Human Capital
and
Economic Development in India

Dissertation Proposal

Submitted by
Surendrakumar Bagde

Dissertation Committee:
Dennis Epple (Co-chair)
Lowell Taylor (Co-chair)
Raj Reddy

July 2008
## Contents

List of Figures \hspace{2cm} iv  
List of Tables \hspace{2cm} v  
Introduction \hspace{2cm} vi  

1 Private investment in human capital and Industrial development:  
   - The case of the Indian software industry \hspace{2cm} 1  
     1.1 Non technical summary \hspace{2cm} 2  
     1.2 Introduction \hspace{2cm} 4  
     1.3 Literature review \hspace{2cm} 6  
     1.4 Indian software industry \hspace{2cm} 8  
     1.5 Undergraduate engineering education in India \hspace{2cm} 11  
     1.6 Model and empirical specification \hspace{2cm} 16  
     1.7 Data \hspace{2cm} 19  
     1.8 Results \hspace{2cm} 21  
     1.9 Discussion and conclusions \hspace{2cm} 31  
     1.10 References \hspace{2cm} 32  
     1.11 Appendix \hspace{2cm} 35  

2 Initial Conditions and Post-entry Performance: the Case of Indian Software Industry \hspace{2cm} 37  
   - Introduction \hspace{2cm} 38  
   - Literature review \hspace{2cm} 38  
   - Model \hspace{2cm} 40  
   - Data \hspace{2cm} 45  
   - Results \hspace{2cm} 48  
     2.5.1 Initial size and firm survival \hspace{2cm} 48  
     2.5.2 Founders’ Human Capital and Firm Survival \hspace{2cm} 48  
   - Conclusion \hspace{2cm} 52  
   - References \hspace{2cm} 53
List of Figures:

1.2A Software Exports and Engineering baccalaureate capacity, 1990 -1996  
1.2B Software exports and engineering baccalaureate capacity, 2002  
1.3 Change in software exports 2003- 1990, 
   by year of policy change allowing private engineering colleges  
1.4 State share of software exports and engineering baccalaureate capacities, 
   1990-2003  
2.1 Optimal Value of $M_0$ and $Q_0$  
2.2 Entry Threshold ($X^*$) and Entry Size ($Q_0$)  
2.3 Optimal Value of $M_0$ and $X = (p - c)$
List of Tables:

1.1 Employment growth in the Indian software industry     9
1.2 Engineering baccalaureate capacity in India, 1951-2004    11
1.3a Sanctioned engineering baccalaureate capacity in ‘00s, by state and year  13
1.3b Software Exports by State, 1990 -2003      14
1.4 Summary statistics                      21
1.5 Initial state level engineering baccalaureate capacity and software exports 25
1.6 Annual change in software exports, 1990-2003     26
1.7a IV estimates - second-stage results        28
1.7b IV estimates - second-stage results        28
1.1A Number of engineering baccalaureates by state and year, 1990-2003  36
2.1 Firm Type and Background of Founders        46
2.2 Timing of Entry and Type of Firms           46
2.3 Exits by Type of Firm                      47
2.4 Educational Background of Founders of De novo firms  47
2.5 Initial Size by Type of Firm                48
2.6 All entrants: Cox Proportional Hazard Ratio Regression  49
2.7 De novo entrants: Cox Proportional Hazard Ratio Regression  51
2.8 Regression of Log of Initial Size for de novo firms  51
3.1 Number of Students Passing Entry Examination for Admission     65
to Engineering Colleges                          66
3.2 Caste-wise Distribution of Rank in Entry Examination       66
3.3 Caste-wise Distribution of Admitted Students              67
3.4 Admissions by Management of College                   68
3.5 Discipline-wise Admissions                          68
3.6 Effective rank for SC Students                    72
3.7 Effect of Rank on First-year Achievement in Engineering  74
**Introduction**

The theoretical models of economic growth have underscored the role of human capital. The empirical analysis of growth for a broad group of countries shows that the school attainment has positive effect on growth (Barro, 1992). Many studies have found that a region’s growth is influenced by the initial level of human capital. Glaeser et al. (1995) find that human capital level in 1960 influences growth of the cities between 1960 and 1990. Similarly, Simon et al. (2002) found that cities that have higher level of human capital initially grow faster in the long run. The regional differences in level of human capital also explain geographic differences in firm formation rates with regions endowed with higher level of human capital having higher firm formation rates (Acs, 2004).

In my dissertation I explore the role of human capital in India’s economic development.

In Chapter 1 my focus is regional dimension of the spectacular growth of software and services industry in India. The Indian software industry has grown rapidly over the last two decades, growing at an average of about 30% per year. The importance of skilled manpower, of engineers in particular, to Indian software exports is widely recognized (Lakha, 1994; Arora and Athreye, 2002). The capacity in undergraduate engineering colleges has increased many folds in last two decades. For example, between 1985 and 2003, undergraduate engineering baccalaureate capacity increased from about 45,000 to about 440,000. The growth in software exports and engineering education capacity appear to be closely linked.

In Chapter 1 I study the regional link between the growth of software exports and the growth in undergraduate engineering baccalaureate capacity. I empirically investigate how software exports by the fourteen major states of India, for the period 1990-2003, are conditioned by state’s undergraduate engineering capacity. We empirically show that some states achieved faster growth of software exports because they had higher stocks of human capital, and they had higher stocks of human capital because they allowed private engineering colleges to operate earlier than other states. I tested this hypothesis using a new, hand collected data set of state level software exports and state level undergraduate engineering capacity in India for a fourteen year period between 1990 and 2003. I address issue of endogeneity of engineering undergraduate capacity and that of unobservable state characteristics that may be correlated with software exports and engineering capacity.

In Chapter 2 I study entrepreneurship in the Indian software industry. I use hand collected data on entry and exit of firms, their export revenues, year of entry, and background of founders to study survival of software firms. The dataset is rich has firms of diverse background, e.g., foreign firms, firms founded by the Indian diaspora, existing Indian firms, and de novo startups (startups and spinoffs, as appropriate),
have entered Indian software exports industry. I find that initial size conditioned on the current size negatively influences firm survival whereas current size has positive effect on firm survival. I also study the role of founder’s education on the firm survival for de novo firms. I find that human capital as measured by the founder’s education significantly reduces annual hazard of exit, and in specific founders who have undergraduate degree in engineering and MS or MBA from foreign country have huge benefit.

In Chapter 3 I study affirmative action in engineering colleges in one major state in India. India has a unique social structure. Indian society is socially stratified and historically many people from lower strata of society had problems in accessing education. In the human capital led economic development providing opportunity to every section of the society, both from different social and economic strata, to increase their human capital is must for achieving equitable economic development in India. The federal and different state governments have adopted affirmative action as one of the policy instrument to achieve this objective. In addition to several initiatives for improving access to school and college education, the most important affirmative action policy has been to reserve a certain percent of seats in institutions of higher education for the socioeconomically disadvantaged groups.

The affirmative action policy places students from the socioeconomically disadvantaged groups, who on an average have lower test scores, in the most selective college. Some critics have argued that this creates “mismatch” between these students’ academic preparedness and academic requirement of the most selective colleges. In Chapter 3 my focus is understating the impact of affirmative action on student’s academic achievement, especially how improvement in college selectivity affects academic performance of the students from the socioeconomically disadvantaged groups. To answer this question I assemble a unique dataset on admissions to 245 undergraduate engineering colleges. The students were admitted in 2006 based on their performance in entry examination. More than 85 percent of these colleges are affiliated to the government university that conducts common examination for these students at the end of first-year. I use the students’ performance in their first-year examination to measure their academic achievement. I find no support for the “mismatch” hypothesis in my data. I find that college selectivity has positive effect on the students’ academic achievement.

References:


Chapter 1

Private investment in human capital and Industrial development: The case of the Indian software industry

Ashish Arora & Surendrakumar Bagde*
Heinz School of Public Policy and Management
Carnegie Mellon University
Pittsburgh PA 15213
{ashish, sbagde}@andrew.cmu.edu

Abstract

Though previous studies have noted the role of skilled labor in the growth of the Indian software industry, they have not empirically investigated its importance. In this study we study the effect of the supply of engineers, measured by engineering baccalaureate capacity, on the regional growth of the software exports between 1990 and 2003. We find significant effect of engineering baccalaureate capacity on the growth of software exports even after controlling for other relevant factors. This conclusion is especially interesting because much of this capacity is due to private, rather than publicly supported colleges, and testifies to the private willingness to invest in human capital even in poor countries.

* This paper is part of the dissertation research of Surendrakumar Bagde. Please direct all correspondence to sbagde@andrew.cmu.edu. Bill Vogt, Dennis Epple, Steve Klepper and Lowell Taylor provided helpful comments and suggestions. We gratefully acknowledge the cooperation and assistance provided by NASSCOM, and the MIDC, Mumbai, Government of Maharashtra. We thank Dr. Sudhir Reddy and Mr. Anil Kumar, NTMIS, Ministry of Human Resource Development, Government of India, for their help in obtaining data on engineering education in India. All remaining errors are our responsibility alone.
1.1 Non Technical Summary

The Indian software industry has grown rapidly over the last two decades, growing at an average of about 30% per year. India is not the only country to have succeeded in software exports. Israel and Ireland are two other countries that have also achieved software success (Arora and Gambardella, 2005).

One common element in all three countries is the role of human capital supply. The importance of skilled manpower, of engineers in particular, to Indian software exports is widely recognized (e.g., Lakha, 1994; Arora and Athreye, 2002). Between 1985 and 2003, undergraduate engineering baccalaureate capacity increased from about 45,000 (59 per million) to about 440,000 (405 per million), even as the total population increased from 765 million to 1086 million. Indeed, one might argue that the plenitude of engineers has created a comparative advantage for India in software service exports.

But how did a poor country, with a perpetual problem of financing public expenditure, create so many engineers? An important part of the answer is that it was the private sector that fuelled this increase. For institutional reasons, most of the additional engineering education capacity created in India was in the form of new colleges, and the vast bulk of these colleges were private colleges, privately financed principally from student tuition revenues. The number of engineering colleges in India increased from 246 in 1987 to 353 in 1995 and over 1100 in 2003. Eighty percent of new colleges added between 1987 and 1995 were in the private sector and the share of private colleges was even higher at 94 percent for colleges added between 1995 and 2002. Software exports were also growing rapidly during this period, and software exports and engineering education capacity appear closely linked.

In this paper, we study the regional link between the growth of software exports and the growth in undergraduate engineering baccalaureate capacity. In this paper we empirically investigate how software exports by the fourteen major states of India are conditioned by local levels of human capital, as measured by the state level engineering baccalaureate capacity. The simple point of the paper is that some states were favored locations for software because they had higher stocks of human capital, and they had higher stocks of human capital because they allowed private engineering colleges to operate earlier than other states. We test this hypothesis using a new, hand collected data set of state level software exports and state level engineering baccalaureate capacity in India for a fourteen year period that coincides with the rise of India as a software power. Our identifying assumption is that initially demand for engineers from the software industry was small, and changes in the number of engineers produced was independent of the current or anticipated growth of the software industry.
Our research covers fourteen states of India, for the period 1990-2003. These fourteen states accounted for 83.47 percent of the country's population in 2003, 78 percent geographic area of India, and 79.2 percent of the net domestic product in 2001-02. As well, the available data require that our measure of software exports is a broad one, including not only the export of software, but also affiliated IT services, including the so-called IT enabled service. IT enabled services are relatively unimportant in terms of revenues for much of our sample and only become significant after the turn of the century. As well, such services, particularly low-end services such as call centers, rarely employ engineers. Their exclusion would therefore only strengthen our results. These, and other issues related to our data are discussed in more detail in the appendix.

There are empirical challenges in such research. First, there may be unobservable state characteristics that are correlated with software industry and engineering baccalaureate capacity. Second, there is issue of endogeneity of engineering baccalaureate capacity. We address this by using a panel dataset and controlling for state fixed effects. We also develop an instrument, discussed in greater detail below, for engineering baccalaureate capacity.

---

1 Uttar Pradesh and Madhya Pradesh reflect the geographic boundaries as in 1990 and not the current boundaries.
2 The source for population and area share is the Census of India, and for the GDP share is Government of India, Ministry of Statistics and Programme Implementation (http://mospi.nic.in/9_nsdp_const_9394ser.htm)
1.2 Introduction

The Indian software industry, which was almost non-existent till late 1980s, grew at tremendous pace after early 1990s. The Indian software exports were about $128 millions in 1990-91 and grew to $485 millions by 1994-95. By 2003-04 the software exports had increased to $12.2 billions. Though differences in definition imply that US government figures show much lower level of software exports from India, there is no denying that they have grown dramatically over the last two decades, growing at an average of about 30% per year.

India is not the only country to have succeeded in software exports. Israel and Ireland are two other countries that have also achieved software success (Arora and Gambardella, 2005). One common element in all three countries is the role of human capital supply. The importance of skilled manpower, of engineers in particular, to Indian software exports is widely recognized (e.g., Lakha, 1994; Arora and Athreye, 2002). Between 1985 and 2003, undergraduate engineering baccalaureate capacity increased from about 45,000 (59 per million) to about 440,000 (405 per million), even as the total population increased from 765 million to 1086 million. Indeed, one might argue that the plenitude of engineers has created a comparative advantage for India in software service exports.

But the unasked question is – whence did plenitude arise? How did a poor country, with a perpetual problem of financing public expenditure, create so many engineers? An important part of the answer is that it was the private sector that fuelled this increase. For institutional reasons, most of the additional engineering education capacity created in India was in the form of new colleges, and the vast bulk of these colleges were private colleges, privately financed principally from student tuition revenues. The number of engineering colleges in India increased from 246 in 1987 to 353 in 1995 and over 1100 in 2003. Eighty percent of new colleges added between 1987 and 1995 were in the private sector and the share of private colleges was even higher at 94 percent for colleges added between 1995 and 2002. Software exports were also growing rapidly during this period, and software exports and engineering education capacity appear closely linked.

One way to study this relationship is through its regional dimension. Arora et al. (2001) suggest that large share of south and west region in engineering baccalaureate capacity, spurred by the growth of private engineering colleges, was one of main reasons for growth of software industry in those regions. In this paper we empirically investigate how software exports by the fourteen major states of India are conditioned by local levels of human capital, as measured by the state level engineering baccalaureate capacity. The simple point of the paper is that some states were favored locations for software because they had higher stocks of
human capital, and they had higher stocks of human capital because they allowed private engineering colleges to operate earlier than other states. We test this hypothesis using a new, hand collected data set of state level software exports and state level engineering baccalaureate capacity in India for a fourteen year period that coincides with the rise of India as a software power. Our identifying assumption is that initially demand for engineers from the software industry was small, and changes in the number of engineers produced was independent of the current or anticipated growth of the software industry.

Our research covers fourteen states of India, for the period 1990-2003. These fourteen states accounted for 83.47 percent of the country's population in 2003, 78 percent geographic area of India, and 79.2 percent of the net domestic product in 2001-02. As well, the available data require that our measure of software exports is a broad one, including not only the export of software, but also affiliated IT services, including the so-called IT enabled service. IT enabled services are relatively unimportant in terms of revenues for much of our sample and only become significant after the turn of the century. As well, such services, particularly low-end services such as call centers, rarely employ engineers. Their exclusion would therefore only strengthen our results. These, and other issues related to our data are discussed in more detail in the appendix.

There are empirical challenges in such research. First, there may be unobservable state characteristics that are correlated with software industry and engineering baccalaureate capacity. Second, there is issue of endogeneity of engineering baccalaureate capacity. We address this by using a panel dataset and controlling for state fixed effects. We also develop an instrument, discussed in greater detail below, for engineering baccalaureate capacity.

The remainder of this paper is organized as follows. Section 2 briefly reviews the relevant literature dealing with the Indian software industry as well as the literature on agglomeration. Section 3 describes the character, size and regional spread of software exports industry. Section 4 describes the size, growth of technical education in India, its regional dimension and importance of private sector. Section 5 develops a simple model linking the baccalaureate capacity and software production and motivates the empirical specification. Section 6 describes the data. We present results and alternative explanations in section 7. Section 8 summarizes the policy implications and concludes.

3 Uttar Pradesh and Madhya Pradesh reflect the geographic boundaries as in 1990 and not the current boundaries.
4 The source for population and area share is the Census of India, and for the GDP share is Government of India, Ministry of Statistics and Programme Implementation (http://mospi.nic.in/9_nsdp_const_9394ser.htm)
1.3 Literature review

The first contribution of our paper is to document and explore the role of education policy in the growth of Indian software exports. Specifically, it points to the importance of a human capital producing sector that responds to market demands. In India, access to engineering education was rationed for several years because expanding engineering education capacity was a lower priority for state and federal governments. The slow economic growth in the 1970s and 1980s meant that the social return to such investments was thought to be low. The private return, however, was high, especially for those engineers that went to work overseas. Indeed, it is widely believed that there was excess demand for admission to engineering colleges, with periodic political and legal battles about the ability of privately run colleges to charge market level tuitions. A key policy innovation was to allow privately funded colleges to satisfy this latent demand. As the success of this policy became clearer, other states followed suit. This variation in timing is the key variation that identifies the effect we seek to estimate.

Our findings also speak to the issue of high tech clusters. The bulk of Indian software industry is concentrated in a few clusters; indeed Bangalore has often been branded as the Silicon Valley of India in press accounts. Following Marshall, Ellison, Glaeser and Kerr (2007) argue that ultimately, firms agglomerate to save the costs of transporting either goods (inputs and outputs), people, or ideas, and find support for all three. The literature on clusters and agglomeration is huge and we refer the reader to overviews such as Fujita and Thisse (1996), Fujita, Krugman and Venebles (2001), and Rosenthal and Strange (2004). Our paper is not about clustering or agglomeration in general, but rather about the agglomeration of the Indian software industry. Thus, we are interested in understanding not simply whether and why Indian software exporting firms cluster, but principally where they cluster and the reasons for that.

In seeking to understand where high tech industries cluster, stories inspired by the Silicon Valley experience stress the role of an anchor university, labor mobility, venture capitalists and networks of specialized firms (cf. Saxenian, 1994). Others highlight the superior availability of infrastructure (Kapur, 2002). In 1990s many state governments in India set up information technology (IT) parks, which provided physical infrastructure such as building space and electrical power. Though infrastructure was undoubtedly important, our results suggest

---

5 Since tuitions were regulated and fixed by the government, private colleges circumvented this by charging “capitation fees”, which were, in effect, the capitalized value of the tuition they would have liked to charge.
6 Ellison, Glaeser and Kerr (2007) use patterns of co-agglomeration to quantify the importance of these three transaction costs. Audretch and Feldman (1996) seek to explain the extent to which R&D intensive industries are more likely to agglomerate.
7 The Software Technology Parks (STP) scheme in 1991 provided reliable internet connectivity and single window clearance for various government permissions to software export firms. There were other schemes like export
that software exports are regionally agglomerated in significant measure because of clustering in a key input, namely engineers.

Klepper (2007) argues that industries cluster in regions which are home to the early industry leaders. He argues that the semiconductor producers cluster in the Silicon Valley because many are spin-offs of some of the early leaders. In particular, he documents that Fairchild, a firm that pioneered the planar process and the integrated circuit, two of the fundamental semiconductor innovations, spawned a number of the firms that would later dominate the Silicon Valley and the semiconductor industry, including Intel, AMD, National Semiconductors, Micron Technology and VLSI Technology. Interestingly, Klepper (2007) argues that a similar process explains the concentration of the automobile industry around Detroit and the tire industry around Akron Ohio. Our results are consistent with Klepper’s explanation for industrial agglomeration, but our focus is not on the identities of the firms.

Berry and Glaeser (2006) develop a theoretical model to explain their finding that US metropolitan areas have diverged in terms of skill intensity over time. In their model, entrepreneurs arise randomly from among those with high human capital. These entrepreneurs are assumed to create firms in the cities where they live. If these firms are disproportionately likely to hire high human capital workers, then if workers are mobile, over time cities with higher initial levels of human capital will also disproportionately attract high human capital workers, leading to a divergence across cities in the share of high human capital workers. We too develop a simple model that motivates our empirical analysis. As in Berry and Glaeser, graduating engineers in our model have varying levels of preference for the city in which they go to college (which, in India, is often close to their birthplace). We ignore amenities and assume that firms are price takers in both the product and the input market. It follows that cities with more graduating engineers will attract or create more firms, resulting in these cities having higher software exports.

There are alternative explanations offered for the rise of the software industry in India, some of which have implications for its regional location. The first is the presence of public sector R&D units. Balakrishnan (2006), for instance, has argued that the presence of public sector R&D institutions, including nine defense related laboratories, in Bangalore accounts for the location of the software industry in Bangalore. Once in place, increasing returns would reinforce the initial lead. In our estimations, state fixed effects should account for such processing zones which offered similar incentives to firms locating in such zones. However, STP scheme offers much higher level of flexibility to firms in their location choices and was targeted to software export firms. Firms could locate anywhere and were required to register with designated STP office to avail various incentives.
differences. We also control for electronics production to take into account potential knowledge spillovers. Moreover, as we show below, the factual premise, namely that entrants were initially mostly located in Bangalore is simply wrong.

Srinivasan (2006) singles out telecommunication reforms and the creation of Software Technology Parks (STPs) as key pieces of the puzzle. In our empirical analysis, we control for industrial production and teledensity to control for physical and telecommunication infrastructure. Further, initial software exports consisted of software programmers being sent overseas on short term assignment, and physical and telecommunication infrastructure was less critical than when such work began to be shifted to India. We cannot control for STPs directly. However, as we discuss below, STPs were not important locations for software export prior to 1998 or so.

Other explanations have to do with the role of the diaspora, and the role of entrepreneurship. As we argue in greater detail later, these are complementary to the human capital based explanation analyzed here. Simply put, if much of the diaspora also consists of engineers, as is plausible, then regions abundant in engineers are also likely to be the source of the diaspora, and hence, disproportionately likely to benefit from the diasporic connections. Similarly, if entrepreneurs are more likely to have an engineering background, then regions that produce more engineers will be home to more entrepreneurs.

1.4 The Indian Software Industry:
According to an early study, Indian software exports were a mere $4 millions in 1980 and rose to $27.7 millions in 1985 (Heeks, 1996). Exports reached $128 millions in 1990. The industry grew very rapidly in the 1990s and exports were over $12 billion in 2003-04.

An important feature of Indian software exports has been the very high human capital intensity, relative to other inputs. Initially, the bulk of the exports consisted of sending software developers to work at the client site in America, on short term assignments. Later, teams of software developers were sent overseas, and only by the mid 1990s, was there significant software activity taking place locally. Initially, physical and communication infrastructure was far less important than commonly believed for the growth of the Indian software exports. Instead, the keys to growth were contacts with potential clients in America and Western Europe, and access to high quality engineers. Indeed, the survey conducted by Arora et al. (2001) referred to earlier found that 57% firms reported manpower shortages as among their most important problems. The next highest problems were employee attrition (44%), market access (42%), and
getting visas (33%). Bringing up the rear were physical infrastructure (12%) and lack of government support (10%).

Thus, employment has closely tracked revenues in this industry. The number of professionals was merely 6800 in 1985 and increased more than eight fold in the next five years to 56,000 (Table 1.1). The growth was at smaller pace in next decade and number of professionals rose to 841,500 in 2003. The number of professionals in the software exports sector has increased more slowly in recent years in comparison to those in the IT enabled services sector (ITES-BPO), from 110,000 in 1999 to 270,000 by 2003.

Table 1.1: Employment growth in the Indian software industry, in '000s

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Software-export sector</td>
<td>110</td>
<td>162</td>
<td>205</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software-domestic sector</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software- in-house captive staff</td>
<td>115</td>
<td>178</td>
<td>260</td>
<td>290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITES-BPO</td>
<td>42</td>
<td>70</td>
<td>180</td>
<td>253</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6.8</td>
<td>56</td>
<td>140</td>
<td>284</td>
<td>420</td>
<td>661</td>
<td>841</td>
</tr>
</tbody>
</table>


Much of what Indian software exports consist of does not require an engineering background, yet software exports from India rely very heavily on engineering graduates. A survey of over 100 Indian software firms in 1997 indicate that 80% of the software professionals employed had engineering degrees, while 12% only had diplomas from private training institutes (Arora et al., 2001). A large fraction of these engineers were not electrical or computer engineers. Instead, these included civil, chemical, textile, and industrial engineers with a 4 year undergraduate degree, though often followed by specialized, non-diploma training in software tools.  

There are some important reasons why firms prefer engineers. In interviews conducted in 1997 and 1998, few firms admit to hiring non-engineers, principally due to apprehensions about the signal it might send to potential customers and to other potential hires. The CEO of the fourth largest software firm, interviewed in 1997 said that he hired only engineering graduates from the best possible schools in India. However, this was not because engineering training or knowledge was relevant, but because these students tended to be smart and their backgrounds were useful in signaling quality to potential customers (reported in Arora et al., 2001). Simply put, the undergraduate engineering degree acts a screening device, because of the intense competition for admission to engineering colleges (Spence, 1984).

---

8 In recent years, software firms have turned to non-engineers as well, particularly for serving the domestic market, and for IT enabled services.
To be sure, this is not a signaling story alone. Engineering graduates are also exposed to the fundamentals of computers and learn basic programming and sometimes even advanced programming language, reducing need for longer duration training. As well four years of engineering education imparts a set of problem solving skills, methods of thinking logically and learning tools that help quick adaptation to changes in technology, domains and tasks (Arora et al., 2001). Perhaps equally importantly, initially the bulk of software exports consisted of software professionals working on client’s site in the US on temporary work permits, or H-1 B visas. US visa requirements meant that it was (remains) easier for engineers to qualify for H-1 B visas.

The software industry also tends to recruit younger workers. Older engineers, settled in other jobs, were less willing to accept jobs with (then) unknown firms, and spend extended periods away from their families.9 Perhaps equally important was their unwillingness to take on the mundane and tedious tasks that were initially required, the most well known of were the Y2K related projects, where someone had to go through the long lines of code, finding and changing how dates were entered. The upshot of the foregoing discussion is that finding and recruiting engineers was critical for a successful software exporter, and that most of the engineers would be newly graduated, rather than experienced ones.

- Uneven regional growth of the software industry:

In the very early period of the 1980s, the software industry was concentrated in Mumbai, the capital of the state of Maharashtra (Heeks, 1996) and also the leading commercial center of the country. As exports grew, the industry spread to other cities and states. Bangalore attracted many multinational companies after Texas Instruments set up its development center in 1985.10 By 1990 the states of Maharashtra (Bombay), Karnataka (Bangalore), Tamil Nadu (Chennai) and Delhi were the ones with large share of exports and states of Uttar Pradesh (NOIDA), Andhra Pradesh (Hyderabad) and West Bengal (Kolkata) also had software exports, albeit at lower levels.

Many multinational companies (MNCs) set up their subsidiaries after foreign investment norms were liberalized by the federal government in 1991 (Athreye, 2005a). These MNCs’ locations were typically in the leading software centers such as Bangalore, Hyderabad, Chennai, Bombay and Delhi (including NOIDA and Gurgaon). Indian software firms typically also had a single Indian location, at least till the early 1990s. Therefore the states which had a

9 Indeed, a recent report indicates that nearly two thirds of the IT workers in India have five or fewer years of experience. [http://www.ciol.com/content/services/register/register.asp?fid=1](http://www.ciol.com/content/services/register/register.asp?fid=1) (accessed 05/25/07)
head-start continued to grow rapidly in 1990s. This resulted in very heavy regional concentration of industry. Seven states contributed 95% of total software exports in 2002-03, but only 48% of the country’s population, 47% of the net state domestic product (NSDP) and 57% of the industrial production in the country. In the other seven states software exports are growing very rapidly but the absolute size of software exports from these states is still small.

1.5 Undergraduate Engineering Education in India

In this section we discuss how India’s undergraduate engineering education sector has evolved in past couple of decades. There are three main points to be made. First, there is substantial regional variation in engineering baccalaureate capacity, especially at the birth of the software export industry in the late 1980s. Second, this regional variation is mainly due to differences in private engineering colleges. Finally, the differences in private engineering baccalaureate capacity are are significantly affected by when the private colleges were allowed in the state.

Table 1.2: Engineering baccalaureate capacity in India, 1951-2004

<table>
<thead>
<tr>
<th>Year</th>
<th>Population in millions</th>
<th>Engineering baccalaureate capacity</th>
<th>Engineering baccalaureate capacity per million of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>361</td>
<td>4788</td>
<td>13</td>
</tr>
<tr>
<td>1985</td>
<td>765</td>
<td>45136</td>
<td>59</td>
</tr>
<tr>
<td>1995</td>
<td>928</td>
<td>105000</td>
<td>113</td>
</tr>
<tr>
<td>2004</td>
<td>1086</td>
<td>439689</td>
<td>405</td>
</tr>
</tbody>
</table>

Source: Our compilations from diverse sources including Ministry of Human Resources Development, Government of India, AICTE, NTMIS.

In India, higher education, particularly technical education, had been provided mostly by the government run institutions, except in last two decades. The majority of the institutions were set up and funded by various state governments. The number of institutions offering undergraduate degree in engineering has increased over the years as also the total intake capacity of these institutions\(^\text{11}\). Table 1.2 shows the growth in engineering baccalaureate capacity between 1951 and 2004. In 2004 the engineering baccalaureate capacity was 91 times that of 1951. Even accounting for population growth, the engineering baccalaureate capacity per million of population grew thirty fold, from 13 in 1951 to 405 in 2004.

\(^\text{11}\) The engineering college capacity of a college/institution is the number of students it can admit in a given academic year in all the disciplines. There are discipline-wise upper limits fixed for each academic year by AICTE. Any increase in the engineering college capacity requires approval of AICTE. The All India Council for Technical Education (AICTE), set up in 1945 as an advisory body, was given statutory status in 1987 through an Act of Parliament. The AICTE grants approval for starting new technical institutions and for introducing new courses or programs.
• **Regional Variation in engineering baccalaureate capacity**

Table 1.3a shows the sanctioned engineering baccalaureate capacity by state and year, in hundreds of undergraduate engineers. The last row of the table denotes the year in which private colleges were first permitted (and entered). Table 1.3a shows large inter-state variation in capacity. In fact share of four states of Andhra Pradesh, Karnataka, Maharashtra and Tamil Nadu was almost 75% in 1990-91\(^{12}\) as compared to 29% of the population. As other states added capacity, the share of these states has declined, but is still around 63% in 2003. As Table 1.3a shows, the growth in capacity has varied over time and across states. Consider the period from 1990 to 1993. Only three states, Karnataka, Maharashtra and Tamil Nadu were adding capacity. In other states the capacity did not increase perceptibly during these years. Also some states have experienced a sudden jump in the capacity, albeit in different years. These variations are important for our empirical analysis. Finally, states that have significant engineering capacity are those where private colleges enter early, though there are some exceptions. Orissa, for instance, permitted private colleges in 1986 but did not witness significant growth in capacity.

• **Role of private self-financed colleges:**

The inter-state disparities in engineering baccalaureate capacity are mostly due to differences in the timing and growth of the private sector colleges. Engineering baccalaureate capacity in a state can increase by two ways: either by expanding capacity in existing institutions or by opening new institutions. The new institutions can be in the public sector or the private sector. Much of the actual increase has been through new private colleges.

In 1981, the vast majority of engineering colleges were in the public sector i.e., funded by the federal or state governments and bound by their rule regarding admissions, salary, promotion and tenure. Tuition fees were very low and the vast bulk of the expenses were met from the budgets of the respective state governments, with the exception of the few institutes and colleges directly supported by the central (federal) government. Budget constrained state governments faced severe limits on increasing capacity. Therefore capacity expansion in the public sector has been infrequent, and mostly limited to accommodating new disciplines such as computer science in 1990s and information technology in early 2000s.

---

\(^{12}\) Their contribution to engineering college capacity was similar even in 1987-88.
Table 1.3a: Sanctioned engineering baccalaureate capacity in ‘00s, by state and year.

<table>
<thead>
<tr>
<th>Year</th>
<th>AP</th>
<th>Delhi</th>
<th>GJ</th>
<th>HR</th>
<th>KA</th>
<th>KL</th>
<th>MH</th>
<th>MP</th>
<th>OA</th>
<th>PN</th>
<th>RJ</th>
<th>TN</th>
<th>UP</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>58</td>
<td>9</td>
<td>33</td>
<td>5</td>
<td>170</td>
<td>27</td>
<td>192</td>
<td>17</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>92</td>
<td>31</td>
<td>23</td>
</tr>
<tr>
<td>1991</td>
<td>55</td>
<td>10</td>
<td>33</td>
<td>6</td>
<td>180</td>
<td>28</td>
<td>199</td>
<td>19</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>92</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>1992</td>
<td>55</td>
<td>10</td>
<td>34</td>
<td>8</td>
<td>188</td>
<td>29</td>
<td>238</td>
<td>19</td>
<td>11</td>
<td>5</td>
<td>13</td>
<td>94</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>1993</td>
<td>55</td>
<td>11</td>
<td>36</td>
<td>8</td>
<td>172</td>
<td>30</td>
<td>256</td>
<td>19</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>118</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>1994</td>
<td>56</td>
<td>10</td>
<td>38</td>
<td>8</td>
<td>193</td>
<td>35</td>
<td>280</td>
<td>19</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>141</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>1995</td>
<td>80</td>
<td>13</td>
<td>44</td>
<td>9</td>
<td>202</td>
<td>45</td>
<td>309</td>
<td>32</td>
<td>12</td>
<td>19</td>
<td>14</td>
<td>185</td>
<td>37</td>
<td>25</td>
</tr>
<tr>
<td>1996</td>
<td>86</td>
<td>12</td>
<td>50</td>
<td>9</td>
<td>203</td>
<td>47</td>
<td>333</td>
<td>34</td>
<td>17</td>
<td>19</td>
<td>15</td>
<td>222</td>
<td>44</td>
<td>26</td>
</tr>
<tr>
<td>1997</td>
<td>130</td>
<td>13</td>
<td>54</td>
<td>33</td>
<td>238</td>
<td>49</td>
<td>344</td>
<td>48</td>
<td>33</td>
<td>22</td>
<td>15</td>
<td>238</td>
<td>49</td>
<td>26</td>
</tr>
<tr>
<td>1998</td>
<td>196</td>
<td>16</td>
<td>64</td>
<td>33</td>
<td>244</td>
<td>51</td>
<td>397</td>
<td>43</td>
<td>45</td>
<td>22</td>
<td>20</td>
<td>273</td>
<td>68</td>
<td>40</td>
</tr>
<tr>
<td>1999</td>
<td>241</td>
<td>21</td>
<td>73</td>
<td>47</td>
<td>262</td>
<td>67</td>
<td>429</td>
<td>71</td>
<td>62</td>
<td>22</td>
<td>27</td>
<td>366</td>
<td>85</td>
<td>45</td>
</tr>
<tr>
<td>2000</td>
<td>277</td>
<td>23</td>
<td>91</td>
<td>67</td>
<td>282</td>
<td>88</td>
<td>429</td>
<td>102</td>
<td>62</td>
<td>34</td>
<td>50</td>
<td>505</td>
<td>153</td>
<td>52</td>
</tr>
<tr>
<td>2001</td>
<td>440</td>
<td>30</td>
<td>106</td>
<td>86</td>
<td>356</td>
<td>113</td>
<td>446</td>
<td>109</td>
<td>88</td>
<td>44</td>
<td>63</td>
<td>655</td>
<td>213</td>
<td>62</td>
</tr>
<tr>
<td>2002</td>
<td>624</td>
<td>34</td>
<td>106</td>
<td>98</td>
<td>381</td>
<td>183</td>
<td>470</td>
<td>160</td>
<td>88</td>
<td>86</td>
<td>82</td>
<td>702</td>
<td>231</td>
<td>107</td>
</tr>
<tr>
<td>2003</td>
<td>658</td>
<td>35</td>
<td>103</td>
<td>101</td>
<td>389</td>
<td>199</td>
<td>475</td>
<td>194</td>
<td>107</td>
<td>107</td>
<td>115</td>
<td>707</td>
<td>242</td>
<td>107</td>
</tr>
</tbody>
</table>

**Year Pvt. Coll.**

| '77 | '99 | '95 | '95 | '57 | '92 | '83 | '86 | '86 | '93 | '98 | '84 | '95 | '96 |


Year Pvt. Coll. – The year privately financed colleges first enter in the state.
Table 1.3b: Software Exports by State, 1990 - 2003 (in millions of Rupees, 1993-94 constant prices)

<table>
<thead>
<tr>
<th>Year</th>
<th>AP</th>
<th>DL</th>
<th>GJ</th>
<th>HA</th>
<th>KA</th>
<th>KL</th>
<th>MH</th>
<th>MP</th>
<th>OA</th>
<th>PN</th>
<th>RJ</th>
<th>TN</th>
<th>UP</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>127</td>
<td>301</td>
<td>0</td>
<td>0</td>
<td>626</td>
<td>2</td>
<td>1,571</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>374</td>
<td>129</td>
<td>80</td>
</tr>
<tr>
<td>1991</td>
<td>158</td>
<td>430</td>
<td>0</td>
<td>0</td>
<td>1,189</td>
<td>6</td>
<td>1,774</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>489</td>
<td>188</td>
<td>94</td>
</tr>
<tr>
<td>1992</td>
<td>242</td>
<td>387</td>
<td>0</td>
<td>119</td>
<td>1,595</td>
<td>13</td>
<td>2,521</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>654</td>
<td>248</td>
<td>143</td>
</tr>
<tr>
<td>1993</td>
<td>277</td>
<td>986</td>
<td>2</td>
<td>170</td>
<td>2,235</td>
<td>17</td>
<td>3,836</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1,124</td>
<td>385</td>
<td>251</td>
</tr>
<tr>
<td>1994</td>
<td>511</td>
<td>2,849</td>
<td>10</td>
<td>248</td>
<td>3,079</td>
<td>19</td>
<td>4,771</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1,800</td>
<td>541</td>
<td>319</td>
</tr>
<tr>
<td>1995</td>
<td>857</td>
<td>3,459</td>
<td>12</td>
<td>458</td>
<td>4,386</td>
<td>29</td>
<td>6,321</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>2,725</td>
<td>770</td>
<td>410</td>
</tr>
<tr>
<td>1996</td>
<td>1,813</td>
<td>5,493</td>
<td>30</td>
<td>745</td>
<td>7,609</td>
<td>62</td>
<td>9,764</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>4,563</td>
<td>1,288</td>
<td>465</td>
</tr>
<tr>
<td>1997</td>
<td>2,127</td>
<td>9,458</td>
<td>40</td>
<td>1,442</td>
<td>12,630</td>
<td>196</td>
<td>12,751</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>7,502</td>
<td>1,644</td>
<td>743</td>
</tr>
<tr>
<td>1998</td>
<td>4,587</td>
<td>17,643</td>
<td>93</td>
<td>7,763</td>
<td>24,347</td>
<td>374</td>
<td>14,114</td>
<td>106</td>
<td>565</td>
<td>56</td>
<td>26</td>
<td>9,174</td>
<td>7,057</td>
<td>1,411</td>
</tr>
<tr>
<td>1999</td>
<td>7,037</td>
<td>26,027</td>
<td>183</td>
<td>6,448</td>
<td>29,158</td>
<td>442</td>
<td>18,703</td>
<td>239</td>
<td>726</td>
<td>99</td>
<td>99</td>
<td>13,171</td>
<td>8,255</td>
<td>2,434</td>
</tr>
<tr>
<td>2000</td>
<td>12,103</td>
<td>23,869</td>
<td>641</td>
<td>9,108</td>
<td>48,681</td>
<td>706</td>
<td>26,853</td>
<td>314</td>
<td>1,256</td>
<td>314</td>
<td>188</td>
<td>24,435</td>
<td>21,935</td>
<td>2,946</td>
</tr>
<tr>
<td>2001</td>
<td>18,052</td>
<td>14,215</td>
<td>754</td>
<td>17,923</td>
<td>71,786</td>
<td>909</td>
<td>37,866</td>
<td>544</td>
<td>1,545</td>
<td>433</td>
<td>278</td>
<td>36,465</td>
<td>15,451</td>
<td>4,363</td>
</tr>
<tr>
<td>2002</td>
<td>22,001</td>
<td>17,121</td>
<td>609</td>
<td>20,377</td>
<td>81,834</td>
<td>958</td>
<td>40,754</td>
<td>604</td>
<td>1,741</td>
<td>406</td>
<td>269</td>
<td>43,529</td>
<td>17,992</td>
<td>7,545</td>
</tr>
<tr>
<td>2003</td>
<td>28,152</td>
<td>19,412</td>
<td>782</td>
<td>27,732</td>
<td>107,598</td>
<td>1,212</td>
<td>54,921</td>
<td>693</td>
<td>1,803</td>
<td>1,009</td>
<td>277</td>
<td>44,925</td>
<td>19,689</td>
<td>8,874</td>
</tr>
</tbody>
</table>

Furthermore, capacity expansion in existing institutions requires approval of All India Council for Technical Education (AICTE). AICTE limits intake capacity of a college. For example, according to the rules in force in 2005, the maximum capacity per discipline was 60 and a college could have maximum of 4 disciplines in first year of its operation. The total capacity could increase by 60 to 300 in the second year and finally to a maximum of 420 in the fourth year. Any increase beyond 420 requires that the institution meet very stringent quality standards, which few do.\textsuperscript{13}

The effect of regulation on capacity expansion combined with states’ (public sector) constraints in adding capacity means new private colleges have been the main source of growth. Analysis of college-level data between 1981 and 2004 support this conclusion. The number of colleges in the entire country increased from 246 in 1987 to 353 in 1995 and over 1100 in 2003. Eighty percent of new colleges added between 1987 and 1995 were in the private sector and the share of private colleges was even higher at 94 percent for colleges added between 1995 and 2002.

Karnataka was among the first state to permit the private sector in undergraduate engineering education. The first such college opened in Karnataka in 1957\textsuperscript{14}. Thereafter one in 1962 and two in 1963 started their operation in the state. Then a large number of private colleges entered, beginning 1979, with nine colleges opening in 1979 and eleven in 1980. The first private college started in 1977 in Andhra Pradesh and in 1983 in Maharashtra after the government introduced policy permitting such institutions to operate. By 1986, only six states had such institutions. Of these, only Madhya Pradesh and Orissa failed to develop leading software clusters. Of the remaining eight states, only Delhi managed to develop a leading software cluster.

As a result, Andhra Pradesh, Karnataka, Maharashtra and Tamil Nadu accounted for almost 75 percent of total engineering baccalaureate capacity in the entire country in 1990. Beginning in 1992, other states began to allow private self-financed institution and by 1999 all fourteen states had allowed private engineering colleges. As a result the share of private colleges has steadily increased over the years, from 62% in 1995 to more than 82% in 2002. In software specific disciplines (principally, electrical and electronic engineering, and computer science and computer engineering), the share is more than 90%.

It is only to be expected that education quality should have suffered greatly during this great expansion in capacity. Many of the new colleges are not up to the task of training

\textsuperscript{14} Manipal Institute of Technology (http://www.manipal.edu/mit/aboutus/overview.htm).
engineers, and their graduates frequently need extended periods of training by employers before they can be put to work. However, as briefly noted earlier, actual engineering skills may be only part of the attraction of engineering graduates, especially in the 1990s. Innate capabilities, a willingness to work hard, and structured problem solving abilities, all of which are likely higher among the graduates of even poor quality engineering colleges, may be more important.\textsuperscript{15} As well, in recent years, large Indian firms have undertaken substantial investments in in-house training, in some cases spending 3-4\% of revenues on training.

Adding to the problem has been a marked decline in the production of engineering PhDs. The number of engineering PhDs produced fell from 629 in 1991 to 298 in 1996, (AICTE, 1999). This decline in PhD production suggests that while there are strong private incentives to invest in an engineering baccalaureate, these do not extend to investing in research degrees. It also points to the limits of the private sector model for education.

1.6 Model and empirical specification

This section presents a simple model of a two region economy to motivate our empirical specification. The model shows that if firms are mobile but workers have idiosyncratic preferences for a region, so that workers are imperfectly mobile, then regions with a higher stock of workers will also have greater volume of production.

Many studies have found that a region’s growth is influenced by the initial level of human capital. Glaeser et al. (1995) find that human capital level in 1960 influences growth of the cities between 1960 and 1990. Similarly, Simon et al. (2002) found that cities that have higher level of human capital initially grow faster in the long run. Thus initial level of human capital seems to advantage cities and regions, perhaps by attracting knowledge-intensive industries. The regional differences in level of human capital also explain geographic differences in firm formation rates with regions endowed with higher level of human capital having higher firm formation rates (Acs, 2004).

- A simple model of supply of engineers and size of industry:

Our objective is to sketch out a simple model to structure our empirical analysis that follows. In our model, both firms and workers choose where to locate. However, whereas firms are profit maximizers and locate in the most profitable location, workers are assumed to have idiosyncratic preferences for their existing location. Workers are homogenous in quality and price taking behavior by firms implies that in any equilibrium where production is not

\textsuperscript{15} It is also a tribute to the superior management capability of Indian firms that they were able to use such inexperienced and poorly trained (but bright) young men and women.
concentrated in a single location, wages must be equalized across regions. In such a model it is easy to see that regions with a greater stock of workers will also have more production activity.\footnote{Blanchard and Katz (1992) analyze how shocks to labor demand and supply affect short term and long term employment dynamics in the United States. Robak (1982) develops a model of the long run equilibrium with local land and labor markets, with fixed location specific amenities and scarce land, and where firms and workers are mobile. We focus on the long run equilibrium and assume that workers stochastically prefer their existing location.}

Formally, we consider two regions, indexed as 1 and 2. Let $N_1$ and $N_2$ be the engineering stock in region 1 and 2 respectively. We assume that the elasticity of supply is zero, i.e., that everybody joins the labor market and is willing to work at prevailing wages, $w_1$ in region 1 and $w_2$ in region 2. We further assume that the utility for engineer $i$ from region 1 if she works in region 1 is $w_1$. The utility in region 2 for $i$ would be $w_2 - C_i$, where $C_i$ is the migration cost (the migration cost includes whatever utility loss there is from moving). Similarly, for engineer $j$ educated in region 2, the utility is $w_2$ when working in region 2 and $w_1 - C_j$ when working in region 1. We assume that $C_i$ and $C_j$ are all drawn from a distribution $F$. We do not specify a lower bound for $C$ so that it can take negative values as well. Then, the fraction of workers moving from region 1 to 2 is $F(w_2 - w_1)$, and the fraction moving from 2 to 1 is $F(w_1 - w_2)$. Let $x = w_2 - w_1$. The total labor supply in region 1 is $N_1(1 - F(x)) + N_2F(x)$, and $N_2(1 - F(-x)) + N_1F(-x)$ in region 2.

Labor demand: There are $M$ firms, which are price takers. Since the good in question is software for export, we also assume free transport of output. We assume that firms can locate anywhere they want. This is sensible since software is a new industry and most firms are de novo startups. Furthermore, a substantial fraction of the software exports from India are accounted for by American firms and by firms set up by people of Indian ethnicity living in America. All firms have the same production function $Q(L)$. It is immediate that with output price taking firms and labor as the only input into production, we must have $w_2 = w_1$ in equilibrium. So, total labor supply in 1 is $N_1(1 - F(0)) + N_2F(0)$.

Since the output price is determined in the export market and therefore the same across regions, labor demand and supply will be equilibrated will be through the distribution of firms across regions. If $y$ percent of firms locate in region 1, and if we normalize the price of the
output $p$ to 1, then each firm employs $m(w)$ workers, given by $Q'(m) = w/p = w$ (Recall that $w$ is same in each region). The labor demand is $Mym$ in region 1 and $M(1 - y)m$ in region 2.

This yields two equilibrium conditions for the labor market to clear in both regions:

$$Mym(w) = N_1(1 - F(0)) + N_2 F(0) \tag{1}$$

$$M(1 - y)m(w) = N_2(1 - F(0)) + N_1 F(0) \tag{2}$$

By adding (1) and (2), we get

$$Mm(w) = N_1 + N_2 \tag{3}$$

Equation (3) gives total demand. Substituting for $Mm(w)$ in (1), we get

$$y = \frac{N_1(1 - F(0)) + N_2 F(0)}{N_1 + N_2} \tag{4}$$

If we let $N_1 = \theta N_2$, then (4) becomes

$$y = \frac{(1 - F(0)) \theta + F(0)}{1 + \theta} \tag{5}$$

In order to understand how share of firms, $y$ respond to changes in capacity imbalance in two regions we differentiate (5) w.r.t. to $\theta$, which equals

$$\frac{dy}{d\theta} = \frac{1 - 2F(0)}{(1 + \theta)^2} \tag{6}$$

This means

$$\frac{dy}{d\theta} > 0 \quad \text{if } F(0) < 1/2 \tag{7}$$

**Empirical Specification**

The simple model developed above suggests that as long as there is some “stickiness” in the labor market, local endowments of human capital will condition the volume of software production in a region. In other words, it suggests that we specify a model with software exports as a function of the stock of engineers. We lack a measure of the stock of engineers in a state over time. However, our measure of engineering baccalaureate capacity is arguably closely

---

*17 This simple model can be easily extended in a number ways. One way is that the idiosyncratic regional preferences may not be symmetric across regions. Regional variations in employment opportunities, career growth prospects and cost of living differences may result in the distribution of $C_i$ being different for each region. Let $C_i$ be distributed with distribution function $F(.)$ in region 1 and $G(.)$ in region 2. In the equilibrium, the wages in two regions are same. Thus, $\frac{dy}{d\theta} = \frac{1 - (F(0) + G(0))}{(1 + \theta)^2}$, so that $\frac{dy}{d\theta} > 0$ if $F(0) + G(0) < 1 \quad (7')$. It is obvious from (7') that holding $F(0)$ and $G(0)$ constant, an increase in $\theta$ would increase the share of firms $y$ in region 1 provided $G(0) + F(0) < 1$. 
related to changes in the stock. Specifically, if there were no mobility of engineers across states, then the growth in the stock of engineers in state $i$ would be equal to the (lagged) engineering baccalaureate capacity in the state. Since the available anecdotal evidence suggests that such mobility is in fact small, we use this as a proxy for a change in the stock of engineers. To the extent that there is mobility, we have measurement error. As discussed in greater detail below, we also explore the results of instrumenting for engineering baccalaureate capacity to address biases due to measurement error, as well as problems posted by potential endogeneity. In other words, if $S_{it}$ are software exports in year $t$ for state $i$, and $K_{it}$ is the corresponding stock of human capital, we have

$$S_{it} = a_{it} + g_{it} + \beta K_{it} + \epsilon_{it}. \quad (8)$$

By taking first differences over time (represented by $\Delta$) we have

$$\Delta S_{it} = \Delta a_{it} + \Delta g_{it} + \beta \Delta K_{it} + \Delta \epsilon_{it}. \quad (9)$$

Note that (8) allows for each state to have a different time trajectory for exports, so that the state effect varies by time. For feasible estimation, we assume that $\Delta a_{it} = \alpha_{i}$ i.e., the change in exports per year (for a given state) does not systematically vary over time. Letting $\Delta g_{it} = \gamma_{t}$ yields

$$\Delta S_{it} = \alpha_{i} + \gamma_{t} + \beta \Delta K_{it} + \Delta \epsilon_{it}. \quad (10)$$

In other words, the expected annual increase in software exports is equal to a state fixed effect, a year effect and $\beta$ times the engineering baccalaureate capacity. This is our benchmark specification. Later, we also report estimates from a related specification where we use as the dependent variable the level of software exports rather than the change. Although the latter specification is not theoretically grounded in our model, it is plausible that with rapidly growing demand, the number of firms may depend not merely on the level of human capital stock but also its growth. Both specifications find support in the data, as discussed below. Also, in both specifications, we exploit the variations in state policy allowing private engineering colleges to develop an instrument for engineering baccalaureate capacity.

1.7 Data

We obtain data on engineering baccalaureate capacity from the “Annual Technical Manpower Review (ATMR)" reports published by National Technical Manpower Information System NTMIS. These reports are prepared by a state-level nodal center of NTMIS and give details of sanctioned engineering baccalaureate capacity and outturn for all undergraduate technical institutions in the state. The Handbook of Engineering Education, a publication of the
Association of Indian Universities has also been used as a supplementary source.\footnote{In a few cases, the data from these reports are inconsistent, typically involving decreases or large increases in capacity or where capacity is markedly inconsistent with the number of graduating engineers. In such cases the other sources have been used to arrive at the acceptable figures of sanctioned intake.} Data on software exports are obtained from the Electronics and Computers Software Export Promotion Council (ESC), which is the apex government trade promotion organization for this sector, for the years 1998-2003. For 1997 and earlier, the ESC does not provide state level export data. Accordingly, we used export revenues by NASSCOM member firms, allocating the export revenues of each firm to the state where its headquarters are located. Till 1995, virtually all firms were located in a single state. Thus, this approximation is a reasonable one. As further described in the data appendix, we verified NASSCOM figures where possible from Dataquest, a trade magazine that has covered the Indian IT industry since 1982, and provides data on sales, exports and employment for the leading firms. For two leading firms, which operated in multiple states, we were able to obtain data on employment by state and allocated export revenues in proportion.

Combining data from two separate sources can lead to problems. For instance, the growth of software exports between 1998 and 1997 yields odd results for some states, particularly for the Delhi, because around this time, firms located in Delhi moved their operations to Gurgaon in the state of Haryana, and Noida, in the state of Uttar Pradesh. Exploratory analysis suggests that these problem are modest, at best. For instance, confining oneself to data from 1998 onwards yields qualitatively similar results. The STPI is another potential source of state level export data. However, for the earlier years only a small fraction of software exports appear to be by companies registered through the STPI. For instance, in Mumbai, in the state of Maharashtra, many of the leading firms were located in SEEPZ, an export promotion zone, and apparently did not report their software exports through STPI. Towards the end of the period, however, exports reported through the STPI are about over 90% of the software exports as calculated by NASSCOM or reported in official Indian statistics.

Carrying out the analysis at the level of the state raises some additional issues. In particular, Delhi is bordered by the states of Haryana and Uttar Pradesh. Software exports from the latter two are concentrated very near their border with Delhi, in Gurgaon and Noida respectively. Since firms can move across the three locations, this results in large jumps and dips in software exports. We chose not to smooth the jumps and dips, principally because doing so does not affect the results.
Data on control variables like population, per capita power consumption, industrial output, teledensity, per capita income and number of students graduating from high school (passing the 12th grade) is obtained from various publications and websites of concerned departments of Government of India as detailed in the data appendix. Table 1.4 shows the descriptive statistics for the variables used in the regressions. Software exports, industrial output and per capita net state domestic product (NSDP) are in constant 1993-94 prices.

The unit of analysis is state even though the industry we are analyzing is mostly located in urban centers in India. To a considerable extent, our hand is forced by the availability of data, since creating measures of the supply of engineers by the relevant metropolitan area, though feasible, is very costly. Moreover, the major software exporting centers are locating at some distance from each other and likely draw upon colleges in the state, at least until the late 1990s. The one exception to this is Delhi, which, as noted already, draws upon Delhi, Haryana, and the western part of Uttar Pradesh.

Table 1.4: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software export (Rupees Million, 1993-94 constant prices)</td>
<td>6662</td>
<td>14422</td>
<td>0</td>
<td>107598</td>
</tr>
<tr>
<td>Change in Software export (Rupees Million, 1993-94 prices)</td>
<td>1724</td>
<td>4081</td>
<td>-9654</td>
<td>25764</td>
</tr>
<tr>
<td>Intake Capacity (number)</td>
<td>11507</td>
<td>14462</td>
<td>525</td>
<td>70660</td>
</tr>
<tr>
<td>Outturn (number of graduating engineers)</td>
<td>4923</td>
<td>5731</td>
<td>235</td>
<td>28107</td>
</tr>
<tr>
<td>Population ('000s)</td>
<td>57017</td>
<td>36867</td>
<td>9082</td>
<td>183205</td>
</tr>
<tr>
<td>Teledensity (no. of telephone lines/100 persons)</td>
<td>3.69</td>
<td>4.78</td>
<td>0.235</td>
<td>41.79</td>
</tr>
<tr>
<td>Per Capita Power Consumption (Kilo Watt hours/year)</td>
<td>429</td>
<td>197</td>
<td>148</td>
<td>921</td>
</tr>
<tr>
<td>Per Capita Incomea (Rupees, 1993-94 prices)</td>
<td>7692</td>
<td>3161</td>
<td>3752</td>
<td>17682</td>
</tr>
<tr>
<td>Industrial Outputb,c (Rupees Million, 1993-94 prices)</td>
<td>53608</td>
<td>48285</td>
<td>5739</td>
<td>269843</td>
</tr>
<tr>
<td>Electronics Production (Rupees Million, 1993-94 prices)</td>
<td>11426</td>
<td>11455</td>
<td>238</td>
<td>47633</td>
</tr>
<tr>
<td>No. of Students Graduating 12th Grade</td>
<td>173213</td>
<td>150213</td>
<td>4521</td>
<td>799916</td>
</tr>
</tbody>
</table>

N = 196. (14 states x 14 years. Some states have missing observations for some variables.)

1.8 Results

We begin with some simple descriptive relationships. Figure 1.1, which shows the log of software exports by state for three years, points the persistence of export leadership: states which were the early export leaders retain their leadership even after nearly a decade and a half. Figure 1.2 shows the relationship between software exports and engineering baccalaureate capacity. For all each of the three years, we see a positive correlation between a state’s engineering baccalaureate capacity and its software exports (in logs). Finally, figure 1.3 shows the change in software exports by state in various years against the year in which the state first allowed private engineering colleges. As can be seen, states which allowed private
colleges to enter earlier are also those where the software exports have increased the most. Delhi, and its two neighboring states, Haryana and Uttar Pradesh, are outliers in that though they were late in allowing private colleges they have shown rapid growth in software exports. Virtually all the software export growth is due to two suburbs of Delhi, namely Gurgaon in Haryana, and Noida, in Uttar Pradesh. It is possible that this is because Delhi, as the only center in the north of India, may have been able to grow even without a large engineering supply of its own, by tapping engineering graduates from virtually all parts of India except the south and the west.

Overall, these figures suggest that states which allowed early entry by private engineering colleges were also favored early destination of software exporters, and their advantage appears to have persisted even as engineering baccalaureate capacity in other states has rapidly expanded. Finally, figure 1.4 plots the share of software exports and engineering baccalaureate capacity (not lagged), over multiple years. It shows a marked positive relationship between these two shares.

Figure 1.1: Software exports by state, 1990, 1995, 2003.
Figure 1.2A: Software Exports and Engineering baccalaureate capacity, 1990-1996

Figure 1.2B: Software exports and engineering baccalaureate capacity, 2002.

Legend: see fig 2a.

Fig. 1.3: Change in software exports 2003-1990, by year of policy change allowing private engineering colleges.
Figure 1.4: State share of software exports and engineering baccalaureate capacities, 1990-2003.

Notes: From tables 3a and 3b. Delhi, UP and Haryana are combined.

We further explore these patterns through regression analysis. Consider first the long term impact of initial engineering baccalaureate capacity. In Table 1.5 we show the impact of engineering baccalaureate capacity in 1987 on the increase in software exports between 1990 and 2003. It is worth pointing out that the total software exports in 1987 were $54 million dollars (Athreye, 2005b) and therefore engineering baccalaureate capacity in a state in 1987 was unlikely to be influenced by software exports industry. As Table 1.5 shows, initial levels of college capacity in the state have a significant and sizable effect upon software exports in 2003, nearly a decade and a half later. The limited number of observations rules out the use of more controls. It is possible, therefore, that this long lasting influence is merely a reflection of unobserved state characteristics. Accordingly, we exploit the within state-variation in capacity over time in Table 1.6, which uses both year and state fixed effects. In addition, we control for per capital income, and teledensity, and use the state’s population to control for size effects.

<table>
<thead>
<tr>
<th>Table 1.5: Initial state level engineering baccalaureate capacity and software exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software exports 2003 minus 1990</td>
</tr>
<tr>
<td>Eng. Baccalaureate Capacity 1987</td>
</tr>
<tr>
<td>Electronics Production 1990</td>
</tr>
<tr>
<td>Lagged Industrial Output 1987</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>$^{R^2}$</td>
</tr>
<tr>
<td>No. of obs. 14.</td>
</tr>
<tr>
<td>Software exports measured in constant 1993 rupees, millions.</td>
</tr>
</tbody>
</table>
In subsequent analysis, we control for several other factors that might have facilitated growth of the software exports in the state. *Electronics production in 1990* is the size of hardware electronics industry in 1990, before the software industry achieved significant size. We include it in our regression as it has been argued that in the initial years of its growth, the software industry also relied on experienced professionals working in the electronics industry to meet its manpower requirement (Lateef, 1997). This also controls for a variety of other influences. For instance, Klepper (2007) has argued that related industries are more likely to spawn successful firms. However, since firms in many sectors, such as banking, finance and manufacturing are also significant producers of software (primarily for their own use), we control for industrial production as well.

Table 1.6 reports on the specification implied by equation (10). We lag engineering baccalaureate capacity by four years as it takes four years to complete an undergraduate engineering degree. This makes it unlikely that our effects reflect the feedback effect of software growth, except possibly towards the end of our sample period. Other controls such as electronic production, industrial output, per capita income and teledensity are also lagged, albeit by one year. Further, the standard errors are cluster corrected to account for the non-independence of errors within a state.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng. Baccalaureate Capacity (4 yr lag)</td>
<td>0.34 (0.1)</td>
<td>0.20 (0.07)</td>
</tr>
<tr>
<td>Lagged Electronics Production</td>
<td>0.40 (0.24)</td>
<td></td>
</tr>
<tr>
<td>Lagged Industrial Output</td>
<td>0.007 (0.023)</td>
<td></td>
</tr>
<tr>
<td>Lagged Per Capita Income</td>
<td>-0.55 (0.61)</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>-0.28 (0.16)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-371 (1308)</td>
<td>22981 (11914)</td>
</tr>
<tr>
<td>State-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.49</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note: Cluster corrected std. errors in parenthesis. No. of obs. 182.

Table 1.6 shows two specifications, with and without time varying state characteristics such as per-capita income, population and electronics and industrial production. In specification 1, a unit increase in capacity increases exports by Rs 340,000 or about $8,000 per year. To put this in perspective, the average revenue per employee in the software industry in India in the mid 1990s was of the order of $15,000. If one takes into account the less than full capacity utilization, students leaving prior to graduation, employment in industries other than software exports, and migration to other states and overseas, the quantitative impact is highly plausible.
As can be seen in column 2 of Table 1.6, controlling for time varying characteristics reduces the impact of engineering baccalaureate capacity on software exports, but the impact remains both economically and statistically significant. Further, with the exception of electronics production, none of the time varying characteristics added have statistically significant coefficients.

- **Potential endogeneity of engineering baccalaureate capacity**

  The identification thus far relies on the fact that the vast bulk of the growth in engineering baccalaureate capacity after 1990 in a state is privately financed, and that differences in the extent of privately financed colleges is overwhelming reason for variation in engineering baccalaureate capacity, both across states and over time. The principal source of variation in the extent of privately financed colleges is when a state permits such colleges -- the earlier the state permitted colleges, the more quickly capacity could increase. It was not until the 1990s that Indian states actively began to compete to attract businesses to locate. Before that, states frequently viewed private business with some suspicion. Though more business friendly states might, prior to the 1990s, have offered tax concessions or regulatory relief, they were unlikely to make significant policy changes in education policy to address business concerns. Moreover, recall that these results control for state fixed effects, which implies that only time variation in the extent to which a state is business friendly would be a source of problems.

  Despite this, it is possible that capacity is correlated with unobserved time varying effects that condition software exports from a state. For instance, a growing software industry in a state may create the expectation of growth in future demand for engineers. This will bias our estimate of the coefficient of lagged baccalaureate capacity upwards. On the other hand, it is likely that capacity is an imperfect measure for the change in the state level stock of human capital, and the resulting measurement error would imply an attenuation bias to zero. To probe the robustness of our results to both sets of concerns we present the estimates using an instrument for engineering baccalaureate capacity.

  Our instrument is based on the year when neighboring states first allow private engineering colleges to operate. Specifically, we create a dummy variable, “policy”, for each state, which is one if private engineering colleges are operating in that state in that year, and zero otherwise. Our instrument is the average of “policy” for neighboring states. Thus, the key assumption underlying our identification strategy is that though a state’s decision to allow private colleges may respond to anticipated demand for software workers in that state, the decision of neighboring states is independent of the anticipated demand in the reference state. Put differently, we assume that although state governments may respond to local software firms, they are not responsive to software firms in neighboring states. However, states do
respond to policy changes in neighboring states. Thus, this is a plausible instrument, though we acknowledge that such instruments have a “reflection” problem (cf Manski, 2000). Thus, we view this as a way of probing the robustness of our results, rather than our benchmark specification.

**Table 1.7a: IV estimates - First-stage results**
Dependent Variable: Lagged Eng. Baccalaureate capacity by state

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average neighboring state policy</td>
<td>-6713 (2618)</td>
<td>-5144 (2429)</td>
</tr>
<tr>
<td>Lagged electronics production</td>
<td>0.37 (0.18)</td>
<td></td>
</tr>
<tr>
<td>Lagged industrial output</td>
<td>0.06 (0.03)</td>
<td></td>
</tr>
<tr>
<td>Lagged per capita income</td>
<td>0.1 (0.82)</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td>-0.24 (0.11)</td>
</tr>
<tr>
<td>Constant</td>
<td>24204 (4918)</td>
<td>33081 (11098)</td>
</tr>
<tr>
<td>State-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>F-statistic for instrument</td>
<td>6.33</td>
<td>4.49</td>
</tr>
<tr>
<td>R²</td>
<td>0.89</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Note: Cluster corrected std. errors in parenthesis. No. of obs. 182.

**Table 1.7b: IV estimates. Second Stage Results**

<table>
<thead>
<tr>
<th></th>
<th>Change in Software Exports (2SLS) (1)</th>
<th>Change in Software Exports (2SLS) (2)</th>
<th>Software Exports (OLS) (3)</th>
<th>Software Exports (2SLS) (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng. Baccalaureate Capacity (4 year lag)</td>
<td>0.62 (0.36)</td>
<td>0.74 (0.50)</td>
<td>1.31 (0.27)</td>
<td>2.02 (1.95)</td>
</tr>
<tr>
<td>Lagged Electronics Production</td>
<td>0.21 (0.23)</td>
<td>1.60 (0.80)</td>
<td>1.35 (0.56)</td>
<td></td>
</tr>
<tr>
<td>Lagged Industrial Output</td>
<td>-0.03 (0.05)</td>
<td>-0.11 (0.08)</td>
<td>-0.16 (0.20)</td>
<td></td>
</tr>
<tr>
<td>Lagged Per Capita Income</td>
<td>-0.67 (0.67)</td>
<td>0.98 (1.68)</td>
<td>0.83 (1.54)</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>-0.15 (0.14)</td>
<td>-0.73 (0.56)</td>
<td>-0.56 (0.35)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-4773 (4489)</td>
<td>9397 (11527)</td>
<td>43376 (37591)</td>
<td>25642 (34762)</td>
</tr>
<tr>
<td>State-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year-fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.45</td>
<td>0.44</td>
<td>0.75</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Note: Cluster corrected std. errors in parenthesis. No. of obs. 182.
Table 1.7a shows the results of the first-stage regression of college capacity on average neighbor policy, with state and year fixed effects, with and without time varying controls. Though the neighbor policy measure is significant, the F statistic is only around 4.5 with time varying controls, and around 6 without them, after cluster correction. This implies that the instrument is not very powerful. With this caveat, we proceed with the estimation, in part to probe the sensitivity of our results.

Table 1.7b presents the corresponding estimates where we instrument for lagged engineering baccalaureate capacity using its predicted value, using a two stage least squares procedure. Columns 1 and 2 present results where the dependent variable is change in software exports over the previous year. Note that the estimated coefficient increases three fold as compared to the OLS estimate, suggesting that upward bias was unlikely and that measurement error is more likely. However, the coefficient is imprecisely estimated, possibly because the instrument is weak. We also present analogous results where we use the log of software exports, rather than the annual change in software exports, in columns 3 (OLS results) and 4 (2SLS). The estimated coefficient of capacity in this case also increases upon instrumenting for it, although the increase is not as large. Once again, the estimated coefficient has a large standard error. The other noteworthy point is that lagged electronics production also has a positive and significant impact on the level software exports, but not on the annual change in software exports. This may point to either knowledge spillovers from electronics, or some sort of time varying state characteristics related to IT production. Other time varying controls are statistically insignificant. Specifically, per capita income and industrial output do not play any role in explaining software exports growth.

- **Alternative explanations**

  **Infrastructure**: To further probe the robustness of our findings, we briefly discuss possible alternative explanations for our findings. The first is that states such as Karnataka and Maharashtra were better endowed with telecommunication or physical infrastructure (cf., Srinivasan, 2006). We control for state fixed effects, and also lagged state telecommunication density, lagged state per capita income and lagged industrial production, which should control for variation over time in these effects. Despite this, the ability of the Delhi region to emerge as a significant software exporter despite only belatedly allowing private engineering colleges (and thus having only a small stock of engineering baccalaureate capacity) may point to the importance of an adequate infrastructure. It is likely, however, that this also points to the importance of a commercial infrastructure, including the supply of entrepreneurs.
Early mover advantage and self-reinforcing effects: It is popularly believed that software production is marked by significant agglomeration economies. Thus, states which get an early start in software production are more likely to persist as leaders. Arguably, the governments of these states would also be more sympathetic to the need to produce more engineers and the academic entrepreneurs more willing to create private colleges to meet that demand. Such dynamic explanations are not easy to test and in any event, this one has some measure of truth. As figure 1 shows, states that were early software exporters continue to lead in software exports. However, the advantages of an early start do not overwhelm other factors. As can be seen from tables 3a and 3b above, in 1990, Maharashtra had three times the exports of Karnataka, yet by 2003, Karnataka had twice the exports of Maharashtra. Moreover, even if the time series results are driven by increasing returns, if human capital supplies provides the basis for early leadership, an agglomeration economies based explanation need not be inconsistent with a human capital based one.

It is unlikely that the agglomeration economies are due to knowledge spillovers, given the simple nature of Indian software exports. It is possible that a region such as Bangalore, well endowed with research institutes and natural amenities, enjoyed a good reputation with potential entrants, particularly multinational firms. This entry, directly or indirectly, increased demand for labor and also the supply of engineering baccalaureate capacity. Since Karnataka was also the leader in engineer baccalaureate capacity from the very beginning of the software industry, and we lack a direct measure for reputation, we cannot definitively distinguish this explanation from the engineering capacity based one. Insofar as Bangalore’s reputation related largely to its supply of engineering talent, this explanation too ultimately supports the human capital story, though not in all the details regarding the role of private engineering colleges.

An under-explored but potentially important source of regional agglomeration is entrepreneurship (Klepper, 2007). Athreye (2005: p 12) estimates that entrepreneurial firms accounted for over a third of the employment and revenue in the Indian software industry in 2001. Among the leading Indian software exporters, more than half were either de novo startups or spawned from other leading software producers (including multinationals). If a state got a head start and was home to successful firms, these firms are likely to spawn other firms, which may reinforce the initial advantage of that state. Insofar as entrepreneurs in the software industry are more likely to be trained as engineers themselves, the location of engineering colleges may be an important source of variation in the supply of entrepreneurship, particularly in the early years of the industry. Therefore, the human capital effects may work through the supply of entrepreneurs.
Diaspora: As noted earlier, the other key factor for export success is contacts with potential clients. Kapur (2002) has pointed to the importance of the Indian diaspora in facilitating such contacts, and Arora, Gambardella and Klepper (2005) provide some evidence of the role of the diaspora in creating firms, including some of the top software exporters. It is possible that the successful states were disproportionately the source of the Indian diaspora. The diaspora explanation is indeed consistent with our results. Anecdotal evidence suggests that a large fraction of the people of Indian ethnicity living in America have engineering undergraduate degrees. If so, state with larger engineering baccalaureate capacity, particularly in the 1980s and early 1990s, were more likely to have produced engineers who emigrated. Thus, though in our discussion thus far we have focused upon the role of engineers as software developers, this is not inconsistent with some fraction of these engineers also forming the diaspora which connected Indian software exporters (in their home states) with their customers in America, or themselves setting up software firms to service American clients.

1.9 Discussion and conclusions:

The importance of human capital -- skilled and creative workers -- to a “high-tech” industry is routinely acknowledged but often public policy discussions tend to focus on more trendy prescriptions such technology parks, venture capital, incubators and university industry centers. Software, perhaps more than any other high-tech industry, relies more intensively upon human capital. Software services, the engine of the Indian software sector, is arguably even more human capital intensive than software products. Thus, few would question the role of human capital stocks in the rise of the Indian software industry. What is less clearly appreciated is that there are significant variations across Indian states in stocks of the relevant human capital, engineers, and that these differences have played an important part in conditioning where the software industry has flourished. Even less well understood are the reasons for this regional disparity in human capital stocks.

We find that these variations exist even after controlling for factors such as how rich or large the state is, and measures of industrial production, electronics production or telecommunication investment. Since engineering education has been controlled and, in the main, provided by state funded colleges, differences in the willingness of states to invest in engineering colleges could, but do not, explain the bulk of the inter-state variation. Instead, it is the role of private engineering colleges which is the key the puzzle. Simply put, states which allowed private engineering colleges to enter early were able to get a head start and, this early advantage has persisted for nearly a decade and a half.
Permitting privately financed colleges helped mitigate the adverse effects of the lack of public investments in higher education. It did not completely ameliorate the problem because, as noted earlier, there has been a marked fall in the production of engineering PhDs, even as baccalaureate capacity has increased. Knowledgeable observers of the Indian software industry, and the leading firms themselves, are increasingly concerned about the divergence, which also points to the limits of relying solely upon private financing for human capital development.

1.10 References:


Software Technology Parks of India, New Delhi, “First Annual Report”.


Appendix A

Private Self-financed Institutions:
These institutions are privately founded and operated, with no financial support from the government. The government exercises no control on their day-to-day functioning though there are various regulations by regulatory body AICTE. The creation of new institution, an increase in the capacity of a discipline or the addition of a new discipline in an existing institution requires the approval of the AICTE (and prior to AICTE, approval from the relevant state government). Their principal activity is undergraduate education, and their operations are principally financed from tuition revenues.

Tuition fees were set by state governments. The fee is same throughout state though there are inter-state variations, though the final structure awaits the resolution of court challenges. These institutions are affiliated to universities (universities are set up by an act of state legislature or by an act of parliament), which prescribes syllabi for various disciplines and conducts examinations. The degrees are awarded by the universities.

Appendix B: Data Source:
The main source of data is the “Annual Technical Manpower Review (ATMR)” reports published by National Technical Manpower Information System NTMIS. These reports are prepared by a state-level nodal center of NTMIS and give details of sanctioned engineering baccalaureate capacity and outturn (numbers graduating) for all undergraduate technical institutions in the state. The ATMR has information on sanctioned intake and outturn. The NTMIS publication “Directory of Technical Institutions” has institution level details of intake and outturn. Sometimes the data from these reports are inconsistent. That is the trend in the intake or outturn appear anomalous. In such cases the other sources have been used to arrive at the acceptable figures of sanctioned intake. The Handbook of Engineering Education, a publication of the Association of Indian Universities has also been used as a supplementary source. We have also used printed publications and web-published data of AICTE whenever needed. In certain cases the website of the institution has been useful in providing relevant information.

The data on software exports are obtained from various reports published by NASSCOM and ESC. ESC compiles state-wise software exports data for the post 1997 period. For the earlier period, we rely upon NASSCOM, which publishes the “Indian Software Directory” which have details of software exports, employment and company location. The Dataquest magazine carries a detailed survey of software companies and provides useful information about various software companies, which were used to verify NASSCOM data, and fill in where data on software exports were missing. Our dataset also includes variables like population, per capita power consumption, industrial output, teledensity, per capita income and number of students passing 12th grade for each state and for each year. The information on these variables is obtained from various publications and websites of concerned departments of Government of India.

19 ESC (Electronics and Computer Software Export Promotion Council) is India’s apex trade promotion organization. ESC is actively engaged in the promotion of India’s export of computer software and services, computer hardware, consumer electronics, telecom equipments and cables.

20 Dataquest covers information technology industry in detail since 1982. They provide details on exports, turnover and employees of software companies.

21 Industrial output is the net value additions by all manufacturing units in a given state in a given financial year. The data is taken from Annual Survey of Industries conducted by Government of India.
<table>
<thead>
<tr>
<th>Year</th>
<th>AP</th>
<th>Delhi</th>
<th>GJ</th>
<th>HA</th>
<th>KA</th>
<th>KL</th>
<th>MP</th>
<th>MH</th>
<th>OA</th>
<th>PN</th>
<th>RJ</th>
<th>TN</th>
<th>UP</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>3927</td>
<td>630</td>
<td>2137</td>
<td>235</td>
<td>9015</td>
<td>2325</td>
<td>1940</td>
<td>9575</td>
<td>723</td>
<td>331</td>
<td>1054</td>
<td>5861</td>
<td>2412</td>
<td>1857</td>
</tr>
<tr>
<td>1991</td>
<td>4368</td>
<td>846</td>
<td>2164</td>
<td>465</td>
<td>8663</td>
<td>2319</td>
<td>1798</td>
<td>10900</td>
<td>824</td>
<td>307</td>
<td>1109</td>
<td>6147</td>
<td>2412</td>
<td>1959</td>
</tr>
<tr>
<td>1992</td>
<td>4385</td>
<td>900</td>
<td>2372</td>
<td>547</td>
<td>9169</td>
<td>2161</td>
<td>2268</td>
<td>12339</td>
<td>805</td>
<td>337</td>
<td>1131</td>
<td>6349</td>
<td>2412</td>
<td>1587</td>
</tr>
<tr>
<td>1993</td>
<td>4367</td>
<td>994</td>
<td>2742</td>
<td>533</td>
<td>7665</td>
<td>2246</td>
<td>1823</td>
<td>14323</td>
<td>845</td>
<td>429</td>
<td>1111</td>
<td>6595</td>
<td>2502</td>
<td>2106</td>
</tr>
<tr>
<td>1994</td>
<td>4405</td>
<td>847</td>
<td>2852</td>
<td>625</td>
<td>11494</td>
<td>2157</td>
<td>1651</td>
<td>14742</td>
<td>870</td>
<td>522</td>
<td>1265</td>
<td>6669</td>
<td>2502</td>
<td>2304</td>
</tr>
<tr>
<td>1995</td>
<td>5610</td>
<td>940</td>
<td>3132</td>
<td>621</td>
<td>11611</td>
<td>2547</td>
<td>2123</td>
<td>15283</td>
<td>851</td>
<td>554</td>
<td>1338</td>
<td>6660</td>
<td>2610</td>
<td>2301</td>
</tr>
<tr>
<td>1996</td>
<td>6298</td>
<td>910</td>
<td>3087</td>
<td>683</td>
<td>12182</td>
<td>2441</td>
<td>1849</td>
<td>13772</td>
<td>901</td>
<td>679</td>
<td>1429</td>
<td>7835</td>
<td>2886</td>
<td>2241</td>
</tr>
<tr>
<td>1997</td>
<td>5900</td>
<td>1160</td>
<td>3158</td>
<td>662</td>
<td>11977</td>
<td>2795</td>
<td>1647</td>
<td>16812</td>
<td>913</td>
<td>813</td>
<td>1217</td>
<td>9111</td>
<td>2749</td>
<td>2439</td>
</tr>
<tr>
<td>1998</td>
<td>5390</td>
<td>1097</td>
<td>3168</td>
<td>657</td>
<td>12036</td>
<td>3001</td>
<td>1763</td>
<td>19516</td>
<td>994</td>
<td>816</td>
<td>1443</td>
<td>11941</td>
<td>3294</td>
<td>2432</td>
</tr>
<tr>
<td>1999</td>
<td>7817</td>
<td>1085</td>
<td>3851</td>
<td>1004</td>
<td>12259</td>
<td>3571</td>
<td>2287</td>
<td>20534</td>
<td>1181</td>
<td>679</td>
<td>1469</td>
<td>13452</td>
<td>3323</td>
<td>2518</td>
</tr>
<tr>
<td>2000</td>
<td>8102</td>
<td>1103</td>
<td>4723</td>
<td>1120</td>
<td>12526</td>
<td>3877</td>
<td>2158</td>
<td>19706</td>
<td>1498</td>
<td>1365</td>
<td>1445</td>
<td>15524</td>
<td>3552</td>
<td>2644</td>
</tr>
<tr>
<td>2001</td>
<td>12171</td>
<td>974</td>
<td>4762</td>
<td>1788</td>
<td>14173</td>
<td>4126</td>
<td>2727</td>
<td>26341</td>
<td>2950</td>
<td>1991</td>
<td>1693</td>
<td>16670</td>
<td>4822</td>
<td>2754</td>
</tr>
<tr>
<td>2002</td>
<td>14680</td>
<td>1160</td>
<td>4902</td>
<td>2225</td>
<td>14195</td>
<td>3764</td>
<td>2050</td>
<td>26791</td>
<td>3259</td>
<td>2081</td>
<td>1901</td>
<td>20550</td>
<td>6703</td>
<td>3459</td>
</tr>
<tr>
<td>2003</td>
<td>20099</td>
<td>1089</td>
<td>6944</td>
<td>1950</td>
<td>14550</td>
<td>3944</td>
<td>3439</td>
<td>27157</td>
<td>4316</td>
<td>2944</td>
<td>1976</td>
<td>28107</td>
<td>8083</td>
<td>3834</td>
</tr>
</tbody>
</table>

Chapter 2

Initial Conditions and Post-entry Performance: the Case of Indian Software Industry

Ashish Arora, Surendrakumar Bagde
Heinz School, CMU, Pittsburgh

Abstract: In this paper we study entrepreneurship and firm survival, using as our empirical backdrop the Indian software service industry. We develop a model that has contrasting predictions about the impact of the scale of entry on firm performance, after conditioning on lagged output. We test predictions of the model. We find that current size positively influences the firm survival but initial size has negative effect on firm survival. We also find that founders’ human capital measured by their education reduces annual hazard of exit. The firms with access to market have advantage as also the firms with experience in the software industry.
2.1 Introduction

We believe that there are observable differences in the “heritage” of the firm. Existing firms diversifying into software bring with them resources of the parents. In particular, foreign firms setting up operations in India have the advantage of a certain (relatively) demand for software development.

A second dimension is the expertise of the founder (for de novo entrants). Here the literature has identified the human capital, both general as well as industry specific experience. Spinoffs from existing firms bring experience and customer and supplier relationships. Firms by the diaspora similarly bring the experience of operating in the market, along with other expertise as well. In this paper, we have measures of resources (existing firm or de novo entrant; foreign or domestic) and of the human capital (education and experience of founder), and we test how these affect survival.

Other than these two, there are unobserved sources of differences as well. These differences imply that firms with similar heritage nonetheless perform differently. These differences may manifest in differences in firm size at the time of entry. In this paper, we lay out a model that nests different explanations for how initial size and unobserved differences are linked. We test these contrasting predictions with a hand collected data set of software exporters registered with the Software Technology Parks (STP) NOIDA, part of the second largest software export cluster in India, between the year 1991 and 2003.

This paper is organized as follows. In the next section we briefly review the literature. In section 3 we develop a stylized model of entry and exit and highlight important predictions of the model. In section 4 we describe the data collection and show some main features of the dataset. In section 5 we present the empirical findings and section 6 summarizes our findings and concludes.

2.2 Literature review

There have been many studies trying to understand the role of initial size in predicting the likelihood of survival. Based on the analysis of several empirical works for manufacturing industries, Geroski stated ‘a positive relationship between the firm size and likelihood of firm survival’ as one of the stylized fact (Geroski, 1995). However, the effect of initial conditions especially the entry size has been found to have contradictory effect on the firm survival. The positive relationship between firm’s start-up size and likelihood of survival is not robust enough to be true in every country and for every industry. Audretsch (1999b) find that the relationship holds for US manufacturing industry but not for Italian manufacturing industry. The firm’s start-up size was found to positively influence the firm survival in the Italian tourist industry (Santarelli, 1998).
Jovanovic (1982) develops theoretical model of ‘noisy’ selection that has implication for the relationship between the entry size and likelihood of survival. In this model firms do not know their efficiency at the time of entry but learn about it during their post-entry performance in the marketplace. Whatever they learn about their efficiency in the marketplace is incorporated in the current size. This means that the current size is sufficient to predict survival of firms and once we control for the current size we should find no effect of initial size. Mata et al. (1995) empirically show that the coefficient on initial size is positive even after controlling for current size. They add coefficients on current size and initial size to conclude that the overall effect of initial size is to increase survival probability. However, they do not provide theoretical model to support their empirical findings.

In this paper we address the conflicting role of entry size on firm survival by developing a simple model, with forward looking firms that invest in marketing, and where current sales condition the distribution of future sales (thereby building in a direct advantage to large scale entry). Our model nests several theories of firm entry and growth that have contrasting predictions about the impact of the scale of entry on firm performance, after conditioning on lagged output. If efficient firms choose larger output at the time of entry, then it follows that, in a model where current sales also condition future sales, both initial and current output should reduce exit. On the other hand, a model in which initial output is exogenously given (for instance, by the size of the initial export contract the firm was able to get), the prediction is the opposite: More efficient firms would be willing to enter even with a small scale of initial output.

This paper also contributes to the understanding of the role of founders’ human capital in the survival of the firm. We specifically investigate the role of general human capital which we measure by the entrepreneur’s education. Our data allows us to measure the founders’ human capital by the discipline of undergraduate degree and graduate degree. This method of measuring founder’s education has benefits as undergraduate degree in liberal arts does not have the same labor market outcome as undergraduate degree in engineering at least in India. The effect of general human capital on firm survival and growth has not been consistent. It was found to influence both survival and growth (Cooper et al., 1994, and Gimeno et al., 1997) but in case of Spanish service industry companies it has no effect on firm survival time (Arribas, 2007). In another study college education was found to have greatest effect on business continuance (Bates, 1990). However, in our study we find that for de novo firms certain type of college education has positive influence on firm survival.

Many such studies have focused mainly on manufacturing industries, and very few to date are for services industry (Audretsch, 1999a). This study enriches the literature by adding one more study of
services industry. As our empirical setting is services industry in India, we can also compare our results with similar studies for services and manufacturing industry in other countries. Not only this, this study is also a study of new industry. Though software firms have existed in other countries, but these were new to India and replication from other countries to India involved different environment, people, etc.

2.3 Model

We develop a simple model of firms’ entry in the export market for software services and exit from it. The model has predictions that guide our empirical analysis. In this model firms invest in marketing and current sales condition the distribution of future sales (thereby building in a direct advantage to large scale entry). Firms enter with initial size $Q_0$. After entering, firms invest $M_0$ in marketing, which conditions the first period output, $Q_1$. The firm’s output is assumed to be function of its marketing effort and its ability to retain some of its customers to the next period, along with a stochastic shock. Thus, firm’s first period output equals $\lambda Q_0 + \alpha \psi(M_0) + \varepsilon_1$, where $\varepsilon_1$ denotes iid shocks that has zero mean, distributed $\Phi(\cdot)$. All firms face the same price, $p$. Whenever firm produces, it incurs a fixed cost of $F$. The marginal cost of firm is $c$. We assume that the marginal cost varies across firms. More precisely, we assume that the price cost margin, $X = p - c$, varies across firms.

The firms are not fully forward looking. Instead, we assume that firms, though potentially long lived, make decisions by looking forward only one period. This assumption is made to keep the model tractable, since our objective here is to understand the impact of initial entry conditions (see below also) on subsequent firm performance. In other words, we model the dynamics of firm investment and growth as a series of two-period decisions, where the firm takes its current output, $Q_t$, as given, decides on an optimal investment, $M_t$, observes the stochastic shock, $\varepsilon_t$, and then decides whether to exit, or to remain in the market and produce $Q_{t+1} = \lambda Q_t + \alpha \psi(M_t) + \varepsilon_t$. We assume that the firm exits if its current profits are negative i.e., if $(p-c)(Q_t + \varepsilon_t) - F < 0$. Once again, this myopic decision ignores the option value of staying on in the market. The advantage of the assumption is that it greatly simplifies the analysis and we conjecture that our principal result would not be materially changed if the assumption were relaxed.

We next analyze in detail the initial two period decision problem immediately after a firm has entered the market. After entering, the firm chooses $M_0$ at the beginning of first period. Having entered, firms produce in first period if $(p-c)(\lambda Q_0 + \alpha \psi(M_0) + \varepsilon_1) - F \geq 0$. This means firms will produce in first period only when $\varepsilon_1 \geq K$, where $K$ equals $\frac{F}{(p-c)} - \lambda Q_0 - \alpha \psi(M_0)$. The firm chooses optimal value
of $M_0$ by maximizing first period profit, but before observing the stochastic shock $\varepsilon_1$. The optimality condition for $M_0$ can be written as

$$E(\pi_1|\varepsilon_1 > K) = \arg\max_{M_0} \int_{\varepsilon_1} \left((p - c)(\lambda Q_0 + \alpha \psi(M_0) + \varepsilon_1) - F\right) \phi(\varepsilon_1) d\varepsilon_1 - M_0$$

(1)

We solve equation (1) for first order condition to find the optimum value of $M_0$ (see appendix 1). The optimum value $M_0^*$ can be expressed as

$$M_0^* = M_0(Q_0, \alpha, (p - c), F)$$

(2)

**Determination of initial size:**

We analyze three separate cases. We analyze the traditional case where firms optimally choose $Q_0$. We also analyze the case where $Q_0$ is exogenously determined (though it may vary across firms). In the latter, we separately analyze where the firm observes $Q_0$ before it decides to enter and when $Q_0$ is observed only after entry.

Since the assumption of exogenously determined output unconventional, it merits further discussion. Traditionally firms are thought to choose their output. The output, in software exports, consists of contract software development for overseas clients. However, firms typically enter the industry with a single contract, and frequently, merely with potential leads, which could materialize into orders. Moreover, most software development contracts relate to software used by firms to run their internal operations, such as billing, human resources, production, sales and so on. Although cost is a consideration, potential problems resulting from failure or delays are a much bigger concern. In such cases, clients are unlikely to increase the size of the contract in response to lower prices. Simply put, under-bidding for a software development contract is unlikely to be very fruitful, and may even be counterproductive in that it may lead the client to doubt the ability of the vendor. More generally, the nature of software development is such that the clients’ requirement frequently changes during the execution of the project. These changes cannot be contracted upon properly ahead of time and typically require re-negotiation (Banerjee and Duflo, 2000).

The incompleteness of the contract, the possibility of re-negotiation and concerns about the final quality of the delivered output made it difficult for newly started Indian firms to increase the size of contracts by underbidding. This is especially true of a newly started firm. Indeed, there is ample
anecdotal evidence that American firms outsourcing to India would typically outsource a small project to a new vendor, and only outsource larger projects if the initial experience was satisfactory (Arora et al., 2001).

Increasing the number of contracts is also difficult. Newly started Indian firms, particularly early in the history of the industry, faced considerable challenges in terms of gaining access to clients. Frequently, clients were obtained through personal connections, either directly or through diasporic Indians living overseas (Kapur, 2000; Athreye, 2005). Price cutting is unlikely to be a viable way for a startup to increase business, without very credible signals about its ability to deliver good quality in a timely fashion on a large scale. Such conditions may well apply to other industries as well, particularly those where firms cannot rely upon mass marketing or large lumpy investments to provide such signals to potential buyers. At the very least, it seems reasonable to entertain the possibility, with a view to testing its implications for subsequent firm outcomes.

The reason the assumption about initial conditions is potentially important has to do with what entry size reveals about the unobserved firm efficiency (proxied here by the price-cost margin, X). If firms choose $Q_0$, then firms that have large size at entry will also be more efficient. Exit depends upon firm efficiency, but also upon the growth of the firm, with the latter not perfectly correlated with X. Thus, even controlling for current output, exit will be lower for firms that entered with higher initial size, $Q_0$. The reverse is true if initial output is exogenous. In this case, if firms observe their initial output prior to entry, a small entry size implies a higher average efficiency. Simply put, only a very efficient firm would choose to enter at a small scale, because only then could it expect to be profitable. (This effect would carry over even with fully forward looking firm, because in our model, current output conditions future output as well, and hence, future profits.) Finally, if firms do not observe $Q_0$ prior to entry, there is no association between X and $Q_0$, and hence initial entry size is not associated with the future performance of the firm. This intuition (formalized below) can be summarized as follows:

*Conditional on lagged output, i) if initial size is exogenous but not observed prior to entry, there is no association with the probability of survival :ii) if initial size is exogenous and observed before entry, initial size is negatively associated with firm survival. iii) if initial size is endogenously chosen, it is positively associate with firm survival.*
Case I: $Q_0$ is exogenously determined and observed prior to entry.

We first analyze the case where $Q_0$ is exogenously determined. We assume here that the firm observes $Q_0$ before deciding whether to enter or not. We assume the firm enters if given its cost structure, its profits are positive. We derive the expressions for the effect of initial size on the exit probability in subsequent periods, conditioning on the lagged output. For simplicity, consider the probability of exit in period 2, given $X$ and $Q_1$. (Probabilities of exit in period $t$, given $X$ and $Q_{t-1}$ can be analogously derived.) In what follows we assume that $X$ is distributed with distribution function $G()$ over some compact domain.

Probability of exit in period 2, given $X, Q_1$ is

$$Pr(Exit|X, Q_1) = \Phi\left(\frac{F}{X} - \lambda Q_1 - \alpha \psi(M_1)\right)$$

To take expectation over $X$, we need to introduce the conditioning even,

$$X \geq \frac{F}{\lambda Q_0 + \alpha \psi(M_0) + \epsilon_1}, X \geq \frac{F}{Q_0}$$

This conditioning events tells us the firm enters if entry size is big enough, given its price cost margin, $X$, to cover its fixed cost, $F$, and will similarly produce in period 1, if its output is big enough to cover the fixed cost.

Let $Z = Max(X \geq \frac{F}{\lambda Q_0 + \alpha \psi(M_0) + \epsilon_1}, X \geq \frac{F}{Q_0})$

The $Z$ is random variable distributed $H(\cdot)$. Then,

$$H(\theta) = \begin{cases} 
0, & \theta < 0 \\
1 - \Phi\left(\frac{F}{\theta} - \lambda Q_0 - \alpha \psi(M_0)\right), & \theta \geq \frac{F}{Q_0}
\end{cases}$$

Also, let minimum value of $Q = Q$, by assumption.

Then exit probability in period 2 given $Q_1$ is
\[
\frac{F/Q}{F/Q_0} = \int_{F/Q_0}^{X} \Phi \left( \frac{F}{X} - \frac{\lambda Q_1 - \alpha \psi(M_1)}{X} \right) g(X | X \geq Z) dX h(Z) dZ
\]

\[
\frac{F/Q}{F/Q_0} = \int_{F/Q_0}^{A(Z; Q_1)} h(z) dZ
\]

We can write this as nothing depends on \(Q_0\) inside the inner integral. To find the effect of initial size \(Q_0\) on the exit probability, we differentiate above equation w.r.t. \(Q_0\) using Leibniz integral rule. This equals

\[
= -A \left( \frac{F}{Q_0}; Q_1 \right) h \left( \frac{F}{Q_0} \right) \left( - \frac{F}{Q_0^2} \right)
\]

\[
= A \left( \frac{F}{Q_0}; Q_1 \right) h \left( \frac{F}{Q_0} \right) \left( \frac{F}{Q_0^2} \right) > 0
\]

**Case II: \(Q_0\) exogenous but not observed:**

The firm does not observe its initial size \((Q_0)\) before entry. After entry firm realizes its \(Q_0\). In this case \(Q_0\) gives no information about firm’s efficiency, i.e., \(Q_0\) and \(X(= p - c)\) are unrelated. The exit probability is given by

Probability of exit in period 2, given \(X, Q_1\) is

\[
Pr(Exit | X, Q_1) = \Phi \left( \frac{F}{X} - \frac{\lambda Q_1 - \alpha \psi(M_1)}{X} \right)
\]

\[
Pr(Exit | Q_1) = \int_{F/Q_1}^{X} \Phi \left( \frac{F}{X} - \frac{\lambda Q_1 - \alpha \psi(M_1)}{X} \right) g(X | X > F / Q_1) dX
\]

\[
\frac{\partial Pr(Exit)}{\partial Q_0} \bigg|_{Q_1} = \int_{F/Q_1}^{X} \frac{\partial \Phi \left( \frac{F}{X} - \frac{\lambda Q_1 - \alpha \psi(M_1, Q_1, X)}{X} \right)}{\partial Q_0} g(X | X > F / Q_1) dX
\]

(4)

\[
\frac{\partial Pr(Exit)}{\partial Q_0} \bigg|_{Q_1} = \int_{F/Q_1}^{X} g(X | X > F / Q_1) dX = 0
\]

(5)
Thus if the firm’s initial size is exogenous and not observed by firms before entry, then conditioned on lagged output, we will find no effect of $Q_0$ on firm survival.

**Case III: $Q_0$ is endogenously chosen:**

In this case firm endogenously chooses its initial size ($Q_0$). It can be shown that if firm faces no entry cost, then the profit $\pi$ will be convex in $Q_0$. This means firms should be choosing the highest value of $Q_0$. We use entry cost $E(Q_0)$ to make profit $\pi$ concave in $Q_0$. The profit $\pi$ is concave in $Q_0$ as long as $\Psi'' < 0$, where $\Psi = \pi - E$.

The probability of exit in period 2, given $X$, $Q_1$ is

$$\Pr(Exit|X) = \Phi\left(\frac{F}{X} - \lambda Q_1 - \alpha \psi(M_1(Q_1, X))\right)$$

(6)

We are interested in finding the effect of $Q_0$ on exit probability conditioned on the lagged output, $Q_1$. In this case we cannot differentiate exit probability with respect to $Q_0$ as $Q_0$ is endogenously determined in the model. However, we can find the relationship between $Q_0$ and exit probability by determining the sign of covariance between them. In this case firms choose their initial size $Q_0$ depending upon their $X$. We have shown in appendix I that $Q_0$ is increasing function of $X$, i.e., $\partial Q_0 / \partial X > 0$. We have also shown in appendix I that $\partial \psi(M_0) / \partial X > 0$. The firm uses same optimality condition to make investment of $M_1$ in the beginning of second period in the second window, and therefore $M_1$ is increasing in $X$. Thus it is obvious from equation (6) that conditioned on $Q_1$, $\Phi(\cdot)$ is decreasing function of $X$. According to Theorem 236 of Hardy, Littlewood and Polya, the covariance of an increasing and a decreasing function of a random variable is negative’ i.e. $\text{cov}(\Pr(Exit|X), Q_0(X)|Q_1) < 0$ (Eckel, 1992).

### 2.4 Data

We hand collected data on software exporters registered with Software Technology Park of India (STPI)\(^1\) NOIDA, part of the second largest software export cluster in India, between the year 1991 and

---

\(^1\) The Software Technology Parks (STP) scheme started in 1991 by Government of India provided reliable internet connectivity and single window clearance for various government permissions to software export firms. There were other schemes like export processing zones which offered similar incentives to firms locating in such zones. However, STP scheme offers much higher level of flexibility to firms in their location choices and was targeted to software export firms. Firms could locate anywhere and were required to register with designated STP office to avail various incentives.
2003. We track these firms from the year they register till 2006. Each firm submits an application to the STPI at the time of registration and subsequent to this STPI maintains detailed dossier on each firm. I had to go through each dossier manually for collecting information on founding background of these firms. We collected data on year of entry, export revenues, background of founders, type of activity, etc.

Table 2.1 shows classification of firms based on the background of founders.

<table>
<thead>
<tr>
<th>Firm Type and Background of Founders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign branch</td>
</tr>
<tr>
<td>Diaspora</td>
</tr>
<tr>
<td>Diversifying</td>
</tr>
<tr>
<td>Spinoff</td>
</tr>
<tr>
<td>Startup</td>
</tr>
</tbody>
</table>

The types of entrants have changed over time. We divide them into three entry cohorts, years before dot.com boom, and years after that. These entry cohorts are: 1991-1998, 1999-2000, and 2001-2003. Table 2.2 shows number of firms of each type for three entry cohorts. Foreign branch are largest group in cohort 1991-1998, diversifying firms are larges in 1999-2000, and spinoffs are largest in 2001-2003, though foreign branch and diversifying entrants are about the same in number. If we combine diaspora and foreign branch, as they are essentially software development platforms of foreign firms, then they are the largest group in cohort 1991-1998 and 2001-2003. The foreign branch and diaspora together have 40 percent share of total firms.

<table>
<thead>
<tr>
<th>Timing of Entry and Type of Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry cohort</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1991-1998</td>
</tr>
<tr>
<td>1999-2000</td>
</tr>
<tr>
<td>2001-2003</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

We treat the firm has having exited if it does not report the export revenue, as the regulations require that any export made be certified by the STPI. Entrants in a year \( t \) are those firms that registered
with STPI in that year for the first time. Similarly, exits in a year $t$ are those firms that report export in year $t-1$ but not in year $t$. The Table 2.3 shows number of firms exiting for different types of firms. The startups have the highest fraction of firm exiting followed by diversifying firms. The diaspora have the lowest fraction of firms exiting the industry.

Table 2.3
Exits by Type of Firm

<table>
<thead>
<tr>
<th>Type</th>
<th>Firms</th>
<th>Exits</th>
<th>Percent Exiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign branch</td>
<td>101</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Diaspora</td>
<td>55</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Spinoff</td>
<td>74</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>Diversifying</td>
<td>70</td>
<td>29</td>
<td>41</td>
</tr>
<tr>
<td>Startup</td>
<td>30</td>
<td>13</td>
<td>43</td>
</tr>
<tr>
<td>Others/non-traceable</td>
<td>44</td>
<td>18</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>374</td>
<td>110</td>
<td>29</td>
</tr>
</tbody>
</table>

The educational background of founders of de novo firms is shown in Table 2.4. There are 99 founders who have at least undergraduate degree in engineering and 62 of these founders have enhanced their education by obtaining graduate degree either in science or business administration. There is interesting subset of these 62 founders who have graduate degree in science or management from foreign country. As one would expect the largest number of such founders are from Indian diaspora. We have coded education as missing if we could not obtain information on educational background of founders.

Table 2.4
Educational Background of Founders of De novo firms

<table>
<thead>
<tr>
<th></th>
<th>Undergrad in Engineering</th>
<th>Undergrad in Engineering, and MS or MBA</th>
<th>Undergrad/grad in liberal arts/science</th>
<th>Education missing</th>
<th>Total from foreign country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diaspora</td>
<td>8</td>
<td>35</td>
<td>8</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>Spinoff</td>
<td>25</td>
<td>25</td>
<td>18</td>
<td>6</td>
<td>74</td>
</tr>
<tr>
<td>Startup</td>
<td>4</td>
<td>2</td>
<td>19</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>62</td>
<td>45</td>
<td>15</td>
<td>159</td>
</tr>
</tbody>
</table>

The initial size of the firm varies by type of the firm (see Table 2.5). The foreign branch firms on an
average enter with biggest size. The diversifying firms’ mean initial size is reduced to 215 if an outlier is excluded. The startups enter with lowest mean size.

<table>
<thead>
<tr>
<th>Table 2.5</th>
<th>Initial Size by Type of Firm (in '000 of US $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. Mean Std. dev. Min Max</td>
<td>Foreign branch 88 1074 1840 0 13072</td>
</tr>
<tr>
<td></td>
<td>Diaspora 44 457 737 0 3386</td>
</tr>
<tr>
<td></td>
<td>Spinoff 59 349 467 0 2339</td>
</tr>
<tr>
<td></td>
<td>Diversifying 58 975 5802 0 44293</td>
</tr>
<tr>
<td></td>
<td>Startup 25 149 195 0 613</td>
</tr>
<tr>
<td></td>
<td>Others 36 487 703 0 2853</td>
</tr>
<tr>
<td></td>
<td>ALL firms 310 687 2730 0 44293</td>
</tr>
</tbody>
</table>

Notes: 1. This table is only for firms which do not have missing initial size. 2. Mean size of diversifying firms is 215 if an outlier firm is excluded.

2.5 Results

2.5.1 Initial Size and Firm Survival:

We test the prediction of our model using Cox proportional hazard ratio regression. We treat exit as failure. All types of firms are at risk of failure from the time they enter software exports industry. We control for timing of entry and allow for interaction between time and age of the firm.

The result is shown in Table 2.6. We find that conditioned on current size, initial size has negative influence on firm survival. A 1% increase in initial size increases annual hazard of exit by 15%. However, a 1% increase in current size has the effect of decreasing annual hazard of exit by 28%. The foreign branch, diaspora and spinoff face lower annual hazard of exit. There are no benefits of being either a diversifier or startup.

These results confirm the prediction of the model that the initial size of the firm is exogenous and observed by firm before they enter the software exports industry. Our results are similar to those found by Mata et al. (1995). However, they add coefficients on current size, which is negative and initial size, which is positive, and interpret this sum of two coefficients. They find that overall effect of initial size is beneficial for firm survival. Our model tells under what condition it is possible to obtain positive coefficient on initial size. This gives us clear idea about what it means to find positive coefficient on initial size. The results mean that firm’s initial size is exogenous and is observed by firm before they enter. Our results also tell us about the nature of the industry. What it means is that in this industry firms know their true efficiency but they are not able to communicate it to the market. Firms get contracts and the size of the contract is decided by the contracting party and they cannot influence its size.
Table 2.6
All entrants: Cox Proportional Hazard Ratio
Regression

|                      | Coeff. | Haz. Ratio | P>|z| |
|----------------------|--------|------------|-----|
| Initial Size         | 0.14   | 1.15       | 0.03|
| Initial size missing dummy | -0.54  | 0.58       | 0.18|
| Current Size         | -0.33  | 0.72       | 0.00|
| Current size missing dummy | -0.40  | 1.50       | 0.43|
| Foreign branch dummy | -0.94  | 0.39       | 0.02|
| Diaspora dummy       | -1.04  | 0.35       | 0.02|
| Spinoff dummy        | -0.69  | 0.50       | 0.07|
| Diversifier dummy    | -0.05  | 0.94       | 0.87|
| Startup dummy        | -0.09  | 0.91       | 0.80|

Control for entry timing, age YES
N 329

Cluster corrected standard errors, Non-traceable/others is omitted type

The larger firms face lower annual hazard of exit. There are many benefits of scale especially in the software services exports sector (Arora et al., 1999). It is easier for larger firm to get bigger overseas contracts.

The diaspora, foreign branch, and spinoffs face lower annual hazard of exit. Though diaspora have lowest annual hazard of exit, the difference between these three types is not significant. The diaspora and foreign branch have better access to market as these firms are mainly located in overseas country where market is and is therefore easier for them to get contracts. On the other hand spinoffs have experience in the industry. Similarly, diaspora and spinoffs have local knowledge of business environment and labor market in India where they are setting up operations.

In other industries spinoffs have been found to exhibit better performance than the startups (Klepper ICC, Agarwal et al., 2004). However, our finding that diversifiers\(^2\) do not have any advantage relative to startups or spinoff is contrary to findings of many studies, though there is some support in literature. In the disk drive industry it was found that only spinoffs have higher survival probability compared to diversifiers and incumbent firms (Agarwal et al., 2004). The diversifiers in the information security industry survive longer than both related startups (spinoffs) and unrelated startups (Arora et al., 2007).

\(^2\) Though we do not report here, even the established diversifying firms who have some international exposure, were not found to have any advantage.
But why diversifiers do not perform better even if they have access to capital, experience of running business, etc.? Perhaps because the software exports business is so different from the typical Indian manufacturing experience. It requires hiring and managing highly trained people, many of whom will work at the client’s site. Plus, it may require marketing to foreign clients, meeting tight timelines and quality standards, neither of which is characteristic of the Indian firm experience.

We also find that the earlier entrants do not have any advantage over the later entrants. This is consistent with the finding that in the earlier years Indian firms were experimenting with business models of delivery of services to overseas clients before mixed delivery model was accepted (Athreye, 2005).

2.5.2 Founders’ Human Capital and Firm Survival:

We now focus on de novo firms only. We are interested in finding the effect of founders’ general human capital on the survival of firms. We use Cox proportional hazard ratio regression. We treat exit as failure and again as with all entrants all types of firms are at risk of failure from the time they enter the software exports industry. We control for timing of entry and allow for interaction between time and age of the firm.

The result is shown in Table 2.7. Similar to the case for all entrants, the current size has positive effect on firm survival and initial size have negative effect on firm survival. What is now important is that once we control for founders’ human capital, the effect of firm type disappears. It does not matter whether firm type is diaspora or spinoff. The firms with foreign educated founders face 77% lower annual hazard of exit.

We can better appreciate the results of Table 2.7 if we analyze initial size of de novo firms. The results of OLS regression of natural log of initial size on type, education indicator variables with control variables are shown in Table 2.8. We use only those firms for which initial size is not missing. The omitted category for firm type is startup and is undergrad in liberal arts/science for education. These results show that spinoffs and diaspora on an average enter with bigger initial size compared to startups. These results also show that the firms with founders who have undergraduate degree in engineering and MS or MBA from a foreign country enter on an average with the same size as the startups with undergrad in liberal arts/science. However, it is important to note that 47 out of 49 firms with foreign educated founders are either diaspora or spinoffs. If we had found their size to be smallest, then given the choice of initial size we could have said that they are the most efficient. Perhaps these firms with foreign educated founders are doing something different which explains their superior performance.
In the regression for all entrants we found that diaspora and spinoffs had advantage over startups. The advantage of spinoff lies in the fact that they have direct experience of working in the industry. The diasporas have experience of relevant market. This suggests that experience matters. However, these results suggest that the acquiring higher education in the foreign country, where there is market for software services, is similar to having direct experience.
2.6 Conclusions:

This study contributes to the literature by developing a model that rationalizes empirical findings. We are able to tell the conditions under which it is possible to get positive coefficient on initial size conditioned on current size. The findings imply that the software exporters in India though knew about their true efficiency but were not able to communicate it to the market. These results are robust even when we limit our analysis to de novo firms and include founders’ human capital as additional control. These results thus tell us about the very fundamental nature of the industry. There are limits to contracting in this industry (Banerjee, 2000).

Even though this is technology intensive industry, firms whose founders have only undergraduate degree in engineering do not have any benefit. However, the firms whose founders have undergraduate degree in engineering and MS or MBA from a foreign country have huge benefit. Once we control for the founder’s human capital, then firm types do not have independent effect on survival of firm. This suggests that engineer founders with MS or MBA from foreign country are not handicapped because of lack of experience in the relevant industry. This type of human capital facilitates access to market by helping overcome some limits of contracting inherent in this industry.

Our finding that spinoffs have lower annual hazard of exit is similar to the findings in many studies both for manufacturing and hi-tech industry. However, that diversifying firms do not have any advantage is contrary to the many empirical findings. It suggests that the advantages like better access to capital, experience of managing business which diversifiers have is not relevant for success in this industry. What is relevant is the direct experience of working in this industry.

We find that it is advantageous to be foreign branch or diaspora. This suggest that access to the markets is important in the software export industry. However, since diaspora firms perform better than foreign firms, local knowledge is also beneficial.
2.7 References


2.8 Appendix 1:

1. Optimum value of $M_0$

$$E(\pi_0 | \varepsilon_i > K) = \frac{\arg \max_{M_0} \int_K^{\infty} ((p - c)(\lambda Q_0 + \alpha \psi(M_0) + \varepsilon_i) - F)\phi(\varepsilon_i) d\varepsilon_i - M_0}{M_0}$$

Applying Leibniz integral rule\(^3\) to above equation we get the condition for optimum value of $M_0^*$

$$(p - c)\alpha \psi'(M_0)\int_K^{\infty} \phi(\varepsilon_i) d\varepsilon_i - 1 = 0$$

$$(p - c)\alpha \psi'(M_0)(1 - \Phi(K)) = 1$$

$$\psi'(M_0) = \frac{1}{1 - \Phi\left(\frac{F - \lambda Q_0 - \alpha \psi(M_0)}{p - c}\right)}\left(\frac{1}{a(p-c)}\right)$$

We plot left hand side and right hand side of above equation to show that there exists a unique value of $M_0$ that is increasing in $\varepsilon$ (see Fig. 1).

2. Next we show that $\frac{\partial X^*}{\partial Q_0} < 0$, when $Q_0$ is exogenous but observed.

Let profit of firm $\pi = \pi_0^* + V_1$, where $V_1$ is expected present value of future profits, and $\pi_0^* = X^*Q_0 - F$. Firm will enter if profit $\pi$ is greater than or equal to the entry cost $E$. That is

$$X^*Q_0 - F + V_1 - E = 0$$

On differentiating equation (4') with respect to $Q_0$

$$X^* + Q_0 \frac{\partial X^*}{\partial Q_0} + \frac{\partial V_1}{\partial Q_0} = 0$$

$$\frac{\partial X^*}{\partial Q_0} = \frac{-(\frac{\partial V_1}{\partial Q_0} + X^*)}{Q_0} < 0$$

We plot $X^*Q_0$ and $F + E - V_1$ as a function of $X$ (see Fig. 2). As entry size increases, the entry threshold decreases, i.e., $\frac{\partial X^*}{\partial Q_0} < 0$.

3. We now show $\frac{\partial Q_0}{\partial X} > 0$, when $Q_0$ is endogenously chosen.

From entry condition $\pi - E \geq 0$, we get

$$\frac{\partial \pi}{\partial Q_0} - \frac{\partial E}{\partial Q_0} = 0$$

\(^3\) Please note that the integrand reduces to zero at lower limit of integration and is zero for upper limit of integration.
Taking \textit{partial derivative} of above equation with respect to price-cost margin \( X = (p - c), \) we get

\[
\left( \frac{\partial^2 \pi}{\partial Q_0^2} - \frac{\partial^2 E}{\partial Q_0^2} \right) \frac{\partial Q_0}{\partial X} + \frac{\partial^2 \pi}{\partial Q_0 \partial X} = 0
\]

\[
\frac{\partial Q_0}{\partial X} = \frac{\partial \pi}{\partial Q_0 \partial X} = -\left( \frac{\partial^2 \pi}{\partial Q_0^2} - \frac{\partial^2 E}{\partial Q_0^2} \right)
\]

We first determine \( \partial \pi / \partial Q_0 \) and then \( \partial^2 \pi / \partial Q_0 \partial X. \) The profit \( \pi \) can be written as:

\[
\pi = \pi_0(Q_0, X) + \pi_1(Q_1(Q_0, M_0(Q_0)), M_0(Q_0), X) +
\]

\[
\frac{\partial \pi}{\partial Q_0} = \frac{\partial \pi_0}{\partial Q_0} + \frac{\partial \pi_1}{\partial Q_0} \frac{dQ_1}{dQ_0} + \frac{\partial \pi_1}{\partial M_0} \frac{dM_0}{dQ_0}
\]

\[
\frac{\partial \pi}{\partial Q_0} = \frac{\partial \pi_0}{\partial Q_0} + \frac{\partial \pi_1}{\partial Q_0} \left( \frac{\partial Q_1}{\partial Q_0} + \frac{\partial Q_1}{\partial M_0} \frac{dM_0}{dQ_0} \right) + \frac{\partial \pi_1}{\partial M_0} \frac{dM_0}{dQ_0}
\]

\[
\frac{\partial \pi}{\partial Q_0} = \frac{\partial \pi_0}{\partial Q_0} + \frac{\partial \pi_1}{\partial Q_0} \frac{dQ_1}{dQ_0} + \left( \frac{\partial \pi_1}{\partial M_0} + \frac{\partial \pi_1}{\partial Q_1} \frac{dM_0}{dQ_0} \right) \frac{dM_0}{dQ_0}
\]

\[\text{FOC for } M_0:\]

\[
\frac{\partial \pi_1}{\partial M_0} = 0 \quad (5')
\]

\[
\frac{\partial \pi_1}{\partial Q_0} + \frac{\partial \pi_1}{\partial Q_1} \frac{dQ_1}{dM_0} = 0
\]

Using FOCs for \( M_0, \) the above equation simplifies to:

\[
\frac{\partial \pi}{\partial Q_0} = \frac{\partial \pi_0}{\partial Q_0} + \frac{\partial \pi_1}{\partial Q_0} \frac{dQ_1}{dQ_0}
\]

\[
\pi_0 = XQ_0 - F
\]

\[
\pi_1 = E(XQ_1 - F) - M_0
\]

\[
\frac{\partial \pi}{\partial Q_0} = X + E(\lambda X)
\]

\[
= X + \lambda X (1 - \Phi(K))
\]

\[
= X \{ 1 + \lambda (1 - \Phi(K)) \}
\]

\[
K = \frac{F}{X} - \lambda Q_0 - \alpha \psi(M_0)
\]
\[
\frac{\partial^2 \pi}{\partial Q_0 \partial x} = \left\{ 1 + \lambda \left( 1 - \Phi (K) \right) \right\} \cdot \left( X \left[ - \lambda \phi \left( - \frac{F}{X^2} - \alpha \frac{\partial \psi (M_0)}{\partial x} \right) \right] + \lambda \frac{F}{X} \phi (K) + X \alpha \phi (K) \frac{\partial \psi (M_0)}{\partial x} \right\}
\]

**We now show that** \(\frac{\partial \psi (M_0)}{\partial x} > 0\)

The first order condition for \(M_0\) can be obtained by differentiating \(\pi_1 (M_0, X)\) with respect to \(M_0\). This condition is

\[
T = \frac{\partial \pi_1}{\partial M_0} = 0
\]

\(T_{M_0} = \frac{\partial T}{\partial M_0} < 0\), **being the second order condition for** \(M_0\).

\[
\frac{\partial T}{\partial X} = \frac{\partial^2 \pi_1}{\partial M_0 \partial x}
\]

\(\pi_1 = E (X Q_1 - F) - M_0\)

\[
\frac{\partial \pi_1}{\partial x} = E (Q_1) = Q_1 (1 - \Phi (K))
\]

\[
\frac{\partial^2 \pi_1}{\partial x \partial M_0} = \left( 1 - \Phi (K) \right) \frac{\partial Q_1}{\partial M_0} + Q_1 (-\phi) (-\alpha \psi (M_0))
\]

\[
= \left( 1 - \Phi (K) \right) \alpha \psi (M_0) + \alpha \psi (M_0) \phi (K_{01}) Q_1
\]

\[
= \alpha \psi (M_0) \left[ 1 - \Phi (K) + \phi (K) Q_1 \right]
\]

Thus

\[
T_X = \frac{\partial T}{\partial X} \frac{\partial^2 \pi_1}{\partial M_0 \partial X} > 0
\]

\(dT = T_X \, dX + T_{M_0} \, dM_0 = 0\)

\(dT = T_X \, dX + T_{M_0} \, dM_0 = 0\)

\[
\frac{\partial M_0}{\partial x} = -T_X > 0
\]

\[
\frac{\partial \psi (M_0)}{\partial x} = \frac{\partial \psi (M_0)}{\partial M_0} \frac{\partial M_0}{\partial x}
\]

We know that \(\frac{\partial \psi (M_0)}{\partial M_0}\) is positive and \(\frac{\partial M_0}{\partial x}\) is also positive, that makes \(\frac{\partial \psi (M_0)}{\partial x}\) positive.
This makes
\[
\frac{\partial^2 \pi}{\partial Q_0 \partial X} = \left\{ 1 + \lambda \left( 1 - \Phi(K) \right) \right\} + \lambda \frac{F}{X} \phi(K) + X \alpha \phi(K) \frac{\partial \psi(M_0)}{\partial X} > 0
\]

\[
\frac{\partial Q_0}{\partial X} = \frac{\frac{\partial^2 \pi}{\partial Q_0 \partial X}}{\frac{\partial^2 E}{\partial Q_0^2}}
\]

The denominator in above is positive as \( \pi - E \) concave in \( Q_0 \). We have also shown that numerator is positive. This makes \( \frac{\partial Q_0}{\partial X} > 0 \).

\[Z = \frac{1}{1 - \Phi} \frac{1}{\alpha (p - c)}\]

![Figure 2.1: Optimal Value of \( M_0 \) and \( Q_0 \)](image)
Figure 2.2: Entry Threshold ($X^*$) and Entry Size ($Q_0$)

Figure 2.3: Optimal Value of $M_0$ and $X = (p - c)$
Chapter 3

Affirmative Action and Academic Achievement in India:
An Empirical Investigation

3.1 Introduction

Several important issues are at the heart of the affirmative action debate. One of the most salient issues is whether students who receive preferential admissions because of affirmative action, benefit from these policies. Because of affirmative action, students from disadvantaged groups, such as Blacks in the USA and Scheduled Castes in India, are given admission to colleges even when they have lower test scores than non-affirmative action students. The differences in scores may be quite large, particularly in selective colleges where the affirmative action really matters.

One criticism of the affirmative action is that it leads to a mismatch between the scholastic requirement of the colleges and the academic preparation of the students from groups benefitting from these affirmative action policies (Alon et al., 2005, Rothstein et al., 2006). This hypothesis has been called the “fit” or “mismatch” hypothesis. Critics of the policy argue that students with lower test scores and GPAs in high schools cannot compete in the academically challenging environment of selective institutions. This mismatch results in lower graduation rates and GPAs in colleges for such students. The critics argue that these students may be better off in colleges where their academic preparedness matches the average peer quality (Bowen et al. 2000).

In this paper, we evaluate the “fit” or “mismatch” hypothesis by analyzing the first-year academic outcomes of students in 210 engineering colleges in a major Indian state. For this analysis, we have assembled a unique dataset, with largest number of colleges and students, and diverse socio-economic background of students (to our knowledge), that has information on students’ performance in entrance examinations, high-school grades (junior college), caste, if eligible for affirmative action benefit, whether the student was admitted through to affirmative action policies and details on college and discipline to which they are admitted. The unique nature of our dataset and setting allows us to specifically identify students who were admitted due to affirmative action policies, which is difficult in many other countries.

In the United States, the study of affirmative action effects poses some difficulties and some of these make empirical analysis difficult. Firstly, each university, or college has its own admission policy. It is difficult to quantify factors influencing admission decisions, e.g., recommendation letters, etc.
Secondly, it is difficult to identify beneficiaries of affirmative action policies. We can only able to guess that a particular student is a beneficiary of the affirmative action policies because of his/her race or gender. Thirdly, since each university also has its own evaluation system, any comparison amongst students of different universities is difficult.

Our empirical setting solves many of these empirical problems. For our study, we use data on admissions to over 200 engineering colleges. The students in all colleges are admitted through a fully transparent common admission system laid out by the state government. Each college admits up to 80 percent\(^1\) of its capacity through this common admission system. The affirmative action policies mandate that 50 percent of these seats in each college be set aside for students belonging to specific castes. Thus, we know for each student his or her caste or social group, the name of the college, and whether the student was admitted under affirmative action. Secondly, all these colleges are affiliated to one university\(^2\), which sets a common curriculum for all the disciplines and administers a common examination to all students. This enables us to compare academic performance of all students in these colleges.

We find that college selectivity helps students of disadvantaged social groups as well as those not eligible for affirmative action benefits. Most students from disadvantaged social groups are admitted to more selective colleges due to affirmative action. This “jump” from a less selective college to a more selective one helps improve test scores in the first year. Although we do not explore reasons for such improvement, we conjecture that this may be due to better peer quality and college resources.

This paper is organized as follows. In the next section we briefly provide some background on affirmative action policies in India. In section 3 we briefly review the literature. In section 4 we describe the data collection and show some main features of the dataset. In section 5, we formulate our research question. In section 6 we describe our empirical strategy followed by a discussion of our results in section 7. Section 8 summarizes our findings and concludes. Finally, we provide some possible directions for future work in section 9.

### 3.2 Affirmative action in India\(^3\):

---

\(^1\) Some colleges even willing may not get requisite number of students, and some minority colleges limit of percent of seats that can be filled to this admission system.

\(^2\) There are some colleges, which admit students under common admission system, but are not affiliated to this university. We do not include these students in our analysis.

\(^3\) Right now I do not go into the historical backgrounds and origin of caste-based affirmative action programs. I intend to do it in my dissertation.
The Indian caste system divides society into closed hereditary groups (Shah et al., 2006). The numerous castes in India can be classified into five hierarchical groups, with graded inequality being a fundamental principal of the system (Ambdekar, ). The set of castes known as Scheduled Castes (SC) are at the bottom of the caste hierarchy, suffering the most discrimination in terms of social exclusion and access to educational opportunities. In addition, there are tribal communities, known as Scheduled Tribes (ST), who have lifestyle and religious practices quite distinct from mainstream Indian society (Deshpande, 2005). STs often live in remote and inaccessible places, which make access to education difficult. Finally, there are certain communities or castes that are classified as social and educationally backward classes (BC). This is mainly due to their educational backwardness, their position within the hierarchy of castes, and the occupations that members of these communities have traditionally pursued. The BCs are further divided into four distinct groups: BC-A, BC-B, BC-C, BC-D depending on vocation, their degree of backwardness, and variety of other pertinent factors. Historically, the BCs have also had difficulty accessing educational opportunities. Relatively speaking, SCs/STs are considered more socially and educationally backward compared to BCs.

To address the widespread inequality that has resulted from the caste system, the Indian Constitution prohibits discrimination based on caste and has provided measures of positive discrimination for ameliorating the social and economic conditions of backward castes. To this end, the federal and state governments have taken various initiatives for improving the access to secondary and higher education for these groups. For instance, some of these initiatives include special residential schools, free accommodation, scholarships, concessions in tuition, as well as other initiatives that may improve their access to higher education. The most important aspect of the affirmative action programs is the reservation of a certain percent of capacity in higher educational institutions for students from these social groups4. The Indian state we are studying reserves following percentage of capacity in each college for these social groups: 15 percent for SCs, 6 percent for STs, 7 percent for BC-As and BC-Ds, 10 percent for BC-Bs, and 1 percent for BC-Cs. Moreover, one-third of capacity in each discipline is reserved for female students from all social groups.

3.3 Previous Studies:

There have been relatively very few empirical studies on the affirmative action’s impact on academic achievement. One comprehensive study by Kirpal et al. (1999) analyzes the academic performance of SC/ST students at the Indian Institutes of Technologies (IITs). They have 436 OPEN

4 The affirmative action program also operates in matters of employment in public sector, recruitment into civil services, etc.
(non-affirmative action), 115 SC, and 21 ST students in their sample from five main campuses of IITs. They find that OPEN category students have highest mean cumulative performance index (CPI) of 7.88, SC students have mean CPI of 6.2, and ST students have lowest mean CPI of 5.9. These mean CPI differences are significant. The graduation rate is also lower for SC/ST students compared to OPEN category students. Authors also showed that academic performance is correlated with socio-economic background of students. However, the sample is very small and is not representative since it is drawn from the most selective colleges in the country.

The most recent empirical analysis of affirmative action in India is study by Bertrand et al. (2008). They study affirmative action in engineering colleges in one of the state in India. They find that affirmative action benefits lower-caste students who are beneficiaries of this policy. They analyze the labor market outcome of the engineering graduates.

Another comprehensive study in the US setting by Bowen et al. (2000) finds that overall graduation rate for African-American is 79 percent and for Whites is 94 percent. Interestingly they find that graduation rates are positively associated with the degree of college selectivity and this association holds for different bands of combined SAT scores. They find no support for the ‘fit’ hypothesis. Instead they find that college selectivity improves graduation rates considerably even at the lowest SAT band. There is better match between the African-American students’ SAT score and the test-score profile of the least selective schools. Instead, Bowen et al. (2000) find that the African-American graduated with the lowest graduation rates from these least selective colleges. In terms of grade performance, the differences between African-American and Whites are significant. The average cumulative grade-point average (GPA) is 2.61 for African-Americans and 3.15 for Whites.

Alon et al. (2005) extend the work of Bowen et al. (2000) by including non-selective colleges in their analysis and again find no support for the “mismatch” hypothesis. Other studies such as those by Sander (2004) finds support for “mismatch” hypothesis (i.e. selectivity harms), whereas Rothstein et al. (2006) using the data of Sander find no support for “mismatch” hypothesis.

The most important criticism of Bowen et al. (2000) is that they draw conclusions about impact of affirmative action from ‘a highly atypical sample of minority students attending highly atypical colleges and universities’ (Sowell, 2004). Other studies have shown that difference in graduation rates

---

5 Lower graduation rates for black students can be due to ‘mismatch between black students’ preparation and academic level of the schools that admit them’. ‘Black students will be more likely to graduate, the argument goes, if they enroll in a school where their own SAT score matches with the school’s test-score profile than if they “reach too high” and go to a school where most fellow students have higher test scores’ (Bowen et al., 200).
between black and white students can be explained by the difference in the combined SAT score. The smaller the gap between SAT scores, narrower will be the differences in graduation rates (Sowell, 2004).

This study is for very large sample of students in large number of college. The result is that we have full spectrum of college selectivity in our dataset. This helps us overcome the problem of atypical sample of students attending highly atypical colleges.

3.4 Data:

The focus of this study is limited to first-year students of four-year undergraduate engineering colleges in the Indian state of Andhra Pradesh. This sample of engineering colleges includes both public and private\(^6\) colleges. Private colleges, though established with the approval of federal government, do not receive any financial support from government. The private colleges can be either non-minority institutions or minority\(^7\) institutions. Almost all these colleges offer undergraduate degree in more than one discipline. The private colleges are required to admit 80% of the sanctioned capacity in each discipline based on rank obtained in the common entry examination and the management of a particular college can fill remaining 20% of the capacity. All non-minority private college and public college admit students based on entry examination rank. The minority institutions can conduct separate entry examination for admitting students in these colleges.

With some exceptions, colleges are affiliated to one of several universities in the state. As only universities can grant degrees, affiliation is required for colleges to operate. The university sets the curriculum, conducts examinations and evaluates students of its affiliated colleges. Each affiliating university sets norms/standards for granting affiliation. The colleges have to meet specified resource requirement: faculty, building, laboratories, and other relevant factors for maintaining expected level of quality. The colleges have to seek affiliation when they are established and periodically afterwards. The periodicity of affiliation depends on the past track record of the college in meeting the norms of affiliation. For example, a college may be required to seek affiliation annually unless it achieves high on these affiliation norms.

Entry examination:

We use dataset of all students who have taken a common entry examination in May 2006. A student should have appeared for their 12\(^{th}\) grade state-wide examination or already passed this

\(^6\) Private colleges are also known as unaided private colleges.

\(^7\) The minority college can be set up by religious or linguistic minority. For example college can be Muslim or Christian minority. However, the most of the colleges are non-minority private unaided.
examination for being eligible to appear for the common entry examination. The entry examination is objective type three hour long test with examination in mathematics, physics and chemistry. The students are required to give information about their caste/social group and some other personal details in the application form for entry examination.

The minimum score for qualifying this test, i.e., for getting a rank assigned is 48 out of maximum score of 160. However, the SC/ST students are assigned rank even if their score is zero. Total of 142,095 students took this examination. 105,831 students qualified the entry test and were assigned ranks, and their distribution amongst different social groups is shown in Table 3.1.

<table>
<thead>
<tr>
<th>Social group</th>
<th>Rank assigned</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheduled Caste (SC)</td>
<td>12564</td>
<td>11.9</td>
</tr>
<tr>
<td>Scheduled Tribe (ST)</td>
<td>3739</td>
<td>3.5</td>
</tr>
<tr>
<td>Backward Class – group A (BC-A)</td>
<td>6226</td>
<td>5.9</td>
</tr>
<tr>
<td>Backward Class – group B (BC-B)</td>
<td>14063</td>
<td>13.3</td>
</tr>
<tr>
<td>Backward Class – group C (BC-C)</td>
<td>828</td>
<td>0.8</td>
</tr>
<tr>
<td>Backward Class – group D (BC-D)</td>
<td>10365</td>
<td>9.8</td>
</tr>
<tr>
<td>Open category</td>
<td>58046</td>
<td>54.8</td>
</tr>
<tr>
<td>Total</td>
<td>105831</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The students are given rank based on the performance in the entry examination. The student with highest score is given a rank of one, with second highest score given a rank of two, and so on. The ranking of a student in the entry examination does not depend on caste or social group of the student. A tie-breaking procedure is followed when there is more than one student with the same score. If students have same total score, then the one with higher score in mathematics is given higher rank. If students have same total and mathematics score, then with higher score in physics is given higher rank. If they have same total, mathematics, and physics score then they are given the same rank. In such case, ties are broken by the student’s score on the 11th and 12th grade examinations.

---

8 At this point we are not considering the students who were later qualified after minimum score was reduced from 48 to 40.
Table 3.2  
Caste-wise Distribution of Rank in Entry Examination

<table>
<thead>
<tr>
<th>Percentile</th>
<th>OPEN</th>
<th>SC</th>
<th>ST</th>
<th>BC-A</th>
<th>BC-B</th>
<th>BC-C</th>
<th>BC-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th</td>
<td>3900</td>
<td>20890</td>
<td>28852</td>
<td>9099</td>
<td>5435</td>
<td>9152</td>
<td>6353</td>
</tr>
<tr>
<td>25th</td>
<td>21011</td>
<td>61602</td>
<td>70508</td>
<td>34617</td>
<td>25876</td>
<td>33324</td>
<td>27285</td>
</tr>
<tr>
<td>50th</td>
<td>44458</td>
<td>93059</td>
<td>98266</td>
<td>59224</td>
<td>49918</td>
<td>56252</td>
<td>50286</td>
</tr>
<tr>
<td>75th</td>
<td>70421</td>
<td>101622</td>
<td>102156</td>
<td>79376</td>
<td>74116</td>
<td>75305</td>
<td>73283</td>
</tr>
<tr>
<td>No. of students</td>
<td>58046</td>
<td>12564</td>
<td>3739</td>
<td>6226</td>
<td>14063</td>
<td>828</td>
<td>10365</td>
</tr>
</tbody>
</table>

The performance of students of different caste/category is shown in Table 3.2. The students in the OPEN category have best average performance and students in ST have lowest average performance in the test. The students of BC-B category have superior performance compared to other sub-categories of the backward castes.

Admissions:

All public and non-minority private colleges in the state admit students through a government monitored admission process. The minority private colleges can also choose to admit some students through this admission process. In this study we focus on students who were admitted to public and private colleges through a government monitored admission process. The students are requested to attend counselling session for admission in order of the rank in the entry examination, with the highest rank holder attending the session first. On the day of counselling, the students meet admission officer at counselling counters. At the counselling counters, the candidate is shown all the vacant seat where he/she can get admitted. Allotment is made when student’s choice matches with the vacant seat available (see Appendix A for more details).

The first phase of counselling is called ‘general counselling’, in which all students can participate and every such student is admitted treating him/her as OPEN category students. The maximum of 50 percent of slots in each discipline can be filled through OPEN category students in each discipline in each college during the general counselling. This is followed by counselling for students of reserved castes (caste-wise counselling), with students of different castes being called separately. During this counselling maximum admissions in each discipline can be equal to the quota for different castes. For example, SC candidates can fill maximum of 16 percent of the slots.

---

9 Large fraction of 80 percent of the capacity is admitted by the minority colleges through their admission process that is similar to government monitored admission process, but government has no direct control on this process.
In the general counselling, some backward caste students also participate, however they are treated as OPEN category students with no benefit of affirmative action available to them. Those with superior ranks will be able to obtain admission in the discipline and college of their choice and those with average or below average rank may still secure admission but may not be able to get admission to a discipline and college of their choice. However, if the student has a low rank or is not sure of getting admission as per his/her expectation, then they may not at all show up for the general counselling. Some students participate in the general counselling with the intension of getting something, knowing that they have the opportunity of improving the choice of the discipline and college during the caste-wise counselling.

For instance, suppose a student from a backward caste takes admission during general counselling. This student is also able to participate in the caste-wise counselling. And if he/she can obtain better college or discipline during the caste-wise counselling, then he/she will take admission during the caste-wise counselling. This student then vacates his/her slot secured during the general counselling. However, due to the specific regulation, such vacated slots are again filled by students of respective caste only and not by OPEN category students. This has the effect of increasing the number of admits from backward caste beyond the quota fixed for that caste. For example, the percent of candidates admitted from BC-B may be more than a quota of 10 percent.

<p>| Table 3.3 Caste-wise Distribution of Admitted Students |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Caste-wise Category</th>
<th>SC</th>
<th>ST</th>
<th>BC-A</th>
<th>BC-B</th>
<th>BC-C</th>
<th>BC-D</th>
<th>OPEN</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6300</td>
<td>2106</td>
<td>4759</td>
<td>9320</td>
<td>618</td>
<td>6765</td>
<td>27640</td>
<td>391</td>
<td>57899</td>
</tr>
<tr>
<td>Admitted in same caste category</td>
<td>6180</td>
<td>2050</td>
<td>4581</td>
<td>8534</td>
<td>574</td>
<td>6025</td>
<td>26427</td>
<td>391</td>
<td>54762</td>
</tr>
<tr>
<td>Admitted as OPEN/CAP/SP/NCC</td>
<td>120</td>
<td>56</td>
<td>178</td>
<td>786</td>
<td>44</td>
<td>740</td>
<td>1213</td>
<td></td>
<td>3137</td>
</tr>
<tr>
<td>% total</td>
<td>10.9</td>
<td>3.6</td>
<td>8.2</td>
<td>16.1</td>
<td>1.1</td>
<td>11.7</td>
<td>47.7</td>
<td>0.7</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Notes: 1. OPEN students can be admitted as OPNE as well as CAP/NCC/SP. 2. CAP: Children of Armed Forces Personnel, NCC: National Cadet Corps, and SP: Sports and Games. 3. 54762+3137=57899.

Table 3.3 shows the details of students admitted through government monitored admission process. A total of 57,899 students from different castes were admitted. First row of the table shows number of students from different castes who were admitted. As explained before, some students from backward castes got admissions without benefit of affirmative action. For example, 6,300 students from SC were admitted. As shown in the second row, 6,180 of 6,300 got admission under affirmative action.
program. Third row shows number of students from different caste that got admission as either OPEN or under different category. We can also see that for some castes, the percent of total exceeds the statutory quota for that caste (e.g. BC-B).

### Table 3.4
**Admissions by Management of College**

<table>
<thead>
<tr>
<th>Type of college management</th>
<th>No. of colleges</th>
<th>Capacity</th>
<th>Students admitted</th>
<th>Admitted as percent of capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>11</td>
<td>2655</td>
<td>2508</td>
<td>94.5</td>
</tr>
<tr>
<td>Minority private</td>
<td>27</td>
<td>9995</td>
<td>2063</td>
<td>20.6</td>
</tr>
<tr>
<td>Non-minority private</td>
<td>207</td>
<td>76120</td>
<td>53328</td>
<td>70.1</td>
</tr>
<tr>
<td>Total</td>
<td>245</td>
<td>88770</td>
<td>57899</td>
<td>65.2</td>
</tr>
</tbody>
</table>

These students were admitted in both public and private colleges (see Table 3.4). The largest numbers of students, 53,328 were admitted in 207 non-minority private colleges. The minority colleges admitted 2,063 students in 27 colleges. The 11 public colleges admitted 2,508 students. As minority private colleges have their own admission process, they only admit around 20 percent of their capacity through the government monitored admission process. However, there are few colleges which seem to have admitted students through this admission process only. Each non-minority private colleges generally admit up to 80 percent of their capacity through the government monitored counselling process. The non-minority private colleges, on an average, admit around 70 percent their capacity as some colleges could fill very small fraction of their capacity.

### Table 3.5
**Discipline-wise Admissions**

<table>
<thead>
<tr>
<th>Discipline</th>
<th>No. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics and communication engineering</td>
<td>16015</td>
</tr>
<tr>
<td>Computer science and engineering</td>
<td>14809</td>
</tr>
<tr>
<td>Electrical and electronics engineering</td>
<td>8930</td>
</tr>
<tr>
<td>Information technology</td>
<td>8169</td>
</tr>
<tr>
<td>Mechanical engineering</td>
<td>4636</td>
</tr>
<tr>
<td>Civil engineering</td>
<td>1450</td>
</tr>
<tr>
<td>Electronics and instrumentation engineering</td>
<td>1081</td>
</tr>
<tr>
<td>Others</td>
<td>2809</td>
</tr>
<tr>
<td>Total</td>
<td>57899</td>
</tr>
</tbody>
</table>
The students have choice of variety of disciplines (see Table 3.5) but largest number of students chose electronics and communication engineering (27%). The computer science and engineering (25%) is second popular choice. The electrical and electronics engineering is at third place with 15% of students, followed by information technology with 14% students. The immense growth in information technology industry has pushed mechanical engineering to fifth place with 8% students.

**Dependent variable:**

The dependent variable is the measure of academic achievement. It is test score in common examinations taken by all these students at the end of their first year in engineering college. The students write approximately three hour long comprehensive test for each of the seven theory subjects at the end of the year. The question paper for a given theory subject is same for all students in different colleges affiliated to this university. The answer papers are evaluated not by faculty members of the college but by independent faculty members drawn from different colleges. The students also take practical examinations but our focus on theory papers alone ensures most common standard for evaluation of students across different colleges. The students’ subjects of this year-end examination depend on the discipline of engineering to which he/she has been admitted. The disciplines like electronics and communication engineering, computer science and engineering, electrical and electronics engineering, and information technology, which account for nearly 81% of admissions, also have six out of seven theory subjects common.

The performance any subject is generally measured as percentage of marks, which is obtained by dividing the test score in each subject by the maximum score obtainable in that subject. The performance in the examination is mean of all such percent marks in seven subjects. We can also measure performance in the examination as sum of test scores in all seven subjects. Or we can convert it into a rank variable, with rank of one assigned to the student with highest test score.

**Sample:**

We construct the database by tracking 105,831 students who were assigned rank in the entry examination, which they took for admissions to engineering colleges in May 2006. Out of these 57,899 were admitted to 245 public and private colleges through government monitored common admission
in July/August 2006. Further we retain only those students which were admitted in private colleges affiliated to one university. We then collected results of the performance in the year-end examination in May/June 2007 from the university which administers common examination to all students of different colleges affiliated to it. We then matched admission records with the results records. We further match these records with records of performances in 12th grade. This reduces our sample to 210 colleges and 27,123 students.

3.5 Research Question:

In each discipline in each college fifty percent of capacity is reserved for students who are eligible for benefits of affirmative action policies. The remaining fifty percent seats are filled through OPEN category students. The priority in which students get to select into the college and choose discipline in that college depends on the rank in the entry examination. The students eligible for affirmative action benefits, in general, have lower performance (i.e. rank) in the entry examination and therefore in the absence of affirmative action policies the capacities in the top-ranked colleges would have been filled in mostly by OPEN category students. Only few of caste-students who had better rank in entry examination would have got admission in the top-ranked colleges. The share of caste-students in the top-ranked colleges would have been nowhere near fifty percent in the absence of affirmative action. However, due to quota of seats for the caste-students, it is possible for these students to get admission in the top-ranked colleges even with relatively lower rank in the entry examination. Thus, affirmative action improves ranking of colleges to which the caste-students are admitted. We are interested in finding the effect of affirmative action on the academic performance. Or stated differently, how does improved ranking of the college for the caste-students affects their academic performance?

3.6 Research Methodology:

Our main interest is in finding the effect of affirmative action on the students’ academic achievement. We model academic achievement as function of student’s ability and college quality. The rank in the entry examination \(r\) is the rank order in which individual chooses college. This rank order is presumed to be the same as the ordering of college quality. At first blush one is then tempted to use rank in the entry examination \(r_i\) as a measure of college quality and therefore to measure the effect of

---

10 There are some students, up to 20 percent of the capacity of the college, directly admitted by colleges, and presently we do not focus on their academic achievement.
11 Even students eligible for affirmative action benefits can get admission as OPEN candidates if they have good rank.
12 It is quite likely that in one or two top-ranked colleges, students of the relatively advanced castes amongst the backward castes would have had share of seats near their quota. For such students, affirmative action helps improve the choice of discipline.
affirmative action on the academic performance. The First-year score is $y_i$, then can be modeled as

$$y_i = \beta_0 + \beta_i r_i + \varepsilon_i$$

The error term $\varepsilon_i$ impounds, among other things, the ability of the student. However, note that rank ($r_i$) potentially plays two roles in this regression. One role is as an indicator of ability. The other role of rank is in determining the priority in which students get to select into colleges. But, the effect of affirmative action is measured by the priority in which caste-students get to select into colleges. The fundamental challenge then is isolating these two effects of rank, the one that is measure of ability and the other one that determines the priority in which students get to select into colleges. We attempted several specifications to isolate these two effects of the rank. For example, in one specification we controlled for the ability component of rank by including fixed-effects for every possible value of total score, and scores in mathematics and physics components of the entry examination. However, the important feature of affirmative action policies and availability of detailed data on students’ caste helps us construct a variable that measures opportunity/access for students of different categories. This variable, called effective rank, is defined for each caste and measures the priority in which students of different categories get to select into colleges. Thus using effective rank along with rank ($r_i$) one would be able to isolate two potential roles of the rank. We now explain construction of the effective rank.

**Ranking within caste and effective rank:**

The entry examination ranks all students irrespective of their caste. However, due to affirmative action certain percent of seats are set aside or reserved for caste students. Therefore the students in different caste essentially constitute separate groups as far as priority in which students get to select into colleges is concerned. For example, if top three SC students have a rank of 135, 700, and 789. Then these students would be first three to be called for filling seats reserved for SC students. They would still be top three even if their ranks were 120, 379, and 679. Thus what matters is rank within a caste group. We address this issue by ranking individuals in each caste separately. The numerical ranking a student in a caste receives is based on all members of the caste who took the entry examination.

The capacity in each discipline in each college is divided amongst seven category/caste groups (OPEN, SC, ST, BC-A, BC-B, BC-C, BC-D) in proportion to their allocated quota. Thus value of a given ranking within a caste depends on the proportion of seats allocated to the caste. For example, a ranking of 50 for a caste assigned 500 seats has the same value as a ranking of 30 for a caste assigned 300 seats. The ranking within caste will not account for this feature of the affirmative action. We solve this problem by
dividing rank within the caste by the seat share for that caste, which we call effective rank. Let $r_{w,i}$ be ranking of an individual $i$ in caste $c$. Let $s_c$ be allocated seat share of caste. Then effective rank $r_{c,i}$ is defined as

$$r_{c,i} = \frac{r_{w,i}}{s_c}$$

For example, the rank within caste is divided by 0.15 for SC students, as 15 percent seats are reserved for these students (see Table 3.6). As explained in the admission section, due to specific regulation of the state government, the realized share of quota for some caste exceeds their allocated quota. We divide ranking within caste by the realized quota except for SC and ST. In the top-ranked colleges we observe the seat share of SC and ST being equal to their allocated share and therefore if we divide ranking within caste by the realized quota for SC and ST, then this will have effect of reducing their seat share below the allocated quota in the top-ranked colleges. The effective rank measures the priority in which students of different caste get to select into colleges. This construction of effective rank permits direct comparison of the coefficients across castes.

<table>
<thead>
<tr>
<th>Rank in entry examination</th>
<th>Ranking within caste</th>
<th>Effective Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>1</td>
<td>6.7</td>
</tr>
<tr>
<td>383</td>
<td>2</td>
<td>13.3</td>
</tr>
<tr>
<td>512</td>
<td>3</td>
<td>20.0</td>
</tr>
<tr>
<td>542</td>
<td>4</td>
<td>26.7</td>
</tr>
<tr>
<td>554</td>
<td>5</td>
<td>33.3</td>
</tr>
<tr>
<td>587</td>
<td>6</td>
<td>40.0</td>
</tr>
<tr>
<td>638</td>
<td>7</td>
<td>46.7</td>
</tr>
</tbody>
</table>

**Table 3.6 Effective Rank for SC Students**

**Empirical model/specification:**

We adopt a simple strategy for identifying the effect of affirmative action. Let $y_i$ be the performance in first-year examination, $r_i$ be the rank in entrance examination, $r_{c,i}$ be the effective rank for student $i$ and belonging to caste $c$. We model performance in the first-year examination as

1. $y_i = \alpha_0 + \alpha_1 r_{c,i} + \alpha_2 r_i + \epsilon_{p,i}$
In the above regression, $\varepsilon_{p,i}$ includes all factors other than rank and effective rank that influence $p_i$. In particular, ability is included in the error. Rank $r_i$ is determined by total score in the entry test (E), and score on the mathematics (M) and physics (P) component of entry test:

2. $r_i = f(E_i, M_i, P_i)$

A key observation is that this function does not depend on the caste of the students. We expect that $r_i$ will be correlated with $\varepsilon_{p,i}$ because the components of $r_i$ are measures of ability.

The $r_{c,i}$ govern access to colleges, and thus provide a measure of college quality attended. The $r_{c,i}$ contain no information about ability that is not already impounded in $r_i$. Similarly, the $r_i$ contain no information about college quality that is not captured by the $r_{c,i}$. Hence the coefficient of $r_{c,i}$ provides a measure of the effect of college quality that is not affected by the correlation of rank with ability once rank in entry examination $r_i$ is included in the regression.

Thus, the effect of affirmative action is identified by the regression in (1). This regression can be estimated by OLS. Identification is obtained from three features of the college choice process. First, the procedures determining rank $r_i$ are invariant across caste. Second, the assignment of seats to castes introduces substantial differences between caste access to colleges, $r_{c,i}$, and rank. Indeed, this is the intent of affirmative action. Third, idiosyncratic variation in performance of individuals on the components of the entry exam coupled with the reservation of seats to castes results in substantial variation of $r_{c,i}$ relative to $r_i$.

The effective rank, $r_{c,i}$, conditional on rank in entrance examination, $r_i$, is not correlated with ability, whether measured or unmeasured. This means that the estimate of $\alpha_i$ should not be biased by the inclusion of ability measures such as performance in the 12th grade examination (H). In fact, including H should improve the precision of the estimate of $\alpha_i$ because H reduces the unexplained variance.

We also include caste fixed-effects in our estimation (why?). Thus we estimating this regression

3. $y_i = \alpha_0 + \alpha_1 r_{c,i} + \alpha_2 r_i + \alpha_3 H_i + \delta \ast (\text{caste} - \text{fixed} - \text{effects}) + \varepsilon_{p,i}$
In (1) our dependent variable is test score and the independent variables are rank variables. In order to get right functional form we convert dependent variable into a rank variable, with rank 1 given to the highest test score, and the second-highest test score getting rank 2, and so on. Similarly we convert achievement in the 12th grade examination (H) also into a rank variable. When converting these variables into rank variables, we use information on all the students who take these examinations.

Thus we estimate (3) with dependent and independent variables being rank variables.

3.7 Results:

The students take examination at the end of their first year in engineering college in seven theory subjects and these subjects differ by discipline. This analysis is limited to disciplines\(^\text{13}\) which have six of seven theory subjects common so that the students’ performance can be compared. We match admission records with the first-year result records, and then with result records for entry examination and 12th grade, we are left with 27,123 students in our dataset.

We normalize all our variables. That means the normalized value of rank of test score for the best student is 1. Similarly, the effective rank has the value of 1 for best student in each caste/category.

<table>
<thead>
<tr>
<th>Table 3.7</th>
<th>Effect of Rank on First-year Achievement in Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Rank of test scores in theory subjects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Rank in entry examination</td>
<td>0.31</td>
</tr>
<tr>
<td>Effective rank within caste</td>
<td>0.27</td>
</tr>
<tr>
<td>Rank of 12th grade score</td>
<td>0.77</td>
</tr>
<tr>
<td>SC</td>
<td>-0.15</td>
</tr>
<tr>
<td>ST</td>
<td>-0.18</td>
</tr>
<tr>
<td>BC-A</td>
<td>-0.10</td>
</tr>
<tr>
<td>BC-B</td>
<td>-0.05</td>
</tr>
<tr>
<td>BC-C</td>
<td>-0.07</td>
</tr>
<tr>
<td>BC-D</td>
<td>-0.06</td>
</tr>
<tr>
<td>Constant</td>
<td>0.36</td>
</tr>
</tbody>
</table>

N 27123 27123

Notes: The dependent variable and independent variables have been normalized. For example, the best student in each category has the effective rank of 1. Similarly, best students’ score on first-year examination is 1.

\(^{13}\) These disciplines are electronics and communication engineering, computer science and engineering, electrical and electronics engineering, and information technology.
Table 3.7 shows OLS estimates for equation (3). As we see in specification (2), once we include student’s 12th grade score the coefficient on rank in the entry examination become insignificant. As one would expect, the positive sign on 12th grade score means students with higher ability have higher performance in the first-year.

The effect of affirmative action is measured by the coefficient on the effective rank. As explained before, the effective rank is also a measure of college quality and therefore the normalized value of 1 means that the student is admitted to the most selective college. Similarly, a normalized value of 0 means student is admitted to the least selective college. Therefore the positive coefficient on the effective ranks means college selectivity improves academic performance for students of all castes/category. We are specifically interested in how improved ranking of college for reserved caste-students impacts their academic achievement in the first-year. For example, consider the SC student who is at the 25th percentile of rank amongst all SC students. The rank in entry examination for this student is 61,602. The normalized value of effective rank for this student is 0.8275. In the absence of affirmative action, this student would have been in OPEN category and the normalized value of effective rank would have been 0.3330. This reduction in effective rank means reduction in college quality, which will have negative effect on the academic performance. If this student switches from being SC to being OPEN the reduction in the academic performance would be equal to coefficient on effective rank times change in effective rank, which equals 0.34*(0.3330-0.8275). That is equivalent to fall in performance by 16.8 percent. Thus SC student gains in academic achievement by 16.8 percent due to affirmative action. This gain equals 11 percent for SC student who is at 10th percentile of rank distribution amongst SC students. The gain due to affirmative action equals 4.2 percent for BC-B student who is at 25th percentile distribution.

The caste fixed-effects are all negative and significant. The reference group is OPEN category. Everything else being same, being member of the caste groups pulls performance down.

3.8 Conclusion:

The empirical findings of this paper suggest that we do not find support for the “fit” or “mismatch” hypothesis in our data. We find that the college selectivity has positive effect on academic achievement in the first year of engineering. The affirmative action policies by placing the students of socioeconomically disadvantaged groups in selective colleges improve their academic achievement. The gains in the academic achievement are higher for more disadvantaged castes amongst those eligible for affirmative action benefits.
The negative caste fixed-effects suggest that being member of SC, ST or BC is disadvantageous relative to the OPEN category. The quantum of the disadvantage suffered is highest for ST students and lowest for BC students. In general SCs and STs are socially and economically more backward than the BCs. It may mean that these caste students have different distribution of scores and the mean of this distribution mimics the socioeconomic hierarchy of these castes and that of the OPEN category. It may also be due the fact that the caste students suffer from negative stereotype threat.

3.9 Future work:

It is intriguing to find the huge negative fixed-effects for some castes. We have not attempted to explain these fixed-effects and we plan to dedicate some of our future efforts for understanding of these effects. The ordering of these effects reflects the socioeconomic ordering of the castes in Indian society and this aspect of Indian society may shed some light on causes of these effects.

The issue of attrition arises as not all those that are qualified in the entry examination seek admissions to engineering colleges. To see how attrition depends on score in entry examination, we break down the possible scores in the entry examination in classes with equal class width. We define attrition rate as the ratio of number of students not seeking admission to the number of students in that class interval taking the entry examination. We find that attrition rate varies by score and by caste. The attrition rate is high for low score and it decreases as the score rises. But at top end of score, attrition is high as students have option of attending other institutions.

Thus, the students who obtain a high rank are likely to go to college. As rank declines, students are more likely to decide not to attend the college. Moreover, the ones who know they are less able are more likely not to attend. So, as measured rank declines, the proportion of low-ability types who drop out rises. Hence, students with low measured rank who remain are more able than the overall students of that rank. This biases the estimates of affirmative action. In our future work, we would address the effect of attrition on the affirmative action.
3.10 References:


3.11 Appendix A:

Counselling Process

Here is what happens during counselling session:

Registration: Candidates will be called as per ranks at the Registration Counter. Candidates have to pay the prescribed registration fee at the counter and obtain a receipt. They have to sign in the registration book at the registration counter. After signing in the register, they have to proceed to the waiting hall at the verification counters.

Verification: Candidates have to submit all the required original certificates at the verification counter, when called. After verification of certificates, candidates have to wait in the counselling hall. They have to fill the details in the option forms provided to them (except the options).

Counselling: Candidates will be sent to the counselling counters as per ranks. At the counselling counter, candidates shall ensure that their details are correctly displayed on the counselling terminal before viewing the vacancies. Candidates shall fill their options in the Option forms provided to them as per their priorities in the order of preference. Candidates shall make sure that the counselling staff has entered the options correctly as per their order of preference. After giving options, candidates have to wait till their name is called at the Allotment counter.

Allotment: Allotments will be made as per merit in a sequence. If the options given by the candidates are not available at the time of allotment, they have to exercise options afresh. Once allotment is made, they are called to the Allotment counter. Candidates shall verify the allotment order for the correctness of the seat allotted and sign on the allotment order. After signing in the Allotment Order, candidates shall proceed to the bank counter for payment of fee.

Fee Payment: Candidates can pay the fee either in the form of D.D. or in cash at the Bank Counter when called. Candidates shall proceed to the Allotment Order Issue counter after paying the fee.

Issue of Allotment Order: Wait at the Allotment Order Issue counter till called. Candidates shall collect the Receipt of Certificates (receipt for the certificates they submitted), Allotment Order and Fee receipt at the counter. After taking the Allotment order and other receipts, candidates shall leave the premises and make room for others.