Public Disclosure Authorized

Public Disclosure Authorized

WPS4576

POLICY RESEARCH WORKING PAPER 4576

Human Capital and the Changing Structure of the Indian Economy

Mohammad Amin
Aaditya Mattoo

The World Bank
Development Research Group
Trade Team
March 2008

Policy Research Working Paper 4576

Abstract

Using panel data for the fourteen major states of India over the 1980-2000 period, the authors estimate the effect of human capital endowment on the performance of the state economies. They find that greater availability of skilled workers had a positive and significant impact on output in the service sectors. They do not find any such effect for the manufacturing sectors. The paper shows that the differential effect on services and manufacturing arises because service sectors are more skill intensive.

This paper—a product of the Trade Team, Development Research Group—is part of a larger effort in the department to understand the determinants of services trade and output. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at mamin@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Human Capital and the Changing Structure of the Indian Economy

By: Mohammad Amin[*] and Aaditya Mattoo[**]

The World Bank

JEL: O11, O15, O53, L60, L80
Keywords: Human Capital, Institutions, India, Skill intensity, Institutional dependence

The views expressed in the paper are those of the authors and should not be attributed to the World Bank. We are grateful to Jens Arnold for useful discussions and to Utsav Kumar, Deepak Mishra and Arvind Subramanian for sharing their data with us. This research is part of a World Bank project on services and is supported in part by the U.K. Department for International Development.

[*]Development Economics Research Group. 1818 H Street NW, World Bank, Washington DC 20433. E-mail: mamin@worldbank.org, phone: 202-473-1915.
[**]Development Economics Research Group, World Bank, 1818 H Street NW, Washington, DC 20433. Email: amattoo@worldbank.org, phone: 202-458-7611.

1. Introduction

The Indian economy has undergone remarkable changes since the early 1980s. India's GDP per capita grew at an annual rate of 3.8% during the 1980-2000 period as compared to 1.7% during the 1950-80 period. Much of this expansion is attributed to the services sector which grew at 5.1% per annum (in per capita terms) increasing its share in GDP from 38% in 1980 to 49% in 2000. Some of the key development policy questions today are whether India's improved performance is sustainable and whether other developing countries can emulate India. While growth is a complex and elusive phenomenon, it may be helpful to identify specific factors that help explain India's growth pattern.

The present paper focuses on the role of the highly skilled. Using panel data for the fourteen major states of India over the 1980-2000 period, we examine the performance of per capita GDP and value added in agriculture, manufacturing (registered and unregistered) and services. Controlling for other factors, and using the system-GMM and the traditional instrumental variable methods to address endogeneity concerns, we find that greater availability of skilled workers had a statistically significant and positive impact on per capita output of the aggregate services sector, but no significant impact on agriculture, and more surprisingly, no impact on manufacturing. One reason for the much stronger impact on services could be that, relative to manufacturing sectors, the service sectors are more skill intensive and therefore more likely to benefit from an expansion in human capital. We test for this idea using data at the disaggregated level (within services and registered manufacturing sectors) and find strong evidence in its favor.

That skilled labor has played an important role in India's overall performance and especially so in the services sector has provoked much discussion but only limited

rigorous analysis. Rodrik and Subramanian (2004) note that India's productivity growth in the last two decades has benefited from its stock of the highly educated but provide no formal evidence. Kochar et al. (2006) show that India's share of output in skill-intensive industries is higher than that of China and comparable to that of much richer countries like Malaysia and Korea. Gordon and Gupta (2004) highlight a number of factors behind India's "services revolution" but human capital is not one of them. Bhide and Shand (2000) and Ahluwalia (2000) do include measures of human capital to analyze variations in growth rates across Indian states in the post 1980 period but both these studies rely entirely on literacy rates. A second problem with these two studies is that they do not address the problem that endowments of human capital could be endogenous.

The plan of the remaining paper is as follows. Section 2 describes the data and the empirical methodology. In section 3 we provide our main results at the aggregate level. Sections 4 contains results at the disaggregate level. Some extensions of the main results are contained in section 5. We summarize our findings in the conclusion and suggest scope for future work.

2. Empirical methodology and data

2.1 *Empirical methodology*

Our econometric analysis is based on estimating a standard reduced form equation (see, for example, Besley and Burgess 2004):

$$Y_{st} = \alpha Y_{st-1} + \beta H_{st} + \gamma \theta_{st} + u_{st} \quad \ldots \quad (1)$$

where the subscript s denotes the state and t the year, H_{st} is the total stock of the skilled workers in state s year t, θ_{st} is a vector of other control variables (discussed below) and u_{st} is the error term. The equation also includes lagged values of the dependent variable as an explanatory variable to capture possible convergence or divergence from the mean. In separate regressions, the outcome variable, Y_{st}, equals GDP and value added in agriculture, total manufacturing, registered manufacturing, unregistered manufacturing and the services sector. In section 3 the dependent variable is value added in the various sectors within registered manufacturing and services sector (pooled together).[1] Our data covers the fourteen major states of India over the 1980-2000 period.[2]

One problem with estimating (1) directly is that data on the stock of skilled labor (H_{st}) at the state level are not available. What we do have is data on the number of students enrolled in higher education (colleges and universities) which is a flow measure since current students add to the stock of human capital in the future. To use this flow measure we make two assumptions and modify equation (1) to an appropriate form.

First, we assume that, from the date of enrollment, it takes about four years for a student to complete her higher education and find a job. This is a reasonable assumption since the data on enrollments that we use below covers students at the undergraduate level (Bachelor of Arts, Science and Commerce) which is a three year program in all the states.[3] The second assumption relates to inter-state migration of the highly skilled. If such migration were significant then it would weaken the case for using enrollments as a

[1] In the regressions, the dependent variables are appropriately deflated by the total population of the states.
[2] List of the 14 major states is provided in Table 3.
[3] Further, enrollments in a given year relate to the academic year September to August while data on the outcome variables (GDP per capita, etc.) relates to the financial year beginning in April implying an additional lag of about 4 months before the full impact of new graduates is felt on output. If we assume that it takes about 5-6 months to search for a job we get that it takes about four years from the date of enrollment before a student's skills have any significant effect on output.

measure of human capital flows at the state level. Using available evidence from the five-yearly rounds of the National Sample Survey, we estimated the likely magnitude of inter-state migration and these are reported in Table 3.[4] As a percentage of students completing their higher education and entering the labor market, inter-state migration of the highly skilled varies between 0.48% (Andhra Pradesh) and 3.01% (Haryana). Since annual data on interstate migration are not available, we cannot control for migration related effects in our regressions. However, given its relatively small magnitude, the assumption that there is negligible interstate migration is unlikely to affect our results significantly.[5]

Using the two assumptions above, we can rewrite (1) as a "4th difference per capita" equation as follows:

$$\frac{Y_{st} - Y_{st-4}}{P_{st}} = \alpha \frac{Y_{st-1} - Y_{st-5}}{P_{st}} + \beta \frac{H_{st} - H_{st-4}}{P_{st}} + \gamma \frac{\theta_{st} - \theta_{st-4}}{P_{st}} + w_{st} \quad \ldots \quad (2)$$

where P_{st} is the total population of state *s* in year *t*, $H_{st} - H_{st-4}$ is the addition to the stock of the highly skilled in the labor market over *t, t-4* period in state *s* which is equal to the total number of students enrolled in Bachelor of Arts, Science and Commerce courses in year *t-4* and state *s*. In the remainder of the paper, we will use the terms human capital, enrollments and skilled workers interchangeably for $H_{st} - H_{st-4} / P_{st}$.

[4] These estimates are based on ongoing work on interstate migration in India by Utsav Kumar, Aaditya Mattoo and Arvind Subramanian.

[5] Other studies also suggest that inter-state migration in India is quite small. See, for example, Cashin and Sahay (1995) and Topalova (2004).

2.2 *Description of main variables*

A formal definition of all the variables used in the regressions is provided in Table 1 along with the data sources. Summary statistics are reported in Tables 2 and 3. All nominal variables used in the regressions are expressed in real terms (1993-94 prices).

Enrollments at the national level averaged 4.2×10^{-3} over the entire period with the standard deviation of 1.9×10^{-3}. Over time, the level of enrollments varies between 2.9 and 6.6 per thousand population. The best performing states here are Karnataka, Maharashtra and Gujarat while the worst is Rajasthan (column 1, Table 3). The figure for Karnataka is more than twice that for Rajasthan. In general, the traditionally laggard states are far behind the others in the level of enrollments.

Figure 1 shows the relationship between enrollments per capita and change in service sector value added per capita over *t*, *t-4*. The relationship is positive, as our regression results will later confirm. Figure 2 shows the relationship between enrollments and output in the registered manufacturing sector while figure 3 does the same for agriculture. Both these relationships appear to be weak as is confirmed later by our empirical results.

The relationship between human capital and output measures depicted in the figures discussed above cannot be interpreted as truly causal due to possible reverse causality and omitted variable bias problems. Although these problems cannot be ruled out completely, we believe that they are less severe with our estimation than is otherwise the case for the following reasons.

First, our estimation strategy involves regressing *changes* in the stock of human capital on *changes* in the outcome variables (equation 2). Hence, we are insulated from a

number of endogeneity concerns that arise in regressing *levels* of human capital on *levels* of output, a common practice in the literature based on cross-country regressions. Second, current enrollment levels in higher education are largely determined by past investments in educational infrastructure (number of colleges, universities, etc) and are unlikely to be affected by current changes in income and sectoral output.[6] Finally, in the panel data estimation, we use the system GMM method which utilizes lagged values of the potentially endogenous variables as instruments. In the cross-section, we use the traditional instrumental variables (IV) estimation strategy. Hence, reverse causality is unlikely to be a significant problem for our estimation.

A relatively more serious concern could be the omitted variable bias problem. For example, states that are more developed due to historical or other reasons are likely to have better educational infrastructure and also better institutions, physical infrastructure, etc. This could create an identification problem by making it difficult to distinguish between the effects of human capital and these other factors on output. In the spirit of Rajan and Zingales (1998), we use data at the disaggregate level (services and manufacturing sub-sectors) to show that the effect of human capital is much stronger (more positive) in those sub-sectors that are relatively more skill intensive. As an additional layer of defense, we include a number of controls such as state and year fixed effects as well as measures of infrastructure, quality of institutions and other time varying state characteristics.

The controls in our specification include measures of infrastructure availability, quality of institutions, size of the states measured by total population, development

[6] In section 5 we look at the relationship between current enrollment levels and sufficiently lagged values of the number of colleges at the state level. Our findings show a strong positive relationship between the two. Further, we do not find any evidence of a direct effect of colleges on current changes in output levels.

expenditure and state and year fixed effects. The choice of controls is motivated by the findings in existing studies (see, for example, Besley and Burgess 2004). Except for the fixed effects, we use the 4th difference per capita values of these controls as shown in equation 3 below.

For infrastructure we use total installed power generating capacity ($Power_{st}$) and total road length ($Roads_{st}$) at the state-year level. There is no readily available direct measure of the quality of institutions at the state-year level for India and we follow Kochar et al. (2006) in using a proxy measure which is transmission and distribution losses of state electricity boards as a percentage of total power generated (*TDL*). India has one of the highest rates of *TDL* in the world averaging around 20% (see Table 3) and it is widely known that these losses are largely due to theft and pilferage reflecting poor quality of governance at the state level.[7] However, our results for *TDL* should be treated with caution because more work is needed to assess how well the variable proxies for the quality of the institutional environment. Total population of the state ($Population_{st}$) captures possible increasing returns to scale and also the availability of (unskilled) labor. States that invest more in human capital are also likely to spend more on other development activities such as the provision of health care, etc. These factors could have a direct effect on output which we filter out by controlling for development expenditure (Dev_{st}). Lastly, state fixed effects (*SFE*) control for all state specific and time-invariant determinants of output while year fixed effects (*YFE*) control for all annual economy wide shocks to output such as trade liberalization, privatization, etc.

[7] Also, the losses impose a financial burden on the state governments further reducing their ability to provide good quality of public services. A recent white paper titled "Maharashtra Power Sector Reforms" issued by the state government of Maharashtra in 2002 noted that: "Every 1% reduction in the T&D loss levels is equivalent to additional revenues of about Rs 120 crores."

Based on the discussion above, the equation we estimate is as follows:

$$\frac{Y_{st}-Y_{st-4}}{P_{st}} = \alpha\frac{Y_{st-1}-Y_{st-5}}{P_{st}}+\beta\frac{H_{st}-H_{st-4}}{P_{st}}+\alpha_1\frac{Power_{st}-Power_{st-4}}{P_{st}}+$$

$$\alpha_2\frac{Roads_{st}-Roads_{st-4}}{P_{st}}+\alpha_3\frac{Dev_{st}-Dev_{st-4}}{P_{st}}+\alpha_4\frac{P_{st}^2-P_{st-4}^2}{P_{st}}+$$

$$\alpha_6(TDL_{st}-TDL_{st-4})+SFE+YFE+\alpha_0+u_{st} \qquad \ldots \qquad (3)$$

Our measure of human capital is most highly correlated with power availability (correlation of .373) followed by development expenditure (.233).[8]

3. Regression results for the aggregate sectors

In this section we report regression results for the aggregate sectors. That is, in separate regressions, the dependent variable in equation 3 equals per capita change in: GDP, value added in services sector, valued added in total manufacturing, registered manufacturing and unregistered manufacturing, and value added in the agriculture.

Regression results for the services sector are reported in Table 4. Without any additional controls, the estimated effect of human capital on services output is positive and significant at less than 1% level (column 1, Table 4). Controlling for power and development expenditure, which are most highly correlated with the human capital variable, we find that the estimated coefficient of human capital declines only marginally but remains positive and significant at less than 1% level (columns 2 and 3, Table 4). Controlling for the remaining variables in the specification does not change our results

[8] The correlation between human capital and other variables in equation 3 are as follows: -.191 with roads, .015 with transmission and distribution losses and -.133 with population.

qualitatively, although the estimated coefficient of the human capital variable does decrease in magnitude. The estimated coefficients of the remaining variables do not show any robust effects on changes in services output with the exception of the lagged dependent variable. Its coefficient is positive (.715) and significant at less than 1% level implying a sharp tendency towards divergence.

Table 4 also reports on test statistics for Sargen-Hansen overidentifying restrictions and second order serial correlation. For the validity of the instruments, it is important that both these test statistics are small (statistically insignificant from zero). Our estimation results show no evidence of second order serial correlation but the Sargen-Hansen test statistic is weakly significant at close to 10% level (p-value of .096) without any additional controls (column 1, Table 4). However, as evident from Table 4, even this weak significance level disappears when we add our basic controls to the specification.

We experimented with a number of alternative specifications but found that the positive and significant impact of human capital on services sector output remained intact. First, it is possible that the state and year fixed effects may filter out much of the variation in our enrollments variable. If this is indeed the case then our final results above may not be robust to small alterations in the sample or the specification. We estimated equation 3 without the state and year fixed effects. Second, in the specifications above we used *change* in the quality of institutions as an explanatory variable. To the final specification (column 4, Table 4), we added the level of institutional quality as another explanatory variable under the assumption that change in output may be affected by the level of institutional development rather than annual changes in it. Third, expansion of

the services sector has been particularly strong in the state of Maharashtra. Change in services output per capita equaled Rs 403 per annum in the state compared to the next highest figure of Rs 291 per annum for the state of Gujarat. We dropped the state of Maharashtra from the sample to ensure that our results above are not driven by a single state. Our main findings reported above were robust to all these changes.

Regression results for the remaining outcome variables are reported in Table 5. Panel A contains results for the effect of human capital on the outcome variables without any additional controls. In Panel B we report results with the various controls discussed above included in the specification. Without additional controls, we find a positive effect of human capital on GDP, total manufacturing and registered manufacturing, significant at less than 5% level. However, these statistically significant effects do not survive our robustness checks as evident from Panel B of Table 5 (columns 1-3). For unregistered manufacturing and agriculture we do not find any significant effect (at 10% level or less) of human capital with or without additional controls (columns 4 and 5, Table 5).

As for the services sector, we find the estimated coefficient of the lagged dependent variable is positive and statistically significant at less than 1% level for the registered manufacturing sector. GDP, unregistered manufacturing and agriculture show only weak significance of the lagged dependent variable (significant between 5-10% level). None of the remaining explanatory variables show any significant effect (at 5% level or less) on the outcome variables in a robust way. The only exception here is the positive and significant effect of development expenditure on output in unregistered manufacturing sector.[9] Our main finding in Table 5 of a statistically insignificant effect

[9] One reason for this could be that development expenditure is primarily targeted towards the poorer sections that are more likely to seek employment in the unregistered manufacturing sector.

of human capital on the outcome variables survives a number of robustness checks such as dropping the state of Maharashtra from the sample, removing the state and year fixed effects from the specification and adding the *level* of institutional quality as an explanatory variable.

4. Disaggregated results

In this section we look at the disaggregated sectors within registered manufacturing and services. These sectors are quite diverse and fitting a single regression at the aggregate level does not allow us to exploit the diversity to better understand the effect of human capital on output. For example, we found above that the highly skilled had a strong effect on the services sector but a weak one on the other sectors. The most natural explanation of this is that the service sectors are the most skill intensive sectors and hence likely to benefit more from a greater availability of human capital. Testing for this requires differentiating across sub-sectors by their respective skill intensities which is the main goal of this section.

We pool data for the 5 main services sectors and 16 registered manufacturing sectors at the ASI 2 digit level. These sectors are listed in Table 6. Consistent data on output of these sectors are available for the time period 1980-1997.

Diversity across sectors is captured using factor intensity ratios which reflect the relative importance of the explanatory variables across sectors. These factor intensity ratios are computed using the input-output matrix for India. To appropriately map the sectors in our data onyo the ones in the input-output matrix we had to group our initial set

of 16 manufacturing sectors into 12 which left us with a total of 17 sectors.[10] The equation we estimate is as follows:

$$\begin{aligned}\frac{Y_{ist}-Y_{ist-4}}{P_{st}} &= (\beta+\beta_{11}Skill_Int_i)\frac{H_{st}-H_{st-4}}{P_{st}}+(\alpha_1+\alpha_{11}Power_Int_i)\frac{Power_{st}-Power_{st-4}}{P_{st}}\\ &+(\alpha_2+\alpha_{22}Road_Int_i)\frac{Roads_{st}-Roads_{st-4}}{P_{st}}+(\alpha_3+\alpha_{33}Serv)\frac{Dev_{st}-Dev_{st-4}}{P_{st}}\\ &+(\alpha_4+\alpha_{44}Serv)\frac{P_{st}^2-P_{st-4}^2}{P_{st}}+\alpha_5\frac{TDL_{st}-TDL_{st-4}}{P_{st}}\\ &+\frac{Y_{ist-1}-Y_{ist-5}}{P_{st-1}}+SFE+YFE+IFE+\alpha_0+u_{ist} \qquad \text{.....} \qquad (4)\end{aligned}$$

where subscript *i* denotes the sector, *IFE* are the industry (sector) fixed effects, *Serv* is a dummy variable equal to 1 if the sector is a services sector and 0 otherwise. Equation (4) is analogous to equation (3) with industry fixed effects and the interaction terms added. Diversity across sectors is captured by β_{11} and α_{jj} coefficients, *j*=1, 2,3, 4.

Variation in skill intensity across sectors is captured by *Skill_Int* which equals the total remunerations of the highly skilled as a percentage of the total remunerations of all the workers (skilled plus unskilled). These skill intensities are reported in Table 7. The coefficient β_{11} measures how the effect of human capital varies across sectors depending on their skill intensity. As mentioned above, we expect the coefficient to be positive implying a greater impact of human capital on the relatively more skill intensive sectors. We also experimented with a number of additional interaction terms and some of the important ones are included in equation (4) and the remaining ones are discussed briefly below. The intensity ratios for roads (*Road_Int*) and power (*Power_Int*) are defined as the

[10] The exact mapping is stated in Table 6.

total input use of the factors as a percentage of the total output of the sector (reported in Table 7). We pay special attention to how the effect of population and development expenditure varies across sectors. It is possible that scale economies may be more important for manufacturing relative to services sectors. Development expenditure (on health, basic education, etc) may complement skilled labor in affecting output. Due to data limitations, it is difficult to assess the importance of these two variables across the narrow sectors. Hence, we take a general approach here by controlling for their differential effect on output across the broad group of manufacturing and services sectors. This is done by interacting the population and development expenditure terms by the dummy (*Serv*) for the services sectors.

Regression results for equation (4) are reported in Table 8. Without any additional controls, the estimated coefficient of our main interaction term (β_{11}) is positive and significant at less than 1% level (column 1, Table 8) implying that the effect of human capital on output increases sharply with the skill intensity of the sectors. The result is robust to controls for population, population, power, roads, development expenditure and the quality of institutions (column 2, Table 8) and the remaining controls discussed above (columns 3 and 4, Table 8). For all the specifications reported in Table 8, the Sargen-Hansen and the second order correlation test statistics are statistically insignificant implying that the exogeneity of the instruments cannot be rejected.[11]

We experimented by allowing the effect of institutions to vary across the group of manufacturing and service sectors. This was done by including the interaction of the *TDL*

[11] The Sargen-Hansen test for overidentifying restrictions is weakly significant at close to 10% level (p-value of .097) with the human capital variables alone (column 1, Table 8). However, even this weak significance disappears when we control for power, roads, etc (columns 2-4, Table 8). This result is similar to what we found for the aggregate service sector in section 3.

term in equation (4) with the dummy for the service sectors (*Serv*) but this did not change our main results much either here or elsewhere in the paper. Further, we did not find any significant difference in the effect of institutions across service and manufacturing sectors.

The overall effect of human capital on output ($\beta + \beta_{11} Skill_Int_i$) varies significantly across sectors. It is negative but statistically insignificant at 10% level or less for the least skill intensive sectors (all manufacturing) and positive and significant at less than 5% level for the most skill intensive sectors (banking and insurance, telecommunication services, business services and real estate services).

Lastly, the coefficient of the lagged dependent variable is positive and significant at less than 1% level implying a strong tendency for output changes to diverge from the mean. The divergence is similar to the findings in the previous section. For the remaining variables, we do not find any significant effects with all the controls in place.

5. Long-run effects

In this section, we briefly report on regression results using the traditional two stage instrumental variables (IV) estimation strategy. We apply this strategy to equation (4) after averaging both sides of the equation over time.[12] Motivation for averaging is twofold. First, it is difficult to instrument for year-to-year variations in the explanatory variables. Averaging removes the time dimension from the equation completely making our task (of finding appropriate instruments) relatively simpler. Second, averaging allows

[12] We take averages over annual changes instead of differences at 4 year intervals because we are looking at the long-term relationship where the lag between the year of enrollment and entry into the labor market is irrelevant. Further, our results are unchanged if we take average of changes across 4 year intervals instead.

us to analyze the long-run impact of human capital (and other explanatory variables) which is helpful if some of our explanatory variables affect output slowly over time.

We note that averaging over time implies that we are left with a pure cross-section at the state-sector level. It also implies that all the explanatory variables above apart from the interaction terms, are absorbed into the state and industry fixed effects. Thus, all state-year and sector-year variants are fully controlled for by these fixed effects.

As above, we treat all our variables as potentially endogenous (except the fixed effects) and instrument them as follows. For *Education* ∗ *Skill_Int* we use $Education_{70}$ ∗ *Skill_Int* and $Colleges_{70}$*Skill_Int* as instruments where $Education_{70}$ is the annual enrollment rate (per capita) over the 1970-79 period and $Colleges_{70}$ is the number of colleges per capita averaged over the 1970-79 period. For all the remaining variables we use their lagged values (averages over 1970-79).[13] For *Roads*Road_Int* we include $Highway_{70}$*Road_Int* as an instrument where $Highway_{70}$ is the per capita length of state and national highways averaged over the 1970s.

IV regression results are reported in Table 9. Without any additional controls, the estimated coefficient of our main interaction term is significant at less than 1% level (column 1, Panel A, Table 9). In column 2 of Table 9, we report regression results controlling for state and sector fixed effects to filter out all the non-interaction terms. The estimated coefficient of the human capital – skill intensity interaction term here remained positive and significant at less than 1% level. Adding the remaining controls to the specification, we find the estimated coefficient of the main interaction term remains

[13] For *Roads*Road_Int* we use *Highway*Road_Int* as an instrument where Highway is value of total length of Highways per capita (average over the 1970-79 period). Data on road length (other than highways) for the period shows significant year to year jumps for some of the states due to changes in data coverage. Excluding roads completely from the specification does not change our main results much.

significant at less than 5% level (column 3, Table 9). For the remaining variables, we find that the effect of development expenditure is much higher for the services sectors than for the manufacturing sectors (significant at less than 1% level). The remaining terms do not show any significant effect on output change. In all the specifications reported in Table 9, our instruments performed reasonably well as reflected by the overdidentification test statistics reported in Panel A and the Shea partial correlations in Panel B.

6. Conclusion

We sought to identify some of the causal factors in the impressive performance of the Indian economy since early 1980s. We find that the endowments of the highly skilled had a significant effect on the services output of individual states. However, manufacturing and agriculture have been largely immune to the availability of the highly skilled. The results also show that the key reason for this differential impact is the skill intensity of the sectors with the more skill intensive services sectors being the primary beneficiaries of greater skill availability.

A key issue is the relative role of history and human capital. It is increasingly believed that the root cause of backwardness in many parts of the world lies in their history. For example, the quality of a country's institutions may be related to its (colonial) history. Many institutions are slow to change which prompts pessimism about the likelihood of coming out of the poverty trap. An interesting question is whether building effective educational institutions is feasible in the medium term with a combination of domestic policy action and external assistance. If this is indeed the case, then, our

findings suggest that investments in higher education could provide a way of alleviating the constraints of inadequate institutional development in the medium term.

References

[1] Ahluwalia, Montek S. (2000), "Economic Performance of States in Post-Reforms Period," *Economic* & *Political Weekly*, May 6, pp. 1637-1648.

[2] Arellanom M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, pp. 277-97.

[3] Arellanom M. and O. Bover (1995), "Another Look at the Instrumental Variables Estimation of Error Components Models," *Journal of Econometrics*, 68(1), pp. 29-52.

[4] Banerjee, Abhijit and Lakshmi Iyer (2005), "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review*, 95(4), pp. 1190-1213.

[5] Besley Timothy and Robin Burgess (2004), "Can Labor Regulation Hinder Economic Performance? Evidence From India," *Quarterly Journal of Economics*, 119(1), pp. 91-134.

[6] Bhide, S and Ric Shand (2000), "Sources of Economic Growth: Regional Dimensions of Reforms," *Economic* & *Political Weekly*, October 14, pp. 3747-3757.

[7] Bramber, Thomas, William R. Clark and Matt Golder (2006), "Understanding Interaction Models: Improving Empirical Analyses," *Political Analysis*, 14, pp. 63-82.

[8]Cashin, Paul and Ratna Sahay (1995), "Internal Migration, Center-State Grants and Economic Growth in the States of India," *IMF Working Paper*, WP/95/66-EA.

[9] Gordon James and Poonam Gupta (2004), "Understanding India's Services Revolution," *IMF Working Paper*, No. WP/04/171, International Monetary Fund.

[10] Hulten, Charles R., Esra Bennathan and Sylaja Srinivasan (203), "Infrastructure, Externalities, and Economic Development: A Study of Indian Manufacturing Industry," *mimeograph*.

[11] Krishna, K.L. (2004), "Patterns and Determinants of Economic Growth in Indian States," *Working Paper* No. 144, Indian Council for Research on International Economic Relations, India.

[12] Kochar, Kalpana, Utsav Kumar, Raghuram Rajan, Arvind Subramanian and Ioannis Tokatlidis (2006), "India's Pattern of Development: What Happened, What Follows?," *IMF Working Paper*, WP/06/22.

[13] Li, G. (1985), "Robust Regression," in D.C. Hoaglin, F. Mosteller and J.W. Turkey edited **Exploring Tables, Trends and Shapes**, New York, Wiley, pp. 281-340.

[14] Nagaraj, R., Aristomene Varoudakis and Marie-Ange Veganzones (2000), "Long-Run Growth Trends and Convergence Across Indian States," *Journal of International Development*, Volume 12.

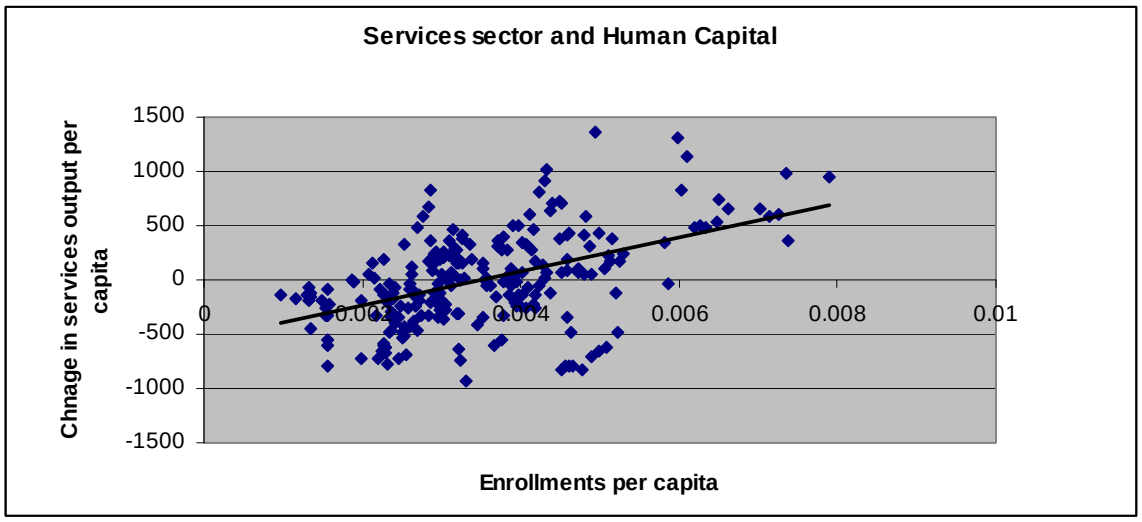
[15] Rajan, Raghuram G. and Luigi Zingales (1998), "Financial Dependence and Growth," American Economic Review, 88(3), pp. 559-586.

[16] Rodrik, Dani and A. Subramanian (2004), "Why India can Grow at 7 Percent a Year or More: Projections and Reflections," *IMF Working Paper*, WP/04/118.

[17] Roodman, David (2004), "The Anarchy of Numbers: Aid, Development and Cross-country Empirics," *Working Paper* No. 32, Center for Global Development.

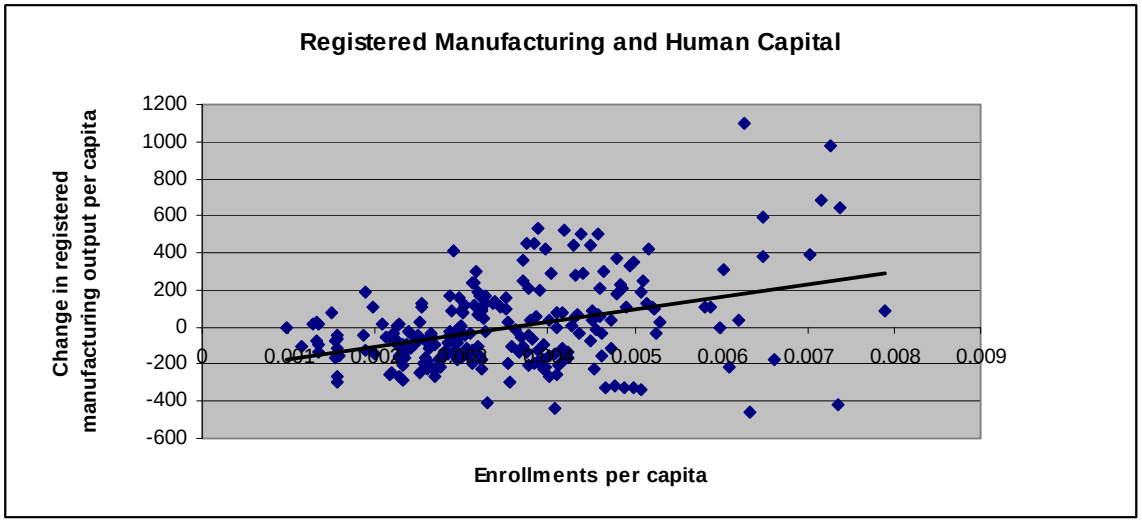
[18] Topalova, Petia (2004), "Factor Immobility and Regional Effects of Trade Liberalization: Evidence from India," *mimeograph*, Department of Economics, Massachusetts Institute of Technology.

Figure 1



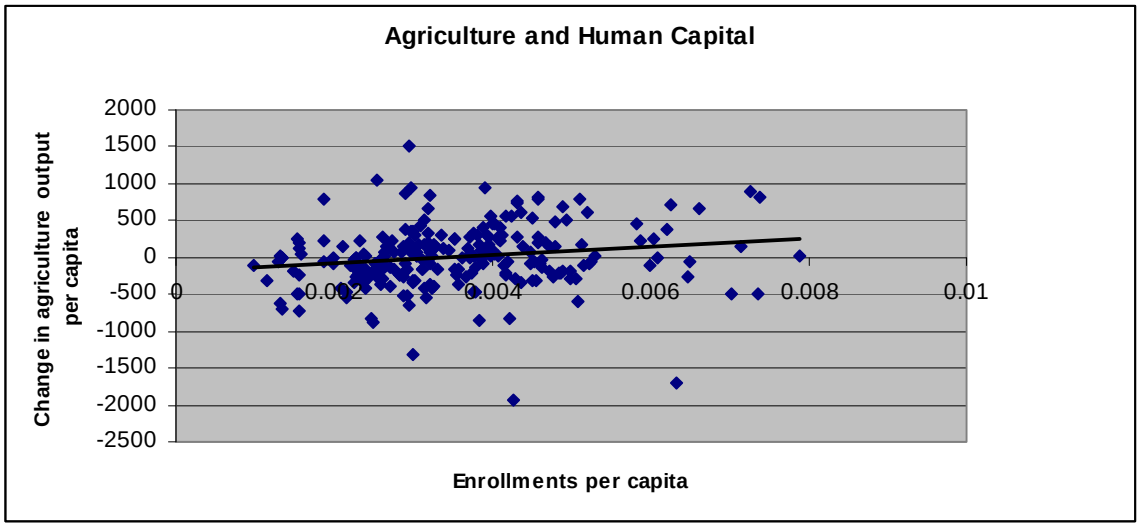
In the figure above, the horizontal axis plots enrollments per capita ($H_{st} - H_{st-4}/P_{st}$) for various state-years in the sample. The vertical axis plots values of $Y_{st} - Y_{st-4}/P_{st}$ after controlling for a linear annual trend where Y_{st} is the aggregate output of the services sector in year *t*, state *s*.

Figure 2



In the figure above, the horizontal axis plots enrollments per capita ($H_{st} - H_{st-4}/P_{st}$) for various state-years in the sample. The vertical axis plots values of $Y_{st} - Y_{st-4}/P_{st}$ after controlling for a linear annual trend where Y_{st} is the aggregate output of the registered manufacturing sector in year *t*, state *s*.

Figure 3



In the figure above, the horizontal axis plots enrollments per capita ($H_{st} - H_{st-4}/P_{st}$) for various state-years in the sample. The vertical axis plots values of $Y_{st} - Y_{st-4}/P_{st}$ after controlling for a linear annual trend where Y_{st} is the aggregate output of the agriculture sector in year *t*, state *s*.

Table 1: Description of variables

[1] Pop and outcome variables: Data on output for broad sectors and GDP (used in section 2) are taken from CSO, Government of India. These are in constant 1993-94 prices (New Series). Output levels of the disaggregated sectors (used in sections 3 and 4) are taken from the Annual Survey of Industry, Government of India. Real values of these output levels are obtained using sector-specific deflators constructed from the ASI data at the 2 digit level. Data source for Population figures is CSO.

[2] Education (E_{st}): Data for these are mostly from the Annual reports of the University Grants Commission, Government of India. These data are available at www.indiastat.com. The website did not provide data for the 1970s and there were a few missing data for 1980-2000 period. These missing data were compiled from various issues of Statistical Abstract, India; Annual reports of the Ministry of Education and Social Welfare, Government of India; annual reports of the Department of Secondary and Higher Education, Ministry of Human Resource Development, Govt. of India; Centre for Monitoring Indian Economy (CMIE) (special edition on the social sector of India).

[3] Road length (*Roads*): Total length of surfaced and unsurfaced roads. Most of the data were taken from various issues of Statistical Abstract, India. For missing data we used data from CMIE (Infrastructure series); various issues of India: Economic Information Year Book.

[4] Electricity capacity (*Power*): The variable measures the installed capacity of electricity generation. Data sources for this variable are the same as for *Roads* plus annual reports of the Central Electricity Authority, Ministry of Energy, Govt. of India.

[5] Development expenditure (*Dev*): This is the real development expenditure (at 1993-94 prices). Nominal values of the variable were taken from various issues of the Reserve Bank of India Bulletin. These were deflated using the GDP deflator calculated from CSO data on real and nominal GDP.

[6] *TDL*: Statistical Abstract, India, various issues.

[7] Factor intensity ratios (Skill intensity, road intensity and power intensity): Calculated using the input-output table for India from Global Trade Analysis Project (GTAP), Version 6. All the factor intensities are defined at the industry level. Skill intensity equals remunerations of the skilled workers as a percentage of total labor remunerations. Road intensity is defined as the total expenditure incurred on transportation as a percentage of total value added in the industry. Similarly, power intensity is expenditure on electricity as a percentage of total value added in the industry.

Table 2: Descriptive Statistics - All states and years*

	Mean	Standard deviation
GDP per capita	7683	1628
Agriculture output per capita	2326	163
Manufacturing output per capita	1465	347
Registered Manufacturing per capita	957	240
Unregistered Manufacturing per capita	508	110
Services output per capita	3022	910
Population	78.5×10^{7}	9.8×10^{7}
Installed electricity capacity (KwH) per capita	.058	.009
Road length (Kms) per capita	.002	.0001
Enrollments in B.A, B.Sc, B.Com per capita	4.2×10^{-3}	1.9×10^{-3}
TDL (%)	20	7.5
Development expenditure per capita	936	158

*The data cover the 14 major states over the 1980-2000 period. All nominal variables are at constant 1993-94 prices.

Table 3: Descriptive statistics by States

State (Abbreviations)	(1) Enrollments per capita	(2) Inter state migration (%)	(3) T&D Loss (%)	(4) Services output per capita
Andhra Pradesh (AP)	3.7×10^{-3}	0.48	21.0	3250
Bihar (BH)	4.2×10^{-3}	2.78	20.5	1454
Gujarat (GJ)	5.8×10^{-3}	1.07	19.2	4117
Haryana (HY)	4.8×10^{-3}	3.01	22.0	3866
Karnataka (KN)	6.1×10^{-3}	1.43	19.9	3271
Kerela (KR)	4.0×10^{-3}	2.56	19.3	3801
Madhya Pradesh (MP)	3.0×10^{-3}	2.77	18.4	2240
Maharashtra (MH)	5.8×10^{-3}	1.30	15.2	5342
Orissa (OR)	2.4×10^{-3}	2.27	24.0	1949
Punjab (PJ)	5.2×10^{-3}	1.51	17.7	4128
Rajasthan (RJ)	2.8×10^{-3}	1.98	24.4	2507
Tamil Nadu (TN)	3.8×10^{-3}	1.56	17.2	4095
Uttar Pradesh (UP)	3.1×10^{-3}	2.67	21.0	2071
West Bengal (WB)	4.3×10^{-3}	2.06	21.2	3121
Number of Observations	21	21	21	21

Inter-state migration: The estimates shown are average rates for the 1980-2000 period and based on ongoing work on interstate migration in India by Utsav Kumar, Aaditya Mattoo and Arvind Subramanian. The interpretation of these numbers is that if inter-state migration of the skilled were taken into account then the net addition to the stock of human capital as measured in our data would change (in absolute value) by an amount equal to the migration rates listed above. These estimates are derived using NSSO data for the number of skilled migrants in various states.

Table 4: Effect of human capital on the aggregate service sector

	(1)	(2)	(3)	(4)
Enrollments per capita	278431***	262639***	235113***	128886**
	(.000)	(.000)	(.000)	(.017)
Power		4387	84.4	3523
(change per capita)		(.493)	(.882)	(.187)
Development expenditure			.341*	.077
(change per capita)			(.084)	(.315)
Roads				-24703
(change per capita)				(.516)
Population (millions)				-30.9
(change per capita)				(.337)
TDL				2.57
(change per capita)				(.471)
Lagged dependent variable				.715***
				(.000)
State and year fixed effects				Yes
Sample size	238	238	238	238
Sargan-Hansen test: p-values	(.096)*	(.339)	(.360)	(.998)
AR(2): p-values	(.485)	(.441)	(.579)	(.303)

p-values in brackets. Significance levels are denoted by *** (1% level or less), ** (5% level or less) and * (10% level or less).

Table 5: Effect of human capital on GDP, manufacturing and agriculture					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	GDP	**Total Mfg.**	**Registered Mfg.**	**Unregistered Mfg.**	**Agriculture**
			Panel A		
Enrollments per capita	375747**	89078**	57391**	31688	-21541
	(.001)	(.023)	(.031)	(.127)	(.492)
Sample size	238	238	238	238	238
Sargan-Hansen test	(.449)	(.498)	(.246)	(.786)	(.293)
AR(2)	(.748)	(.439)	(.583)	(.963)	(.380)
			Panel B		
Enrollments per capita	290205	82137	48163	9902	-17696
	(.239)	(.170)	(.432)	(.777)	(.904)
Power	-3663	-3185	-2969	-1972	-4171
(change per capita)	(.773)	(.621)	(.645)	(.386)	(.667)
Development expenditure	-.236	.028	-.070	.106**	-.272
(change per capita)	(.533)	(.762)	(.558)	(.049)	(.327)
Roads	-9280	-54699*	-31246	-26517*	973
(change per capita)	(.919)	(.075)	(.115)	(.100)	(.989)
Population (millions)	-142.3	-35.9	-31.4	-2.11	8.41
(change per capita)	(.183)	(.401)	(.374)	(.887)	(.838)
TDL	3.28	-.912	.420	-.836	.664
(change per capita)	(.662)	(.818)	(.891)	(.627)	(.910)
Lagged dependent variable	.315	.534*	.697***	.412*	.467*
	(.179)	(.066)	(.008)	(.086)	(.093)
State and year Fixed effects	Yes	Yes	Yes	Yes	Yes
Sample size	238	238	238	238	238
Sargan-Hansen test	(.998)	(.999)	(.997)	(.999)	(.998)
AR(2)	(.745)	(.199)	(.102)	(.483)	(.339)

p-values in brackets. Significance levels are denoted by *** (1% level or less), ** (5% level or less) and * (10% level or less). "Mfg" is Manufacturing.

Table 6: Disaggregated sectors		
Short name	**Sector Name: Manufacture of**	**Mapped to ASI code**
Food	Food products	20, 21
Beverages etc.	Beverages, Tobacco and Related products	22
Textiles	Cotton, Wool, Silk, Man-made, Jute & Other Vegetable Fibre Textiles	23, 24 & 25
Textile products	Textile products	26
Wood	Wood and Wood products	27
Paper	Paper, Paper products, printing	28
Leather	Leather & Leather products	29
Chemicals	Basic chemicals and chemical products, Rubber, Plastic, Petroleum & Coal products	30, 31
Minerals	Non-metallic mineral products	32
Basic metals	Basic metals & alloys	33
Metal products	Metal products & parts	34
Transport parts	Transport equipment & parts	37
Trade	Trade, Hotels & Restaurants	
Banking	Banking & Insurance	
Business etc.	Real estate, ownership of dwellings & business services	
Telecom	Communication	
Transport services	Transport services	

Table 7: Factor intensity ratios (%)

	(1)	(2)	(3)
Sector: Short name	**Skilled labor**	**Power**	**Roads**
Food	17.5	2.4	28.5
Beverages etc.	11.8	2.8	10.0
Textiles	13.4	5.3	32.0
Textile products	12.2	0.7	17.9
Wood	10.9	11.8	7.7
Paper	15.3	47.3	24.3
Leather	13.1	1.6	13.4
Chemicals	17.0	24.2	22.1
Minerals	13.3	24.5	19.2
Basic metals	14.5	47.0	35.4
Metal products	14.7	19.6	17.6
Transport parts	15.4	7.0	8.7
Trade	17.1	3.9	4.2
Banking	42.2	1.7	0.4
Business etc.	32.7	1.6	1.4
Telecom	42.2	6.1	0.8
Transport services	17.1	3.2	5.5

Source: Authors own calculations based on input-output matrix of India from Global Trade Analysis Project (GTAP), Version 6.

Skilled labor (column 1) denotes remunerations of skilled labor in the sector as a percentage of total labor remunerations. Intensity ratios for power (column 2) and roads (column 3) equal total expenditure incurred by firms in the industry on the two respective inputs and expressed as a percentage of total value added in the sector..

Table 8: Disaggregate level results				
	(1)	**(2)**	**(3)**	**(4)**
Enrollments per capita	1255	1016	2089	-17498
	(.850)	(.874)	(.766)	(.154)
*Enrollments*Skill_Int*	854***	793***	729**	1448**
	(.006)	(.004)	(.040)	(.044)
Power		-480	-616	-151
(change per capita)		(.172)	(.202)	(.813)
Roads		3755	7075	4994
(change per capita)		(.228)	(.335)	(.556)
Population (millions)		-41.0	-1.51	-1.90
(change per capita)		(.552)	(.011)	(.616)
Development expenditure		-.007	.012	.002
(change per capita)		(.624)	(.296)	(.853)
TDL		-.474*	-.449	-.079
(change per capita)		(.083)	(.106)	(.858)
*Power*Power_Int*			9.96	.251
			(.518)	(.989)
*Roads*Road_Int*			-227	-275
			(.481)	(.420)
*Population*Serv*			3.76	-2.38
			(.030)	(.753)
*Development expenditure*Serv*			-.067	-.025
			(.182)	(.478)
Lagged dependent variable				.702***
				(.001)
State, sector and year fixed effects				Yes
Sample Size	3303	3303	3303	3303
Sargan-Hansen test (p values)	(.097)*	(.776)	(.591)	(.112)
AR(2) p values	(.375)	(.509)	(.390)	(.674)

p-values in brackets. Significance levels are denoted by *** (1% level or less), ** (5% level or less) and * (10% level or less).

Table 9: Disaggregate level results – Long run (Sample=236)			
	(1)	(2)	(3)
Panel A: Second stage regression results			
*Education*Skill_Int*	196***	337***	286***
	(.000)	(.000)	(.027)
*Power * Power_Int*			58.3
			(.214)
*Population*Serv*			496
			(.975)
*Dev*Serv*			.471**
			(.003)
*Roads*Road_Int*			-2923
			(.254)
State & sector fixed effects	No	Yes	Yes
R^2	.203	.819	.826
Overidentification test (p values)	(.891)	(.680)	(.122)
Panel B: Shea partial correlations from the first stage regressions			
*Education * Skill_Int*	.781	.329	.162
*Power * Power_Int*			.137
*Population*Serv*			.551
*Dev * Serv*			.395
*Roads * Road_Int*			.074
p-values in brackets.			