

Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry*

Ramana Nanda [†]
Harvard Business School

Tarun Khanna
Harvard Business School

HBS Working Paper 08-003
December 15, 2008

Abstract

This study explores the importance of cross-border social networks for entrepreneurship in developing countries by examining ties between the Indian expatriate community and local entrepreneurs in India's software industry. We find that local entrepreneurs who are based in cities with weaker formal institutions rely significantly more on diaspora networks for business leads and financing. Relying on these networks also has a greater impact on firm performance for entrepreneurs located in weaker institutional environments. Our results provide micro-evidence consistent with a view that cross-border social networks play an important role in helping entrepreneurs to circumvent the barriers arising from imperfect local institutions in developing countries.

JEL Classification: F22, L14, L26, L86, O17, O19

Key Words: Diasporas, Informal Networks, Institutions, Entrepreneurship.

*We are extremely grateful to Kiran Karnik and Sunil Mehta at NASSCOM for allowing us to survey NASSCOM members for this research. This paper has benefited from very helpful discussions with Abhijit Banerjee, Rodrigo Canales, Sylvain Chassang, Bob Gibbons, William Kerr, Asim Ijaz Khwaja, Karim Lakhani, Josh Lerner, Rafel Lucea, Antoinette Schoar, Jordan Siegel and especially Kevin Boudreau and Nicola Lacetera. We also wish to thank the participants of the MIT Development and Organizational Economics Lunches and the HBS International Seminar for their comments on early stages of this research. All errors are our own.

[†]Corresponding author: RNanda@hbs.edu

1 Introduction

Ethnic and social networks have played an important role in promoting international trade for centuries, by helping to overcome weaknesses in the information and contracting environment faced by buyers and sellers across nations (Curtin, 1984; Greif, 1993; Rauch, 2001)

Recent research examining expatriate communities from developing countries suggests that even today, they may play an important role in increasing bilateral trade between their country of origin and the country in which they are based (Gould 1994; Rauch 2001; Rauch and Trindade 2002). Despite the wealth of cross-country research on diaspora networks, however, there is little empirical research directly examining ties between the diaspora and local entrepreneurs in developing countries. For example, do entrepreneurs in developing countries who face greater transaction costs or barriers to trade rely more on their diaspora contacts for help with their business? Anecdotal accounts of the links between local entrepreneurs and the expatriate community suggest that in fact the opposite may be true (Saxenian 2002, 2006; Saxenian and Li 2003), implying that perhaps these networks may be an outcome of positive assortative matching rather than a means to overcome weak domestic institutions.

In this paper, therefore, we depart from the prior literature studying diaspora networks at the macro-economic level to examine the extent to which entrepreneurs *within* a given country vary in their reliance on expatriate networks. In particular, we address two questions about the relationship between local entrepreneurs and the diaspora: First, are entrepreneurs who are based in cities where matching with prospective clients, new referrals and access to finance is easier, less likely to rely on the diaspora than entrepreneurs based in cities where this is harder? Second, do entrepreneurs who rely on the diaspora have better performing firms than those who do not tap into the expatriate community?¹

We outline a model of diaspora networks to examine how the local institutional environment for an entrepreneur and her prior career history might affect both her propensity to rely

¹These questions presume that there are frictions associated with choosing firm locations, and being able to tap into the diaspora that preclude all entrepreneurs from accessing them equally. We elaborate on, and provide evidence of these in more detail in Sections 2 and 3.

on the diaspora, as well as her firm’s revenue. We then use original data, collected through a survey sent to the CEO’s of all member firms of NASSCOM – India’s primary software association² – to examine the predictions of the model. To our knowledge, this is the first such systematic study of individual entrepreneurs in India’s software and services industry and therefore our findings on the backgrounds of the entrepreneurs and performance of their firms should also be of broader interest to those studying software and services firms in India.

We find that entrepreneurs located outside software hubs – in cities where access to finance and information flow on prospective clients is harder – rely significantly more on diaspora networks for business leads and financing. Relying on diaspora networks is associated with better performing firms, even more so for entrepreneurs based outside software hubs. We show that these results are consistent with a framework in which diaspora networks serve as important intermediaries for cross border trade, and are particularly helpful for domestic entrepreneurs in environments where formal institutions are weak and hence the informal barriers to trade are higher. However, we also find that the benefits from the diaspora accrue most to entrepreneurs who have previously lived abroad and returned to India, compared with those who have not lived abroad, suggesting that there may be important frictions in the ability for domestic entrepreneurs to tap into diaspora networks. Although we cannot rule out all sources of endogeneity, we provide a number of tests to rule out several alternative explanations. In particular, our instrumental variables coefficients suggest that our results do not seem to be driven by unobserved individual ability or by endogeneity in the location choice of entrepreneurs.

This study is part of a growing line of research documenting the important role that cross-border diaspora networks play in helping innovation and entrepreneurship in developing countries (Agarwal, Kapur and McHale, 2008; Rauch and Trindade, 2002; Saxenian, 2002; Kerr, 2007; Kapur, 2001) Our results complement prior cross-country work on the role of diaspora networks in international trade, by providing micro-evidence that is consistent with cross-border social networks serving as important substitutes to missing formal institutions

²NASSCOM (the National Association of Software and Service companies) is the primary business association for the Software and Services Industry in India and estimates that its members account for about 90% of industry revenues (www.nasscom.org)

in developing countries.

2 Diasporas and Domestic Entrepreneurs

Institutions that facilitate the formation and growth of new businesses are either weak, or completely missing in developing countries. Entrepreneurs based in developing countries therefore use a number of strategies to overcome these weaknesses, including a greater reliance on informal networks to help conduct business (McMillan and Woodruff, 1999; Rauch and Casella, 2001; Banerjee and Munshi, 2004). This paper examines diaspora, or cross-border networks, constituted by ties between expatriates from developing countries who are based abroad and entrepreneurs who live at ‘home’. Many studies have argued that expatriate networks seem to be vital in overcoming information barriers in cross-border business and also an important channel for driving knowledge and capital transfer across countries (Saxenian 2002, 2005, 2006; Kerr, 2007).

The focus of our study is the link between entrepreneurs in India’s software industry and the Indian Diaspora. The Indian software industry provides a good setting to study diaspora networks for several reasons. First, the vast majority of software business is conducted for clients outside India. Since output of software products and services is often hard to specify in advance or verify easily, and cross-border formal contracts are extremely hard to enforce, ‘relational contracting’ is especially important to generate business in this industry. While firms in the Indian software industry have been documented to use a number of formal mechanisms to overcome hurdles to business generation – such as the use of quality certifications (Arora et al 2001) or choice of contract structure (Banerjee and Duflo, 2000) – anecdotal accounts suggest that expatriate networks continue to play an important role in generating business and getting access to capital for entrepreneurs in India, specially because the industry is highly export oriented.³ Our own discussions with entrepreneurs in India support this view, with many individuals telling us that particularly in the early years of their company’s

³Devesh Kapur (2001) provides numerous examples where the Diasporas from developing countries have played a role in either enhancing or vouching for the reputation of businesses in developing countries.

existence, their network of Indians living abroad was invaluable in generating new business for their firms.

Second, software firms in India are spread across a number of cities with varying quality of local institutions. Software hubs lie at one end of this spectrum, where the high density of proximate firms in the same industry facilitate matching, referrals and better-monitoring of clients. Firms that don't directly compete with each other collaborate on marketing efforts, potential clients can stop by to visit local firms located close to other companies they have business with, and it is easier for firms to stay abreast with the latest trends and customer needs in the market (Chin et al, 1996). In addition, firms in hubs can avail of several formal institutional arrangements that reduce information asymmetries and promote matching with prospective clients. For example, one of the primary modes of formal networking and information exchange available to India's software entrepreneurs and foreign clients are conferences and seminars organized by NASSCOM. As can be seen from Table 1A, these conferences are run across a number of cities in India, but a large fraction of them are situated in one of the software hubs. This gives firms based in hubs an important advantage in terms of exposure to new business opportunities and to the "buzz" on new developments and trends in the market (Gertler, 2008).

Firms located outside hubs have far less access to these domestic networking channels and entrepreneurs located in these cities must look to other channels to compensate for the lack of formal and institutional networking opportunities available in hubs. Given the export intensity of this industry, one such channel might be the diaspora network. The variation in the local institutional environment for domestic entrepreneurs thus provides us with a natural testing ground to examine whether the difficulty of matching, referrals or monitoring within a city is related to entrepreneurs' reliance on diaspora networks to overcome hurdles to their business.

Third, India provides a good setting for such a study because the Indian diaspora is both extensive and varied, estimated at over 18 million people spanning 130 countries. A significant portion of the diaspora is composed of highly-skilled immigrants who maintain

strong ties to their home country. For example, Saxenian’s survey of Chinese and Indian immigrant professionals in Silicon Valley found that 80% of the Indian respondents exchanged information on American jobs or business opportunities with people in India, 67% served as an advisor or helped to arrange business contracts and 18% invested their own money in start-ups or venture funds in India (Saxenian 2002). Our study examines which entrepreneurs in India seem to rely most on these diaspora networks.

3 Local Institutions and Diaspora Networks

In order to guide the interpretation of our results, we first develop a simple model to examine how the institutional environment of the city where entrepreneurs are based might affect the extent to which they rely on informal channels such as expatriate networks and how this in turn would impact their firm’s performance.

In our model, revenue for entrepreneurs’ firms is based on the extent to which they ‘network’ for their business. Entrepreneurs choose the ‘optimal mix’ of networking using local institutions and diaspora networks in order to maximize firm revenue, a choice that is based on (1) each entrepreneur’s own costs of networking in their respective city, (2) their cost of accessing diaspora networks and (3) the extent to which local institutions and diaspora networks serve as complements rather than substitutes.

As we show in our model, the optimal investment in diaspora networks for a given entrepreneur varies considerably based on the extent to which the two serve as complements rather than substitutes. This allows us to generate specific predictions about the reliance on diaspora networks as well as firm performance for entrepreneurs, based on different levels of complementarity between informal networks and the local environment.

We consider a static economic environment consisting of I entrepreneurs who are located among J cities. Each city j is characterized by its ‘cost of local networking’ C_L which captures the ease with which individuals based in that city are able network to match with new clients, gain critical information for their business, and effectively contract with their

counterparties. The lower C_L is, the easier it is to effectively ‘network’. We assume that all individuals in a given city j face the same cost of local networking, so that the cost of local networking for an individual i , $C_{L_i} \in [0, 1]$ is identical within cities, but differs for individuals located in different cities.

While the environment of a city imposes some constraints on the ability of an entrepreneur to network locally, their cost of networking is also affected by their prior career histories. In particular, some entrepreneurs may already have an established informal network of contacts that can help with leads to new business and other critical information for their startup. In the highly export oriented software industry, one such very useful informal network is that of the expatriate community. If, for example, an entrepreneur has lived abroad at some point prior to starting their business, they will have built direct ties with the expatriate community and hence find it easier to sustain, and rely on, such a network for their business. They may be also connected to certain communities that make it easier for them to network abroad. We therefore also model individuals based on how hard it is for them to access the expatriate network. Let an individual’s type be defined by their cost of accessing the expatriate community $C_{E_i} \in [0, 1]$. In this framework, therefore, those whose cost of accessing the expatriate network is lower (say because they have lived abroad) will have a lower C_{E_i} .

Revenue for entrepreneur i ’s firm, Y_i is determined by (i) the extent to which she networks locally and with the diaspora⁴ and (ii) by the firm’s production function. We model firm revenue using the Constant Elasticity of Substitution (CES) production function. The CES production function has the attractive property that inputs are treated as either complements, or as substitutes depending on the parameters of the model. We can therefore model the optimal combination of local and diaspora networking for a given entrepreneur depending on the parameters of the model and generate testable hypotheses about how the relationship between the local environment and diaspora networks would vary based on whether these serve as complements or substitutes. Note that we assume that in this static framework C_{L_i} and C_{E_i} are fixed for a given individual⁵. Entrepreneurs therefore choose to allocate their

⁴For the purposes of this model, we normalize all other factors contributing to revenue to 1. In the empirical analysis we control explicitly for several firm-specific attributes.

⁵Clearly both the location decisions and career paths of individuals are endogenous in the long run and thus

networking time between the local and expatriate market in such a way that it maximizes firm revenue given the cost of accessing the local and expatriate network. By solving the entrepreneur's maximization problem, we can derive the optimal level of local and expatriate networking for individual i , D_i^* given costs C_{L_i} and C and γ (indexing the degree to which inputs are substitutes).

3.1 Relative Strength of diaspora Network

Our first two claims arise naturally from the constrained optimization problem faced by the entrepreneur:

Claim 1 D_i^* is decreasing in C_{E_i} . That is, those with a lower cost of accessing the diaspora will rely more on these networks

Claim 2 D_i^* is increasing in C_{L_i} . Those living in hubs will rely less on diaspora networks

However, we also examine whether an individual's attributes *interact* with their local environment in determining their reliance on the diaspora. That is, we examine whether those with a lower cost of accessing the diaspora in hubs rely differently on expatriate networks than those with a lower cost of accessing the diaspora who live outside hubs. The answer to this depends on whether diaspora networks serve as complements or substitutes to the local networking environment. The intuition behind this result is that if domestic and expatriate networks tend towards being perfect complements (say because they bring different and complementary knowledge, contacts and benefits to the entrepreneur), entrepreneurs want to combine them in equal proportions regardless of their cost of accessing these networks. Hence even if entrepreneurs have a very high cost of accessing the diaspora they will attempt to have some diaspora ties. Note, however, that although the *relative* strength of the networks will be the same in each of these cases, the *absolute* strength of both expatriate and local networks will be much greater for those in Hubs (driven by the lower cost of networking in hubs). This will lead to better firm performance for entrepreneurs based in Hubs. We return to this in more detail in Claim 6 below.

can be chosen by entrepreneurs. However, we treat them (and hence C_L and C_E) as fixed for the purposes of this static model. We examine the implication of relaxing this assumption in Section 4.

As E and L function more as substitutes, a change in C_E has a stronger effect on strength of diaspora networks when C_L is large than when it is small. That is, differences in the cost of accessing the expatriate network (such as when an individual has lived abroad vs. not) have a much greater impact on the relative strength of the diaspora network in small cities than in hubs. The intuition for this result is that when these networks serve substitutes, the relative mix between local and expatriate networks has less of an impact on firm performance. Hence entrepreneurs will choose their optimal mix of local and expatriate networks based on their cost of accessing these networks. Since the cost of local networking is higher outside hubs, those who have a low cost of accessing the expatriate network but live outside hubs will rely much more on expatriates than those who have a low cost of accessing expatriates but live in hubs (since the latter entrepreneurs can also access the local network at a low cost).

Claim 3 *If γ is such that E and L function as substitutes, being able to access the diaspora cheaply raises D_i^* more outside hubs than in hubs.*

In order to show these difference graphically, we plot simulated values of $\frac{E_i^*}{L_i^*}$ in Figure 1A and 1B. The charts plot the ratio of the external to the local network as a function of the ease of local networking when the networks serve as complements vs. substitutes. It can be seen that when E and L are substitutes (Figure 1B), there is a strong effect of the interaction between an individuals background and their local networking environment; when there are complements (Figure 1A), the effect is close to zero.

3.2 Networking Strategy and Firm Revenue

We now turn to the effect that the relative prices, C_E and C_L and the parameter γ have on firm revenue. Although the optimal mix of diaspora and local networking for the entrepreneur (i.e. the ‘networking strategy’) maximizes firm revenue for a given entrepreneur, different costs of accessing the local and expatriate networks will lead to different optimal combinations of the networks, and hence will also affect the absolute level of firm revenue. We therefore examine this relationship between networking strategy and Firm Revenue in this section. The next two claims also stem from the absolute advantage gained from lower costs of local and expatriate networking.

Claim 4 *Entrepreneurs who have a low cost of accessing the diaspora have higher firm revenue*

Claim 5 *Firms in hubs have higher revenue.*

Again, we also examine whether an individual's attributes *interact* with their local environment in determining firm revenue.

Claim 6 *If γ is such that E and L are substitutes, being able to access the diaspora cheaply increases revenue more outside hubs than in hubs.*

When networks serve as substitutes, being able to rely on the diaspora can help overcome a poor local networking environment. So entrepreneurs outside hubs who have a low cost of accessing the diaspora can make up for this weakness by relying more on the diaspora. The relative gains to relying on the diaspora are much greater outside hubs than in hubs, where the strong local networking conditions don't give those with a low cost of diaspora access a much greater advantage.

The opposite is true if the networks serve as complements. To see why this is the case, recall that when the networks serve as complements, entrepreneurs want to combine the networks in close to equal proportions. In hubs, those who have a low cost of accessing the diaspora have a low cost of accessing *both* networks and hence their 'stock' of inputs is very high – those who live in hubs but have a high cost of accessing the diaspora still need to network with the diaspora in order for the complementarity of the networks to yield benefits. Thus although their relative strength of the diaspora network is the same as those who have lived abroad, their stock of both networks is lower. Hence their revenue is lower. The same is true for those who live outside hubs, since they have a high cost of accessing at least one of the networks. Because of this, the *relative difference in the stock of networks* for those who live in hubs is greater than for those who live outside hubs. Hence a lower cost of accessing the diaspora has a stronger effect on revenue for entrepreneurs based in hubs than those based outside hubs. Again, in order to show these difference graphically, we plot simulated values of Y_i^* in Figure 3A and 3B. The charts plot simulated firm revenue as a function of different costs of local networking. It can be seen that when E and L are substitutes (Figure 3B), the

difference between firm revenue for those who have lived abroad and those who have not, is greater in cities with high costs of local networking, while if E and L are complements (Figure 3A), then the difference is greater in hubs where there is a low cost of local networking.

3.3 Empirical Strategy

Given the Claims in the Section above, we run the following two regressions to operationalize the model and test the relationships related to reliance on the diaspora network and firm revenue. First we look at how firm location and individuals' attributes are related to their reliance on the diaspora by estimating the regression:

$$D_i = \left(\frac{E_i}{L_i}\right) = \alpha_0 + \alpha_1 \frac{1}{C_{Li}} + \alpha_2 \frac{1}{C_{Ei}} + \alpha_3 \frac{1}{(C_{Li} * C_{Ei})} + \Psi X_i + \varepsilon_i \quad (1)$$

where C_{Li} and C_{Ei} are defined as before. Hence $\frac{1}{C_{Li}}$ is the degree to which the city in which the individual lives has strong local networking opportunities, $\frac{1}{C_{Ei}}$ is a variable that captures the ease with which an individual can access expatriate networks and X_i is a matrix of other individual-, firm- and city- level controls. The fourth term in equation (1) is the interaction between $\frac{1}{C_{Li}}$ and $\frac{1}{C_{Ei}}$ and therefore captures whether easier access to the diaspora has a different effect on diaspora reliance for entrepreneurs based in hubs compared to those who are not.

In the second regression, we look at how these same variables are related to the entrepreneur's firm revenue, by estimating:

$$Y_i = \beta_0 + \beta_1 \frac{1}{C_{Li}} + \beta_2 \frac{1}{C_{Ei}} + \beta_3 \frac{1}{(C_{Li} * C_{Ei})} + \Phi X_i + \xi_i \quad (2)$$

Based on the *Claims* from the model in the section above, we can make specific predictions about the coefficients in equations (1) and (2). In particular, note that since $\frac{\partial D_i}{\partial C_E} < 0$ (*Claim 1*) and $\frac{\partial D_i}{\partial C_L} > 0$ (*Claim 2*) we should expect $\alpha_2 > 0$ and $\alpha_1 < 0$. Similarly since $\frac{\partial Y_i}{\partial C_E} < 0$ (*Claim 5*) and $\frac{\partial Y_i}{\partial C_L} < 0$ (*Claim 6*) we should expect $\beta_2 > 0$ and $\beta_1 > 0$. Our predictions on the coefficients α_3 and β_3 depend on the value of γ – that is, whether the diaspora functions as a substitute or complement to local networking opportunities.

Recall from Claims 3 and 6 that the interaction between the cost of local and foreign networking has a different effect on the relative strength of the diaspora network and firm revenue for different values of γ (that is, whether inputs are complements or substitutes). Since γ is unobserved to us, but we have proxies for the other parameters in the model, the values of α_3 and β_3 can therefore help to shed light on extent to which γ treats the inputs as complements rather than substitutes. Our regressions (1) and (2) therefore also impose two sets of checks on the consistency of our theoretical model. First, we have specific predictions for the coefficients $\alpha_1, \alpha_2, \beta_1$ and β_2 that provide a consistency check on the framework of our theoretical model. Second, we look at whether two different estimations of the extent to which the networks serve as complements vs. substitutes –through the coefficients α_3 and β_3 – are also consistent with *each other*. That is, we want to make sure that if equation (1) implies that the networks function as substitutes, then equation (2) implies the same as well.

Since the conceptual model we outline is static, it does not explicitly deal with the endogeneity in the location choice of entrepreneurs. In a dynamic setting, however, we should expect that those with a high cost of accessing the diaspora may move to hubs precisely because they want to avail of the better institutional environment. In order to deal with this source of endogeneity, we also present the results from regressions using instrumental variables to account for the endogeneity in location choice, as well as discuss reasons why there may be considerable frictions to changing one’s location as an entrepreneur. Before turning to the results in Section 5, we outline the data that we use for this study.

4 Data

4.1 Survey Design and Implementation:

In November 2004, we administered a survey to the CEOs of all member-firms of the main industry associations for Indian Software Industry: the National Association of Software and Service companies, or NASSCOM. NASSCOM has approximately 900 members that represent over 90% of the revenues of the Indian software industry, making it a very attractive sample of firms to study. Moreover, since statistics on India’s software industry are generally based on

data gathered from NASSCOM’s member firms, this sample also provides a useful comparison and complement to other studies on the software industry in India (Athreya, 2005).

The survey was administered online, after significant work in designing and pre-testing both the questions and the web-interface. It included a number of questions relating to the respondents’ back-ground, such as their prior education, work experience and the time they had spend living or working out-side India. In addition, the survey included questions relating to their sources of funding and their most important business contacts in India and abroad.

We received 218 responses from the 920 emails sent out, which is a response rate of approximately 24%. After removing expatriate Indians and foreign CEOs were left with 207 responses of which we have complete data for 182⁶. In Appendix 1, we report the breakdown of firms by their city of location, firm age and firm size (number of employees), and compare these to data we have on entire population of NASSCOM member firms. As can be seen from these tables, the firms in our sample are very representative of the population of NASSCOM members along these observable metrics.

4.2 Main Variables:

As shown in equations (1) and (2), our main dependent variables of interest are (1) The strength of the diaspora network and (2) Entrepreneurs’ firm Revenue. Operationalizing the strength of an individual’s reliance on diaspora is difficult. In order to do so, we asked the respondents to list up to top 5 business contacts (not in their firm or paid consultants) who they had consulted in the previous three months for client leads, business generation and matters relating to their firm’s business. For each of these 5 contacts, we asked the respondents to list the city in which the contact was based, and whether the person was of Indian origin. We then coded those members of the network who were of Indian origin but lived outside India as being part of the Indian diaspora. Although this measure, which we call *DIASPORA_i* does not capture the strength of the entire diaspora network, it does proxy for reliance on the diaspora through the importance that CEOs place on their diaspora network.

⁶However, due to the fact that private firms often do not share their revenue data, we have revenue data for only 111 firms.

We also asked founder-CEOs about their sources of start-up capital, and the fraction of this that came from abroad. As an alternative measure of reliance on the diaspora therefore, we also look at the share of start-up capital for these entrepreneurs' firms that came from abroad. We call this variable *FOREIGNFRAC_i*. Many, but not all firms, report their revenue to NASSCOM as part of secondary data that the association collects from its members. We use revenue data that NASSCOM collected from its member firms for fiscal 2004 for this study. Our dependent variable for equation (2), is the log of revenue in Million Rupees, and is coded as *LOGREV_i*.

As shown in equations (1) and (2), our main right-hand-side variables are (1) the ease of local networking opportunities available to entrepreneurs in each city and (2) the ease with which entrepreneurs can access the diaspora. We proxy local networking opportunities by looking at networking events organized by NASSCOM for their members in the two years prior to our study, and look at the share of these events that were held in each of the cities in our sample. We call this variable *NETWORKSHARE* and use it to operationalize the ease of local networking in each city. In order to operationalize the ease of accessing the diaspora, we create a dummy variable that takes a value of 1 if the respondent had lived abroad for at least one year *prior* to their current job (either as a student or for work). Our premise here is that since individuals who have lived abroad will have developed direct links to expatriates based abroad, this would make it easier for them to network with the diaspora. We call this variable *LIVED ABROAD*.

We have a number of variables to control for unobserved heterogeneity at the individual, firm and city level. At the individual level, we control for the CEO's age, their educational background (as a proxy for human capital and 'ability') and whether they are currently working in the same city as they grew up. At the firm level, we control for the firm's age and size (in terms of number of employees), its business line(s), whether the firm is a subsidiary of an Indian or Multinational firm, whether it has a foreign headquarter. Finally, at the city level, we control for the city's population density and the share of total software exports from India that are constituted by the firms in that city. In addition, we control for the share of all export-oriented software firms that are based in the city, to control for both market structure

as well as informal sources of "buzz" that arise from local agglomeration economies.⁷ Each of these variables, and their sources are outlined in Appendix 2.

5 Results

5.1 Descriptive Statistics

In Table 2 we report t-test of how reliance on the diaspora and some of the main control variables vary by firms located in hubs vs those located outside hubs. As can be seen from Table 2, respondents and firms across hubs and non-hubs are very similar along demographic and educational characteristics. However, CEOs based outside hubs are much more likely to have one of their top contacts based outside India (55% compared to 44%). In addition, they are more likely to have one of their top contacts from the diaspora (36% compared to 23%). These numbers show another interesting fact – that *within* the group of contacts outside India, CEOs based outside hubs are more likely to rely on the diaspora. (65% of the their top foreign contacts are of Indian origin, compared to 52% for CEOs located in hubs).

In Figures 2, and 4, we break provide more detail on the relationship between the city in which entrepreneurs are located and both their reliance on the diaspora and their firm's revenue. Figure 2 plots the share of top contacts that are from the diaspora for each city, comparing these fractions for entrepreneurs who have lived abroad vs. those who have not. As can be seen from Figure 2, entrepreneurs who have lived abroad and now live outside hubs use the diaspora much more than those who have not lived abroad and live outside hubs. This difference is not present for entrepreneurs living in hubs. Comparing Figure 2 to Figures 1A and 1B, it can be seen that it maps closely to the lower panel (Figure 1B) – suggesting that in fact diaspora networks function as substitutes to poor local networking institutions rather than as complements to good ones.

Figure 4 plots firms revenue for each city, based on whether the entrepreneurs have lived abroad or not. As can be seen from Figure 4, the difference in firms performance between those who have lived abroad and not is much greater outside hubs than in hubs. Again,

⁷The share of exports and share of software firms is based on data from the The Software Technology Parks of India, which is a government body that oversees all software companies that have any export business.

comparing this results with the simulated results in Figures 3A and 3B, it can be seen that the findings are consistent with the lower panel – suggesting that in this case too the networks are seen to function as substitutes rather than complements to the local institutional environment.

5.2 Main Results

Although suggestive of our findings, the results shown in Figures 2 and 4 are only bivariate comparisons. In Tables 3 and 4, we therefore, report results from multivariate regressions, controlling for observable covariates at the individual, firm and city level. Table 3 reports the results of OLS Regressions where the dependent variable is the share of the CEO’s top 5 contacts that are from the diaspora. That is, we operationalize equation (1) by running the regression:

$$\begin{aligned} DIASPORA_i = & \alpha_0 + \alpha_1 NETWORKSHARE_i + \alpha_2 LIVEDABROAD_i \\ & + \alpha_3 (NETWORKSHARE_i * LIVEDABROAD_i) + \Psi X_i + \varepsilon_i \end{aligned} \quad (3)$$

where ε_i are clustered at the city level. As can be seen from Table 3, consistent with *Claim 1*, $\alpha_2 > 0$, so that CEOs who have lived abroad rely on the diaspora more. Moreover, consistent with *Claim 2*, $\alpha_1 < 0$, so that CEOs based in hubs rely less on the diaspora (although this is not statistically significant in the later models). Finally, similar to the results in Figure 2, CEOs who have lived abroad and are based outside hubs use the diaspora the most, that is $\alpha_3 < 0$ consistent with the view that diaspora networks function as substitutes to the local networking opportunities of entrepreneurs. These results are robust to the inclusion of Individual, Firm-level and City-level covariates into regressions.

In Table 4, we report results from OLS regressions where the dependent variable is the Log of Firm Revenue in 2004. That is, we operationalize equation (2) by running the regression:

$$\begin{aligned} LOGREV_i = & \beta_0 + \beta_1 NETWORKSHARE_i + \beta_2 LIVEDABROAD_i \\ & + \beta_3 (NETWORKSHARE_i * LIVEDABROAD_i) + \Phi X_i + \xi_i \end{aligned} \quad (4)$$

Again, as with Figure 4 and consistent with the view that local and diaspora networks are substitutes, we find that $\beta_3 < 0$. That is, the benefit from ‘cheaper’ access to the diaspora is greater for firms outside hubs than in hubs. We also find that *Claims 5 and 6* are supported, in that both β_1 and β_2 are positive so that firms based in hubs have higher revenue, and the firms for CEOs who have lived abroad have higher revenue.

5.3 Robustness Checks

5.3.1 Alternate Measure of Reliance on the diaspora

One of the concerns with our networking measure is that the top 5 contacts may not be representative of all the diaspora contacts an entrepreneur has. We therefore also cross-checked this measure with another other question inserted in the survey that asked them the fraction of their overall network that was composed of Indians based outside India. The correlation with this measure was 0.39 and significant at 0.1%. In addition, we look at a second measure of the reliance on the diaspora – the share of foreign funding that the founders received from abroad. That is, we re-run equation (3) with $FOREIGNFRAC_i$ as the dependent variable, and the number of commercial bank branches in that city ($BANKS_i$) as the main measure of the strength of local financial institutions. These results are reported in Table 5. As with Table 3, we find that those who have lived abroad but live outside hubs have the highest fraction of foreign funding. Moreover, those living in cities with worse formal financial institutions are more likely to rely on the diaspora for capital, and even more so if they have lived abroad. This result suggests that the reliance on diaspora for capital follows a similar mechanism as the reliance on the diaspora for leads. It is the entrepreneurs based in small cities who have access to diaspora who tap into it for both leads and capital.

5.3.2 Unobserved Entrepreneur and Firm Attributes

One concern with the results that we have shown so far is that those who have been abroad are different in a number of ways (such as ability or wealth) and that the returns to these attributes are systematically different in hubs and non-hubs. For example, “weaker entrepreneurs” might settle in smaller cities and also rely more on the diaspora. A second concern is that entrepreneurs choose their firm location based on their abilities to access domestic and

expatriate networks that are most suited to their firms, so that in fact we are just picking up unobserved differences in the need to access diaspora networks across firms.

We provide a number of checks that suggest that our results are not being driven by such spurious correlations. First, we control for individual ability using a dummy of whether the CEO went to one of the elite engineering colleges in India – the Indian Institutes of Technology or the Indian Institutes of Management. This seems to be a good measure of individual ability, in that entrepreneurs who went to one of these universities have firms with higher revenue per employee. Yet, entrepreneurs who went to IIT don't use the diaspora less, or place less importance on their diaspora networks. In fact, the coefficients suggest that if anything they use the diaspora more (though not significant)⁸. We also control for whether the individual is based in the same city in which they went to high school, and find that those who relocated to a given city (perhaps in order to make the most of the networking opportunities for the firm they want to start) do not seem to rely differently to diaspora networks or external finance than those who remained in the same city.

Finally, we re-run the models in Tables 3, 4, and 5, but instrument for the endogeneity in the entrepreneur's location decision and the decision to go abroad. We use three sets of instrumental variables. First, we use a dummy variable that takes a value of 1 if the entrepreneur went to high school in one of the cities we coded as hubs. Individuals have a strong preference for locating close to their family home, or the city in which they grew up even if this is not ideal for their business. We find this to be true in our data as well. Second, we use the encyclopaedia of Indian Surnames to code the ethnicity of the entrepreneur, which is associated both with domestic location choice, and with the probability of going abroad. We heard quotes such as the following in our discussions with entrepreneurs: "being from South India, I wanted to start my business here because of the familiarity" or "people prefer to start their business in their home town – it gives them a sense of familiarity". This pull towards the family home often led entrepreneurs to locate outside hubs. It was also associated with the probability of going abroad. For example, using the telephone directories

⁸We also do not find that the distribution of individuals who went to IIT varies consistently by their location suggesting that at least on this observable measure of individual ability, there is no sorting by cities

in the US, we find significant variation in the propensity that these surnames are found in the US. Finally, we use a dummy variable that takes a value of 1 if the respondent’s parents had a family business in the past, since individuals may choose to return to their home town to work with their families and also be less likely to go abroad if their family business requires their participation. We report the results of these instrumental variable estimates in Table 6. The columns in Table 6 compare the OLS coefficients model (6) of Tables 3, 4, and 5 with their respective Instrumental Variables counterparts. The point estimates on the IV regressions suggest that indeed, there did seem to be some amount of endogeneity in the choice of location and going abroad; once corrected for, the effects seem to be stronger than those reported in the OLS regressions. However, due to the much larger standard errors associated with the IV regressions, the coefficients are not estimated as precisely and hence are not statistically significant.

It is possible that our results may be driven by selection: that is, since it is harder to do business in small cities, firms in small cities may be less likely to survive relative to firms in hubs, unless they have access to diaspora networks. Since we only surveyed the CEOs of surviving firms, the firms outside the hubs might be more likely to be ones where the CEOs relied on the diaspora. While this explanation is plausible, and cannot be ruled out, it is equivalent to a strong-version of model that we develop in that it is the entrepreneurs in small cities without connections to the diaspora do so poorly that they are forced to shut down.

The fact that we are finding consistent differences between entrepreneurs’ location and firm performance raises two important questions. First, what is it that makes the cost of local networking for entrepreneurs based outside software hubs so high? Our discussions with the entrepreneurs revealed substantial ‘frictions’ in networking opportunities of entrepreneurs based outside hubs. Many entrepreneurs said they found it hard to break into the social networks in hubs. On the other hand, those in hubs such as Bangalore told us that it was very easy to network locally. ‘People just swing by’ and ‘walking into a hotel in Bangalore is just like walking into a hotel in the US’. The second question our results raise is why entrepreneurs do not all either locate their firms in hubs or use the diaspora more intensively? Our IV results

are consistent with previous studies on entrepreneurs' location decisions that have found that there is significant 'inertia' in choosing where to locate their business (Buenstorf and Klepper 2005; Figueiredo, Guimaraes and Woodward, 2000) and that individuals often choose the location for their businesses for reasons other than the networking and financing needs. Why, then, do entrepreneurs in small cities not all rely more on the diaspora when the benefits seem so large? Consistent with the estimates in the regressions, we find that entrepreneurs who do not have strong ties to the diaspora find it hard to break into the diaspora networks. Some entrepreneurs living in the smaller cities explicitly told us that they had a hard time getting Indian expatriates to help them with business, and that they wished they had more connections with the diaspora to help them sell business more aggressively.

6 Conclusions

While several recent studies on cross-border ethnic networks have highlighted the important role that they might play in facilitating entrepreneurship in developing countries, little is known about the extent to which domestic entrepreneurs rely on the diaspora and whether this varies systematically by the characteristics of the entrepreneurs or their local business environment. In this paper, we develop a conceptual model of diaspora networks to examine whether diaspora networks serve as substitutes to the functioning of the local business environment, thereby helping entrepreneurs to circumvent the barriers to trade arising from imperfect institutions in developing countries.

Our results suggest that entrepreneurs who live in hubs, where the local institutional environment is stronger, are able to avail of these benefits and do not necessarily gain significantly from relying more on diaspora networks. Entrepreneurs based in smaller cities, however, are faced with a weaker institutional environment, where information asymmetries create barriers to trade. Those located in such cities who have lived abroad are much more likely to tap into diaspora networks for help with their business; relying on these diaspora networks is also associated with significantly better firm performance relative to entrepreneurs who do not avail of their help. Moreover, our findings suggest that frictions preventing all entrepreneurs from locating in hubs or from being able to access diaspora networks allow

these differences to persist over time. They also suggest that despite the numerous formal contracting mechanisms to overcome the barriers to international trade, there is still scope for informal networks to impact strategies and outcomes for entrepreneurial firms.

Our results are also consistent with the recent research by Agarwal, Kapur and McHale (2006) who use firm patenting data to argue that “co-location and co-ethnicity seem to substitute rather than complement each other in terms of knowledge flows”. Our findings shed additional light on the mechanism through which these networks work. Given the fact that it is those who have lived abroad prior to starting their business who are most likely to access the diaspora networks, our findings also suggest that ‘brain circulation’ might be critical for developing countries to tap into their diaspora. That is, these networks are successful not just because of the expatriates who live abroad, but because some of the expatriates have returned back home and know how to effectively tap into the diaspora. These results should also be of relevance to policy makers in developing countries aiming to leverage their diasporas to help with domestic entrepreneurship. The implications of our findings is that they may be better off developing links between the diaspora and smaller cities rather than with hubs.

References

- [1] Acemoglu, D, and S. Johnson. 2005 “Unbundling Institutions” *Journal of Political Economy*, 13, pp.949-995
- [2] Agarwal, A., D. Kapur, and J. McHale. 2006. “Birds of a Feather - Better Together? How Co-Ethnicity and Co-Location Influence Knowledge Flow Patterns.” Working Paper.
- [3] Agarwal, A., D. Kapur, and J. McHale. 2008. “Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation.” Working Paper.
- [4] Arora, A., V. S. Arunachalam, J. Asundi, and R. Fernandes. 2001. “The Indian software services industry.” *Research Policy*, 30:8, pp. 1267-87.

- [5] Athreye, S. S. 2005. "The Indian software industry and its evolving service capability." *Industrial and Corporate Change*, 14:3, pp. 393-418.
- [6] Banerjee, A. and K. Munshi. 2004. "How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur." *Review of Economic Studies*, 71:1, pp. 19-42.
- [7] Banerjee, A. V. and E. Duflo. 2000. "Reputation effects and the limits of contracting: A study of the Indian software industry." *Quarterly Journal of Economics*, 115:3, pp. 989-1017.
- [8] Basant and Chandra. 2001. "Inter Firm Linkages in the IT Industry in India: A Case Study of Telecom Technologies." IIM Working Paper.
- [9] Borjas, G. J. 1994. "The Economics of Immigration." *Journal of Economic Literature*, 32:4, pp. 1667-717.
- [10] Buenstorf, G. and S. Klepper. 2005. "Regional Birth Potential, Agglomeration Economies and Home Bias in the Location of Domestic Firms." Working Paper.
- [11] Chin, K.S., I.J. Yoon and D.A. Smith. 1996. "Immigrant small business and international economic linkage: a case of the Korean wig business in Los Angeles, 1968-1977", *International Migration Review* 30, pp. 485-510
- [12] Casella, A. and J.E. Rauch. 2002 "Anonymous market and group ties in international trade" *Journal of International Economics* 58 pp.19-47
- [13] Curtin, P. D. 1984. "Cross-Cultural Trade in World History" Cambridge: Cambridge University.Press
- [14] Desai, M. A., D. Kapur, and J. McHale. 2004. "Sharing the spoils: Taxing international human capital flows." *International Tax and Public Finance*, 11:5, pp. 663-93.
- [15] Figueiredo, O., P. Guimaraes, and D. Woodward. 2002. "Home-field advantage: location decisions of Portuguese entrepreneurs." *Journal of Urban Economics*, 52:2, pp. 341-61.
- [16] Gertler, M. 2008. "Buzz Without Being There?" working paper

- [17] Gillespie, K., L. Riddle, E. Sayre, and D. Sturges. 1999. "Diaspora interest in homeland investment." *Journal of International Business Studies*, 30:3, pp. 623-34.
- [18] Gould, D. M. 1994. "Immigrant Links to the Home Country - Empirical Implications for United-States Bilateral Trade-Flows." *Review of Economics and Statistics*, 76:2, pp. 302-16.
- [19] Greif, A. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review* 83:3, pp. 525-48.
- [20] Kapur, D. 2001. "Diasporas and Technology Transfer." *Journal of Human Development*, 2, pp. 265-86.
- [21] Kapur, D. and J. McHale. 2002. "Sojourns and Software: Internationally Mobile Human Capital and High-Tech Industry Development in India, Ireland and Israel." Working Paper.
- [22] Keller, W. 2004. "International technology diffusion." *Journal of Economic Literature*, 42:3, pp. 752-82.
- [23] Kerr, W. 2007. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics*
- [24] Kremer, M. 1993. "The O-Ring Theory of Economic-Development." *Quarterly Journal of Economics*, 108:3, pp. 551-75.
- [25] McCormick, B. and J. Wahba. 2001. "Overseas work experience, savings and entrepreneurship amongst return migrants to LDCs." *Scottish Journal of Political Economy*, 48:2, pp. 164-78.
- [26] McMillan, J. and C. Woodruff. 1999. "Interfirm relationships and informal credit in Vietnam." *Quarterly Journal of Economics*, 114:4, pp. 1285-320.
- [27] Rauch, J. E. 1999. "Networks versus markets in international trade." *Journal of International Economics*, 48:1, pp. 7-35.

- [28] Rauch, J. E. 2001. "Business and social networks in international trade." *Journal of Economic Literature*, 39:4, pp. 1177-203.
- [29] Rauch, J. E. and A Casella eds. 2001. *Networks and Markets*. New York: Russell Sage Foundation.
- [30] Rauch, J. E. and V. Trindade. 2002. "Ethnic Chinese networks in international trade." *Review of Economics and Statistics*, 84:1, pp. 116-30.
- [31] Roy, S. and S.H.M. Rizvi. 2002. *Encyclopaedia of Indian Surnames*. New Delhi: B.R. Publishing Corporation.
- [32] Saxenian, A. 2002. "Local and Global Networks of Immigrant Professionals in Silicon Valley." Public Policy Institute of California.
- [33] Saxenian, A. 2005. "From brain drain to brain circulation: Transnational communities and regional upgrading in India and China." *Studies in Comparative International Development*, 40:2, pp. 35-61.
- [34] Saxenian, A. 2006. *The New Argonauts: Regional Advantage in a Global Economy*. Boston: Harvard Business School Press.
- [35] Saxenian, A. and C. Y. Li. 2003. "Bay-to-bay strategic alliances: the network linkages between Taiwan and the US venture capital industries." *International Journal of Technology Management*, 25:1-2, pp. 136-50.
- [36] Sorenson, O. and P. G. Audia. 2000. "The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940-1989." *American Journal of Sociology*, 106:2, pp. 424-61.
- [37] Uzzi, B. 1996. "The sources and consequences of embeddedness for the economic performance of organizations: The network effect." *American Sociological Review*, 61:4, pp. 674-98.

FIGURES 1A, 1B: SIMULATED RELIANCE ON DIASPORA NETWORKS

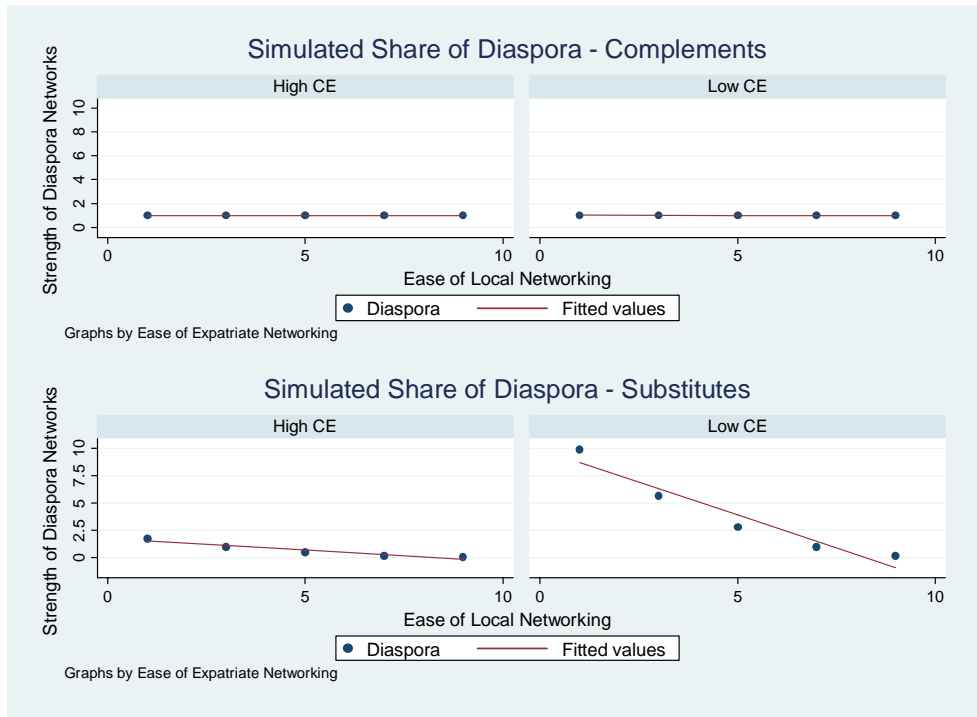
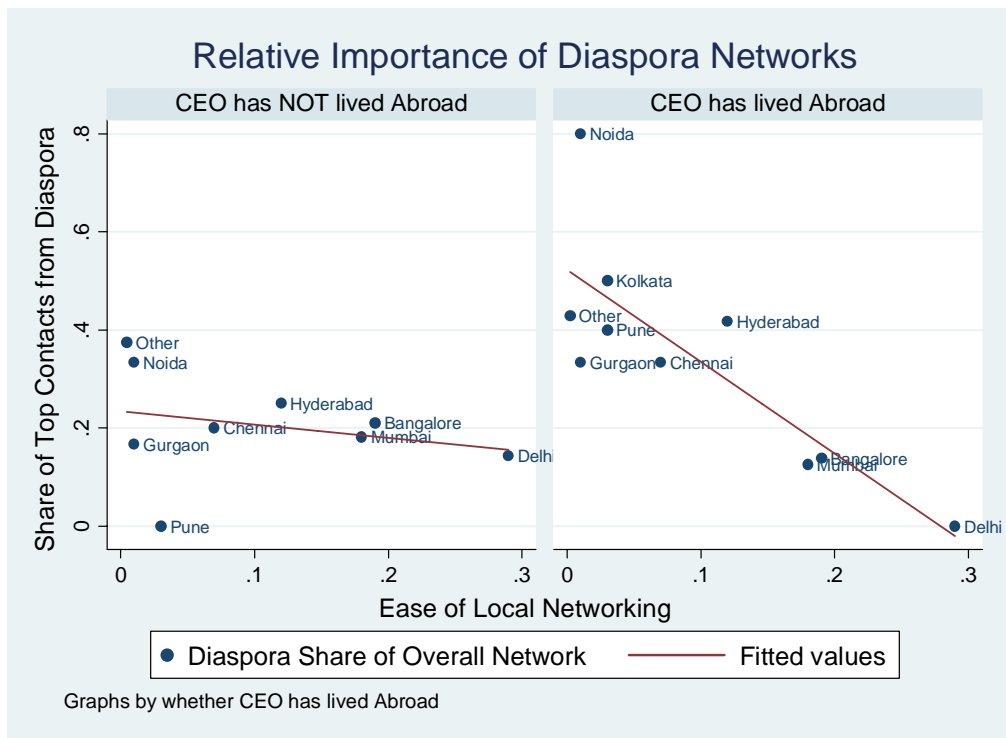


FIGURE 2: ACTUAL RELIANCE ON DIASPORA NETWORKS



FIGURES 3A, 3B: SIMULATED FIRM REVENUE

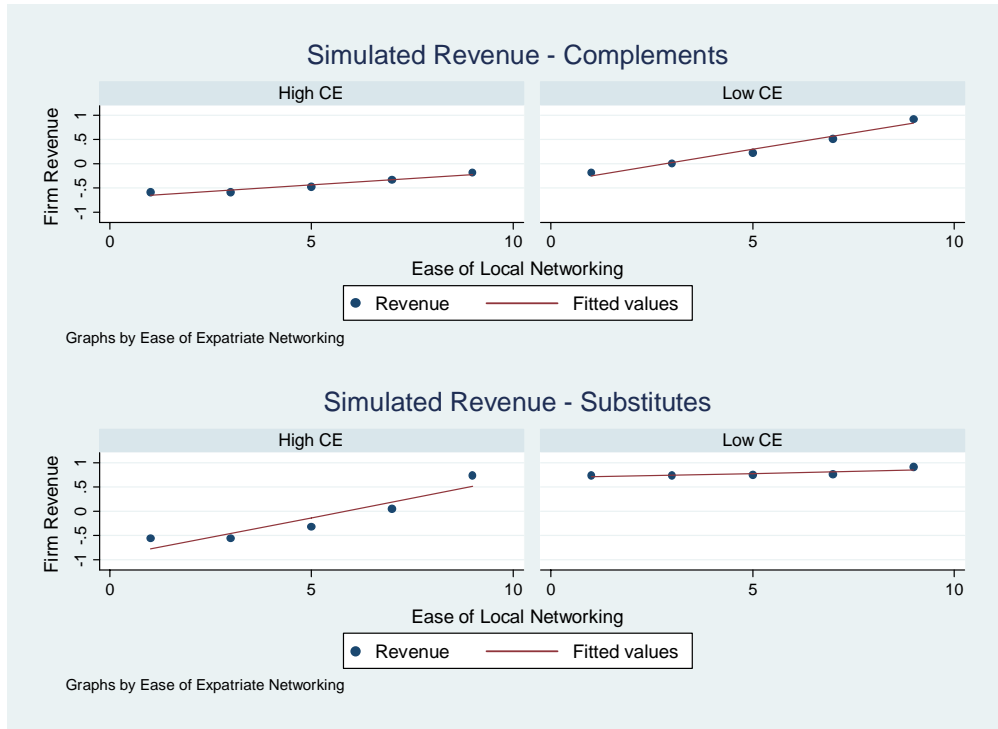


FIGURE 4: ACTUAL FIRM REVENUE

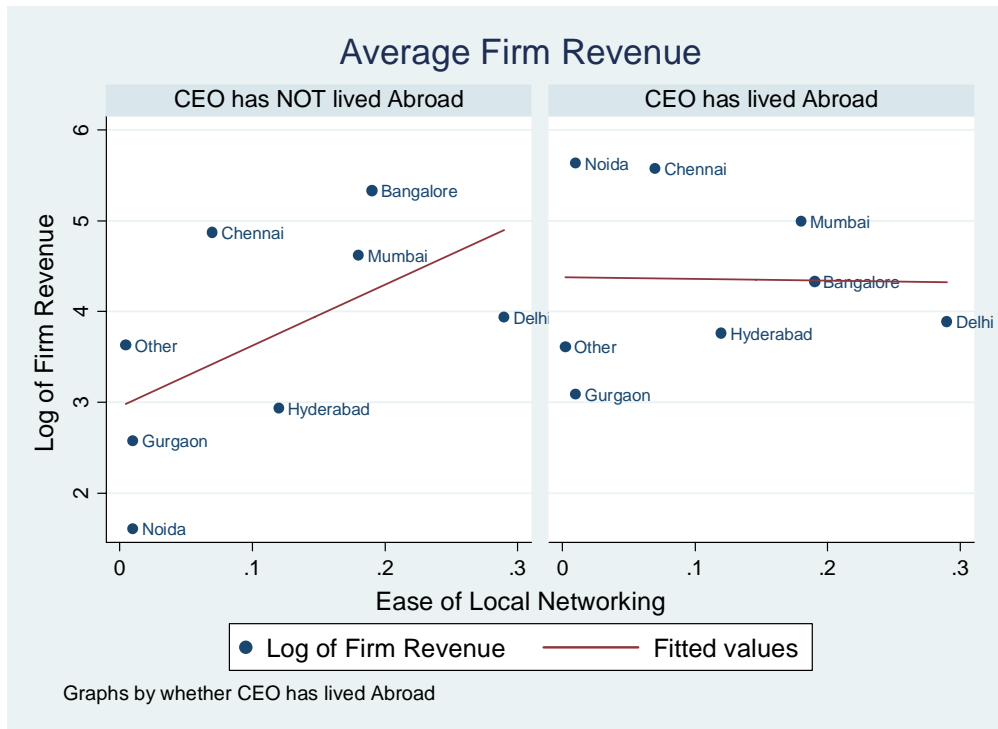


TABLE 1: MEASURES OF NETWORKING AND FINANCING COST ACROSS CITIES

City	Share of NASSCOM Events	Share of all Software Firms	Share of all Software Exports	Number of Commercial Bank Branches in 2000	Population Rank
Delhi	29%	9%	8%	1446	3
Bangalore	19%	20%	35%	806	5
Mumbai	18%	17%	8%	1556	1
Hyderabad	12%	11%	10%	578	5
Chennai	7%	11%	16%	838	4
Kolkata	3%	5%	2%	1188	2
Pune	3%	6%	7%	350	8
Gurgaon	1%	6%	8%	56	152
Noida	1%	5%	4%	51	140
Other (average)	0%	1%	1%	180	30

Note: "Other" cities include Ahmedabad, Bhubaneswar, Chandigarh, Cochin, Coimbatore, Indore, Jaipur, Nagpur, Pondicherry, Raipur, Rajkot, Trivandrum, and Vadodara; Population Rank for these cities is average across all

Source: 2002-2003 NASSCOM Directories; Software Technology Parks of India Directories, Reserve Bank of India, Census of India

TABLE 2: SUMMARY STATISTICS ON CEOs AND FIRMS BY FIRM LOCATION

	Total Sample	Software Hub ¹	Non-Hub City ²	Two-Tailed T-test for Equality in Means
Total Responses	207	140	67	
Complete Responses	182	127	55	
Firm Age (Years)	8.1	7.8	8.8	-0.96
Firm Size (Employees)	733	824	524	0.85
Firm Revenue (Million Rupees)	88	89	87	0.04
Fraction that are Subsidiaries of MNC or Indian Business Group	24%	26%	18%	1.13
Age of CEO (Years)	43	42	44	-1.44
Fraction of CEOs who have lived abroad	58%	55%	64%	-1.07
Fraction with a technical degree	54%	54%	55%	-0.02
Fraction with an MBA	37%	36%	40%	-0.48
Fraction who have studied at an IIT ³	21%	23%	18%	0.69
Fraction who have studied at an IIM ³	14%	13%	15%	-0.2
Fraction of Top 5 Contacts based outside India	47%	44%	55%	-1.99**
Fraction of Top 5 Contacts from Diaspora	27%	23%	36%	-2.94***

* Significant at 10% ** Significant at 5% *** Significant at 1%

1: Coded as Hub if CEO is based in Bangalore, Chennai, Hyderabad, Mumbai or New Delhi

2: Coded as Non-Hub if CEO is based in Kolkata, Pune, Gurgaon, Noida or one of the "Other" Cities

3: IIT (Indian Institutes of Technology) and IIM (Indian Institutes of Management) are elite educational institutions in India

Source: Survey Data; Firm Revenue from NASSCOM

TABLE 3: RELIANCE ON DIASPORA NETWORKS*OLS Regressions: Dependent Variable is Fraction of Top 5 Contacts that are from Diaspora*

	(1)	(2)	(3)	(4)	(5)	(6)
NETWORKSHARE	-0.731*** (0.250)	-0.701*** (0.230)	-0.172 (0.330)	-0.232 (0.270)	-0.208 (0.300)	-0.071 (0.290)
LIVED ABROAD		0.0597 (0.045)	0.182** (0.074)	0.177** (0.074)	0.196** (0.087)	0.203** (0.085)
NETSHARE x LIVED ABROAD			-0.929** (0.380)	-0.903** (0.380)	-0.985** (0.460)	-1.083** (0.430)
City-Level Covariates	No	No	No	Yes	Yes	Yes
Firm-Level Covariates	No	No	No	No	Yes	Yes
Individual-Level Covariates	No	No	No	No	No	Yes
Observations	182	182	182	182	182	182
R-squared	0.05	0.07	0.09	0.09	0.12	0.13

Robust standard errors in parentheses, clustered by 19 cities in the sample

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking and corresponds to the term HUB in equation (12); LIVED ABROAD is a dummy variable that equals 1 if the CEO has lived abroad to study or work for at least a year prior to working at current job, and corresponds to the term ACCESS in equation (12). The interaction term in the table above corresponds to the interaction term in equation (12)

TABLE 4: FIRM REVENUE*OLS Regressions: Dependent Variable is Log Revenue*

	(1)	(2)	(3)	(4)	(5)	(6)
NETWORKSHARE	0.0166 (0.210)	0.0365 (0.200)	0.441 (0.380)	0.177 (0.240)	0.306* (0.160)	0.435** (0.180)
LIVED ABROAD		0.0142 (0.040)	0.101 (0.079)	0.108 (0.086)	0.0578** (0.026)	0.0605** (0.026)
NETSHARE x LIVED ABROAD			-0.644 (0.450)	-0.71 (0.490)	-0.424** (0.170)	-0.469** (0.190)
City-Level Covariates	No	No	No	Yes	Yes	Yes
Firm-Level Covariates	No	No	No	No	Yes	Yes
Individual-Level Covariates	No	No	No	No	No	Yes
Observations	111	111	111	111	111	111
R-Squared	0.01	0.01	0.02	0.07	0.73	0.77

Robust standard errors in parentheses, clustered by city

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking and corresponds to the term HUB in equation (13); LIVED ABROAD is a dummy variable that equals 1 if the CEO has lived abroad to study or work for at least a year prior to working at current job, and corresponds to the term ACCESS in equation (13). The interaction term in the table above corresponds to the interaction term in equation (13)

TABLE 5: FRACTION OF FOREIGN FUNDING RAISED AT STARTUP*OLS Regressions: Dependent Variable is Fraction of Foreign Funding*

	(1)	(2)	(3)	(4)	(5)	(6)
BANKS	-0.112* (0.061)	-0.082 (0.054)	-0.030 (0.056)	-0.178 (0.150)	-0.107 (0.150)	-0.129 (0.150)
LIVED ABROAD		0.192** (0.072)	0.353*** (0.110)	0.311*** (0.100)	0.335*** (0.096)	0.308*** (0.082)
BANKS x LIVED ABROAD			-0.198* (0.110)	-0.158 (0.110)	-0.174* (0.092)	-0.141* (0.071)
City-Level Covariates	No	No	No	Yes	Yes	Yes
Firm-Level Covariates	No	No	No	No	Yes	Yes
Individual-Level Covariates	No	No	No	No	No	Yes
Observations	109	109	109	109	109	109
R-Squared	0.02	0.08	0.1	0.12	0.26	0.31

Robust standard errors in parentheses, clustered by city

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: BANKS (Number of Commercial Bank Branches in City / 1000) measures the development of the formal financial sector; LIVED ABROAD is a dummy variable that equals 1 if the CEO has lived abroad to study or work for at least a year prior to working at current job

TABLE 6: INSTRUMENTAL VARIABLE ESTIMATES

	<i>Reliance on Diaspora</i>		<i>Firm Revenue</i>		<i>Foreign Funding</i>	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
NETWORKSHARE	-0.071 (0.290)	-4.567 (3.250)	0.435** (0.180)	1.222 (3.450)		
LIVED ABROAD	0.203** (0.085)	0.531 (0.490)	0.061** (0.026)	0.287 (0.230)		
NETSHARE x LIVED ABROAD	-1.083** (0.430)	-1.045 (2.650)	-0.469** (0.190)	-1.832 (1.080)		
BANKS					-0.129 (0.150)	-0.483 (0.410)
LIVED ABROAD					0.308*** (0.082)	0.644 (0.660)
BANKS x LIVED ABROAD					-0.141* (0.071)	-0.440 (0.660)
City-Level Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Firm-Level Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Individual-Level Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	182	182	111	111	109	109

Robust standard errors in parentheses, clustered by city

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking , BANKS (Number of Commercial Bank Branches in City / 1000) measures the development of the formal financial sector; LIVED ABROAD is a dummy variable that equals 1 if the CEO has lived abroad to study or work for at least a year prior to working at current job

APPENDIX 1

Table 1a: Distribution of Firms by City

	<i>Number of Firms in sample</i>	<i>Fraction of Firms in Sample</i>	<i>Fraction of all NASSCOM member firms</i>
Bangalore	54	26%	23%
Mumbai	43	21%	19%
Hyderabad	17	8%	8%
Pune	17	8%	7%
New Delhi	15	7%	10%
Noida	14	7%	5%
Chennai	11	5%	10%
Gurgaon	10	5%	6%
Kolkata	4	2%	3%
Others	22	11%	10%
	207	100%	100%

Table 1b: Distribution of Firms by Year of Founding

<i>Year of Founding</i>	<i>Number of Firms in sample</i>	<i>Fraction of Firms in Sample</i>	<i>Fraction of all NASSCOM member firms</i>
before 1990	26	13%	12%
1990-1994	36	18%	17%
1995	4	2%	6%
1996	4	2%	7%
1997	11	5%	8%
1998	18	9%	6%
1999	22	11%	12%
2000	34	17%	15%
2001	14	7%	6%
2002	19	9%	6%
2003	11	5%	4%
2004	5	2%	2%
	204	100%	100%

Table 1c: Distribution of Firms by Number of Employees

<i>Number of Employees</i>	<i>Number of Firms in sample</i>	<i>Fraction of Firms in Sample</i>	<i>Fraction of all NASSCOM member firms</i>
Upto 10	7	3%	2%
11-50	47	23%	17%
51-150	46	23%	27%
151-500	60	29%	30%
501-2500	32	16%	18%
Greater than 2500	12	6%	6%
	204	100%	100%

APPENDIX 2: COVARIATES IN REGRESSIONS

Variable	Description	Source
AGE	Respondent's Age	Survey
AGE2	Respondent's Age - Squared	Survey
IIT/IIM	Respondent studied at one of the Indian Institutes of Technology or Indian Institutes of Management	Survey
SAMEHI	Respondent is based in same city he or she went to highschool	Survey
FIRMSIZE	Firm's Size	Survey
FIRMAGE	Firm's Age	Survey
SUBSID	Firm is a subsidiary of an Indian or Multinational company	NASSCOM/ Company Website
FOREIGNHQ	Firm has a foreign headquarter	NASSCOM/ Company Website
BIZLINE	Dummies for business line of the firm (embedded software, IT-enabled services IT-software, Infrastructure & Support Services, Systems Integrator, and/or Product Development)	NASSCOM
POPDENSITY	Population Density of City	Census of India, Wikipedia
AGGLOMERATION	Share of Total STPI Firms in City	Software Technology Parks of India
SHSOFTEXP	Share of Software Exports from the city	Software Technology Parks of India

APPENDIX 3: RESPONDENT 'ETHNICITY'

Breakdown of CEOs' Ethnicity Based on their Surnames
(Source: Encyclopaedia of Indian Surnames)

	Frequency	Fraction	Fraction in each category that are based in Hubs
North Indian	52	29%	60%
South Indian	36	20%	77%
East Indian	14	8%	64%
West Indian	30	16%	56%
Other or Unknown Ethnicities	50	27%	84%
Total	182	100%	70%

Note: North Indian Includes those from Uttar Pradesh, Madhya Pradesh, Bihar, Rajasthan, Punjab, Haryana, Himachal Pradesh, Uttaranchal and Jammu & Kashmir; South Indian includes individuals from Andhra Pradesh, Karnataka, Tamil Nadu and Kerala; East Indian includes individuals from West Bengal, Orissa, Assam and other North Eastern States; West Indian Includes Gujarat and Maharashtra; Other includes Christians, Muslims, Parsis as well as individuals for whom the ethnicity was ambiguous