3.5 show that the coefficient of TFP is positive and statistically significant. This evidence suggests that also in the hotel industry, more productive firms also engage

more in FDI on average.

Table 3.5: TFP and FDI: linear probability model

	(1) FDI	(2) FDI	(3) FDI	(4) FDI
TFP	0.0425***	0.0423***	0.0362***	0.0355***
	(0.0088)	(0.0090)	(0.0072)	(0.0073)
Stars	-0.0487*	-0.0480*	-0.00298	-0.00128
	(0.0249)	(0.0254)	(0.0287)	(0.0283)
Number of foreign units	0.00470***	0.00498***	0.00307*	0.00381**
	(0.0015)	(0.0016)	(0.0016)	(0.0016)
Number of foreign markets	-0.0513***	-0.0523***	-0.0453***	-0.0482***
	(0.0124)	(0.0128)	(0.0123)	(0.0120)
Number of workers	0.000142***	0.000143***	0.000136***	0.000139***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age	0.0110	0.0111	0.0111**	0.0113**
	(0.0070)	(0.0071)	(0.0053)	(0.0054)
Capital per worker		-0.0174		-0.0466***
		(0.0249)		(0.0124)
Country FE	Yes	Yes	Yes	Yes
N	805	805	805	805
R^2	0.521	0.522	0.510	0.514

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the firm level are in parentheses.

So far, I assumed that leased & managed properties also correspond to the case of FDI. However, the firm is contracting ownership in such a case. The empirical literature on multinational firm boundaries do not distinguish on whether a firm owns or leases a plant. Since my data set discriminates properties depending on ownership, as a robustness check, I re-define the variable FDI_{hcj} to be a dummy variable equal to one only when a hotel property is owned & managed. Columns 3 and 4 of Table 3.5 show that with this alternative classification, the correlation between firm productivity and FDI is positive and statistically significant, on average.

Nevertheless, there might exist substantial differences between properties under franchise contracts and management contracts, as well as, between properties under lease & management and ownership & management. Hence, I also conduct an analysis to evaluate how firm-level productivity varies with each organizational mode by means of a multinomial logit model. To this aim, I define a variable called $mode_{hcj}$ that equals 1, 2, 3 and 4 depending on whether a property h in a country c from hotel firm j is operated under a franchise contract, management contract, lease & management or ownerhip & management respectively. I assume that the log-odds is defined as follows:

$$ln\ \frac{Prob(mode_{hcj}=m)}{Prob(mode_{hcj}=4)} = \alpha_m + \beta_m TFP_j + \delta_m stars_{hcj} + X_j^{'}\gamma + Z_c^{'}\theta + \epsilon_{hcj} \quad (3.8)$$

where I use the same notation as in Equation 3.7 and $m \in \{1,2,3\}$. However, note that I do not include country fixed effects due to the possibility of incidental parameters problem in a multinomial logit setup. As a consequence, $Z_c^{'}$ is defined as country controls that have been suggested to matter when studying the foreign market entry mode of hotel firms (Contractor and Kundu, 1998a,b; Rodríguez, 2002; Chen and Dimou, 2005; Quer et al., 2007; León-Darder et al., 2011; Martorell et al., 2013). Table 3.6 shows the estimated multinomial logit where the base category is ownership & management. The first row of Table 3.6 shows the estimated change in the relative log odds of choosing either a franchise contract (F), a management contract (M) or lease & management (L) versus ownership & management. These coefficients are negative and increase as we move to the right of the first row. Hence, they provide suggestive evidence of the productivity ranking.[12] On average, the most productive firms are more likely to use ownership & management; the second most productive firms are more likely to engage in lease & management; the second least productive firms are more likely to engage in management contracts; and the least productive firms are more likely to engage in franchise contracts when globalizing.

[12] I also estimate an ordinal logit for the variable $mode_{hcj}$. I find a positive and statistically relationship between this variable and firm TFP.

Table 3.6: TFP and business models: multinomial logit

	(1) F	(2) M	(3) L
TFP	-2.022***	-0.290***	-0.231*
	(0.4079)	(0.0275)	(0.1189)
Stars	-1.948***	-0.424	-0.462
	(0.5644)	(0.3508)	(0.6640)
Foreign units	-0.0115	-0.0251***	-0.0123
	(0.0245)	(0.0066)	(0.0206)
Foreign markets	4.554***	0.364***	0.276
	(0.7382)	(0.0306)	(0.1818)
Age	-3.899***	-0.157*	-0.107***
	(0.6341)	(0.0824)	(0.0410)
Number of workers	-0.0140***	-0.000852***	-0.000927
	(0.0019)	(0.0001)	(0.0006)
GDP per capita	-0.000126	-0.0000240	0.000000664
	(0.0001)	(0.0000)	(0.0000)
Rule of Law	4.295*	0.203	0.879*
	(2.4235)	(0.5278)	(0.4728)
Cultural Distance	11.12	13.28***	17.16***
	(35.7864)	(4.0457)	(2.5032)
Constant	3.976	1.195	-1.558
	(6.0514)	(1.8353)	(3.4605)
N		573	
Pseudo R^2		0.3748	

Coefficients are statistically different from zero at the following levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the firm level are in parentheses.

3.4 Conclusions

In this chapter I have documented that the main patterns from the international literature on multinational firms extend to one specific services industry in Spain: the hotel industry. Specifically, I showed that, on average, (1) hotel firms engaged in global business activities are more productive than those who only operate domestically; (2) more productive hotel firms operate in more countries, with more brands and with more properties per brand-country; and (3) more productive hotel firms are more likely to engage in FDI when going abroad instead of contracting with third parties.

3.5 Appendix

Variable	Description and source
TFP	This variable equals the Levinsohn and Petrin (2003) total factor productivity (TFP) estimate. Source: Own elaboration based on SABI data by Informa D&B and Bureau Van Dijk
Number of rooms	This variable equals the total number of rooms of each property. Source: 2015 Spanish hotel census by Alimarket
Workers	This variable equals the average number of workers of each hotel firm between 2008 and 2014. Source: SABI data by Informa D&B and Bureau Van Dijk
Capital per worker	This variable equals the average capital per worker ratio of each hotel firm between 2008 and 2014. Source: SABI data by Informa D&B and Bureau Van Dijk
Age	This variable equals the age of each hotel firm in the year 2014. Source: SABI data by Informa D&B and Bureau Van Dijk
Stars	This variable equals the number of stars of each hotel property. Source: 2015 Spanish hotel census by Alimarket

Variable	Description and source
Foreign units	This variable equals the number of foreign units that each hotel firm has, conditional on having international presence. Source: 2015 Spanish hotel census by Alimarket
Foreign markets	This variable equals the number of foreign markets in which a hotel firm operates, conditional on having international presence. Source: 2015 Spanish hotel census by Alimarket
Brands	This variable equals the number of brands that each hotel property operates abroad, conditional on having international presence. Source: 2015 Spanish hotel census by Alimarket
Density	This variable equals the share of brand-country pairs in which a hotel firm operates, conditional on having international presence. Source: 2015 Spanish hotel census by Alimarket
Average	This variable equals the average foreign properties per brand-country in which a hotel firm operates, conditional on having international presence. Source: 2015 Spanish hotel census by Alimarket
Ownership & management	This variable equals one if a hotel property is operated under an ownership & management mode and zero otherwise. Source: 2015 Spanish hotel census by Alimarket

Variable	Description and source
Lease & management	This variable equals one if a hotel property is operated under a lease & management mode and zero otherwise. Source: 2015 Spanish hotel census by Alimarket
Management contract	This variable equals one if a hotel property is operated under a management contract and zero otherwise. Source: 2015 Spanish hotel census by Alimarket
Franchise contract	This variable equals one if a hotel property is operated under a franchsie contract and zero otherwise. Source: 2015 Spanish hotel census by Alimarket
GDP per capita	This variable equals gross domestic product per capita of each country in which a firm operates in the year 2014. Source: World Bank
Rule of Law	This variable equals the average of the Rule of Law index between 1996 and 2014 for each country in which a firm operates. This index reflects perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Higher values indicate better institutional quality. Source: The Worldwide Governance Indicators, World Bank

Variable	Description and source
Cultural Distance	This variable equals the cultural distance between Spain and each country in which a hotel firm operates. Source: Own elaboration based on World Value Survey 2005 and 2010

Table 3.7: Summary statistics of sections 3.3.2 and 3.3.3

Variable	Mean	Std. Dev.	Min.	Max.	N
TFP	0.946	1.097	0.019	22.082	1803
Number of rooms	374.719	2796.959	4	73317	1803

Table 3.8: Summary statistics of section 3.3.4

Variable	Mean	Std. Dev.	Min.	Max.	N
lnforeignproperties	1.356	1.454	0	5.549	49
lnforeignmarkets	0.795	0.989	0	3.663	49
lnbrands	0.323	0.605	0	1.945	49
lndensity	-0.219	0.421	-1.428	0	49
lnaverage	0.457	0.514	0	1.598	49
Ln(TFP)	0.421	0.793	-1.157	2.819	49
Ln(workers)	4.321	1.859	1.299	8.687	49
Ln(capital per worker)	-1.198	1.606	-4.276	3.068	49
Age	18.776	9.749	4	42	49

Table 3.9: Summary statistics of section 3.3.5

Variable	Mean	Std. Dev.	Min.	Max.	N
TFP	3.242	3.991	0.314	16.764	805
Stars	4.375	0.619	2	6	805
Number of foreign units	134.724	100.647	1	257	805
Number of foreign markets	22.922	13.321	1	39	805
Number of workers	1892.069	2452.405	3.666	5924.833	805
Age	25.503	8.640	4	42	805
Capital per worker	1.353	1.308	0.0138	21.498	805
GDP per capita	22302.9	18831.430	955.141	96732.410	573
Rule of Law	0.242	0.884	-1.07	1.86	573
Cultural Distance	0.206	0.061	0.107	0.435	573

Bibliography

Alesina, A. and E. La Ferrara (2002). Who trusts others? *Journal of Public Economics 85(2)*, 207–234.

Algan, Y. and P. Cahuc (2010). Inherited trust and growth. *American Economic Review 100 (5)*, 2060–92.

Antràs, P. and E. Helpman (2004). Global sourcing. *Journal of Political Economy, 112(3)*, 552–580.

Antràs, P. and E. Helpman (2008). Contractual frictions and global sourcing. *The Organization of Firms in a Global Economy. Cambridge, MA: Harvard University Press*.

Antràs, P. and S. R. Yeaple (2014). Multinational firms and the structure of international trade. *Handbook of International Economics, 4*, 55–130.

Bartel, A. P. (1989). Where do the new u.s. immigrants live? *Journal of Labor Economics, 7(4)*, 371–391.

Batsaikhan, M. (2013). Trust, trustworthiness, and success in business. *mimeo, Georgetown University*.

Bauer, T., G. S. Epstein, and I. N. Gang (2007). The influence of stocks and flows on migrants location choices. *Research in Labor Economics, 26(6)*, 199–229.

Benfratello, L. and T. Razzolini (2008). Firms' productivity and internalisation choices: Evidence for a large sample of Italian firms. *Centro Studi Luca d'Angliano Development Studies Working Paper 236*.

Bernard, A. B., I. V. Beveren, and H. Vandenbussche (2014). Multi-product exporters and the margins of trade. *The Japanese Economic Review, 65(2)*, 142–157.

Bernard, A. B. and J. B. Jensen (1995). Exporters, jobs, and wages in U.S. manufacturing: 1976-87. *Brookings Papers on Economic Activity: Microeconomics, 1995*, 67–112.

Bernard, A. B. and J. B. Jensen (1999). Exceptional exporter performance: Cause, effect or both? *Journal of International Economics, 47(1)*, 1–25.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2007). Firms in international trade. *Journal of Economic Perspectives, 21(3)*, 105–130.

Brambor, T. and W. R. Clark (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis 14(1)*, 63–82.

Buch, C. M., C. T. Koch, and M. Koetter (2009). Margins of international banking: Is there a productivity pecking order in banking, too? *Deutsche Bundesbank Discussion Paper. Series 2: Banking and Financial Studies, 12/2009*.

Butler, J., P. Giuliano, and L. Guiso (2016). The right amount of trust. *Journal of the European Economic Association, forthcoming*.

Cassey, A. J. and B. O. Smith (2014). Simulating confidence for the Ellison-Glaeser index. *Journal of Urban Economics, 81*, 85–103.

Chen, J. J. and I. Dimou (2005). Expansion strategies of international hotel firms. *Journal of Business Research, 58(12)*, 1730–1740.

Chiswick, B. R. and P. W. Miller (2004). Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development, 26(1)*, 1–11.

Contractor, F. J. and S. K. Kundu (1998a). Franchising versus company-run operations: Modal choice in the global hotel sector. *Journal of International Marketing, 6, 28-53*.

Contractor, F. J. and S. K. Kundu (1998b). Modal choice in a world of alliances: Analyzing organizational forms in the international hotel sector. *Journal of International Studies, 29(2)*, 325–357.

Corcos, G., D. M. Irac, G. Mion, and T. Verdier (2013). The determinants of intrafirm trade: Evidence from French firms. *The Review of Economics and Statistics, 95(3)*, 825–838.

Delgado, M. A., J. C. Fariñas, and S. Ruano (2002). Firm productivity and export markets: a non-parametric approach. *Journal of International Economics, 57(2)*, 397–422.

Dunlevey, J. A. (1991). On the settlement patterns of recent Caribbean and Latin immigrants to the U.S. *Growth and Change, 22(1)*, 54–67.

Duranton, G. and H. G. Overman (2005). Testing for localisation using micro-geographic data. *Review of Economic Studies, 72(4)*, 1077–1106.

Ellison, G. and E. L. Glaeser (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy, 105(5), 889-927*.

Ellison, G. and E. L. Glaeser (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review Papers and Proceedings, 89(2)*, 311–316.

Fisman, R. and E. Miguel (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy 115(6)*, 1020–1048.

Gambetta, D. (2000). Can we trust trust? *Trust: Making and Breaking Cooperative Relations. Department of Sociology, University of Oxford*, 213–237.

Gennaioli, N., A. Shleifer, and R. Vishny (2012). Money doctors. *Journal of Finance, 70(1)*, 91–114.

Glitz, A. (2014). Ethnic segregation in Germany. *Labour Economics, 29*, 28–40.

Guimaraes, P., O. Figuereido, and D. Woodward (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics, 85(1)*, 201–204.

Guiso, L., P. Sapienza, and L. Zingales (2004). The role of social capital in financial development. *American Economic Review 94(3)*, 526–56.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *The Journal of Economic Perspectives 20(2)*, 23–48.

Guiso, L., P. Sapienza, and L. Zingales (2010). Civic capital as the missing link. *Working paper 15845*.

Hellerstein, J. K., M. McInerney, and D. Neumark (2008). Measuring the importance of labor market networks. *IZA Discussion Paper 3750*.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004). Export versus FDI with heterogeneous firms. *American Economic Review, 94(1)*, 300–316.

Ioannides, Y. M. and L. Datcher Loury (2004). Job information, networks, neighborhood effects, and inequality. *Journal of Economic Literature, 42(4)*, 1056–93.

Jaeger, D. A. (2000). Local labor markets, admission categories, and immigrant location choice. *mimeo, College of William and Mary*.

Jaeger, D. A. (2008). Green cards and the location choices of immigrants in the United States, 1971-2000. *Immigration: Trends, Consequences and Prospects for the United States. Research in Labor Economics, 27*, 131–183.

Kerr, W. R. and M. Mandorff (2015). Social networks, ethnicity, and entrepreneurship. *NBER Working Paper 21597*.

Khwaja, A. I. and A. Mian (2005). Do lenders favor politically connected firms? Rent provision in an emerging financial market. *The Quarterly Journal of Economics, 120(4)*, 1371–1411.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? A cross country investigation. *The Quarterly Journal of Economics 112(4)*, 1251–1288.

Kohler, W. K. and M. Smolka (2009). Global sourcing decisions and firm productivity: Evidence from Spain. *CESifo Working Paper No. 2903, CESifo Group, Munich*.

La Porta, R., F. L. de Silanes, A. Shleifer, and R. W. Vishny (1997). Trust in large organizations. *American Economic Review Papers and Proceedings 87*, 333–338.

Lampe, K. V. and P. O. Johansen (2003). Criminal networks and trust. on the importance of expectations of loyal behaviour in criminal relations. *Organised Crime, Trafficking, Drugs, 102*.

León-Darder, F., C. Villar-García, and J. Pla-Barber (2011). Entry mode choice in the internationalisation of the hotel industry: a holistic approach. *The Service Industries Journal, 31(1)*, 107–122.

Levinsohn, A. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *RevReview of Economic Studies, 70(2)*, 317–341.

Lin, S. F. and C. Thomas (2008). When do multinational firms outsource? Evidence from the hotel industry. *mimeo*.

Mandorff, M. (2007). Social networks, ethnicity, and occupation. *Thesis Dissertation. The University of Chicago*.

Marshall, A. (1890). *Principles of Economics*. New York: Macmillan and Co.

Martorell, O., C. Mulet, and L. Otero (2013). Choice of market entry mode by Balearic hotel chains in the Caribbean and Gulf of Mexico. *International Journal of Hospitality Management, 32*, 217–227.

Maurel, F. and B. Sedillot (1999). A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics, 29(5)*, 575–604.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics. New York: Academic Press*.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71 (6)*, 1695–1725.

Munshi, K. (2003). Networks in the modern economy. *The Quarterly Journal of Economics, 118(2)*, 549–599.

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public Economics 93*, 950–964.

Olken, B. A. (2012). Corruption in developing countries. *Annual Review of Economics, 4*, 479–505.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64 (6)*, 1263–1297.

Putnam, R. D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Quer, D., E. Claver, and R. Andreu (2007). Foreign market entry mode in the hotel industry: The impact of country- and firm-specific factors. *International Business Review, 16(3)*, 362–376.

Rodríguez, A. R. (2002). Determining factors in entry choice for international expansion. The case of the Spanish hotel industry. *Tourism Management, 23(6)*, 597–607.

Rohner, D., M. Thoenig, and F. Zilibotti (2013). War signals: A theory of trade, trust, and conflict. *The Review of Economic Studies, 80(3)*, 1114–1147.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics, 88(4)*, 641–658.

Satyanath, S., N. Voigtländer, and H.-J. Voth (2013). Bowling for fascism: Social capital and the rise of the Nazi party in Weimar Germany, 1919-33. *NBER Working Paper 19201*.

Schmidheiny, K. and M. Brulhart (2011). On the equivalence of location choice models: Conditional logit, nested logit and Poisson. *Journal of Urban Economics, 69(2)*, 214–222.

Tabellini, G. (2008). Presidential adress: Institutions and culture. *Journal of the European Economic Association, 6(2-3)*, 255–294.

Tabellini, G. (2010). Culture and institutions: Economic development in the regions of Europe. *Journal of the European Economic Association 8(4)*, 677–716.

Tomiura, E. (2007). Foreign outsourcing, export, and FDI: A productivity comparison at the firm-level. *Journal of International Economics, 72(1)*, 113–127.

Zak, P. J. and S. Knack (2001). Trust and growth. *The Economic Journal 111(470)*, 295–321.

Zavodny, M. (1997). Welfare and the location choices of new immigrants. *Economic Review-Federal Reserve Bank of Dallas, 2*, 2–10.