Returns on Human Capital Investments in Offshore IT Services Industry: A Firm Level Analysis

Amit Mehra*, Nishtha Langer*, Ravi Bapna^, Ram Gopal#

(* Indian School of Business, ^ University of Minnesota, # University of Connecticut)

December 2, 2010

Abstract

The revenue growth model of IT services firms has been historically been based on scaling of firm size by adding new employees. Speculation exists that revenue growth is linear in nature; however no rigorous examination of this fact exists in the literature. In this paper, we investigate the nature of growth of firms in the Indian IT services industry. In addition to scaling firms have been looking at other levers, especially employer funded training, that may be positively linked to firm productivity. We use a panel dataset of small-to-medium sized Indian IT services firms’ training investments and performance to examine the nature of growth and the returns to training investments on firm productivity. We use a combination of econometric methodologies to eliminate the endogenous choice of inputs that are common in estimating production functions. We find that the average firm has can be characterized as having sub-linear revenue growth as it adds new employees, and that training is indeed a very important ingredient to overcome the productivity drag through sheer scale. We find that increase in training investments is significantly linked to increase in revenue per employee. Further marginal returns to training are increasing in firm size.

Keywords: IT services, non-linear growth, ROI of training, productivity, human capital
1. Introduction

Global sourcing of IT and IT enabled Services represents a $1.5 trillion industry\(^1\) of which the share occupied by the Indian IT services industry is rapidly growing, experiencing double digit growth rates over the last decade, currently standing at about $71.3 billion employing 2.3 million people\(^2\). Given that human capital is the crucial ingredient for the IT services industry, growing revenue by simply adding more employees is going to be challenging. Historically, the companies followed the linear growth model to increase their revenues by boosting the employee numbers to generate additional headcount based billing translating into higher revenues. However, the problems with this model are becoming evident as the revenue per employee has reportedly fallen significantly over the years in the Indian IT Services companies (Business Today (February 11, 2007)). Among the many explanations put forth by experts, this has happened not only because of the loss of productivity due to “benching” of employees where they do not contribute to revenues and yet are paid their salaries, but also due to supply constraints of skilled IT programmers. It can also be argued that managing the complexity large pools of knowledge workers working together to provide end-to-end services is non-trivial, hence scaling may provided lower than expected returns. While a variety of opinions exist, there is lack of any rigorous research that examines the nature of growth and returns to scale of firms in the Indian IT services industry. We attempt to bridge this gap in the literature in this study.

Is there no recourse to falling productivity per employee with scale? There is hope as global giants like IBM and Accenture maintain similarly large workforces that are almost twice as productive per employee compared to their IT services counterparts in India. Several strategies have been suggested to improve employee productivity for Indian IT services firms.

---
\(^1\) Gartner 2010 sizing of the IT outsourcing industry
\(^2\) [http://www.nasscom.in/upload/SR10/ExecutiveSummary.pdf](http://www.nasscom.in/upload/SR10/ExecutiveSummary.pdf)
These range from growing the business in high margin areas such as consulting, improving the efficiency of code development through techniques such as agile development and solution accelerators (The Economic Times (28 April, 2008)) and investing in productization strategies where the core product is shared across different clients and only the interface is customized based on client requirements. Productization converts IT services into a high margin business by reusing major portions of the code and using cheaper labor to work on standardized templates for the interface layer (The Economic Times (19 March, 2010)). The crux of all these ideas is to improve the productivity per employee. In addition to such innovative business models, one fundamental way to improve the productivity of an employee is to shore up their skill levels. This is extremely important because rapid changes in technology necessitate constant updating and refreshing of technology skills that can be easily developed via formal classroom training in which firms are quite experienced. Further, the procurement of high-margin IT projects is also critically contingent on a workforce that has up to date technical and domain knowledge. That is why companies such as HCL Technologies are realizing the need to hire fresh graduates and train them for implementing a non-linear growth model (Press Trust of India (April 25, 2010)).

The question of returns to scale in this context is intricately linked to the availability of high quality human capital capable of doing global knowledge work. This is especially true for the IT services industry where employee costs and associated productivity are the major determinants of revenue and gross profits. The productivity of human capital in this industry crucially depends upon the technical and other expertise of the employees that require continuous overhaul due to the fast pace of technological change (Lee et al. 1995, Tambe and Hitt 2008), and workplace reorganization (Bresnahan et. al. 2002). However, highly skilled employable graduates are hard to find. Bapna et al 2010 point out that the rapid growth in the last
decade masks the underlying weaknesses of a higher education system in crisis where a majority of college graduates are unemployable (see Irani 2008 for views from the Indian popular press). While studies have shown that there is disillusionment with skills gap even in the US workforce when they join fresh from college, in countries like India, the skills gap is widespread (Kapur and Mehta 2008). The consequence of these challenges is that Indian IT-ITeS firms are forced to make significant investments in providing education and training to their employees (Hatakenaka 2008). While the initial focus of the training investments were to bridge the skills gap of fresh college graduates, IT-ITeS firms have evolved to provide continuous training and education to cope with the dynamic demands of the global clientele and advances in technology and consequently firms spend a lot of resources and effort on training their employees in order to improve their employee productivity. For instance, The Economic Times (18 Jan, 2010) reports that the bellwether company Infosys recently increased their training budgets by 24% over the previous year to a total amount of $230 million for the year 2010-11. While training investments have been increasing rapidly, prior research has not examined the linkage between these firm level revenues and training investments. We present a comprehensive econometric framework to analyze the nature of growth and the concomitant impact of training in the context of the Indian offshore IT services industry.

Our work is motivated by the fact that in absence of credible estimates of returns on training investments (ROI) Indian IT services firms may under invest if they underestimate the ROI and may over invest if they overestimate the ROI of training. This knowledge is also of prime importance to policy makers who may be considering allocation of government resources to subsidize private investments in training. Unfortunately, however, the exact impact of training
on human capital productivity is not well understood. Our analysis is that first one to provide a
glimpse into how training translates into revenues for IT services firms.

To estimate returns to scale and ROI of training investments we develop an econometric
framework that is based on some recent developments in making causal inferences from the
production function literature (Olley and Pakes 1996, Levinsohn and Petrin 2003). In particular,
we base our initial model on Gandhi et al. 2009 wherein we assume that firm’s input decisions,
in our case the number of employees and the level of training investments are endogenous. The
key estimation challenge here is that there are unobservable (to the econometrician) firm specific
production capabilities that influence the level of inputs that go into the production process. For
instance, a firm whose existing organizational capabilities are synergistic to employee training
may find it more attractive to invest in training. Failure to account for this would lead to an
upward bias in any estimate of impact of training because the coefficient for training would
absorb the positive impact of the unobserved variable. Our approach is based on Gandhi et al
2009 who show that the firm's optimization problem (either maximizing profits or minimizing
costs) contains sufficient information to identify gross output production functions through a
suitable transformation of the firm's first order condition.

While the Indian IT services industry is $71.3 billion in size, it is important to note that
not all firms are equal and that the market share is somewhat concentrated. For instance, the top
five firms account for roughly 50% of the revenues. These big firms, such as Infosys and Tata
Consultancy Services have made a long term strategic commitment to invest in training. They
have institutionalized training through large training facilities such as the Infosys Corporate
University campus in Mysore, and thus have huge training related sunk costs. They also
participate heavily in the Campus Connect type of programs that is a backward integration
exercise aimed to improve the quality of curriculum in engineering colleges. These societal level investments have a broader objective, that of improving the image of these companies in the eyes of government and society. In contrast, the smaller IT Services firms invest in training mainly for improving the productivity of their human capital. Further, because of the lack of institutionalization their training programs are less organized and based on training facilities that are easily available in the corporate education market or on niche expertise that already exists within these firms. Clearly, the training programs in small to mid-sized companies are structurally different from those imparted by the big firms. We are mainly interested in identifying the direct impact of training programs run by such firms on their productivity.

Consequently, we use a data set of small and mid-sized firms in our analysis. Given that these firms constitute 25%-50% of the Indian IT services industry based on revenues and the fact that these set of firms is considered increasingly important in providing the next push to the growth of IT services industry\(^3\), we focus on an important subset of firms from this industry.

Our first major finding is that the rate of growth of revenues by increasing the number of employees is sub-linear. To the best of our knowledge ours is the first study to formally establish this. We also show that investing in training helps improve revenue growth at a faster pace than just investing in hiring more employees. We find the presence of scale effects in training. In order to maintain a linear growth model, firms have to invest less in training per employee as the size grows. Hence training is an effective tool to bump up the overall firm output. Further, we find that the return to training is increasing in firm size.

The rest of this paper is organized as follows. In Section 2 we review the existing literature. Section 3 describes our econometric model and estimation approach. In Section 4 we

describe our data, following which in Section 5 we present our analysis results. Section 6 concludes with directions for future research.

2. Literature Review

Our research has its roots in the theory of human capital which has long been argued as a critical resource and a key driver of value for most firms (Becker 1975, Pfeffer 1994). In the context of the growing global IT services industry, this perspective has added importance as the pool of knowledge workers constitutes both the primary tangible as well as the intangible resource. Further, we also draw upon the emerging IS literature on on skills development of IT workers (Lee et al. 1995, Tambe and Hitt 2008, Joseph et al. 2009) as well as the labor economics literature that measures ROI of employer funded training (Bishop 1996, Black and Lynch 1996). Methodologically, our work relates to the production function estimation literature in IS (Brynjolfsson and Hitt 1996, Lee and Barua 1999, Dewan and Kraemer 2000). In order to account for the impact of training on labor productivity we use a variant of the Cobb Douglas functional form suggested by Bartel 1991 where training is considered to moderate the impact of the labor input. We take care of endogeneity of inputs based on the optimization approach suggested by Gandhi et al 2009, as well as the fixed effects models of Wooldridge 2009.

At one level our work builds on the emerging IS literature that points to the increased importance of skilled workers in IT industries. Ang et al. 2002 suggest that this can be attributed to the complexity of IT jobs that arises from the need to master relatively difficult technical concepts such as data modeling, process discipline and systems design theory. The challenge intensifies when we layer onto this the added dimension of offshore outsourcing of IT, where soft-skills and cultural differences playing an equally important role (Langer et al. 2008, Levina and Vaast 2008). This overall complexity raises the need for significant education inputs either
from the education systems of the various countries where offshore outsourcing is taking place or from IT services firms themselves. Bapna et al. 2010, in an employee level analysis within a single firm, study the impact of human capital investments in the context of the Indian IT services industry. Specifically, they examine whether these human capital investments directed towards employee training are effective in improving employee performance and productivity. Controlling for unobservable employee characteristics and possible selection bias, Bapna et al. 2010 find significant positive impact of training on employee performance. Their findings suggest that the value of training is conditional upon a focused curricular approach that emphasizes a structured competency development program.

In the broader labor economics literature, the returns of human resource management practices on worker productivity have been previously studied by Ichnniowski et al. 1997. It is well established that one of the most important HRM practices that impact worker productivity is in-house training provided by the firms. An excellent review paper that covers different aspects of this literature is by Bishop 1996. Most of the earlier work on returns to training was executed by survey data collected from workers. Hence the returns to training were measured by using wage increments as the dependent variable resulting in a measure of productivity at the worker and not the firm level. This issue was dealt with in later papers (e.g. Black and Lynch 1996) when researchers could get firm level data such as the cost of training. Researchers also expanded on the scope of the research by analyzing questions such as the conditions when employer provided training has more impact (Lynch and Black 1995). However, due to limitations of data or methodology the reported returns on training seemed to be systematically underreported (Bartel 2000). Moreover, most of the papers including the recent work are in the context of non-knowledge sectors such as manufacturing (Almeida and Carneiro 2009). Thus
there is considerable uncertainty regarding the returns to training, especially in the IT services sector. This is the gap in literature that we seek to address.

Methodologically speaking, the typical approach to estimating returns from training is by using a Cobb-Douglas production function (Zellener et al. (1966), Hoch (1958) and Mundlak and Hoch (1965)). The issue with this methodology is that the OLS estimates using production functions may turn out to be inconsistent due to various reasons. A good summary of these reasons is provided in Beveren (2007). The most important of these issues is the endogeneity of input choice or the simultaneity bias, first noted by Marschak and Andrews (1944). The reason for this bias is that the inputs in the production function are not independently chosen but are determined by the unobserved productivity of the firm. Since the unobserved variables form part of the error term, the explanatory variables (inputs) are correlated with the error term, thus biasing the OLS estimates. Several approaches have been suggested over the years to deal with this issue. These are the fixed effects method of Marschak and Andrews (1944) and the instrumental variables methods as reviewed in Beveren 2007. However, each of these methods had their drawbacks. Hence some new approaches have recently been suggested by Olley and Pakes (1996), Levinsohn and Petrin (2003), Acerberg et al. (2006) and Gandhi et al. (2009). We utilize a variation of the method suggested in the last paper to deal with the endogeneity problem in our estimates.

In summary, we contribute to the emerging IT human capital literature by establishing that the nature of growth of the small to mid-sized IT services companies in India is sub-linear. We also show that employer funded training is a key lever that firms utilize to boost the productivity of their labor and that in order to maintain a linear growth model, firms have to
invest less in training per employee as the size grows. Hence training is an effective tool to bump up the growth rate. In the next section we describe our econometric estimation procedure.

3. Econometric Model

The basic productivity theory predicts that productivity of a firm depends upon factors of both capital and labor. We differentiate between intangible capital and tangible capital as intangibles like intellectual property and organizational processes are considered to be more effective in improving the productivity of human capital. We use a variant of the Cobb-Douglas production function suggested by Bartel (1991) to capture the impact of training on firm productivity. Specifically, the relationship between a firm $i$’s input and output in period $t$ is represented as:

$$ R_{it} = A_{it} K_{it}^{δ} C_{it}^{δ} P_{it}^{δ} L_{it}^{α} $$

(1)

where $K_{it}$ and $C_{it}$ refer to the tangible and intangible capital inputs and $P_{it}$ is an observed variable that captures the productive efficiency of firm $i$ in period $t$. $L_{it}$ refers to the effective labor of firm $i$ in period $t$ as described by Bartel (1991). This captures the amount of labor services actually supplied by the workers. The extent of this labor depends upon the numbers of employees and their human capital generated through training. Thus $L_{it} = E_{it}(1 + \beta T_{it})$, where $E_{it}$ is the actual numbers of employees and $T_{it}$ measures the training that the firm provides its employees. Note that this definition of labor accounts for the fact that the productivity of an employee is enhanced by the factor $\beta T_{it}$ due to training investments, $T_{it}$. This accounts for the direct benefits (increase in productivity of the employee due to better knowledge and skills) and indirect benefits (increase in productivity of employee due to better support from peers since they too have better skills) of training on employee productivity. $A_{it}$ is defined such that
\[ \ln(A_t) = \delta_0 + \omega_t + \varepsilon_t. \] Here \( \delta_0 \) refers to the mean efficiency level of all firms over time, \( \omega_t \) represents unobserved firm specific productivity that is observed before the firm makes its period \( t \) input decisions and \( \varepsilon_t \) captures unanticipated productivity shocks that firm does not observe before making its period \( t \) input decisions. Any measurement error is also included in the \( \varepsilon_t \) term.

A key challenge in estimating the parameters of interest from the production function using the inputs and outputs of profit maximizing firms is the endogeneity problem (Marschak and Andrews 1944). The endogeneity problem is caused by the presence of productive factors that are unobservable to the econometrician but that are transmitted to the firm's optimal choice of inputs. Recent developments in the productivity literature suggest that firm's optimization problem (either maximizing profits or minimizing costs) contains sufficient information to identify gross output production functions nonparametrically (Gandhi et al 2009). We adopt a variant of the procedure suggested in Gandhi et al. (2009) and Gandhi et al. (2008). They show that the apparent tension between identifying gross output production functions and controlling for the endogeneity problem can be resolved by exploiting information contained in the firm's first order condition (FOC). In particular, we transform the firm's FOC in such a way that it contains information on the underlying production function, does not suffer from an endogeneity problem.

Let the cost per employee to the firm is represented by the parameter \( W_t \). Then the profit function for the firm is as follows:

\[ \Pi_t = R_t - W_tE_t - T_t. \]

Taking the first order conditions of the firm’s profit with respect to the firm’s decision variables \( T_t \) and \( E_t \), we get:
\[
\begin{align*}
\frac{\partial \Pi_{it}}{\partial E_{it}} &= \frac{\alpha R_{it}}{E_{it}} - W_{it} = 0 \\
\frac{\partial \Pi_{it}}{\partial T_{it}} &= \frac{\alpha \beta R_{it}}{(1 + \beta T_{it})} - 1 = 0
\end{align*}
\]

Using the above two equations and accounting for the errors committed by the firm in optimization (Maddala and Lahiri (2009)), we obtain:

\[
R_{it} = \frac{E_{it}W_{it}}{\alpha} + u_{1it} \quad \text{(2)}
\]

\[
R_{it} = \frac{(1 + \beta T_{it})}{\alpha \beta} + u_{2it} \quad \text{(3)}
\]

Here, \(u_{1it}\) and \(u_{2it}\) represent the error terms arising due to errors in optimization. Eliminating \(R_{it}\) from equations (4) and (5), we get:

\[
E_{it} = -\frac{1}{\beta W_{it}} + \frac{1}{W_{it}} T_{it} + u_{it} \quad \text{(4)}
\]

Here, we have the constant term and the coefficient of \(T_{it}\) in addition to the error term, \(u_{it}\), which is equal to \((u_{2it} - u_{1it})\alpha \beta\). Notice that the revenue function, \(R_{it}\), of the firm contains the unobserved firm productivity variable \(\omega_{it}\) which is the cause of the endogeneity problem.

However, this transformation results in elimination of \(R_{it}\) and hence allows us to remove the source of the endogeneity problem in Equation (4). Thus, equation (4) can now be estimated using OLS without getting a biased estimate for the coefficients. However, we still have to recover the estimate for \(\beta\) as well as its standard error using the estimates of the constant and the slope term. We do this by using the Delta method (Oehlert 1992) which is also implemented in Stata. The remaining challenge with this transformation of the FOCs is that it does not allows us to identify \(\alpha\), the scale parameter, given the data that we have. To recover \(\alpha\), we thus adopt a
fixed-effects approach starting with the Cobb Douglas specification duly modified by the training effect.

Recall, that the usual method for estimating the exponents of Equation (1) requires us to take natural logarithms of both sides which yields:

\[
\delta = \delta_0 + \delta_2 c + \delta_3 p + \alpha l + \omega + \varepsilon \tag{5}
\]

As per convention, we represent the lower case letters to represent the natural logarithms of the variables represented by the upper case letters. Again, it should be evident that if the firm’s choice of the input variables depends on \( \omega \), these will be correlated with the error term, \( \omega + \varepsilon \), as it also contains \( \omega \). Consequently, the OLS estimators will be biased. The fixed-effect procedure works provided \( \omega = \omega \), i.e., the unobserved part of the firm specific productivity shock that is observed before firm’s input decisions are made is time invariant. The equation we estimate is therefore:

\[
r_{it} - r = \delta_1(k_{it} - \bar{k}_i) + \delta_2(c_{it} - \bar{c}_i) + \delta_3(p_{it} - \bar{p}_i) + \alpha(e_{it} - \bar{e}_i) + \alpha\beta(T_{it} - \bar{T}_i) + \varepsilon_{it} \tag{6}
\]

The terms with the bars represent the average of the observations for firm \( i \) over all the years in the panel. As one can observe that \( \omega \) is no longer in the right hand side of the equation. Consequently, the endogeneity problem is accounted for and the estimates are no longer biased. Note that we use the approximation \( \ln(1 + \beta T_{it}) = \beta T_{it} \). This approximation is similar to that used in Bartel (1991) and is valid if \( \beta \) is small. Finally, note to the extent we use some reasonable variables to capture \( p_{it} \), or observed variables that represent firm specific and time variant productivity shocks, the assumption of \( \omega = \omega \) is not too strong since we capture some impact of the firm specific time variant productivity shocks as part of the observed variables. The specific variable we use to capture \( p_{it} \) is the earning per share (adjusted for face value per share).
This ratio provides the earning for each dollar of the equity and hence is a measure of the firm’s productivity.

We now come to the issue of controls. Firm revenues may also be impacted by exogenous shocks to the economy that are common to all firms. Hence we use year dummies while estimating Equation (6). Further, while estimating Equation (4), we must account for the fact that the number of employees hired by the firm may depend upon the firm’s location. For instance, it is possible that certain locations are constrained in terms of the availability of programmers and firms with offices in such locations may not be able to ramp up its numbers easily. To control for such exogenous location based factors, we divide the country into four groups and create three dummy variables called West, North and East with South being the base group. Since, no new offices were opened by firms in the sample during the period 2006-2008, the value of location dummies are time invariant.

Note that we estimate equations (4) and (6) separately under the reasonable assumption that the errors due to optimizations and the errors due to the unanticipated productivity shocks are uncorrelated to each other (Maddala and Lahiri 2009). However, if these errors are correlated to each other, we must estimate these equations using a seemingly unrelated regression (SUR) approach. For robustness, we report the results of SUR as well. These results show that there the estimates under OLS and the SUR approach are very similar. This gives us more confidence in our estimates.

Another issue is that since we use a system of two equations, one may think that we must use the econometric techniques to simultaneously estimate the two equations in order to avoid the estimation bias that results when some endogenous variables are also used as explanatory variables. However, our two equations constitute a recursive system of equations in which there
is only a unidirectional dependency among the endogenous variables. Thus, while $e_{it}$ is the endogenous explanatory variable in Equation (6), there is no endogenous explanatory variable in Equation (4). Consequently the two equations can be ordered such that $e_{it}$ is determined only by exogenous variables and $r_{it}$ can be determined by $e_{it}$. In effect, there is no feedback from Equation (6) into Equation (4). This rules out any contemporaneous correlation between error term and the explanatory variables. Hence we can separately estimate the two equations without any identification problems (Kennedy (2008)).

4. Data

One of the primary challenges in addressing the question of returns to scale and the concomitant impact of employer funded training is the lack of available data that systematically captures this across the Indian IT services industry. To overcome this challenge, we invested in collecting primary data by surveying the head of human resources and the chief financial officers of representative small-to-medium sized IT services firms in India. We hired the Indian subsidiary of reputed US based market research firm TNS4 to collect this data for us. The survey methodology was initial personal contact followed by telephonic reminders. The survey response rate was nearly 40%. We collected firm revenues, number of employees and training investment data through this method on a year to year basis from 2006-2008. We ended up with data from 32 IT services firms. Data for all three years for all variables was sometimes not available and so we have some missing values in the data.

We focus on the small-to-medium sized firms (these account for upto 50% of market share of the $70 billion industry) primarily because of two reasons. Firstly, in contrast to

---

4 http://www.tnsglobal.com/global/alm/india/
the top-five Indian IT services firms, these firms are less likely to invest in training for non-productivity based reasons. It is an established fact that top-five view training as societal level investments, and this clouds the basis of econometric study aimed at linking training to productivity. For instance, many college graduates who go through Infosys’ 14 weeks foundation program upon joining the company often get hired by competitors the day they graduate from the foundation course. Infosys views this as a broader contribution to society (Delong 2006). In contrast, the smaller IT Services firms invest in training mainly for improving the productivity of their human capital. Further, because of the lack of institutionalization their training programs are not orchestrated at a corporate-university scale (such as those at the top-five companies) and are based on training facilities that are easily available in the corporate education market. Clearly, the training programs in small to mid-sized companies are structurally different from those imparted by the big firms. Given that we are mainly interested in identifying the direct impact of training programs run by such firms on their productivity, we believe that our focus on the small and mid-sized firms sharpens our analysis. Secondly, our conversations with senior leadership at Nasscomm, the influential industry body, indicated that these set of firms is considered increasingly important in providing the next push to the growth of IT services industry. They already constitute 25%-50% of the Indian IT services industry based on revenues and policy makers seek guidance on how to scale this segment of the industry to the next level.

In addition to our primary survey data, we used the Prowess database maintained by the Centre for Monitoring Indian Economy, an independent economic think-tank headquartered in Mumbai which has been in existence since 1976. Using the database we collected audited information on the firm’s earnings per share, face value per share, intangible assets (inclusive of
goodwill, software and others) and tangible assets (including plant and machinery, land and building, computer equipment etc.). In addition we also collected firm revenues as well as number of employees and training investments whenever available. This served as a cross-check on the validity of the primary data collection process, and also helped us remove some missing values in the data. Finally, we visited the web-sites of each of these firms and studied their corporate histories and office addresses in order to collect location data. Table 1 and Table 2 in the Appendix provide summary statistics and correlation between variables.

We have data from 32 distinct companies over three years. The average firm in our dataset has revenues of $49 million, and an employee base of about 2000. The data is skewed, and we run the analysis by eliminating the two very large companies in our dataset. The average company in our dataset spends roughly $57 per employee per year on training. Figure 1 below, a plot of the natural log of revenue by the number of employees, gives us a first glance at the nature of growth for this segment of the Indian IT-services industry. It appears that as firms grow revenue per employee does not continue to grow at the same pace. We present the results of the more formal analysis in the next section.

![Figure 1 – Initial Evidence of Diseconomies of Scale in a Plot of ln(Revenue) v. Number of Employees.](image-url)
5. Analysis and Results

The results of the estimation are reported in Tables 3, 4 and 5. We find from estimating equation (6) that $\alpha$, the coefficient of scale parameter, corresponding to the labor inputs is positive and significant, with a value of 0.52. Likewise, from estimating equation (4) we find that coefficient of the training input $\beta$ is also positive and significant with a value of 0.0075. Recall that our labor input is modified from the traditional Cobb Douglas specification ($L_t = E_t (1 + \beta T_t)$) to reflect the moderating impact of training on individual output as per Bartel 1991. This makes it trickier to directly interpret $\alpha$ as the elasticity of labor. However, it is reasonable to conclude based on the value of $\alpha$ that this segment of the Indian IT services industry faces sub-linear growth. Given the positive and significant value of $\beta$, it is useful to consider the efficacy of training as a tool to improve towards a linear growth model.

For instance, with some simplification, our estimation reveals that (ignoring capital and other terms)

$$R = \text{Constant} \times (E(1 + 0.0075 T))^{0.526}$$

The results show that the firm growth is sub-linear with labor (number of employees). A question to ask is: how much should firms be investing in training to realize linear growth with the number of employees. In other words, the firms want to see

$$R = \text{Constant} \times E$$

Equating the two, we get

$$E = (E(1 + 0.0075 T))^{0.526}$$

Rearranging the terms, we get
\[1 + 0.0075T = E^{1-0.526} \]

Assuming \( \frac{1}{E} \) is close to 0, we get

\[ T = E^{-0.0989} \]

\[ \frac{T}{E} = \frac{E^{-0.0989}}{0.0075} \]

Thus, in order to maintain a linear growth model, firms need to invest less in training per employee as the size grows. Hence training is an effective tool to bump up the growth rate.

To the best of our knowledge ours is the first study to rigorously quantify the effect of scale in the context of the small-to-mid size firms in the Indian IT services industry. We find, based on \( \alpha \), that increasing the labor force can definitely increase revenues, as the conventional wisdom suggests. However, the revenue per employee is decreasing as the firm scales. This could be because scaling may involve increasing transaction costs of finding qualified employees, increased managerial complexity of making productive use of thousands of young knowledge workers or the effect of scraping the bottom of the barrel recruiting from a lower skilled labor pool. Figure 2 shows that the scaling reduces the revenues per employees from almost 2 times of the firm that is at the median size to about 0.75 as the firm size goes from 100 to 1000.
It is useful to examine whether training can mitigate the reducing impact of scale on revenues. In order to do this analysis, we vary the number of employees and fix the value of the other variables at their respective median values. We then find the marginal returns to training using the derivative of the firm’s revenue function from Equation (1). We see that the returns to training are seemingly very large, but certainly not unheard of in the literature. For instance Bartel 1991 reports a series of case studies that have looked at such returns provide returns to training ranging from 100% to 2000%. In order to judge the impact of training investment on firm revenues, we notice from Figure 3 that the revenues per dollar of training are increasing with increasing scale. Thus bigger firms are at an advantage over small firms in getting benefit from training.
It is also useful to discuss the coefficients corresponding to the control variables. In Table 3, we find that the adjusted EPS turns out to be significant and so we conclude that time varying firm specific productivity shocks have a significant impact on revenue. Thus use of this variable allows us to separate the firm specific effect from the macroeconomic productivity shocks common to all firms. The estimates for other factors of production in show a mixed result. While intangible assets come out to be significant, the tangible assets are insignificant. This is in accordance with previous studies that show that valuation and success of IT services firms depend more on intangible assets rather than tangible assets. Like in previous works on IT productivity, we use year dummies to control for the exogenous economy specific shocks that are common to all firms. Such shocks will have an impact on productivity of firms and so we must isolate their impact from the impact of the factors of production. However, the coefficients
of the year dummies turn out to be insignificant at the 5% level. Thus macroeconomic reasons do not play a significant role in our data.

As mentioned above, the estimate of the slope and the intercept turns out to be significant in Equation (4). We use these estimates in the delta method to recover the point estimate and standard error of $\beta$ which turns out to be significant. However, the coefficients for all the location controls turn out to be insignificant which shows that location may not be a significant driver of a firm’s number of employees. This points out that if certain locations may have a better and bigger programmer pool available, these locations may have larger numbers of potential employers and so there is no significant advantage or disadvantage of location as far as hiring of programmers is concerned.

6. Conclusions and Future Work

We set out to examine the nature of growth and the concomitant impact of training investments in the small-to-medium firm segment of the burgeoning Indian IT services industry. This segment currently occupies 25-50% of the overall market of just over $70 billion and is expected to fuel the future growth of this industry. In contrast to the top-five firms that occupy the other 50% of the market, training investments made by the small-to-medium sized companies are purely for increasing employee productivity, and thus focusing on this segment provides for a sharper analysis. Broadly speaking, while the conventional wisdom in the Indian media is that the industry currently has a linear growth model, that is IT services firms are able to grow revenues linearly by adding more employees, we find that this is not the case at least for the small-to-medium sized firms. We find that growth for such firms is sub-linear, and that investing in training, i.e. improving the capabilities of their human capital is a viable way to move towards
linear growth. We also find that the returns of training are increasing with increasing scale. Thus bigger firms are at an advantage over small firms in getting benefit from training.

We develop a robust econometric framework based on the recent developments in estimating production functions with endogenous labor inputs and training investments. We apply this framework to a unique dataset that captures the training investments made over a three year period by small-to-medium sized Indian IT services firms. To the best of our knowledge this the first attempt at developing a robust methodology for examining returns to scale and training and our approach is general enough to be applied outside our chosen context.

Future work needs to focus on gaining a better understanding of the factors that lead to sub-linear growth in this segment of the industry. Given that human capital is the crucial ingredient for the IT services industry, growing revenue by simply adding more employees is going to be challenging, especially in the background of a deficient higher education system.

We speculate on the nature of the problems with the current model. Among the many explanations put forth by experts for declining revenues per employees are: the loss of productivity due to “benching” of employees where they do not contribute to revenues and yet are paid their salaries; supply constraints of skilled IT programmers; and managing the complexity large pools of knowledge workers working together to provide end-to-end services. While a variety of opinions exist, future work needs to isolate the relative importance of each of these factors.

While our work contributes by finding the returns to training and analyzing how these returns vary at different scales, this analysis can be extended by incorporating:

1) The impact of training at different levels of the human resource pool.
2) The relative impact of the different types of training such as soft skill training vis-à-vis project management.

3) Analyzing the impact of training on IT product firms that may have a different business model compared to IT services firms.

We conducted a preliminary analysis in this regard. Data for 40 IT product firms was obtained from a survey conducted by a consultancy firm. This data is, however, available for a single year and so we don’t have panel structure. Further, the names of the IT product firms were not revealed to us and hence we could not use the Prowess database and direct visits to company websites to augment our data. Consequently, we have data only for firm revenues, number of employees and training expenses for one year for these 40 firms. The summary statistics of the data along with correlation matrix of the variables are in Tables 6 and 7 in the Appendix. In absence of the panel data we cannot estimate Equation (6) using the fixed effects. Hence we cannot find an unbiased estimate of $\alpha$. However, the methodology used in setting up Equation (4) is unique in that it does not depend upon the panel structure to remove the endogeneity problem. Hence we can still estimate Equation (4) using the pooled OLS model. The results of the estimation are provided in Tables 8 and 9 in the Appendix. One can see that the estimate of $\beta$ is insignificant for IT product firms. Thus it appears that training has an impact on revenues for IT services firms but not on the revenues for IT product firms. This suggests intriguing differences between factors that drive productivity of IT services and IT product firms. Further, either the IT product firms have not yet figured out how to do effective training or training does not drive revenues for IT product firms, and there are some other reasons why these firms still invest in training their employees. Thus further research is needed to understand the drivers of training in IT product firms.
In summary, while tremendous strides have been made in the globalization of the production and consumption of IT, a deeper understanding of the nature of growth and key challenges faced by this industry is only now beginning to attract the attention of researchers. We find that the rate of growth of revenues by increasing the number of employees is sub-linear in the case of the small-to-medium sized companies in the Indian IT services industry. To the best of our knowledge ours is the first study to formally establish this. We also show that investing in training helps improve revenue growth at a faster pace than just investing in hiring more employees. We find the presence of scale effects in training. In order to maintain a linear growth model, firms have to invest less in training per employee as the size grows. Among other aspects, future research will dig deeper into the differential returns from different types of training and at different employee levels within the organization.
**References**


Business Today (February 11, 2007) “Reinventing Indian IT”


Press Trust of India (April 25, 2010) “HCL Tech to Focus on Hiring Freshers, Non-Linear Growth”


The Economic Times (28 April, 2008) “IT Companies Eye Non-Linear Model to Prop up Business”

The Economic Times (19 March, 2010) “IT Companies Hire Non-Techies, Cut Costs”

The Economic Times (18 Jan, 2010) “Infosys Bets on Training to Groom Leaders”


### Table 1: Mean and Standard Deviation of the variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of observations</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue (in million $)</td>
<td>85</td>
<td>49.39107</td>
<td>122.740318</td>
</tr>
<tr>
<td>Employee</td>
<td>84</td>
<td>1942.738</td>
<td>5754.559</td>
</tr>
<tr>
<td>Training (in ‘000 of $)</td>
<td>84</td>
<td>111.0048</td>
<td>699.888636</td>
</tr>
<tr>
<td>Adjusted EPS</td>
<td>79</td>
<td>1.453443</td>
<td>2.054511</td>
</tr>
<tr>
<td>Intangible assets (in million $)</td>
<td>55</td>
<td>4.777645</td>
<td>8.64406591</td>
</tr>
<tr>
<td>Tangible assets (in million $)</td>
<td>84</td>
<td>15.30449</td>
<td>38.4894773</td>
</tr>
</tbody>
</table>

### Table 2: Correlation Coefficient among the variables

<table>
<thead>
<tr>
<th></th>
<th>Revenue</th>
<th>Employee</th>
<th>Training</th>
<th>Intangible assets</th>
<th>Tangible assets</th>
<th>Adjusted EPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td>0.9710</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>0.8487</td>
<td>0.9413</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intangible assets</td>
<td>0.7192</td>
<td>0.7237</td>
<td>0.4295</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tangible assets</td>
<td>0.9438</td>
<td>0.9481</td>
<td>0.8324</td>
<td>0.8437</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Adjusted EPS</td>
<td>0.8525</td>
<td>0.7871</td>
<td>0.6801</td>
<td>0.6649</td>
<td>0.7714</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3: Estimating Equation (6)

<table>
<thead>
<tr>
<th>$r_{it} - \tilde{r}$</th>
<th>Coefficient (Standard Error)</th>
<th>OLS</th>
<th>SURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{it} - \tilde{k}_t$</td>
<td>0.0603764 (.170861)</td>
<td>0.0798242 (.1586886)</td>
<td></td>
</tr>
<tr>
<td>$c_{it} - \tilde{c}_t$</td>
<td>0.3007813 (.06336)***</td>
<td>0.2958642 (.058322)***</td>
<td></td>
</tr>
<tr>
<td>$p_{it} - \tilde{p}_t$</td>
<td>0.17113 (.05387)***</td>
<td>0.1779363 (.049995)***</td>
<td></td>
</tr>
<tr>
<td>$e_{it} - \tilde{e}_t$</td>
<td>0.5260654 (.2066337)***</td>
<td>0.4904508 (.1957022)***</td>
<td></td>
</tr>
<tr>
<td>$T_{it} - \tilde{T}_t$</td>
<td>-6.04e-06 (.0000237)</td>
<td>0.0004884 (.0006394)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-value = 0.800</td>
<td>P-value=0.445</td>
<td></td>
</tr>
<tr>
<td>$Y2$</td>
<td>0.0108882 (.0379189)</td>
<td>0.0065782 (.0351022)</td>
<td></td>
</tr>
<tr>
<td>$Y3$</td>
<td>0.0768355 (.0565582)#</td>
<td>0.0782636 (.0517096)#</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>Dropped</td>
<td>Dropped</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Estimating Equation (4)
<table>
<thead>
<tr>
<th>$F_{it}$</th>
<th>Coefficients (Standard Error)</th>
<th>OLS</th>
<th>SURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{it}$</td>
<td>4.920931 (.2077157)***</td>
<td>5.025114 (.2098662)***</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>103.9684 (400.729)</td>
<td>-1232.313 (553.8872)**</td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>326.223 (313.2806)</td>
<td>701.9296 (452.0108)**</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>199.4336 (296.6804)</td>
<td>251.6711 (418.9248)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>653.4977 (251.6789)**</td>
<td>831.6545 (376.7956)**</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Estimates of $\alpha$ and $\beta$

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$ (from Equation 3)</th>
<th>$\beta$ (from Equation 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS SURE</td>
<td>OLS SURE</td>
</tr>
<tr>
<td>Value</td>
<td>.5260654 (.2066337)**</td>
<td>.4904508 (.1957022)**</td>
</tr>
<tr>
<td></td>
<td>.0075301 (.0029577)**</td>
<td>.0060423 (.0027873)**</td>
</tr>
</tbody>
</table>


Table 6: Mean and Standard Deviation of the variables for IT Product firms

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of observations</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue (in $ Mn)</td>
<td>40</td>
<td>888.6802</td>
<td>2080.715</td>
</tr>
<tr>
<td>Employee</td>
<td>40</td>
<td>238.275</td>
<td>361.1669</td>
</tr>
<tr>
<td>Training (in $)</td>
<td>40</td>
<td>60438.09</td>
<td>92267.07</td>
</tr>
</tbody>
</table>

Table 7: Correlation Coefficient among the variables for IT Product Firms

<table>
<thead>
<tr>
<th></th>
<th>Revenue</th>
<th>Employee</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td>0.7240</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>0.6923</td>
<td>0.9262</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8: Estimating Equation (6) for IT Product firms

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>Coefficients (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS (taking normal T)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>.0000806  (5.32e-06)***</td>
</tr>
</tbody>
</table>

Table 9: Estimate of $\beta$

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
</tr>
<tr>
<td>Value</td>
</tr>
<tr>
<td>.0001892</td>
</tr>
<tr>
<td>(.0002656)</td>
</tr>
<tr>
<td>p-value = 0.481</td>
</tr>
</tbody>
</table>