

Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry

**Ramana Nanda
Tarun Khanna**

Working Paper

08-003

Copyright © 2007, 2009 by Ramana Nanda and Tarun Khanna

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

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 (Revised)
February 5, 2009

Abstract

This study explores the importance of cross-border social networks for entrepreneurs 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 have previously lived outside India rely significantly more on diaspora networks for business leads and financing. This is especially true for entrepreneurs who are based outside software hubs – where getting leads to new businesses and accessing finance is more difficult. 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 domestic 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, John McHale 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, the Myron Weiner Seminar on International Migration 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; 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 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, little is known about which entrepreneurs in developing countries rely most on diaspora networks. Is it those who face greater transaction costs and barriers to trade that rely most on the diaspora or are these primarily hub-to-hub ties between entrepreneurs in developing countries and those that live abroad? Anecdotal accounts of the links between local entrepreneurs and the expatriate community suggest that it may in fact be the latter (Saxenian 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 order to examine this question in more detail, 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. We use original data, collected through a survey sent to the CEO's of all member firms of NASSCOM (India's primary software association¹) to examine how the career experiences of entrepreneurs as well as the local institutional environment where they are based might impact their propensity to rely on diaspora networks for business leads and financing. 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

¹NASSCOM (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)

performance of their firms should also be of broader interest to those studying software and services firms in India.

We find that the entrepreneurs who have previously lived abroad (and hence have an easier time accessing the expatriate networks) rely significantly more on diaspora networks for business leads and financing and also have better performing firms. However, the importance of having lived abroad is far greater for entrepreneurs based outside the software hubs – in cities with weak networking institutions or where access to bank finance is limited. We show that these results are consistent with a framework in which diaspora networks serve as important intermediaries for cross border business, but are most helpful for domestic entrepreneurs in environments where networking and financing institutions are weak and hence the barriers to running a successful business are higher.

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; Kapur 2001; Kerr 2008; Rauch and Trindade 2002) 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 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 (Rauch and Casella 2001). 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 (Agarwal, Kapur and McHale 2008; Foley and Kerr 2008; Gillespie et al 1999; Kerr 2008; Saxenian 2006).

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.² 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 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 (Sorenson and Audia 2000). 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

²Kapur (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.

clients are conferences and seminars organized by NASSCOM. As can be seen from Table 1, 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.

2.1 Hypotheses

In order to guide the interpretation of our findings, we develop a simple framework within which to examine the networking strategies of local entrepreneurs. In this framework, revenue for entrepreneurs' firms is based on the extent to which they can successfully generate new business by tapping into their networks. Given the constraints on their time, entrepreneurs

face a choice between the extent to which they should rely on diaspora or local networks in order to generate business and maximize firm revenue, a choice that is based on (1) each entrepreneur’s cost of accessing diaspora networks, (2) their own costs of networking in their respective city and (3) the extent to which local institutions and diaspora networks serve as complements rather than substitutes.³

Using this framework, we therefore classify individuals along two key dimensions when studying their reliance on diaspora networks, as shown in the regression equation below:

$$DIASPORA_i = \alpha_0 + \alpha_1 LIVEDABROAD_i + \alpha_2 HUB_i + \alpha_3 (HUB_i * LIVEDABROAD_i) + \Psi X_i + \varepsilon_i \quad (1)$$

First, we examine whether or not these local entrepreneurs have lived outside India at some point during their career – as a measure of their cost of accessing the diaspora. We hypothesize that those who have previously lived abroad will have a lower cost of accessing diaspora networks, as they are more likely to have developed direct ties with the expatriate community and hence find it easier to sustain, and rely on, such a network for their business. Hence, all else held constant, we would expect that those who have lived abroad will tend to rely more on diaspora networks for their business than those who have not lived abroad. The second dimension along which we categorize individuals is the strength of the local networking (and financing) institutions in the city where they are based. Although hubs benefit firms by facilitating the use of skilled labor and specialized inputs, they are also known to facilitate the acquisition of tacit knowledge, build social ties and expose entrepreneurs to new opportunities (Sorenson and Audia 2000). Those who live in software hubs, where information about business opportunities and access to new clients is easier will therefore find it easier to network locally. Hence, all else equal, we would expect that entrepreneurs who are located in hubs would rely less on diaspora networks for their business.

³As the intuition of our model is quite straightforward, we leave a formal treatment of this simple model to the appendix. More formally, embedding these two different networks as inputs in a CES production function shows that the optimal investment in diaspora networks for a given entrepreneur can vary considerably based on the extent to which the two sets of networks serve as complements rather than substitutes.

In order to study whether diaspora networks help overcome weaknesses in the local networking environment, we examine the interaction between entrepreneurs' reliance on diaspora and local networks. If these networks serve as substitutes for one another, having lived abroad will be much less important for entrepreneurs located in hubs (as those based in hubs can effectively rely on the good local networking institutions to generate new business). On the other hand, if these networks serve more as complements, or are the result of hub-to-hub ties, then those in hubs will find the diaspora networks at least as if not more important than those located outside hubs. The sign of α_3 (the coefficient on $HUB_i * LIVEDABROAD_i$) in equation (1) will therefore shed light on the nature of these networks.

In addition to looking at reliance on diaspora networks, we also look at the startup's revenue as shown in equation (2) below:

$$\begin{aligned} LOGREV_i = & \beta_0 + \beta_1 LIVEDABROAD_i + \beta_2 HUB_i \\ & + \beta_3 (HUB_i * LIVEDABROAD_i) + \Phi X_i + \xi_i \end{aligned} \quad (2)$$

Since those who have a lower cost of accessing a given network will be more efficient at generating business, we expect that those who have lived abroad or those who live in hubs will tend to have better performing firms. However, as with equation (1), we expect that if local and diaspora networks serve as substitutes, then the benefit of having lived abroad will be less for entrepreneurs based in hubs and hence the sign of β_3 will be negative. On the other hand, if diaspora networks complement the hub networks, those who have lived abroad and live in hubs will have the best performing firms so that β_3 will be positive. Again, the sign of β_3 will help to shed light on the nature of these networks.

Our hypothesis is that diaspora networks can serve as substitutes to the local networking and financing environment for entrepreneurs and hence will be most important for entrepreneurs based in cities with weak networking (and financing) institutions. We therefore expect that both α_3 and β_3 will be negative. Moreover, since the signs on these coefficients imply a certain relationship between diaspora and local networks, we expect that the signs on these coefficients should be consistent with each other. In particular, if regression (1) implies that

the networks serve as substitutes, then we expect that this is implied by regression (2) as well. This helps to provide a check on the internal consistency of our framework.

3 Data

3.1 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 182.⁴ 60% of the respondents are one of the cofounders. Of the respondents who are not themselves the founders, half are CEOs of firms under the age of 5 (and 70% are CEOs of firms under the age of 10). This composition of respondents reflects the relatively young and entrepreneurial nature of the Indian software industry.

In Appendix Table 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

⁴However, due to the fact that private firms often do not share their revenue data, we have revenue data for only 111 firms.

of NASSCOM member firms. As can be seen from these tables, the firms in our sample are quite representative of the population of NASSCOM members along these observable metrics. Given the response rate of 24%, however, there still remains a concern that we face a response bias along some dimension we are not able to measure. For example, if CEOs who have lived abroad (or those who are more successful) are likely to respond differently than those who do not, and also more likely to be based in certain cities than others, this may bias our results. We articulate these concerns and a discussion of our checks in more detail in Section 5, after we present our results.

3.2 Main Variables:

As shown in the regression equations above, our main dependent variables of interest are (1) $DIASPORA_i$: Entrepreneurs' reliance on diaspora networks and (2) $LOGREV_i$: Entrepreneurs' firm revenue.

Operationalizing reliance on diaspora networks is difficult since it would require collecting information on the entrepreneur's active network, and the share of it that is constituted by the diaspora. We therefore look at three different proxies that capture related aspects of this ideal measure. First, 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 does not capture the strength of the entire diaspora network, it provides a good proxy for the share of the most recent important people they relied on that are constituted by the diaspora. Our second measure is more broad: we asked entrepreneurs the fraction of their overall network that was composed of Indians based outside India. While this does not provide an indication of how reliant entrepreneurs are on the diaspora, it complements the earlier, more narrow measure, and helps to provide confidence that our results are not driven by any specific measure we use to operationalize reliance on the diaspora. Finally, 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*.

Our main explanatory variables are (1) the ease with which entrepreneurs can access the diaspora and (2) the ease of local networking opportunities available to entrepreneurs 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 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.⁵

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, an indicator for whether they attended one of the elite Indian Institutes of Technology (IIT) or Indian Institutes of Management (IIM) – 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, and whether it has a foreign

⁵As a robustness check, we also use a binary variable, differentiating cities based on whether or not they are a 'Hub' (as outlined in Table 3).

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.⁶

4 Results

4.1 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 1, and 2, we plot the bivariate relationship outlined in our regression equations. Figure 1 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 1, those based in hubs rely little on the diaspora whether or not they have lived abroad. However, the importance of having lived abroad (and hence being able to access the diaspora more easily) is greater for those based outside the hubs. Consistent with our hypothesis, this suggests that diaspora networks may be acting as substitutes for local networking opportunities.

Figure 2 plots firms revenue for each city, based on whether the entrepreneurs have lived abroad or not. As with Figure 1, those based in hubs have similar performing firms, whether

⁶The 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.

or not they have lived abroad. However, living abroad is associated with better performing firms for entrepreneurs who live outside the hubs, suggesting that diaspora networks can help overcome the barriers to doing businesses in smaller cities with weaker networking institutions.

4.2 Main Results

Although suggestive of our findings, Figures 1 and 2 are only bivariate comparisons. We therefore move to a multivariate analysis, where we are able to control for several covariates at the individual, firm and city level. In Table 3, we report the results of OLS regressions where the dependent variable is the share of the CEO's top 5 contacts that are from the diaspora. As can be seen from Table 3, (and consistent with our hypothesis of α_3 being negative) having lived abroad is less important for those based in cities with a high networkshare when it comes to reliance on the diaspora. On the other hand, being able to access the diaspora networks is much more important for those who live in cities with poor networking environments (as can be seen from the coefficient on *LIVEDABROAD* in Table 3). Looking across the columns of Table 3, our results continue to be significant after controlling for firm-, individual- and city-level covariates. In Table 4, we re-run the same regression, but in this case the dependent variable is the share of the respondent's overall network that is constituted by the diaspora. The results using these two different measures of entrepreneurs' reliance on the diaspora are very consistent with each other.

In Table 5, we again run a similar regression to that in Table 3. However, our dependent variable is *FOREIGNFRAC_i*, the share of the entrepreneur's start-up capital that came from abroad. As we only have this data available for those who were one of the cofounders, the results for this table are based on the responses from the 109 founders in our sample. In addition, we replace the variable *NETWORKSHARE* with the variable *BANKS* which measures of number of commercial banks in each city and hence provides a measure of strength of the local financial institutions. Similar to the results in Table 3, we find that the importance of having lived abroad to raise foreign capital is much greater for founders based in cities with fewer formal financing options. As with the prior results, these findings continue to remain significant after controlling for several covariates.

In Table 6, we operationalize equation (2) by examining the factors contributing to firm revenue. Again, (and consistent with our hypothesis of β_3 being negative) we find that the importance of having lived abroad and accessing the diaspora has a smaller impact on firm revenue for those based in hubs. Note that although the coefficients are not significant in column 2 of Table 6, it is not due to a small coefficient, but rather the large standard errors due to the fact that we are not controlling for firm size in the regressions. Once we control for firm size, both the coefficients and the standard errors are attenuated, but β_3 continues being negative once we control for other covariates. Not surprisingly, including firm size in the regressions also increases the R-squared substantially – as seen from the columns 2 and 3. In fact, firm size alone explains just under 70% of the variation in firm revenue.

5 Discussion

Although our findings are all consistent with the hypotheses we outline in Section 3, 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, if those who have lived abroad are more able or less financially constrained and are also more likely to settle in smaller cities, then our results may be biased by this unobserved attribute. A related concern is that if the propensity to respond to our survey varies differentially across cities for those who have lived abroad, this may confound our findings.

We provide a number of checks that suggest that our results are not being driven by such a spurious correlation. First, we control for individual ability using a dummy of whether the CEO went to one of the elite institutions of higher learning 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 (as seen in Table 6). As can be seen from Table 2 however, we also do not find that the distribution of individuals who went to these universities varies consistently by their location suggesting that at least on this observable measure of individual ability, there is no obvious sorting by cities. We also examine whether conditional on having

lived abroad, the share of people who attended IITs or IIMs varies across hubs. The P value for the two-tailed test is 0.88 highlighting that there is virtually no difference in the distribution of these ‘higher ability’ individuals across cities. Lastly, 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. Although none of these tests is conclusive, they all point to the fact that our results are not driven by unobserved returns to ability or wealth across cities.

It is possible that our results may be driven in part 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 the framework that we outline 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? It suggests that there is significant inertia in terms of locating close to one’s prior job (Buenstorf and Klepper 2005; Figueiredo,

Guimaraes and Woodward 2000; Michelacci and Silva 2007) or that individuals choose where to locate their businesses for reasons other than the pure networking and financing needs of their firms. Consistent with this view, 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”. While one interpretation of our results is that it allows entrepreneurs to optimize their location choice based on the composition of their networks, these accounts suggest that location choices may not be as optimal *ex ante*. Although our results cannot directly speak to the efficiency of these networks, the presence of these frictions suggest that cross-border ethnic networks could also play a role in improving efficiency rather than purely impacting the *ex ante* location choices of entrepreneurs.

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 use novel data from a survey sent to the CEOs of Indian software firms to study these questions in more detail.

Our results suggest that entrepreneurs who live in hubs, where the local networking environment is stronger, are able to avail of local networks and do not necessarily gain significantly from relying more on diaspora networks. Entrepreneurs based in smaller cities, however, are faced with a weaker networking and financing environments, and hence are

disadvantaged in effectively generating business and growing their firms. Those located in such cities who have lived abroad are much more likely to tap into diaspora networks for help with their business, suggesting that diaspora networks serve as important intermediaries to overcome the weaker institutional environments where they are based. 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.

7 Appendix: A Model of Diaspora Networks

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. We 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 diaspora and (ii) by the firm’s production function. We model firm revenue using the Constant Elasticity of Substitution (CES) production function. Hence, revenue for the entrepreneur i ’s firm, Y_i is modeled as a function of L_i and E_i – that represent the entrepreneur’s degree of networking locally and with the expatriate community, respectively. γ is a parameter in the model and determines the extent to which the inputs are treated as complements or substitutes in the production function. The entrepreneur aims to maximize firm revenue subject to her ‘budget constraint’ imposed by the amount of time she can spend networking. Thus, the entrepreneur’s maximization problem can be written as:

$$\max_{Y_i, L_i, E_i} Y_i = [L_i^\gamma + E_i^\gamma]^{\frac{1}{\gamma}} \quad s.t. \quad L_i C_{L_i} + E_i C_{E_i} \leq T \quad (3)$$

where, as above, C_{L_i} and C_{E_i} are the cost of networking locally and with the expatriate community respectively, and T is the fixed amount of time each day that can be spent on networking. Note that we assume that in this static framework C_{L_i} and C_{E_i} are fixed for a

given individual⁷. Entrepreneurs therefore choose to allocate their time between the local and expatriate networking 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 in (3), we can derive the optimal level of local and expatriate networking for individual i given γ and costs C_{L_i} and C_{E_i} . Solving this, we get that:

$$L_i^* = \frac{T(C_{L_i})^{\frac{1}{\gamma-1}}}{(C_{L_i}^\gamma)^{\frac{1}{\gamma-1}} + (C_{E_i}^\gamma)^{\frac{1}{\gamma-1}}} \text{ and } E_i^* = \frac{T(C_{E_i})^{\frac{1}{\gamma-1}}}{(C_{L_i}^\gamma)^{\frac{1}{\gamma-1}} + (C_{E_i}^\gamma)^{\frac{1}{\gamma-1}}} \quad (4)$$

Substituting the values of L_i^* and E_i^* from (4) into (3), we can then solve for entrepreneur i 's firm revenue which is:

$$Y_i^* = \left[\left[\frac{T(C_{L_i})^{\frac{1}{\gamma-1}}}{(C_{L_i}^\gamma)^{\frac{1}{\gamma-1}} + (C_{E_i}^\gamma)^{\frac{1}{\gamma-1}}} \right]^\gamma + \left[\frac{T(C_{E_i})^{\frac{1}{\gamma-1}}}{(C_{L_i}^\gamma)^{\frac{1}{\gamma-1}} + (C_{E_i}^\gamma)^{\frac{1}{\gamma-1}}} \right]^\gamma \right]^{\frac{1}{\gamma}} \quad (5)$$

7.1 Relative Strength of diaspora Network

At the optimal level of local and expatriate networking, the ratio of the expatriate to local networks (that is, the relative strength or the reliance on the diaspora network) is derived by dividing the two terms in equation (4). This ratio, which we define as D_i^* is:

$$D_i^* = \frac{E_i^*}{L_i^*} = \left[\frac{C_{L_i}}{C_{E_i}} \right]^{\frac{1}{1-\gamma}} \quad (6)$$

The comparative statics on equation (6) yield the predictions for reliance on diaspora networks. In particular, note that the cross partials are:

$$\frac{\partial^2 D_i^*}{\partial C_E \partial C_L} = \frac{\partial^2 D_i^*}{\partial C_L \partial C_E} = -\frac{1}{(\gamma-1)^2 C_E C_L} \left[\frac{C_{E_i}}{C_{L_i}} \right]^{\frac{1}{\gamma-1}} \quad (7)$$

so that the extent to which the relative strength of diaspora networks for those who have lived abroad differs between cities and depends on the value of γ .

⁷Clearly both the location decisions and career paths of individuals are endogenous in the long run and thus 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.

7.2 Networking Strategy and Firm Revenue

The comparative statics on equation (5) yield the predictions for firm revenue. To see this, note that

$$\frac{\partial Y_i}{\partial C_E} = -T (C_{E_i})^{\frac{1}{\gamma-1}} \left[(C_{L_i})^{\frac{\gamma}{\gamma-1}} + (C_{E_i})^{\frac{\gamma}{\gamma-1}} \right]^{\frac{1-2\gamma}{\gamma}} < 0 \quad (8)$$

$$\frac{\partial Y_i}{\partial C_L} = -T (C_{L_i})^{\frac{-1}{\gamma-1}} \left[(C_{L_i})^{\frac{\gamma}{\gamma-1}} + (C_{E_i})^{\frac{\gamma}{\gamma-1}} \right]^{\frac{1-2\gamma}{\gamma}} < 0 \text{ and} \quad (9)$$

$$\frac{\partial^2 Y_i}{\partial C_E \partial C_L} = \frac{\partial^2 Y_i}{\partial C_L \partial C_E} = -\frac{T(1-2\gamma)(C_E C_L)^{\frac{1}{\gamma-1}} \left[(C_{L_i})^{\frac{\gamma}{\gamma-1}} + (C_{E_i})^{\frac{\gamma}{\gamma-1}} \right]^{\frac{1-3\gamma}{\gamma}}}{(\gamma-1)} \quad (10)$$

so that, as with (7), the extent to which firm revenue for those who have lived abroad differs across cities depends on the value of γ .

References

- [1] 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.
- [2] Agarwal, A., D. Kapur, and J. McHale. 2008. "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation." Working Paper.
- [3] Arora, A., V. S. Arunachalam, J. Asundi, and R. Fernandes. 2001. "The Indian software services industry." *Research Policy*, 30:8, pp. 1267-87.
- [4] Athreye, S. S. 2005. "The Indian software industry and its evolving service capability." *Industrial and Corporate Change*, 14:3, pp. 393-418.
- [5] 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.

- [6] Buenstorff, G. and S. Klepper. 2005. "Regional Birth Potential, Agglomeration Economies and Home Bias in the Location of Domestic Firms." Working Paper.
- [7] Curtin, P. D. 1984. "Cross-Cultural Trade in World History" Cambridge: Cambridge University.Press
- [8] 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.
- [9] Foley, C. Fritz and W. Kerr (2008): " High-Skill Immigration and FDI: A Link Through US Ethnic Scientists and Engineers", HBS Working Paper.
- [10] Gertler, M. 2008. "Buzz Without Being There?" working paper
- [11] 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.
- [12] 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.
- [13] Kapur, D. 2001. "Diasporas and Technology Transfer." *Journal of Human Development*, 2, pp. 265-86.
- [14] Kerr, W. 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics*, 90, pp. 518-537
- [15] Michelacci, C. and O Silva. 2007 "Why so many local Entrepreneurs?" *Review of Economics and Statistics*, 89:4, pp. 615-633
- [16] Rauch, J. E. 2001. "Business and social networks in international trade." *Journal of Economic Literature*, 39:4, pp. 1177-203.
- [17] Rauch, J. E. and A Casella eds. 2001. *Networks and Markets*. New York: Russell Sage Foundation.

- [18] Rauch, J. E. and V. Trindade. 2002. "Ethnic Chinese networks in international trade." *Review of Economics and Statistics*, 84:1, pp. 116-30.
- [19] Saxenian, A. 2002. "Local and Global Networks of Immigrant Professionals in Silicon Valley." Public Policy Institute of California.
- [20] Saxenian, A. 2006. *The New Argonauts: Regional Advantage in a Global Economy*. Boston: Harvard Business School Press.
- [21] 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.
- [22] 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.

FIGURE 1: RELIANCE ON DIASPORA NETWORKS

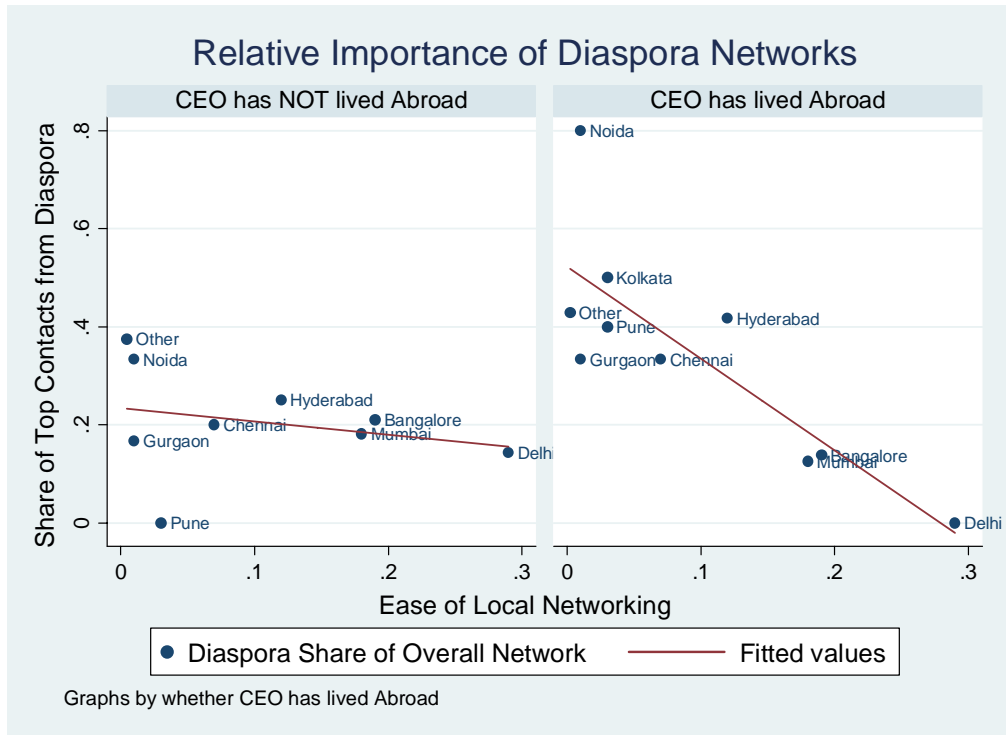


FIGURE 2: FIRM REVENUE

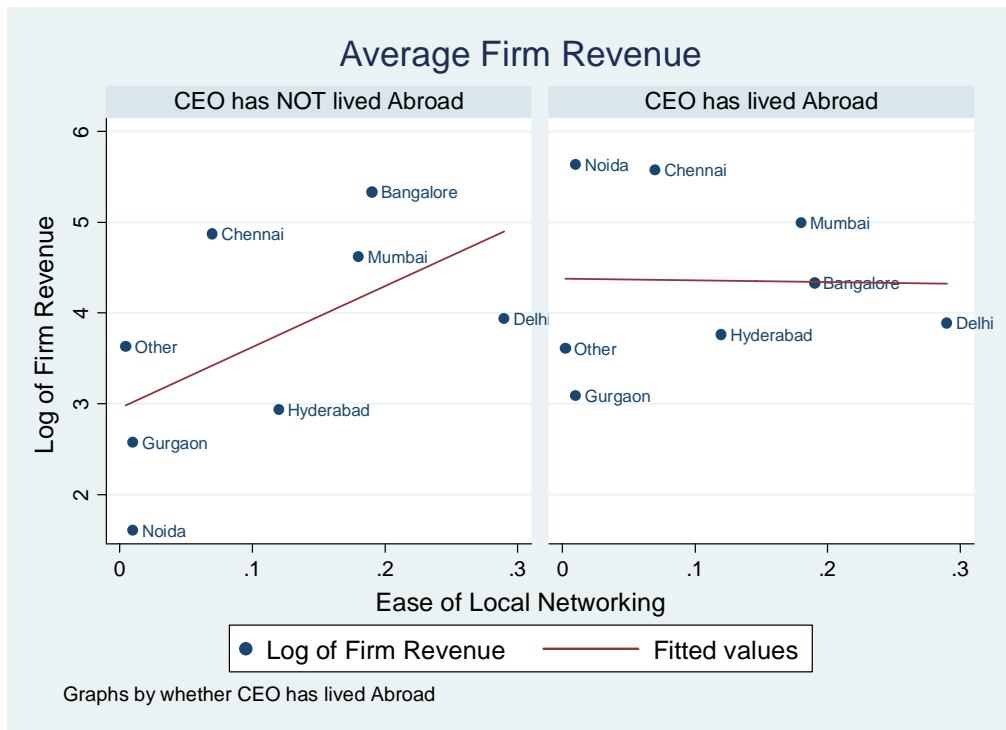


TABLE 1: MEASURES OF NETWORKING AND FINANCING COST ACROSS CITIES

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 who have studied at an IIT or IIM ³	29%	30%	27%	0.72
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 (i.e. one of the top 5 cities in Table 1)

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)
LIVED ABROAD	0.071* (0.040)	0.182** (0.074)	0.200** (0.086)	0.207** (0.084)	0.202** (0.087)
NETWORKSHARE		-0.172 (0.330)	-0.143 (0.380)	-0.020 (0.350)	-0.032 (0.380)
NETSHARE x LIVED ABROAD		-0.929** (0.380)	-1.014** (0.450)	-1.108** (0.420)	-1.082** (0.430)
LOG FIRM SIZE (EMPLOYEES)			-0.010 (0.018)	-0.013 (0.017)	-0.014 (0.017)
FIRM AGE			-0.004 (0.004)	-0.005 (0.004)	-0.005 (0.004)
CEO's AGE				0.035** (0.014)	0.033** (0.014)
CEO WENT TO IIT/IIM				0.003 (0.040)	0.004 (0.042)
SAME HIGHSCHOOL-CITY				0.012 (0.038)	0.005 (0.037)
Firm-Level Covariates	No	No	Yes	Yes	Yes
City-Level Covariates	No	No	No	No	Yes
Observations	182	182	182	182	182
R-squared	0.02	0.09	0.11	0.13	0.13

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

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

Note: 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; NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking across cities; SAME HIGHSCHOOL CITY is a dummy variable with a value of 1 if the CEO is based in the same city s/he went to highschool. Firm- and City-level covariates that are not reported are outlined in Appendix 2 along with their sources.

TABLE 4: RELIANCE ON DIASPORA NETWORKS*OLS Regressions: Dependent Variable is share of overall networks that is from Diaspora*

	(1)	(2)	(3)	(4)	(5)
LIVED ABROAD	0.051 (0.039)	0.140** (0.056)	0.140** (0.060)	0.138** (0.057)	0.135** (0.061)
NETWORKSHARE		0.329 (0.210)	0.348 (0.210)	0.331 (0.200)	0.116 (0.280)
NETSHARE x LIVED ABROAD		-0.686* (0.390)	-0.685* (0.380)	-0.663* (0.330)	-0.647* (0.350)
LOG FIRM SIZE (EMPLOYEES)			0.004 (0.007)	0.001 (0.006)	0.000 (0.007)
FIRM AGE			-0.003 (0.003)	-0.003 (0.003)	-0.003 (0.003)
CEO's AGE				0.000 (0.026)	-0.002 (0.028)
CEO WENT TO IIT/IIM				0.065 (0.047)	0.064 (0.044)
SAME HIGHSCHOOL-CITY				0.044 (0.051)	0.044 (0.055)
Firm-Level Covariates	No	No	Yes	Yes	Yes
City-Level Covariates	No	No	No	No	Yes
Observations	182	182	182	182	182
R-squared	0.01	0.02	0.08	0.09	0.09

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

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

Note: 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; NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking across cities; SAME HIGHSCHOOL CITY is a dummy variable with a value of 1 if the CEO is based in the same city s/he went to highschool. Firm- and City-level covariates that are not reported are outlined in Appendix 2 along with their sources.

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

	(1)	(2)	(3)	(4)	(5)
LIVED ABROAD	0.207** (0.074)	0.353*** (0.110)	0.362*** (0.100)	0.338*** (0.086)	0.308*** (0.082)
NETWORKSHARE		0.030 (0.056)	0.038 (0.055)	0.020 (0.039)	-0.129 (0.150)
NETSHARE x LIVED ABROAD		-0.198* (0.110)	-0.204** (0.091)	-0.177** (0.069)	-0.141* (0.071)
LOG FIRM SIZE (EMPLOYEES)			0.020 (0.024)	0.012 (0.025)	0.013 (0.028)
FIRM AGE			-0.018** (0.006)	-0.017*** (0.006)	-0.016** (0.006)
CEO's AGE				-0.010 (0.049)	-0.021 (0.054)
CEO WENT TO IIT/IIM				0.208* (0.110)	0.201 (0.120)
SAME HIGHSCHOOL-CITY				-0.011 (0.082)	-0.003 (0.086)
Firm-Level Covariates	No	No	Yes	Yes	Yes
City-Level Covariates	No	No	No	No	Yes
Observations	109	109	109	109	109
R-squared	0.07	0.10	0.25	0.30	0.31

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

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

Note: 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; NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking across cities; SAME HIGHSCHOOL CITY is a dummy variable with a value of 1 if the CEO is based in the same city s/he went to highschool. Firm- and City-level covariates that are not reported are outlined in Appendix 2 along with their sources.

TABLE 6: FIRM REVENUE*OLS Regressions: Dependent Variable is Log Revenue*

	(1)	(2)	(3)	(4)	(5)
LIVED ABROAD	0.013 (0.037)	0.101 (0.079)	0.056** (0.024)	0.062** (0.025)	0.061** (0.026)
NETWORKSHARE		0.441 (0.380)	0.309** (0.120)	0.378** (0.140)	0.435** (0.180)
NETSHARE x LIVED ABROAD		-0.644 (0.450)	-0.412** (0.160)	-0.465** (0.170)	-0.469** (0.190)
LOG FIRM SIZE (EMPLOYEES)			0.111*** (0.003)	0.105*** (0.004)	0.104*** (0.004)
FIRM AGE			-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)
CEO's AGE				0.027*** (0.007)	0.026*** (0.007)
CEO WENT TO IIT/IIM				0.057* (0.029)	0.059* (0.030)
SAME HIGHSCHOOL-CITY				0.008 (0.019)	0.007 (0.021)
Firm-Level Covariates	No	No	Yes	Yes	Yes
City-Level Covariates	No	No	No	No	Yes
Observations	101	101	101	101	101
R-squared	0.00	0.02	0.73	0.77	0.77

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

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

Note: 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; NETWORKSHARE (Share of NASSCOM conferences) measures the cost of local networking across cities; SAME HIGHSCHOOL CITY is a dummy variable with a value of 1 if the CEO is based in the same city s/he went to highschool. Firm- and City-level covariates that are not reported are outlined in Appendix 2 along with their sources.

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