



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

**EDUCATIONAL HUMAN CAPITAL AND LEVELS OF INCOME:
EVIDENCE FROM STATES IN INDIA, 1965-92**

Kamakshya Trivedi

Number 97

March 2002

Manor Road Building, Oxford OX1 3UQ

Educational Human Capital and Levels of Income: Evidence from States in India, 1965-92.

Kamakshya Trivedi¹

Nuffield College

University of Oxford

email: kamakshya.trivedi@nuffield.ox.ac.uk

February 2002. Comments Welcome.

Abstract

This paper examines the long-run (steady-state) relationship between levels of educational human capital and levels of income for the 15 major states of India between 1965 and 1992. The relationship is estimated using the Pooled Mean Groups (PMG) technique; which produces common long-run coefficients but allows heterogeneity of the short-run adjustment parameters, and is well-suited to approximately square panels. The results suggest that levels of educational human capital, proxied by secondary school enrollment rates, have a robust positive impact on steady-state levels of income. This is true for male and female education, and the regressions also suggest that states which have larger gender-gaps in education have lower steady-state incomes. The estimated relationship is robust to the inclusion of alternative measures, added controls, and variation in the degree of state coverage.

JEL Classification: O11, O53, I29

Keywords: growth, India, education, panel data

¹This research was generously supported by Nuffield College, Oxford. I have benefitted from the support and advice of various people. I am most grateful of all to Gavin Cameron for constant encouragement and helpful supervision. In addition specific thanks are due to Chris Bowdler, Rui Fernandes, Bryan Graham, Christian List, Misa Tanaka, and Jon Temple, who gave me written comments and numerous suggestions to improve the paper. A draft of this paper was presented at a Gorman Workshop in the Department of Economics at Oxford, and I would like to thank all the participants, especially Steve Bond, Sophocles Mavroedis and John Muellbauer for helpful comments. Lastly, I am very grateful to Tim Besley, Robin Burgess, Rohini Pande, and the Economic Organization and Public Policy Programme, STICERD, LSE, for access to data; and Vijayan Punnathur and the Rai family for their hospitality and help in New Delhi. The standard disclaimer applies: responsibilities for all remaining errors and omissions is mine.

1 Introduction

This paper studies the relation between income levels and the levels of educational capital in India. The aim is to clarify the role of education on incomes in aggregate data, especially within developing countries. Towards this aim, the paper uses a panel of the major Indian states between 1965 and 1992, and demonstrates that higher educational levels are correlated with higher levels of per capita incomes, and therefore higher transitional economic growth.

India is an important developing country to study for a number of reasons: it is a *mover and shaker* of global statistics. For example, it enjoys the dubious distinction of being home to the largest number of people living below a miserly poverty line. For many of these people, even small increases in their income levels can often make the difference between life and death. Another reason for interest is that learning about India might provide important hints about economic processes in other large developing countries. From an econometric perspective too, India is a rare example of a low-income federally organized country where long time series data is available for regional states. Finally, state governments have jurisdiction over many policy variables, including education; and barring a few exceptions over time, the states have been self-governing units with substantial powers of taxation and spending.² Hence, most of the results reported here have natural policy implications.

While there is substantial international consensus among policy makers on the role of education in raising incomes, the econometric evidence is considerably less clear.³ Early growth-accounting exercises such as Denison (1985) simply assumed the significance of education in explaining the level of income. However, recent empirical research in this area has raised many questions about the popular conventional wisdom that education is good for economic growth. Benhabib and Spiegel (1994) was one of the first papers to document the empirical insignificance of educational attainment. In their cross-country specifications the level of education (measured by average years of schooling or literacy rates) does not enter significantly in explaining the level of income after controlling for the level of capital and the population, (Table 3 in their paper). A related literature (Barro and Lee (1994), Barro and Sala-i-Martin (1995), and Barro (1997)) has noted similarly pessimistic findings on the role of female education for economic growth. For example, Barro (1997) comments, “*In earlier results, Barro and Lee (1994) found that the estimated coefficient on female secondary and higher schooling was significantly negative. With revised data on education, the estimated female coefficients are essentially zero.*”⁴

These somewhat counter-intuitive and perplexing results have not gone uncontested, however. The work of Temple (1999) seems to suggest that the assumption of parameter homogeneity in cross-country regressions might be partially responsible. He is able to reverse some of the Benhabib and Spiegel (1994) results using the robust Least Trimmed Squares estimator on their data set. In similar spirit, Lorgelly and Owen (1999) show that Barro and Lee’s (1994) empirical results on female education are

²More precisely, the Indian Constitution specifies *education* as a ‘cocurrent’ subject, i.e. one on which both the state governments and the central government can make policy.

³Temple (2001) shows how difficult it is to make reliable generalizations in this area.

⁴On the other hand, Caselli, Esquivel and Lefort (1996), in their cross-country panel, find a statistically significant positive coefficient on female schooling and a significant negative coefficient on male schooling.

sensitive to whether or not the East Asian newly industrializing countries are included in the sample. The implication is that while it might be the case that changes in educational attainment are not positively correlated with changes in incomes in a few countries or at some points in time, they might well be positively related *on average*.⁵ In other words, the assumption of parameter homogeneity in cross-country work on growth should be made with a good deal of caution.

Apart from demonstrating that education raises incomes in low-income countries, this paper advances the empirical literature on economic growth by addressing directly the issue of parameter heterogeneity. It might be thought that assuming parameter homogeneity would not be as detrimental to a within country study, such as this. However, there exist vast disparities between Indian states on almost every measurable dimension, so it would not be prudent to sidestep the issue. The approach in this paper is to use the recently developed technique of Pooled Mean Groups (PMG) suggested by Pesaran, Shin and Smith (1999) to take account of parameter heterogeneity. Given the relatively small sizes of data sets used in empirical growth analyses – in my case, 15 states and approximately 25 years – allowing full parameter heterogeneity implies estimating many parameters, with the associated imprecision. The Pooled Mean Groups estimator provides a middle path: it allows different short-run adjustment coefficients for each cross-sectional unit, but restricts the long-run steady-state parameters to be constant across states. So in my conditional convergence model, states have the same long-run steady-state relationship between levels of education and levels of income, but each state has different short-run dynamics, i.e. each state converges to its steady state level at a different rate.

A few words about existing state-level studies of India: To the best of my knowledge, this is the first paper that focuses on the relation between education and steady-state incomes using an aggregate state level panel. Nosbusch (1999) uses school enrollment rates in a test of the textbook Solow model augmented with human capital. The only other related paper is Nagaraj, *et al* (1998), who examine the issue of convergence and long-run growth trends. As explained in the next section, owing to measurement error their results on education are less reliable than the rest of the analysis. Both Nosbusch (1999) and Nagaraj, Varoudakis and Veganzones (1998) use fixed effects estimation, but neither addresses the issue of parameter heterogeneity. In passing, note that the findings of both papers are consistent with the results presented here.

Section 2 of this paper starts by briefly explaining the theoretical framework for the empirics; and argues that the empirical conclusions of this paper are broadly consistent with models which allow for balanced growth paths, and a corresponding steady-state level which depends on the level of educational capital. It then goes on to discuss the econometric methodology: specifically the rationale and mechanics of the Pooled Mean Groups (PMG) estimator in some detail. Starting from an unconstrained Autoregressive Distributed Lag model, it explains how to obtain common long-run or steady-state relationships between the dependent and independent variables. So, in the context of this paper, the empirics will show that increases in the levels of educational human capital raise the steady-

⁵ As stated by Temple (1999),

“Simple cross-country regressions do not detect an effect of human capital because of a small number of countries, perhaps ones in which human capital accumulation has had little or no effect.”

state levels of income. Section 3 discusses the data used in the estimation. It outlines the advantages and limitations of different education measures which could be used in the Indian context. In the end high school enrollment is used, but robustness of results is assessed using alternative measures. Section 4 then goes on to discuss the chief results of the regressions, which show a robust and positive impact of the educational variable on incomes. Section 5 concludes. Appendix A details the sources of data used in this paper.

2 Theoretical Framework and Econometric Methodology

2.1 Theoretical Framework

The focus on the impact of the *level* of education on the *level* of steady-state incomes in this paper is in contrast to much of the empirical literature on education and growth. The idea is to use a conditional convergence model – equation (8) derived in the next subsection – to calculate the long-run impact of schooling on the steady-state level of income. In a recent paper, Shioji (2001) uses a similar approach to estimate the effect of public capital on steady state output per capita.⁶ Such an empirical specification is valid if the underlying theoretical growth model allows linearization of dynamics around a steady-state. For example, if per capita log output, y , evolves as in (1):

$$\begin{aligned} \frac{dy}{dt} &= \lambda[y^* - y(t)] \\ \text{where } y^* &= \text{steady state output per capita,} \\ &\text{and } \lambda \text{ is the convergence parameter.} \end{aligned} \tag{1}$$

then a conditional convergence specification is a valid empirical representation.⁷ The empirical results presented in this paper are consistent with growth models which feature transitional dynamics around a steady-state level.⁸ Within this class, the results suggest an important role for models where the corresponding steady-state income level depends on the level of educational human capital. However, given the high values of the conditional convergence coefficients reported in the regressions, typical closed versions of human capital augmented Solow models in the fashion of Mankiw, Romer and Weil (1992)⁹ would most likely be a poor fit.¹⁰ On the other hand, open economy versions of such models would generate higher convergence rates for given capital shares, provided that the degree

⁶Shioji (2001) provides an explicit growth model with public capital. The model generates the standard conditional convergence dynamics which are linear around a steady-state

⁷For more details see Sachs and Warner (1997)

⁸Note that the impact on growth rates is through the transition dynamics of the model. When the steady-state level is raised, given the economy's current position, it is now further behind in relation to its new higher steady state than it was in relation to its former lower steady state. This implies that it has more catching up to do, and it should grow faster in the interim transition phase, *ceteris paribus*.

⁹Note that it is possible to augment the neo-classical Solow (1956) model with human capital in more interesting ways than Mankiw *et al* (1992). For instance, Jones (1998, ch.3) outlines an augmented Solow model where it is assumed that an economy accumulates human capital by spending a constant fraction of time learning new skills instead of working. In this model, the level of steady-state income depends on the ratio of the stocks of skilled and unskilled labour.

¹⁰For a similar argument, see Caselli *et al* (1996).

of capital mobility is sufficiently high.¹¹ Since, in this paper the panel consists of states within a country rather than a cross country sample, the open economy assumption is not at all implausible. Within India, mechanisms of technology transfer, or systems of inter-state federal transfers – some of which are explicitly aimed at reducing regional income inequalities – would constitute mechanisms which speed up the convergence process. Thus, without pushing the theoretical interpretation too far, open economy versions of human capital augmented neo-classical models would be consistent with the results reported here.

As an alternative to the conditional convergence approach pursued in this paper, one could directly estimate a production function, one of the inputs for which would be educational human capital. Shioji (2001) provides some reasons why the conditional convergence approach might be preferable to the direct production function approach. One such argument which would apply with equal force to this study has to do with contemporaneous correlation. In the absence of good instruments, and given the small data set, a production function approach which involves regressing current output on current levels of education may be subject to bias. The conditional convergence approach might avoid these problems by associating current output with past levels of education.

2.2 Econometric Methodology

This subsection derives the precise equation under estimation. Most empirical work on growth assumes that the parameters of interest are the same across all cross-sectional units.¹² This is also true of existing studies on states within India. The implication is that states follow the same underlying model relating income to its determinants. However, from the work of Lee, Pesaran and Smith (1998) and Temple (1998, 1999, 2000), we know that this is by no means a trivial assumption, and allowing parameter heterogeneity can change results of growth regressions drastically. This suggests that we should allow for different coefficients for each state on the independent variables, including measured educational capital. But allowing for complete parameter heterogeneity also has a drawback – the imprecision associated with estimating many parameters, especially in small data sets. Consider the Mean Group (MG) estimator. To compute this, one estimates separate equations for each group. The mean of the estimates, the MG estimator, will provide consistent estimates of the average of the parameters, as shown by Pesaran, Smith and Im (1996). However, the MG estimators are likely to be inefficient when either T or N are small.

One way out of the dilemma would be to ignore the issue and proceed on the assumption of parameter homogeneity; on the grounds that parameter heterogeneity is much less likely to be a serious issue in a within-country study, as opposed to a cross-country study. For example, one could estimate a traditional pooled estimator such as the Dynamic Fixed Effects (DFE) estimator. Such an approach is not without merit, and might well be persuasive in most cases, but for countries like India, it will not do. There is no gainsaying the fact that India comprises enormous differences, as

¹¹See for instance, the open economy model with partial capital mobility in Barro and Sala-i-Martin (1995, ch.3)

¹²The partial exception to this generalization is the ‘fixed effects’ panel data model – which allows the researcher to control for time-invariant factors through intercepts which vary across the cross-sectional units.

pointed out in the introduction as well.^{13,14}

The PMG estimation procedure provides a sagacious middle path between assuming identical coefficients and allowing complete parameter heterogeneity. The PMG estimator allows the intercepts, short-run coefficients and error variances to differ freely across states, but the long-run coefficients are constrained to be the same. This means that the speeds with which states adjust to long-run equilibrium relationships are estimated separately for each state, but the long-run relationships are common across states. Pesaran *et al* (1999) remark that there are “*often good reasons to expect the long-run equilibrium relationships between variables to be similar across groups, due to budget or solvency constraints, arbitrage conditions, or common technologies influencing all groups in a similar way.*” In the present case, it seems sensible to expect that the long-run relation between income and educational human capital – is similar across states within India, once state-specific institutions have been controlled for; even as the short-run adjustment processes are unique to each state. This seems a reasonable way of incorporating heterogeneity in growth regressions, although there still remains the assumption of common long-run relationships.

In addition, the PMG estimator is designed for panels where both, the number of cross-sectional units and the number of time series observations, are quite large and of the same order of magnitude. Since the panel used in this paper consists of 16 states observed annually over approximately 25 years, it is well suited to the PMG estimation procedure.¹⁵ If the long-run relationships between growth and the conditioning variables are indeed similar across states, then by pooling time-series data on different cross-sections, we widen the informational base of the estimator; and therefore expect more precise estimates of the long-run parameters of interest. For example, Pesaran *et al* (1999) find that when compared to Mean Group (MG) and Dynamic Fixed Effects (DFE) estimates, the PMG estimates are more robust to outliers and to choice of lag order.¹⁶

Some researchers split the cross-sectional units into somewhat arbitrary groups and estimate them

¹³Most states in India are larger than small countries around the world. Travelling through the 16 states in the data set (listed in table A-1, in Appendix A), one would encounter astonishing differences: at least 20 different languages with more than 1 million speakers, not to mention 22000 dialects; adult (ages 15-59) literacy rates ranging from 90% in Kerala (state=9) to 36% in Bihar (state=3); and a female-male ratio varying from 1.04 in Kerala (state=9) to 0.88 in Uttar Pradesh (state=20), which is one of the lowest in the world.

¹⁴In Trivedi (2000) I estimated pooled fixed-effects conditional convergence regressions using the same dataset. The assumption of identical error variances between states – groupwise homoscedasticity – was clearly rejected by the data. In other words, parameter heterogeneity is an important concern for any study dealing with states within India, and must be tackled directly.

¹⁵To the best of my knowledge, this is one of the first papers to use the recently developed Pooled Mean Groups procedure in the growth context. Cameron (1999) uses this estimator to examine the issue of productivity convergence between Japan and the USA in a panel of similar dimensions – 11 industries over 27 years. Bassanini and Scarpetta (2001) use the PMG estimator to study human capital and growth in an OECD cross-country panel which is also approximately square – covering 21 countries and spanning 27 years. However, unlike this paper, they do not focus separately on the role of male and female education.

¹⁶In fact, even if the assumption of long-run parameter homogeneity is not satisfied, PMG estimates might still be better than say, MG estimates in terms of Mean Square error. Informally, the argument can be expressed in terms of the traditional bias-variance trade-off. Pooling data in the presence of parameter heterogeneity might provide misleading but more precise estimates than estimators based on unpooled data. Note that PMG provides consistent estimates of the long-run relationships and short term adjustment coefficients under the assumptions specified.

separately. This approach to the problem of parameter heterogeneity, while informative, is eschewed here since the number of states is anyway quite small. However, in section 4.2, robustness of the PMG estimator is assessed by dropping each state in turn, and checking its impact on key coefficient estimates.

Formally, the model to be estimated takes the following general form,

$$\Delta y_{it} = \phi_i y_{i,t-1} + \sum_{j=1}^k \beta_i^j x_{i,t-1}^j + \sum_{j=1}^k \sum_{s=1}^m \delta_{is}^j \Delta x_{i,t-s}^j + \mu_i + \varepsilon_{it} \quad (2)$$

where i indexes the state, j indexes each independent variable in the vector x , and, s determines the lag length for each variable in x . Equation (2) is a straightforward reparameterization of an unrestricted Autoregressive Distributed Lag model. Note that in the case of the panel examined in this paper, the so-called *Nickell bias*¹⁷ is likely to be quite small since T is quite large. (In the worst cases of truncation due to missing values, it is still about 20 years). Pooling the data in (2) and constraining coefficients to be the same across states along with state-specific intercepts produces the dynamic fixed effects (DFE) model,

$$\Delta y_{it} = \phi y_{i,t-1} + \sum_{j=1}^k \beta^j x_{i,t-1}^j + \sum_{j=1}^k \sum_{s=0}^m \delta_s^j \Delta x_{i,t-s}^j + \mu_i + \varepsilon_{it} \quad (3)$$

In most of the estimation that follows m will be restricted to 0, i.e. each variable in x will enter the right-hand side in levels and only once in differences. Assume, for the moment a single regressor, $k = 1$. In this case the DFE model can be rewritten in the equilibrium-correction form as follows,

$$\Delta y_{it} = \phi[y_{i,t-1} - \theta X_{i,t-1}] + \delta \Delta X_{it} + \mu_i + \varepsilon_{it} \quad (4)$$

where $\theta = \left(-\frac{\beta}{\phi}\right)$ is the long-run coefficient of the ‘equilibrium’ or ‘steady-state’ relation between y and X .¹⁸ The PMG estimator, by contrast, allows the adjustment coefficients to differ between states, and only constrains the long-run relationship to be homogenous:

$$\Delta y_{it} = \phi_i[y_{i,t-1} - \theta X_{i,t-1}] + \delta_i \Delta X_{it} + \mu_i + \varepsilon_{it} \quad (5)$$

The PMG estimates of the long-run coefficients discussed below are Maximum Likelihood estimates.¹⁹ The errors ε_{it} are assumed to be stationary and normally distributed. Both these assumptions are

¹⁷Nickell (1981) showed that fixed effects estimators of dynamic panel data models are biased, and consistency relies upon T being large.

¹⁸Note that $\phi_{(i)} \neq 0$ is required to ensure the existence of a long-run relationship. In the estimation results described below, ϕ is always less than zero. The restriction $\phi = 0$ is always rejected. The restriction on $\phi_{(i)}$ is related to the presence of unit roots in the data; for more details see Pesaran *et al* (1999). Table A-6 in the Data Appendix reports results of the panel unit root test suggested by Im, Pesaran and Shin (1997) for levels and differences of the y and x variables. The PMG estimator is consistent in the case of $I(0)$ or $I(1)$ regressors.

¹⁹PMG estimation implies that the long-run coefficient(θ) is a non-linear function of the short term adjustment parameters(β, ϕ). Estimation will therefore involve use of iterative procedures.

tested. The PMG estimates of the adjustment coefficients are the simple unweighted average of the state coefficients, given by

$$\hat{\phi} = N^{-1} \sum_{i=1}^N \hat{\phi}_i, \quad \hat{\delta} = N^{-1} \sum_{i=1}^N \hat{\delta}_i \quad (6)$$

In other words, they are the mean group estimators of the adjustment coefficients. Their estimated standard errors are the standard deviations of the coefficients across states:

$$s.e.(\hat{\phi}) = \sqrt{Var(\hat{\phi})}; \text{ where } Var(\hat{\phi}) = \frac{1}{N(N-1)} \sum_{i=1}^N (\hat{\phi}_i - \hat{\phi})^2 \quad (7)$$

The consistency of the Mean Group estimates and of the standard errors obtained by this method is demonstrated in Pesaran, *et al* (1996).

In practice, lagged values of the x variables are used instead of current values in (5) *a la* Cameron (1999), in order to rule out any contemporaneous correlation with the error term. So finally, the equation under estimation, with only one X variable would look like,

$$\Delta y_{it} = \phi_i[y_{i,t-1} - \theta X_{i,t-2}] + \delta_i \Delta X_{i,t-1} + \mu_i + \varepsilon_{it} \quad (8)$$

Thus a long-run relationship, θ , is posited between the level of income, $y_{i,t-1}$, and the level of education in the previous period, $X_{i,t-2}$. In general, we will also control for the impact of other x variables on the steady-state level of income.

3 Data and Measurement

The emphasis of this paper is on education, rather than on any notion of human capital, which is a much broader concept.²⁰ While some of the early literature in this area used enrollment rates or census based adult literacy rates to proxy for educational human capital, (for e.g. Mankiw, *et al* (1992), Barro (1991), Sala-i-Martin (1997)), the recent trend has been to develop education stock estimates based on mean school years of education in an economy, (for e.g. Pritchett (1999), Benhabib and Spiegel (1994)).²¹ Clearly, none of these measures is perfect,²² but literacy rates and estimates of average years of schooling have the advantage of being more direct proxies for the *stock* of educational human capital than enrollment rates. In this paper, I use high school enrollment rates to proxy for the stock of educational human capital since they are available annually at the state level. The chief

²⁰ While the two concepts might be correlated, it is important to recognize that we are consciously ignoring important aspects of human capital such as health & nutrition, informal education & vocational training, innate or acquired skills and talents, and so on.

²¹ Note that estimates of years of schooling are themselves constructed using either enrollment rates (for eg., Nehru, Swanson, and Dubey, 1995) or census observations on literacy rates and/or years of schooling (for eg., Psacharopoulos and Arriagada, 1986).

²² The important weaknesses of the different measures is well documented. (See for eg., Nehru, Swanson and Dubey (1995)). Probably the most crucial is that none of these measures gauge the quality of education, which affects comparability both over time and across units.

disadvantage of census based literacy estimates is that they are only available at discrete intervals – in the case of India they are available at 10 year intervals. However, census based literacy rates and high school enrollment rates are highly correlated. Table 1 is a correlation matrix for male and female literacy rates (IMLR, IFLR), and male and female high school enrollment rates (MSEC, FSEC). The high overall correlation between the enrollment and literacy data suggests that the enrollment rate is a reasonable proxy for the level of educational capital in a state. The two measures are providing similar information, but the enrollment rate has the advantage of being recorded annually. I also constructed similar correlation matrices for each state. In general, the correlations are very high (between 0.7 and 0.9); except in case of the correlation between IMLR and MSEC in Jammu and Kashmir, which has fewer overall observations.²³

Table 1: Correlation Matrix of Education Measures				
	MLITRAT	FLITRAT	MSEC	FSEC
IMLR	1			
IFLR	0.958	1		
MSEC	0.829	0.82	1	
FSEC	0.9056	0.9579	0.909	1
(obs=476)				

Two conceptual problems with using enrollment rates as a proxy for the level of educational capital remain. The first has to do with migration.²⁴ If a large proportion of individuals migrate to another region to work after completing their education, then lagged enrollment would not be a good proxy of the educational level of the work force in that particular region; and the results from regressing enrollment on income would be misleading. This is not a major problem for this study for the following reason. It is widely recognized that inter-state migration in India is very low, possibly on account of the large linguistic and socio-cultural differences between states; and a significant fraction of that migration has to do with factors unrelated to employment or earning opportunities.²⁵ The second has to do with dropout rates. If dropout rates are high then enrollment rates would exaggerate the true level of educational capital in a given state. Even so, we make no adjustment for dropouts in the estimation below given the absence of reliable detailed information on dropout rates. Note however that data from the PROBE report (1999) suggests that dropout rates are quite low, even in the most educationally backward states.²⁶ Thus while there might be some upward bias, the problem is unlikely to be very serious. Nevertheless, I check the robustness of the main results using literacy rates, (interpolated for each state based on a state-specific growth rate).

There is still the further issue of precisely which enrollment rate to use. Measurement of school

²³ All the regression estimates reported below omit Jammu & Kashmir because there are many missing values.

²⁴ This problem has received relatively little attention in the literature, perhaps because much of it is based on cross-country data-sets.

²⁵ See for example, Skeldon (1986) and Rosenzweig & Stark (1989).

²⁶ Results from the PROBE report suggest that both dropout and enrollment rates are related predominantly to the quality of schooling. Dropout rates are high in states where, on account of poor school quality, enrollment is low. Therefore, in the absence of quality of schooling data, enrollment rates might in fact neatly capture the quality of schooling as well.

enrollment in India is replete with problems.²⁷ The worst case is that of primary enrollment – a figure much used in growth regressions, and by international agencies.²⁸ Official school enrollment rates based on school records are (on average) exaggerated up to 20% over true figures. When matched with household survey data, it turns out that inflated class-1 figures account for the bulk of the discrepancy. Under-age enrollment is the most prevalent method used by teachers to inflate figures; although ‘nominal enrollment’, (enrollment of children at the start of the year who don’t attend school or only attend for a very short period), ‘double enrollment’(simultaneous enrollment in a local government school and an unrecognized private school), and ‘fake enrollment’(teachers including fake names in the school register) are also common.

Teachers have several incentives to inflate enrollment figures: (1) if low enrollment causes a drop in the pupil-teacher ratio below officially prescribed norms, the teacher may be transferred. (2) direct and indirect government pressures to show progress in enrollment over time, or attain ‘universal’ enrollment in lower age groups. (3) officially unrecognized private schools seek ‘double enrollment’, often a mutually advantageous arrangement. (4) official incentive schemes such as food for mid-day meals are an incentive for both teachers and parents to exaggerate enrollment. There are thus several good reasons to avoid primary school enrollment data. By contrast, high school enrollment – boys and girls between 11 and 14 years of age – are far more reliable, and the difference between official Department of Education figures and household survey figures is much less.

Since we are trying to discern the effect of the level of educational capital on the level of steady-state income, basic econometric theory suggests that it is important to control for other variables that might determine steady-state income. This precludes estimated coefficients on the education measure from being contaminated by other important, but omitted variables. Following the robustness studies of Levine and Renelt (1992) and Sala-i-Martin (1997), I include a proxy for non-educational human capital, (the infant mortality rate), and a proxy for the level of physical capital, (a principal components index of physical infrastructure) in all the regression results reported below.²⁹

²⁷The following discussion draws heavily from the PROBE Report (1999). The report includes a comprehensive analysis of the state of basic education in India. It is based on a survey of the 4 *Bimaru* states – Bihar, Madhya Pradesh, Rajasthan, & Uttar Pradesh – and Himachal Pradesh. It is a mine of interesting (and depressing!) information on education in India. I am grateful to Jean Dreze for pointing me in the direction of the report, and also for clarifying doubts about education statistics in India.

²⁸For instance, Nagaraj, *et al* (1998) find that of the different education variables that they use, only primary school enrolment has any significant impact on growth. Moreover, they find that primary school enrollment has the biggest impact on steady state incomes, compared to all the other explanatory variables in their regressions. On the basis of this they conclude that primary education should be prioritized over other levels. As explained in this paper, primary school enrolment is a particularly bad proxy for educational human capital in India.

²⁹In their baseline specification used to test robustness of the many different variables used in the empirical growth literature, Levine & Renelt (1992) choose the initial level of income, the investment rate, the secondary school enrollment rate, and the rate of population growth. In his study, Sala-i-Martin (1997) chooses the initial level of income, life expectancy and primary school enrollment rate. According to Sala-i-Martin (1997), these variables have certain properties that make them the appropriate benchmark: “...they have to be widely used in the literature,... and they have to be...somewhat ‘robust’ in the sense that they systematically seem to matter in all regressions run in the previous literature.”

An alternative method in selecting control variables is to start from economic theory, rather than empirical evidence. One could choose depending on one’s priors about the income growth process, or what model of economic growth one believes in. There are methodological advantages to both approaches, and in practise, they are almost always combined.

I use the infant mortality rate because estimates of the commonly used measure – life expectancy at birth – are not available for states in India until after the 1970s.³⁰ But even with the infant mortality rate there is a big difference in the quality of statistics before and after 1970. As one would expect, the quality of statistics after the 1970s (from the Sample Registration System (SRS)) is far better than those recorded before (from the Civil Registration (CR) system). However, because of substantial overlap it is possible to discern the magnitude of inaccuracy for each state. Using these overlapping observations I have constructed a consistent series on infant mortality using a spliced series for the pre-1970 period and the SRS estimates for the post 1970 observations. This technique causes the largest adjustments for states with the greatest CR deficiency. This construction is described in greater detail in Appendix A.

The other obvious variable to include in a study of steady-state incomes is the stock of capital. Unfortunately there is no good capital formation data available at the state-level in India, until very recently. Hence, I use a principal components measure of physical infrastructure based on 2 series on energy production (installed capacity and generation), 1 series on energy consumption (high voltage electricity consumption by industry), and 1 series on the length of state highways. Moreover, in section 4.3, I test the robustness of the key results using an alternative proxy for capital – fixed capital employed in the factory sector. The construction of the principal components index of physical infrastructure is described in more detail in Appendix A.

4 Estimation and Discussion

I turn now to discussing the regression estimates. All coefficients reported in tables 2 and 3 are the long-run steady-state parameters,³¹ estimated by Maximum Likelihood in TSP. Residuals from all estimated

³⁰The Infant Mortality Rate is defined as

$$IMR = \frac{\text{Number of infant deaths during the year}}{\text{Number of live births during the year}} \times 1000$$

³¹The chief exception is the ϕ coefficient on lagged income, which is a Mean Group estimate computed as in (6). In the presence of parameter heterogeneity, the MG estimates are consistent, but they are likely to be very inefficient for the sample size we are dealing with. I also computed the MG and DFE estimates for the same regressions. These are not reported in the tables below, but, in general, the standard errors of both the PMG and the DFE are smaller than those of the MG estimates. As econometric theory suggests, pooling sharpens the estimates considerably.

models are tested for non-stationarity,³² absence of AR(1) serial correlation,³³ and normality.³⁴

4.1 The Level of Education and Steady-State Income levels

Column 2a in Table 2 has male and female high school enrollment as independent variables in the regression. The initial results are encouraging – both enter the regression with a positive and significant coefficient, although the positive effect of female schooling on steady-state income is approximately double that of the male schooling variable. Column 2b adds two important control variables to the right-hand side – the infant mortality rate, and a principal components measure of physical infrastructure. Unless the impact of educational human capital on steady-state income works entirely through increasing non-educational human capital and physical capital accumulation, we would expect the high school enrollment variables to retain their positive and significant coefficients even after including the relevant proxies. Adding the controls reduces the size of coefficients on both male and female enrollment, but the model seems better specified – the log-likelihood increases, the standard error of regression decreases, and there are fewer states for whom we reject the null of normally distributed residuals. This suggests that the coefficient estimates on the schooling variables in column 2a were at least partly picking up the influence of non-educational human capital and physical capital. As one would expect, the infant mortality rate enters with a negative coefficient and physical infrastructure enters with a positive coefficient. Both are statistically significant. In addition, note that even in column 2b the coefficient on male high school enrollment remains substantially smaller than that on female high school enrollment. This last result – the larger impact of female high school enrollment – is somewhat perplexing, especially in the light of much cross-country evidence such as the aforementioned Barro (1997), and Barro and Lee (1994).

In columns 2c and 2d we allow the male and female enrollment variables to enter separately. Column 2c retains the male school enrollment but drops the female school enrollment. The coefficient estimate on male schooling rises to of 0.009; but the standard error of regression increases and the maximized log likelihood falls. Column 2d repeats this exercise with the female schooling variable, but without the male schooling variable. The results are similar in spirit: female enrollment has a positive

³²The test statistic for the \overline{LM} test of unit roots in heterogenous panels is based on Im *et al* (1997).

$$\overline{LM}_{NT} = \frac{1}{N} \sum_{i=1}^N LM_{iT}$$

where LM_{iT} is the LM statistic from the first stage Dickey-Fuller regressions (with an intercept but without a time trend) performed separately for each state. The errors of the first stage DF regression are assumed serially uncorrelated. Critical values for the \overline{LM} statistic are tabulated in Im *et al* (1997).

³³The F statistic for $AR(1)$ serial correlation is based on Durbin (1970) and is asymptotically equivalent to the LM test suggested by Breusch, Godfrey (1978) for regressions containing Lagged dependent variables. The same statistic can be used to test for higher order Autoregressive processes. In fact, I test for up to $AR(3)$ residual serial correlation. Except for the regressions reported in columns 3d, 3e & 3f, the hypothesis of no serial correlation at these higher orders is not rejected at the 5% level.

³⁴The test of Normality is based on the W statistic of Shapiro and Wilk (1965). It is an analysis of variance test of normality. The test has been shown to have good power against a wide range of non-normal alternatives even for small samples ($n < 20$). Critical values for W are tabulated in Shapiro and Wilk (1965).

and significant estimate of 0.012, which is still substantially larger than the coefficient estimate of male high school enrollment in column 2c.³⁵

In terms of impact, these coefficient estimates (columns 2a to 2d) imply that a 10% rise in the level of male schooling raises steady-state income levels by between 6% and 9%. (Between 1965 and 1992, the mean of the male schooling variable is 56.8%; the standard deviation is 17.1%). An analogous increase of 10% in female schooling raises steady-state incomes between 7% and 12%. (The mean of the female schooling variable is 32% and the standard deviation is 20.3%). In the Indian context, this finding of a larger relative impact of female high school education is consistent with recent microeconomic evidence from labour-market studies. In work based on household survey data between 1983 and 1993-94, Duraisamy (2000) reports that, “...the returns to women’s education exceed that to men’s at the middle, secondary and higher secondary levels. Especially at the secondary level, the returns to additional schooling of women is over twice as large as the corresponding returns for men.” Even at the cross-country level, recent work by Knowles, Lorgelly and Owen (2000) suggests that female education makes a bigger statistically significant contribution to output per worker than male schooling.³⁶

It is also worth thinking about the much larger t -statistics on the education variables in columns 2c and 2d, which suggest a possible ‘multicollinearity’ situation in the regressions in column 2a and 2b. Multicollinearity describes a situation where two or more regressors are very highly correlated. In this case it is difficult to disentangle their respective effects, and this is reflected in high standard errors (so that it is not possible to estimate either regressor precisely), and unstable point estimates. Table 1 shows that the correlation coefficient between male and female school enrollment is 0.91.³⁷

³⁵One important drawback of not including both education variables in regressions 2c & 2d is that it is not possible to rule out that only one variable matters, and including the other simply picks up that correlation. Concretely, it could be argued that female education does not matter; but in a model like 2d, the female high school enrollment variable is positive & significant because it is picking up the effect of the omitted male schooling variable, with which it is correlated, and *vice versa*.

³⁶The difference in the impact of male and female schooling should not be over-emphasized. A simple back-of-envelope *average* elasticity calculation (based on regression estimates in column 2b) yields elasticities of 0.05 & 0.04 for the male and female education variables respectively. So proportional changes in male and female schooling have similar effects on the level of income; and if anything a slightly higher impact from a proportional increase in male high school enrollment. Thus, part of what these regression estimates are capturing is some sort of ‘low level effect.’ The existence or not of such non-linearities in the impact of schooling would be an interesting avenue for further research.

³⁷When calculated separately for each state, the lowest correlation coefficient between male and female high school enrollment is 0.71 (for Uttar Pradesh), and it is below 0.9 for only 4 out of the 15 states (Bihar, Gujarat, Haryana & Uttar Pradesh) in the sample. Thus, the presence of some multicollinearity seems certain.

In Trivedi (2000), I have explored this issue in greater detail. A somewhat more formal method of detecting multicollinearity is to look at Variance Inflation Factors (VIF) from the regressions. A VIF for regressor x^j is given by,

$$VIF(x^j) = \frac{1}{1 - R_j^2}$$

where R_j^2 is the square of the multiple correlation coefficient that results when x^j is regressed against all the other explanatory variables. Intuitively, the VIF measures how the variance of an estimator is *inflated* by the presence of multicollinearity. In Trivedi (2000), I calculated VIFs for all the regressors after simple fixed effects estimation on a similarly specified model. In model 2b, for instance, male and female enrollment had VIFs which were more than double that of the other regressors, again suggesting multicollinearity.

The specification in column 2e is an isomorphic transformation of the general unrestricted model in column 2b. It includes male high school enrollment and the difference between male and female high school enrollment, which can be thought of as a ‘gender-gap’ in education. Using this isomorphic transformation mitigates the problem of multicollinearity since the correlation between male enrollment and the gender-gap is close to zero.³⁸ If the levels of both male and female education matter to steady-state income distinctly from each other, we should see a positive and significant coefficient on male schooling, and a negative and significant coefficient on the gender-gap variable. Moreover, we should expect more precise estimation of male schooling relative to the general specification in column 2b.³⁹ In fact, this is the case. In the results in column 2e, the significance of male schooling holds up much better relative to the more general specification in 2b; it is more precisely estimated even as the rest of the results are unchanged.

³⁸The correlation coefficient between male school enrollment and the gender-gap in school enrollment is approximately only -0.1. As Hendry (1995) states, “*a given model may or may not have multicollinearity depending on which isomorphic representative is considered.*”

³⁹Since 2e is an iso-morphic transformation of 2b, the coefficient estimates can be unscrambled from the estimates of 2b. Suppose θ_m and θ_f are the long-run coefficients on male and female enrollment in 2b. If λ_m and $\lambda_{(m-f)}$ are the coefficients on male enrollment and the gender-gap in enrollment, respectively; then the transformation 2e implies,

$$\begin{aligned}\lambda_m &= (\theta_m + \theta_f) \text{ , and} \\ \lambda_{(m-f)} &= -\theta_f\end{aligned}$$

Table 2: Education & levels of Income: 15 States, 1965-92								
Dependent Variable: Annual Growth Rate of Real State PerCapita Income								
model		2a	2b	2c	2d	2e	2f	2g
		PMG	PMG	PMG	PMG	PMG	PMG	PMG
lagged (t-1) income	ϕ	-0.4574 [-4.620]	-0.5210 [-4.971]	-0.4812 [-5.269]	-0.4792 [-4.948]	-0.5210 [-4.971]	-0.5165 [-5.173]	-0.5079 [-4.537]
lagged(t-2) male high school enrollment	θ	0.0061 [2.961]	0.0059 [2.778]	0.0093 [7.292]		0.0133 [9.143]		
lagged(t-2) female high school enrollment	θ	0.0127 [5.315]	0.0075 [3.376]		0.0117 [9.169]			
lagged (t-2) infant mortality rate	θ		-0.0012 [-2.723]	-0.0013 [-2.739]	-0.0013 [-3.205]	-0.0012 [-2.723]	-0.0012 [-2.786]	
lagged(t-2) physical infrastructure	θ		0.0180 [1.938]	0.0476 [3.458]	0.0147 [1.711]	0.0180 [1.938]	0.0300 [2.764]	0.0391 [2.827]
lagged(t-2) [Male minus Female high school enrollment]	θ					-0.0075 [-3.376]		
lagged(t-2) [Average of Male and female enrollment]	θ						0.0121 [10.281]	
lagged(t-2) [3 lag years MA of infant mortality rate]	θ							-0.0005 [-2.138]
lagged(t-2) [3 lag years MA of male high school enrollment]	θ							0.0136 [8.656]
lagged(t-2) [3 lag years MA of male minus female high school enrollment]	θ							-0.0070 [-2.947]
fixed effects	μ	Yes	Yes	Yes	Yes	Yes	Yes	Yes
number of observations	NT	360	360	360	360	360	360	360
Maximised Log Likelihood		448.85	473.13	455.07	461.49	473.13	461.46	487.00
Standard error of Regression		0.079	0.074	0.078	0.078	0.074	0.077	0.073
LM bar test of non-stationary residuals		10.398	10.049	9.772	9.698	10.049	9.712	11.236
No. of states for which Ho. Of normality of residuals rejected		6	3	3	3	3	3	3
F-test of No AR(1) residual serial correlation		0.356	0.186	0.233	0.173	0.186	0.112	0.491

Notes: t-statistics are reported in parentheses.
Standard error of Regression is the simple average of consistent estimates of equation standard errors.
The LM bar test of Non-stationary residuals is based on Im, Pesaran and Shin (1997). 5% critical values between 3.63 and 3.73.
Residuals for each state were tested for Normality using the Shapiro-Wilk (1965) test.
p-values reported for the F-test of No AR(1) serial correlation in residuals, Durbin (1970), Breusch, Godfrey (1978).

A slightly different approach would be to include a measure of average enrollment, using male and female population shares from the census. This also gets us around any multicollinearity – notice the precisely estimated coefficient on average enrollment in column 2f; but in doing so, it also disregards

valuable information on the possible separate effects of male and female schooling on levels of income. Besides, the diagnostics also suggest that the model in 2e is superior to 2f – so we shall make no further use of this average measure of enrollment, and rely instead on the gender-gap specification to mop up any multicollinearity.

Overall, these results demonstrate a remarkably strong and clear positive relation between male and female education and income levels. This might well seem a disingenuous conclusion, but it is not common outside of samples that comprise developed countries, and even then the focus is usually on male education indices. There is also evidence that the impact of female school enrollment is greater than that of male school enrollment. The results are especially striking because even though we use schooling variables as proxies for educational human capital, one might believe that measures such as high school enrollment or infant mortality would affect income with more than just one or two lags. However, using longer lags in all the regressions would considerably reduce the number of observations available for estimation. Hence, in column 2g, I use moving averages of the variables for the three previous years as regressors in case of the infant mortality rate, the male enrollment rate, and the gender-gap in enrollment.⁴⁰ All the coefficients are correctly signed and significant, the coefficients on male enrollment and the gender-gap are almost unchanged, and the model has better diagnostics compared to 2e.

4.2 Sensitivity Analysis

In the introduction to this paper, I cited a few studies which caution that empirical results on the relation between education and income might be overly sensitive to the inclusion of certain countries (or states). Hence, it is important to check whether the results reported in table 2 above are sensitive to the exclusion of a certain state. In other words, we would want that deleting a certain observation – for which the model may be a particularly bad approximation – should not alter the basic character of the results.⁴¹ Following standard procedure, I recursively estimate the baseline regression model in column 2e after dropping each state one after the other, and present the coefficient estimates on male school enrollment and the gender-gap in enrollment in figures 1 and 2.

Both figures demonstrate the remarkable stability in sign and size of coefficient estimates as state coverage is varied. This should not be entirely surprising, since the PMG procedure already allows for a degree of parameter heterogeneity; and as mentioned in section 2.2, it is more robust to outliers relative to DFE and MG estimates. Specifically, the non-inclusion of Kerala alters coefficient estimates along predictable lines – it is widely recognized that Kerala is an atypical state.⁴² Being a state with exceptionally high enrollment but with only a moderate level of income, its non-inclusion increases the long-run coefficient estimate of male enrollment in figure 1.

⁴⁰ As specified in (8), these moving averages are then further lagged 2 periods. So for example the coefficients estimate the equilibrium relation between Y_{t-1} and a moving average of X in $(t-3)$, $(t-4)$ and $(t-5)$.

⁴¹ Such ‘unrepresentative’ or ‘influential outlier’ observations could perhaps arise because of a coding or transcription error; or – as Temple (2000) points out, from a parameter heterogeneity perspective, the observations may be entirely correct but drawn from a different regime.

⁴² Much has been written about the uniqueness of Kerala’s educational achievements. See for instance, Dreze and Sen (1995, 1997), and the references therein.

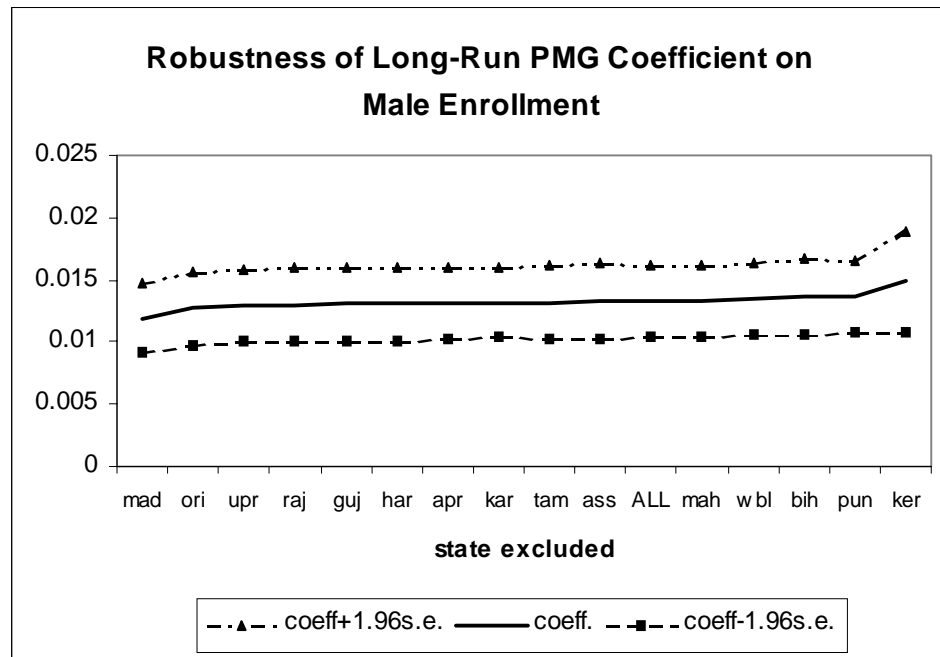


Figure 1: RECURSIVE ESTIMATES – MALE ENROLLMENT

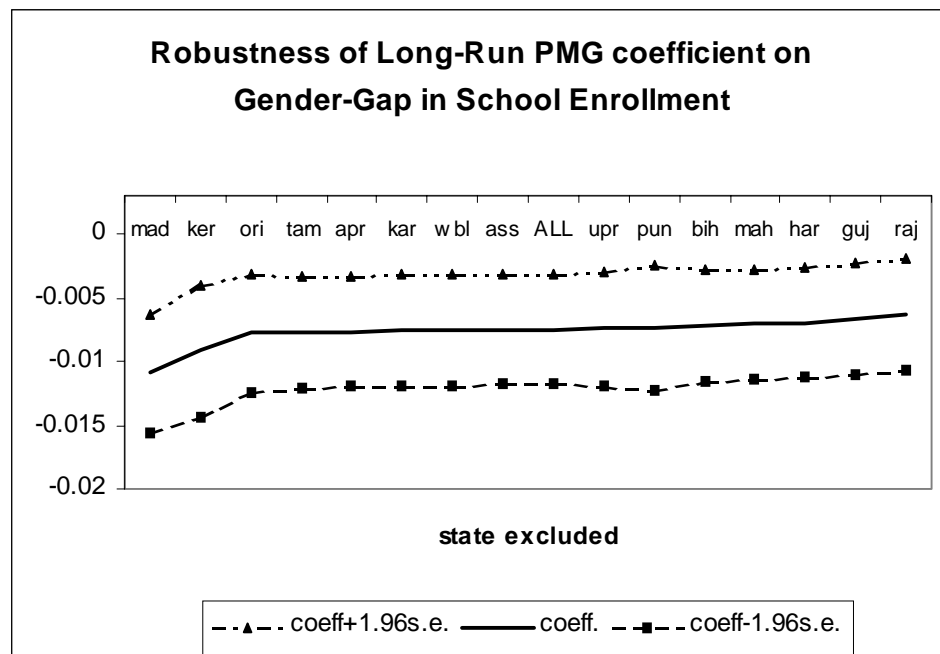


Figure 2: RECURSIVE ESTIMATES – GENDER GAP IN ENROLLMENT

Turning to figure 2, the negative coefficient on the gender-gap – implying an income penalty to states with larger gender-gaps in education – is also robust to alterations in state coverage. The uniqueness of Kerala again stands out: as a state with an exceptionally low gender-gap, but only a moderate income level, its inclusion reduces the estimated penalty in the entire sample. The case of Madhya Pradesh – the other outlier in figure 2 – defies a similarly facile explanation. Dropping Madhya Pradesh from the sample under estimation considerably increases the size of the coefficient; but note that the point estimate of the baseline model with all states falls within the upper confidence bound of the model excluding Madhya Pradesh.

4.3 Alternative measures and Robustness

The positive impact of the level of male and female education on income reported in table 2 is surprisingly clear, and it is also robust to alterations in the sample of states included. In this section we assess the robustness of the results from a slightly different angle: by using additional controls and alternative measures for education, such as the literacy rate.

In table 2, I have measured the gender-gap in enrollment as the difference between the male and female enrollment rates. However, from an economic point of view, measuring the gender-gap as the difference is not ideal, since female enrollment in secondary school in the mid-sixties was still at a very low level. As a result, the difference between male and female enrollment increases over time in many states, even though female enrollment has risen at a much faster rate relative to male enrollment. This undesirable outcome can be finessed by using the ratio of male to female enrollment, instead of the difference, to measure the gender-gap. Measured in this fashion, the gender-gap in enrollment declines over time in all states, albeit of-course at variable rates. Moreover, the simple correlation between male school enrollment and the ratio of male to female enrollment is again low – approximately - 0.5; substantially lower than the correlation between male and female school enrollment; so that multicollinearity should not be as much of a problem. In column 3a, (and in all subsequent results reported in table 3), I measure the gender-gap as this ratio. The negative and significant coefficient on the gender-gap measured thus, again underlines the importance of female enrollment to steady-state incomes. The results in column 3a imply that after controlling for the positive impact of the level of male education on income, states which have a lower level of female education (relative to male education) – or simply, states with a bigger gender gap in secondary school enrollment – will have lower steady-state incomes on average.⁴³

In column 3b, I check the sensitivity of the coefficients on the education variables in this specification by adding a government policy variable on the right-hand side. This is the real (per capita) state government *Development* expenditure, which includes *inter alia*, expenditure on education, health-care, nutrition, social and community services, natural calamities, etc., and is generally the category under which most capital expenditure is accounted.⁴⁴ Both the male education and the gender-gap

⁴³Note that the proxy for non-educational human capital – the infant mortality rate – is not precisely estimated when the gender-gap in education is measured as a ratio.

⁴⁴The accounting counterpart of *Development* expenditure in government budgets in India is *Non-Development* Expenditure – which includes *inter alia*, organs of state government, administrative services, etc., and is generally the head under which most revenue expenditure is classified.

in education are significant and have the right signs. However, physical infrastructure is no longer significant once government expenditure is included – suggesting that much of the action on physical infrastructure in this period is coming from government development expenditure. In other words, adding government policy to the regression reduces considerably the explanatory power of the physical infrastructure variable, but the positive impact of the level of education emerges largely unaffected.

Thus far I have been using the index of physical infrastructure as a proxy for the level of physical capital in each state. However, this measure ignores, to a significant extent, the role of private capital since many if not most investments in physical infrastructure during the sample period were funded by public authorities. While there is no way of fully rectifying this obvious lack of data, it is possible to mitigate it partially by including a measure of (real) fixed capital utilized in the factory sector in the regressions, and checking how this affects the results.⁴⁵ Column 3c includes this variable and reports estimates for the gender-gap specification. The results are encouraging, in that both the male enrollment and the gender-gap in enrollment are of similar size, significant and correctly signed. Interestingly, physical infrastructure remains positive and significant, while real fixed capital in the factory sector has a negative coefficient, but it is not estimated very precisely.⁴⁶

⁴⁵A ‘factory’ is defined by the Factories Act (1948). Given the significant presence of the public sector in industry during the sample period, a measure of fixed capital in the factory sector includes both private and public capital.

⁴⁶The inclusion of fixed capital in the factory sector makes no difference to the results even when male and female enrollment variables are regressed separately.

Table 3: Robustness & Alternative Measures: 15 States, 1965-92							
Dependent Variable: Annual Growth Rate of Real State PerCapita Income							
model		3a	3b	3c	3d	3e	3f
		PMG	PMG	PMG	PMG	PMG	PMG
lagged (t-1) income	ϕ	-0.5223 [-5.144]	-0.6257 [-5.790]	-0.5388 [-5.227]	-0.6501 [-5.370]	-0.6427 [-5.687]	-0.6846 [-5.455]
lagged(t-2) male high school enrollment	θ	0.0089 [7.259]	0.0067 [6.398]	0.0084 [5.105]			
lagged (t-2) infant mortality rate	θ	0.0004 [1.124]	-0.0001 [-0.305]	0.0001 [0.360]	-0.00001 [-0.080]	-0.0001 [-0.675]	0.0005 [1.452]
lagged(t-2) physical infrastructure	θ	0.0559 [5.646]	0.0070 [0.581]	0.0616 [4.282]	0.0011 [0.095]	0.0031 [0.257]	0.0293 [1.282]
lagged(t-2) [Ratio of Male to female enrollment]	θ	-0.1691 [-5.498]	-0.0670 [-2.123]	-0.1531 [-4.421]			
lagged(t-2) government development expenditure	θ		0.0021 [5.075]				
lagged(t-2) real fixed capital in factory sector	θ			-0.0021 [-0.149]			
lagged(t-2) male literacy rate	θ				0.0363 [14.027]		0.0242 [2.967]
lagged(t-2) female literacy rate	θ					0.0328 [10.858]	
lagged(t-2) [Ratio of Male to Female Literacy Rate]	θ						-0.3241 [-2.333]
fixed effects	μ	Yes	Yes	Yes	Yes	Yes	Yes
number of observations	NT	360	360	360	351	351	351
Maximised Log Likelihood		474.53	499.00	487.401	455.92	452.29	463.01
Standard error of Regression		0.075	0.072	0.074	0.075	0.077	0.073
LM bar test of non-stationary residuals		10.280	10.251	10.902	12.595	11.748	11.903
No. of states for which Ho. Of normality of residuals rejected		4	4	6	4	6	2
F-test of No AR(1) residual serial correlation		0.792	0.439	0.805	0.970	0.060	0.963

Notes: t-statistics are are reported in parentheses.
Standard error of Regression is the simple average of consistent estimates of equation standard errors.
The LM bar test of non-stationary residuals is based on Im, Pesaran and Shin (1997). 5% critical values between 3.63 and 3.73 for columns 3a to 3f, and between 3.73 and 3.80 for columns 3d to 3f.
Residuals for each state were tested for normality using the Shapiro-Wilk (1965) test.
p-values reported for the F-test of No AR(1) serial correlation in residuals, Durbin (1970), Breusch, Godfrey (1978).
Regressions in columns 3d to 3f omit the states Haryana & Assam, on account of missing values.

In section 3, I outlined the potential drawbacks of using the enrollment rate as an indicator of the level of human capital in a state. However, it was also explained that for a panel-data study of Indian

states, it is nevertheless the best indicator to use. Even though enrollment rates are an ‘input’ into the educational human capital production process, they are highly correlated with ‘output’ outcomes such as literacy rates. Moreover, given the low rates of inter-state mobility of labour within India, the link between the level of enrollment and the level of education in a state is substantially unbroken. Here, I adopt a different approach to assess the validity of the reported conclusions using enrollment rates. In columns 3d-3f, I use the overall male and female literacy rates to proxy for the level of educational human capital. Literacy rates are measured every 10 years in the census, so they have been linearly interpolated separately for each state in the intervening years. Consequently, I ignore the magnitude of the coefficient estimates, and focus only on the signs and significance of the results as a robustness test.⁴⁷ Columns 3d and 3e include the male and female literacy rates one after another along with the standard controls. Both are significant and correctly signed. Column 3f includes the male literacy rate and a gender-gap in literacy rates. Reassuringly again, both variables are correctly signed and significant. It appears therefore, that the level of secondary school enrollment is, in fact, capturing something fairly close to the level of educational capital in a state; and its beneficent impact on the level of income is robust to the use of alternative measures.

5 Concluding Remarks

This paper has estimated the effect of the level of educational capital on long-run income levels in a panel of Indian states between 1965 and 1992. It was recognized that parameter heterogeneity is an important problem for such an empirical exercise. Addressing this concern was traded off against obtaining results that were generalizable across states. To this end, the empirical strategy of Pooled Mean Groups was employed to estimate common long-run relationships within the framework of an error correction model, while allowing for the short-run adjustment parameters to be estimated separately for each state. The key finding is that the stock of educational capital, proxied by the secondary school enrollment rate, has a significant positive impact on the steady-state level of per capita income; and on attendant growth rates. This effect is robust to the inclusion of control variables proxying for non-educational human capital and physical capital, to alternative measures, and to variation in state coverage. Another important set of findings is that both male and female educational capital are positively related to the steady-state incomes; or that gender-gaps in education reduce long-run incomes. These are striking findings from a low-income country, especially India – a country not normally associated with high incomes or for its prowess in the field of basic education.

Given these findings, and the background of India’s particularly poor educational standards, it would be easy, though justified, to critique yet again public policy neglect of this important area. This statement warrants an important qualification: that education might play a role in augmenting levels and growth rates of state incomes, is at best, a second order reason for public policy intervention. There are many other important reasons why governments should facilitate school education among people. These have to do with the inherently *developmental* role that education plays, such as enhancing

⁴⁷The states Assam and Haryana are dropped from the regressions using literacy rates. This is because census based literacy rates for Haryana are only available after 1971; and the census could not be conducted in Assam in 1981 due to disturbed conditions in the state.

capabilities or improving quality of life – see, for instance, Dreze and Sen (1995). The evidence in this paper does suggest, however, that the dichotomy between such a *developmental* role of education and an exclusively *economic / instrumental* rationale for education may be overdrawn in practice. For example, Sen (1999) states,

“that a country need not wait until it is much richer (through what may be a long period of economic growth) before embarking on rapid expansion of basic education and healthcare.”

The argument in this paper is that, *on average*, a country is unlikely to get much richer without an expansion of schooling or basic education; even though there might be some exceptions at points in time.⁴⁸

⁴⁸In other words, the results of this paper imply that the role of education – especially in low income countries with poor education – might be robust to certain changes in the definition of “socio-economic welfare”. Regardless of whether “socio-economic welfare” is defined in terms of narrow income or growth criteria, or in terms of broader *developmental* criteria, education seems to play an important role in advancing it.

A Data and Sources

The data set covers the 16 major states of India (see table A-1 below), in most cases for a period of 25 years, between 1965 and 1992. Most of these states were reorganized along linguistic lines to currently specified boundaries in 1956⁴⁹, and have existed as such until the year 2000.⁵⁰ In 1960 Bombay State was split into Gujarat and Maharashtra. In 1965-66 the core Punjabi Suba was split up into Punjab, Haryana and Himachal Pradesh; and data for both Punjab and Haryana commences in 1965 for most variables. Data on Jammu and Kashmir is patchy in the early years and in the early 90s for most variables. Hence it is dropped from the sample under estimation. Out of the 16 states, it is the smallest and not of special economic significance either. In this limited statistical sense, it is the least painful to omit.

Table A-1: Sample of 16 States	
State	Code
Andhra Pradesh	1
Assam	2
Bihar	3
Gujarat	4
Haryana	5
Jammu & Kashmir	7
Karnataka	8
Kerala	9
Madhya Pradesh	10
Maharashtra	11
Orissa	14
Punjab	15
Rajasthan	16
Tamil Nadu	18
Uttar Pradesh	20
West Bengal	21

Summary Measures for main variables are tabulated in table A-5. The LM bar statistic for panel unit root tests in levels and differences are tabulated in table A-6.

A.1 Income/Growth

The two sources for the incomes data are

- Ozler, Datt and Ravallion(1996): This data set compiles a consistent set of figures on incomes, price indices, population, *inter alia* for the rural and urban areas of India's sixteen major states

⁴⁹The States Reorganisation act, 1956, specified 14 states within the Indian Union. For more on the linguistic reorganisation of states in India, see Paul Brass (1990).

⁵⁰In the year 2000, three of the states in the sample were bifurcated. Uttaranchal, Jharkhand and Chattisgarh were carved out of Uttar Pradesh, Bihar and Madhya Pradesh, respectively.

spanning the period 1958-1992.⁵¹

- *Estimates of the State Domestic Product*, Central Statistical Organization, various issues. Estimates after 1981 are from diskettes obtained directly from the CSO office, Sardar Patel Bhavan, New Delhi.

Real state per capita income is calculated in the following manner. A deflator is constructed using different price indices for agricultural labourers (SCPIAL1) and industrial workers (STCPIW1) by state and year from the Ozler *et al*(1996) data set, and by weighing them by the respective rural and urban population shares (POP1 and POP2). The population data comes from the decennial census estimates. Between any two censuses it is assumed to grow at a constant rate of growth derived from the respective population totals. Like almost all other variables in the paper the deflator is also time-varying and state-varying.

$$deflator_{i,t} = \frac{POP1}{POP1+POP2} \times SCPIAL1 + \frac{POP2}{POP1+POP2} \times STCPIW1$$

Estimates of the Net State Domestic Product (computed at factor cost and current prices) for each state and all sectors and year are then divided by the total population and the deflator to obtain consistent estimates of real state per capita income. Growth rates are calculated by taking log differences of the real state per capita income, and divided by the number of intervening years.

A.2 Education Measures

- High School enrollment data comes from the serial publication, *Education in India*, Department of Education, Government of India. The data relates to boys and girls between 11 and 14 years of age. The enrollment rates are calculated as the percentage of students enrolled in classes 6 – 8 to the estimated child population in the age group 11 to 14. Schooling in this age group is sometimes also categorized as ‘upper primary’.
- Literacy data is taken from censuses, 1951,1961,1971,1981,1991, *Census of India*, Registrar General and Census Commissioner, Government of India.

A.3 The Infant Mortality Rate

Data on infant mortality rates from the Sample Registration Survey (SRS) was collected from various issues of the *Sample Registration Bulletin*, Office of the Registrar General, Government of India; and from Bose, A., *India’s Basic Demographic Statistics: 177 Key Tables with Graphics*, 1996. Data on infant mortality rates from the Civil Registration (CR) sample is taken from various issues of the publication, *Vital Statistics of India*, Office of the Registrar General, Government of India.

In India, vital statistics are recorded under two alternative systems: the Civil Registration system (CR) and the Sample Registration System (SRS). Civil Registration data are severely deficient

⁵¹I am grateful to the *Economic Organisation and Public Policy Programme*, STICERD, LSE, for access to the Ozler *et al* data set, and the state government finance data. Specific thanks in this regard are due to Tim Besley, Robin Burgess, and Rohini Pande.

primarily due to incomplete coverage, the extent of which varies in different states. In contrast, the SRS is a reliable dual record system which became operational in 1969-70, covering about 3700 sample units.⁵² The number of sample units has been increasing over the years. As of 1995 it stood at 6300 sample units covering well over 2 million people. The SRS estimates are far more accurate than the CR estimates, but the CR estimates have the advantage of being available before 1970. Fortunately, since data from the CR sample are available even after the commencement of the SRS, it is possible to infer the degree of inaccuracy in the CR data for different states.⁵³ Comparing birth and death rates from the two samples in 1988 reveals that the states with the biggest inaccuracies (more than 70%) are Assam, Bihar, Rajasthan, Uttar Pradesh and West Bengal.

A consistent series of the infant mortality rate is constructed by using a spliced series for the pre-1970 observations and using the SRS estimates for the post 1970 observations. This has the advantage of retaining the variation within the original CR data, while appropriately rescaling it to make up for the deficiency in sample coverage. The splicing is done separately for each state, via a scaling factor constructed as an average from the ratios of overlapping observations of CR and SRS data. West Bengal had the fewest number of overlapping observations (5), followed by Jammu and Kashmir, (7). All other state scaling factors were constructed with about 15 years of overlapping observations. There is thus little chance that randomness in any given year would affect the scaling factors and distort the data. Reassuringly, plotting the spliced pre-1970 series against the CR estimates reveals that the biggest adjustments are in states where the CR deficiency is the greatest.

A.4 Physical Infrastructure

The data on 3 electricity measures used to compute the index of physical infrastructure have been collected from various issues of the *Statistical Abstract of India* (SAI), published by the Central Statistical Organization (CSO), Department of Statistics, Ministry of Planning, Government of India.⁵⁴ The measures are:

- PEIPCAP: total installed capacity of electricity generation plants ('000s kilowatts);
- PEIPGEN: total energy generated (crores of kilowatt hours);
- PEIPINDH: sale of high voltage power to industry (crores of kilowatt hours).

Each measure is divided by state population to obtain a per capita number. In addition, the physical infrastructure index includes data on total state highways – PTIPSHW – taken from various issues of the SAI, CSO, Department of Planning, Government of India. The acknowledged primary

⁵² A sample unit in rural areas is a village or a segment of a village if it had a population of 2000 or more. In urban areas, a sample unit is a census enumeration block with a population ranging from 750 to 1000.

⁵³ In fact the official publication using CR data – *Vital Statistics of India* – routinely tabulates the ratio of vital statistics obtained via the CR system and SRS. The discrepancy is not minor. For example, in 1988 the infant mortality rate at the all-India level computed from CR data is 70% below the corresponding figure obtained from the SRS.

⁵⁴ The primary source until 1970-71 was the *Central Water & Power Commission*, Ministry of Irrigation and Power, Government of India. After 1971-72, the source is acknowledged to be the *Central Electricity Authority*, Ministry of Energy, Government of India.

sources change over the years – Ministry of Transport and Communications, Ministry of Transport, Ministry of Shipping and Transport, and finally, Ministry of Surface Transport.⁵⁵ I also collected data on Surfaced State Highways – PTIPSSHW – which are a subset of total state highways, and have the desirable feature of measuring both the quantity and quality of the roads infrastructure of states. However, since in the early part of the sample period there were few surfaced state highways to speak of, it is difficult to get a consistent series over the whole period. Hence we restrict attention to total State Highways,⁵⁶

- PTIPSHW: length of total State Highways (km.) as a proportion of total state land area.

Missing values for each state, the bulk of which are in the early 1960s, are linearly interpolated using within state growth rates.

Table A-2: Correlation Matrix of Physical Infrastructure Measures				
	PEIPCAP	PEIPGEN	PEIPINDH	PTIPSHW
PEIPCAP	1			
PEIPGEN	0.9762	1		
PEIPINDH	0.8469	0.8591	1	
PTIPSHW	0.5695	0.5863	0.5379	1
(obs=521)				

Table A-2 is a correlation matrix of the infrastructure variables. One immediately striking feature is the very high pairwise correlation coefficients between the different electricity variables. This suggests that rather than including the different physical infrastructure variables in each regression, it might be preferable to construct one composite index of physical infrastructure which might proxy for the level of physical capital in each state. Hence, a principal components measure of physical infrastructure is constructed by combining the four infrastructure variables.⁵⁷ Table A-3 calculates the proportion of variation explained by each computed principal component, and it is clear that the 1st principal component is massively dominant – it explains 81% of the sum of the individual variances of the infrastructure measures.

⁵⁵Highways appear to be strategically more important for the economy than conventional road networks. For instance, according to the CMIE (1998) report, *Infrastructure in India*, national highways constitute only about 2% of the total road network, but carry close to 40% of the total road traffic.

⁵⁶In fact, the series on total state highways, PTIPSHW, and total surfaced state highways, PTIPSSHW, are highly correlated, with a correlation coefficient of 0.73. The worry in omitting to use PTIPSSHW was that we would not capture effectively the role of a state which while not expanding the total length of highways, channelled its resources into converting the existing unsurfaced highways into surfaced highways. On the whole, however, it appears as if states that do more to improve the roads infrastructure tend to do more of both.

⁵⁷Principal Components analysis is a statistical technique which linearly transforms an original set of variables into a smaller set of uncorrelated or orthogonal components, which represent most of the variation in the original set of variables. The theory and usage of principal components is well discussed in Duntelman (1989).

Table A-3: Eigenvalues & Explained Variance			
Principal Component	Eigenvalue	Proportion	Cumulative
1	3.22056	0.8051	0.8051
2	0.56934	0.1423	0.9475
3	0.18678	0.0467	0.9942
4	0.02333	0.0058	1
(obs=521)			

Moreover, since its eigenvector yields coefficient weights which are all positive, as reported in Table A-4, it is the principal component which can most easily be interpreted as a general measure of physical capital.⁵⁸ Rather serendipitously, the relative weights for each of the infrastructure measures are more or less equal in the first principal component: the lowest weight (0.40) for state highways is not that different than the highest weight (0.54) for electricity generation.

Table A-4: Eigenvectors & Factor Loadings				
Principal Component	1	2	3	4
Variable				
PEIPCAP	0.53506	-0.23763	-0.43193	0.68606
PEIPGEN	0.53914	-0.21486	-0.36788	-0.72652
PEIPINDH	0.51035	-0.24652	0.82314	0.03482
PTIPSHW	0.40321	0.91466	0.02322	0.01696
(obs=521)				

A.5 State Government Development Expenditure

Data on Government Development Expenditure in State Budgets is from various issues, *Reserve Bank of India Bulletin*, Reserve Bank of India; and from *Report on Currency and Finance*, Reserve Bank of India.

A.6 Fixed Capital - Factory Sector

Data on Fixed Capital in the factory sector has been compiled from two sources: various issues of the *Statistical Abstract of India* (SAI), published by the Central Statistical Organization (CSO), Department of Statistics, Ministry of Planning, Government of India; and *India Database, the Economy: Annual time series data in two volumes*, by H. Chandhok and The Policy Group.

⁵⁸ Principal Components is a statistical technique and so one should not put a strong theoretical interpretation on them, as such. However, one's theoretical priors would suggest that any principal component which is a serious candidate for representing a broadly defined measure of physical capital would be increasing in each measure of physical infrastructure, and would therefore assign positive weights to each measure.

Table A-5: Summary of Main Variables (1965-1992)

	Log Real Per Capita Income	Male High School Enrollment	Female High School Enrollment	Spliced Infant Mortality Rate	Physical Infrastructure (Principal Component)
Andhra Pradesh	6.912 [0.231]	43.135 [11.385]	22.244 [9.016]	105.932 [31.881]	-0.305 [1.127]
Assam	6.791 [0.230]	48.728 [9.653]	31.173 [9.403]	125.171 [47.329]	-1.668 [0.299]
Bihar	6.439 [0.189]	39.539 [6.411]	11.996 [5.225]	173.545 [104.145]	-1.019 [0.283]
Gujarat	7.092 [0.234]	58.294 [10.939]	35.684 [10.243]	130.091 [43.566]	2.087 [1.853]
Haryana	7.262 [0.259]	68.330 [6.551]	30.535 [11.983]	99.378 [21.448]	1.916 [1.440]
Jammu & Kashmir	6.924 [0.228]	57.634 [9.301]	28.794 [9.648]	86.503 [20.969]	-2.182 [0.164]
Karnataka	6.946 [0.200]	50.800 [9.265]	30.619 [10.429]	87.904 [20.189]	0.672 [0.892]
Kerala	6.777 [0.200]	89.954 [12.475]	82.501 [15.756]	42.889 [16.867]	0.490 [0.537]
Madhya Pradesh	6.725 [0.229]	50.002 [16.651]	18.579 [8.675]	141.451 [24.647]	-0.282 [0.946]
Maharashtra	7.167 [0.263]	66.139 [13.103]	39.950 [15.115]	90.409 [27.714]	2.496 [1.816]
Orissa	6.765 [0.220]	44.029 [13.250]	21.630 [12.210]	129.916 [12.890]	-0.302 [0.581]
Punjab	7.451 [0.238]	63.055 [7.315]	44.896 [12.688]	98.653 [35.785]	3.362 [1.906]
Rajasthan	6.663 [0.184]	48.018 [12.490]	12.675 [5.276]	127.258 [32.732]	-0.941 [0.686]
Tamil Nadu	6.919 [0.259]	73.851 [18.547]	48.494 [19.072]	102.842 [32.253]	0.167 [0.702]
Uttar Pradesh	6.782 [0.170]	55.251 [5.168]	18.182 [5.235]	204.350 [124.264]	-0.894 [0.407]
West Bengal	7.094 [0.183]	52.832 [15.377]	34.015 [16.523]	112.560 [36.576]	-0.123 [0.201]
Total	6.920	56.824	32.001	117.036	0.293
S.D. overall	[0.326]	[17.092]	[20.278]	[61.271]	[1.792]
S.D. between	[0.250]	[13.020]	[17.189]	[37.319]	[1.551]
S.D. within	[0.218]	[11.525]	[11.530]	[49.316]	[1.040]

Standard Deviations in parentheses.

Table A-6: Panel Data Unit Root Tests		
Variable	Levels	Differences
Log Per Capita Income	0.99	15.35
Male Secondary School Enrolment	2.23	9.44
Female Secondary School Enrolment	5.52	6.94
Infant Mortality Rate	3.82	13.00
Physical Infrastructure (PC)	1.39	13.00

Notes: 5% critical value based on Im, Pesaran & Shin (1997), table 3, between 3.63 and 3.73 (approx.)
Jammu & Kashmir is omitted from calculations.
Sample period is 1965-1992.

References

- [1] Barro, R. (1991). "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*, 106, 2 (May), 407-443.
- [2] Barro, R. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, London: MIT Press.
- [3] Barro, R., and Jong-Wha Lee (1994). "Sources of Economic Growth," *Carnegie-Rochester Conference Series on Public Policy*, 40 (June), 1-46.
- [4] Barro, R., and X. Sala-i-Martin (1995). *Economic Growth*. New York: McGraw-Hill.
- [5] Bassanini, A., and S. Scarpetta (2001). "Does Human Capital Matter for Growth in OECD Countries? Evidence from Pooled Mean-Group Estimates," OECD Economics Department Working Paper No. 282, OECD.

- [6] Benhabib, J., and M. Spiegel (1994). "The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-country data," *Journal of Monetary Economics*, 34, 2 , 143-174.
- [7] Bose, A. (1996). *India's Basic Demographic Statistics: 177 Key Tables with Graphics*. Delhi: B.R. Pub. Corp.
- [8] Brass, P. (1990). *The Politics of India Since Independence*. Cambridge: Cambridge University Press.
- [9] Breusch, T.S. (1978). "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334-355.
- [10] Cameron, G. (1999). "The Sun Also Rises: Productivity Convergence between Japan and the USA," mimeo, Nuffield College, University of Oxford.
- [11] Caselli, F., G. Esquivel, and F. Lefort (1996). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics," *Journal of Economic Growth*, 1, 3 (September), 363-390.
- [12] Chandhok, H., and The Policy Group (1990). *India Database, the Economy: Annual Time Series Data in two Volumes*. Delhi: LM Books.
- [13] CMIE (1998). *Infrastructure in India*. Bombay: Centre for Monitoring Indian Economy Pvt. Ltd.
- [14] Denison, E. (1985). *Trends in American Economic Growth, 1929-1982*. Washington: Brookings Institution.
- [15] Dreze, J., and A. Sen (1995). *India: Economic Development and Social Opportunity*. Delhi: Oxford University Press.
- [16] Dreze, J., and A. Sen, eds. (1997). *Indian Development: Selected Regional Perspectives*. Delhi: Oxford University Press.
- [17] Dunteman, G. (1989). *Principal Components Analysis*. Newbury Park: Sage Publications.
- [18] Duraisamy, P. (2000). "Changes in Returns to Education in India, 1983-94: By Gender, Age-Cohort and Location," Economic Growth Center Discussion Paper no. 815, Yale University.
- [19] Durbin, J. (1970). "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410-421.
- [20] Godfrey, L.G. (1978). "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293-1302.
- [21] Hendry, D. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- [22] Im, K., M. H. Pesaran, and Y. Shin (1997). "Testing for Unit Roots in Heterogenous Panels," mimeo, Cambridge University.

- [23] Jones, C. (1998). *Introduction to Economic Growth*. New York: W.W. Norton.
- [24] Knowles, S., P. Lorgelly, and D. Owen (2000). "Are Educational Gender Gaps a Brake on Economic Development? Some Cross-country Empirical Evidence," manuscript, University of Otago, (forthcoming in *Oxford Economic Papers*).
- [25] Lee, K., M. H. Pesaran, and R. Smith (1998). "Growth Empirics: A Panel Data Approach - A Comment," *Quarterly Journal of Economics*, 113, 1 (February), 319-323.
- [26] Levine, R., and D. Renelt (1992). "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82, 4 (September), 942-963.
- [27] Lorgelly, P., and P. Owen (1999). "The Effect of Female and Male Schooling on Economic Growth in the Barro-Lee Model," *Empirical Economics*, 24, 3, 537-57.
- [28] Mankiw, N. G., D. Romer, and D. Weil (1992). "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 2 (May), 407-437.
- [29] Nagaraj, R., A. Varoudakis, and M.-A. Veganzones (1998). "Long-run Growth Trends and Convergence Across Indian States," Technical Paper No. 131, OECD Development Centre.
- [30] Nehru, V., E. Swanson, and A. Dubey (1995). "A New Database on Human Capital Stock in Developing and Industrial Countries: Sources, Methodology, and Results," *Journal of Development Economics*, 46, 2 (April), 379-401.
- [31] Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 6 (November) 1417-1426.
- [32] Nosbusch, Y. (1999). "Convergence Across Regions: Evidence from India." Quantitative Economics Project, London School of Economics and Political Science.
- [33] Ozler, B., G. Datt, and M. Ravallion (1996). "A Data Base on Poverty and Growth in India," mimeo, World Bank.
- [34] Pesaran, M. H., R. Smith, and K. Im (1996). "Dynamic Linear Models for Heterogenous Panels," in L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: A Handbook of Theory with Applications*, (2nd rev. ed.). Dodrecht: Kluwer.
- [35] Pesaran, M. H., Y. Shin, R. Smith, (1999). "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of American Statistical Association*, 94, 446 (June), 621-634.
- [36] Pritchett, L. (1999). "Where has all the Education gone?" World Bank Working Paper no. 1581.
- [37] Psacharopoulos, G. and A. Arriagada (1986). "The Educational Composition of the Labour Force: An International Comparison", *International Labour Review*, 125, 5 (September-October), 561-574.

- [38] Rosenzweig, M. and O. Stark (1989). "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *Journal of Political Economy*, 97, 3 (August), 905-926.
- [39] Sachs, J. and A. Warner (1997). "Fundamental Sources of Long-Run Growth," *American Economic Review*, 87, 2 (May), 184-188.
- [40] Sala-i-Martin, X. (1997). "I Just Ran Two Million Regressions," *American Economic Review*, 87, 2 (May), 178-183.
- [41] Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.
- [42] Shapiro, S., and M. Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, 3-4 (December), 591-611.
- [43] Shioji, E. (2001). "Public Capital and Economic Growth: A Convergence Approach," *Journal of Economic Growth*, 6, 205-227.
- [44] Skeldon, R. (1986). "On Migration Patterns in India during the 1970s," *Population and Development Review*, 12, 4 (December), 759-779.
- [45] Solow, R. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70, 1(February), 65-94.
- [46] Temple, J. (1998). "Robustness Tests of the Augmented Solow Model," *Journal of Applied Econometrics*, 13, 4 (July-August), 361-375.
- [47] Temple, J. (1999). "A Positive Effect of Human Capital on Growth," *Economics Letters*, 65, 1 (October), 131-134.
- [48] Temple, J. (2000). "Growth Regressions and what the Textbooks Don't Tell You," *Bulletin of Economic Research*, 52, 3 (July), 181-205.
- [49] Temple, J. (2001). "Generalizations that aren't? Evidence on education and growth," *European Economic Review*, 45, 4-6 (May), 905-918.
- [50] The PROBE Team (1999). *Public Report on Basic Education in India*. New Delhi: Oxford University Press.
- [51] Trivedi, K. (2000). "Economic Growth, Convergence, and Levels of Income: Evidence from States in India, 1960-90." M.Phil. Thesis in Economics, Nuffield College, University of Oxford.