Educational Expenditure of Large States

A Normative View

Utilising the pooled data for 15 large Indian states over the period 1992-93 to 1997-98, this study employs panel data models to estimate the normative (average) levels of expenditure on primary, secondary and higher education. The findings of the study reveal that the actual spending on educational services in low income states is lower than their 'needs'. This finding implies that the existing fiscal equalisation mechanism has not been effective in offsetting the revenue and cost disabilities of the poorer states in India.

ALLEN ROY, B KAMAIAH, M GOVINDA RAO

Provision of basic education has been recognised as a social obligation of the state. Compulsory primary education for all children up to the age of 14 has been enshrined in the Directive Principles of State Policy in the Constitution. Though education is included in the concurrent list, the major responsibility of providing educational facilities rests on the state governments. But the vast differences in literacy rate, variation in enrolment and dropout rates among states in respect of primary, secondary and higher education, pose a question beyond the sincerity with which the states have been pursuing their social obligation. As can be seen from Table 1, the proportion of revenue expenditure spent on education across the 15 large states for the fiscal year 1997-98 is quite uneven (the same is the case with previous years also). The uneven nature might be attributed to the unequal level of development and presence of social pressure groups in these states. It may be necessary and useful in this context to take a normative view of educational expenditure, to assess the extent of disparities in respect of states' expenditure on education. In this context, there is a need to develop a conceptual framework and evolve an appropriate methodology to classify the states on the basis of the deviation of actual expenditure from its normative level in respect of provision of educational services.

In the absence of any yardstick to measure the extent of relative emphasis laid upon the provision of education by governments, the actual expenditures may be considered as a proxy. In other words, the higher (lower) the per capita expenditure on education, the higher (lower) is the emphasis the state lays upon provision of education. But such a view may be misleading when there are significant cost variations in the provision of educational services across the states. Thus, to meaningfully assess of the relative position of the states, it is necessary to ascertain the cost of providing a 'standardised' unit of educational service across Indian states. One such way is to treat each state as an independent entity and estimate educational cost functions separately for each state based on state specific-factors such as student-teacher ratio, enrolment ratio and infrastructure facilities. However, an exercise of this kind eludes the possibility of a common basis for comparison across the states. In this context, estimating an all-India average (normative) cost of providing a standardised unit of educational service might prove to be more meaningful [Rao and Agarwal 1992].

The present paper proceeds from the supply (cost) side and attempts to estimate the normative expenditure levels with regard to expenditure on education for 15 large Indian states for the fiscal year 1997-98. On the basis of normative expenditures, this paper proceeds further to make a comparative analysis of the normative and actual expenditure levels with the objective of classifying states on the basis of the relative emphasis laid on the provision of education. For a meaningful analysis, expenditure on education is categorised into three heads, viz, primary, secondary and higher. The expenditure considered in this study relates to total revenue expenditure, which is the sum total of non-plan and plan revenue expenditures.

In order to get reliable estimates for the expenditure functions, cross-section data pertaining to 15 different states are pooled for six years from 1992-93 to 1997-98. Pooled data, which deals with both the inter-temporal dynamics and the individuality of the entities being investigated in the study, provides qualitatively superior estimates. The analysis involving pooled data allows comparison between dissimilar/heterogeneous units (in our case states). The inclusion of cross-section data adds a lot of variability to the pooled data, thereby reducing the extent of collinearity among the variables. In addition, the degrees of freedom are also greatly enhanced. For the above-mentioned reasons, pooled data significantly contribute towards producing more reliable parameter estimates [Dielman 1989]. In this exercise we have employed the panel data model in respect of groupwise heteroskedasticity, cross-group error-correlation and autocorrelation [for an elaborate description of various models of panel data analysis, see Baltagi 1995].

Table 1: Educational Expenditure as Per Cent of Revenue Expenditure (RE) across 15 Large Indian States, 1997-98.

States	Primary Education	Secondary Education	Higher Education
Andhra Pradesh	10.88	7.48	2.82
Assam	14.01	6.48	2.13
Bihar	13.61	3.84	2.00
Gujarat	21.24	6.81	1.67
Haryana	6.01	4.34	1.72
Karnataka	16.97	8.85	2.25
Kerala	9.30	6.05	2.31
Madhya Pradesh	10.20	3.39	1.87
Maharashtra	16.70	7.95	1.75
Orissa	10.79	5.10	3.18
Punjab	4.03	7.07	1.67
Rajasthan	18.94	7.75	1.62
Tamil Nadu	8.40	6.63	1.72
Uttar Pradesh	9.58	4.69	1.32
West Bengal	6.44	9.61	2.49

The rest of this paper is organised as follows. The first section discusses the selection of variables and panel data models. The second is devoted to discussion of results. And the last section provides some policy implications and concluding remarks. Models of panel data are discussed in Appendix 1.

Methodology

Pooling data allows modelling differences across units. In other words, pooling of time series and cross-section data helps estimation of average responses underlying a given relationship. This can be done under various sets of assumptions and pooling schemes. In this study, we have considered the various possibilities of the simultaneous occurrence of heteroskedasticity, cross-group error correlation and autocorrelation. The details are discussed in Appendix 1.

In the present context, expenditures on education are seen to vary across the states over time. The reason is that the quantity as well as quality (and hence the cost) of services vary according to the capacities and priorities of the states to spend on education. But it is desirable to reduce the interstate disparities in expenditure on (and cost of) education. It is also necessary to ensure that each state provides a minimum level of educational services [Rao 1993]. To compare expenditures across states, it is necessary to have a standard that serves as a yardstick for comparison. However, the choice of a standard measure of comparison is not an easy task. One way of simplifying matters is to take a normative view and to propose the average level of providing educational services as a proxy for the normative level of expenditure. This line of thinking is pursued in this study.

Sticking to the supply side view, cost functions that reflect the per capita cost of providing educational services in the relevant age group are considered. These functions are based on quantity as well as price variables. The formulation of the models and hence the choice of determinants has been largely guided by economic logic. The models have been estimated using pooled data, and the estimated pooled regressions are then used for projecting the normative expenditure levels.

In the cost functions for primary, secondary and higher education the dependent variable for the respective heads is the public expenditure per person in the relevant beneficiary group of population. The independent variable set consists of a vector of quantity and cost determinants of expenditure on relevant heads. Viewing the heterogeneous nature of the Indian states, yearly preliminary regression exercises were first carried out to identify the important determinants. The results of the yearly regressions are not reported here. In line with the principle of parsimony, the numbers of independent variables in the final specifications have been kept at the minimum.

A simple method of estimating normative expenditure levels is to express public services as expenditure per intended beneficiary population group and estimate the cost of achieving an average output level. For example, primary education expenditure is meant to benefit children in the age group 5-9 and therefore expenditure per child in the age group is taken as the starting point for estimation. Similarly, the relevant beneficiary age groups considered in this study for secondary and higher education are 10-18 and 20-24 respectively (Although equalising expenditure per beneficiary (in the relevant population group) has the virtue of simplicity, it is not possible to clearly identify the beneficiary group for an expenditure head like education, since the whole society benefits from education. Expenditure needs of various states with regard to three different expenditure heads pertaining to education were also estimated by considering the total population as the relevant beneficiary group. Owing to poor performance in terms of parameter estimates, the results are not reported here).

As a starting point, the independent variable set chosen for the regression specifications included number of students (ENROL), student-teacher ratio (STRATIO), salary levels of the teachers (SALARY), price differences across the states (PRICEDIF), the degree of urbanisation of the states (UTOTPP), the literacy rates (LITERACY) of the states, proportion of dropout in primary education (PDROPOUT), proportion of SC/ST students in secondary education (SCSTSEC), proportion of SC/ST students in higher education (SCSTHIGH), etc. The choice of regressors in the final specification has been made by taking into consideration the model diagnostic test statistics. All the variables considered in this study are in their natural logarithm. A description of the variables appearing in the final specifications, along with sources of data, is presented in Appendix 2. The baseline (final) specifications for the three expenditure heads are as follows:

Primary education:

$$\ln PRIMARY_{it} = a_0 + a_1 \ln PENROL_{it} + a_2 \ln PDROPOUT_{it} + a_3 \ln LITERACY_{it} + a_4 \ln PSTRATIO_{it} + a_5 \ln PRICEDIF_{it} + \varepsilon_{it}$$

Secondary education:

$$\ln SECONDARY_{it} = a_0 + a_1 \ln SENROL_{it} + a_2 \ln LITERACY_{it} + a_3 \ln SSTRATIO_{it} + a_4 \ln PRICEDIF_{it} + a_5 \ln SCSTSEC_{it} + \varepsilon_{it}$$

Higher education:

$$\ln HIGHER_{it} = a_0 + a_1 \ln HENROL_{it} + a_2 \ln LITERACY_{it} + a_3 \ln HSTRATIO_{it} + a_4 \ln PRICEDIF_{it} + \varepsilon_{it}$$

In the above equations, a are parameters and e are the error terms. After specifying the models for the three expenditure heads, the next step is to estimate the models using the pooled data. Utilising the parameter estimates obtained from the panel data regressions, an in-sample forecasting of normative expenditure levels for the fiscal year 1997-98 is done for 15 states with regard to primary, secondary and higher education. These in-sample projections are considered as the 'normative' expenditures after making the necessary data adjustments to obtain the absolute figures. Finally, to facilitate comparison, the ratio between actual expenditure levels and normative expenditure levels has been calculated.

Empirical Results

Pooled regression models have been estimated in the present study. While estimating the models, problems arising out of the presence of heteroskedasticity, autocorrelation and cross-group error correlation, as discussed in Appendix 1 have been taken care of. Looking at the model diagnostic statistics (in the present case LM, LR and Wald) as well as the statistical significance of the individual coefficients, we have chosen the final models, i e, one for each expenditure head. Since adjusted R^2 or R^2 is not a meaningful summary statistic in the context of pooled data, it has not been used and hence not reported. LR diagnostic statistics being appropriate for pooled data models in respect of heteroskedasticity coupled with cross group error correlation, S_2 is reported in this exercise. Only the estimates of those models that are finally chosen are presented in Table 2. Since all the variables considered in this study are in their natural logarithm, the coefficients facilitate interpretation in terms of elasticity.

In the case of primary education all the variables except the student-teacher ratio are significant. Expenditure per child in the case of primary education (the relevant age group being 5-9) bears an inverse relationship with the proportion of enrolment in primary education, thereby implying a situation where per child expenditure comes down when the proportion of enrolment goes up. Such a relationship may suggest existence of extra capacity in terms of number of schools, classrooms, teachers, etc) in the case of primary education and signify a shift in preference from government education to private education. On the contrary, per child expenditure and dropout proportion in primary education are directly related. A higher dropout ratio effectively leads to erosion of enrolment in primary education. This in turn may lead to increase in the per capita expenditure in the provision of primary education. Therefore, attempts to check dropout rate can help better utilise public expenditure on primary education. Again, per child expenditure in primary education and literacy rates across the states tend to move together. Here it may be argued that high literacy rate leads to higher educational awareness, which in turn acts as social pressure demanding better quality of primary education. Hence, per capita cost of providing primary education tends to be higher with increasing literacy rates. The estimated model for primary education further suggests a direct relationship between per child expenditure and price differences across states. This relationship supports the economic hypothesis that higher prices lead to higher cost of providing a standardised unit of service.

In the case of secondary education, expenditure per beneficiary group (age group 10-18) is positively related to the proportion of enrolment, signifying underprovision of capacity. This implies that an increase in enrolment proportion requires an additional commitment by way of creating capacity. Alternatively, the direct relationship may be interpreted as a situation characterising overcapacity utilisation. As in the case of primary education, per child expenditure in secondary education is positively related to literacy rates and price differences across the states. Furthermore, the estimated model suggests that the impact of varying levels of student-teacher ratio and proportion of SC/ST students in secondary education have an insignificant impact on expenditure per beneficiary group.

Economies of scale operate in the provision of higher education, as the expenditure per beneficiary group (age group 20-24) is found to be inversely related to the proportion of enrolment in higher education. It may be noted that the expenditure head that is considered here consists of non-technical university education, the demand for which has been sliding down in per capita terms of the relevant age group, while the capacity that has been created is enormous and hence underutilised. Thus, the inverse relationship between enrolment proportion and per capita expenditure signifies the presence of excess capacity. Similar to the results obtained for primary and secondary education, expenditure per beneficiary group in the case of higher education bears a significant direct relationship with literacy rate and price differences across the states.

Table 3: Proportion of Actual Expenditure Levels to Normative Expenditure Levels, 1997-98

States	Total Expenditure		
	Primary Education	Secondary Education	Higher Education
AP	1.01	1.22@	1.70@
Assam	1.19	0.91	0.87
Bihar	0.65	0.36#	0.85
Gujarat	1.67@	1.02	0.88
Haryana	0.91	1.01	1.33@
Karnataka	1.11	1.15	1.16
Kerala	0.98	0.96	0.94
MP	1.03	0.46#	0.86
Maharashtra	1.50@	1.19	0.92
Orissa	0.84	0.63#	1.22@
Punjab	0.76	2.04@	1.49@
Rajasthan	1.19	0.98	0.94
Tamil Nadu	0.96	1.01	0.87
UP	0.61#	0.63#	0.72#
West Bengal	0.70#	1.18	1.04

Notes: @ actual expenditure significantly higher than the normative level.
actual expenditure significantly lower than the normative level.
The significance level is set at ± 20 per cent of the normative expenditure level.

Table 2: Parameter Estimates for Different Expenditure Heads: Panel Data Results

Expenditure Head: Primary education
Dependent Variable: Per capita expenditure on primary education

Model		Explanatory Variables					Test Statistics
	Constant	PENROL	PDROPOUT	LITERACY	PSTRATIO	PRICEDIF	LR
S_2R_0	-0.0845 (-0.062)	-0.5937 (-9.974)*	0.2117 (8.916)*	1.793 (10.037)*	-0.0554 (-0.37)	0.7893 (7.975)*	290.3451

Expenditure head: Secondary education
Dependent variable: Per capita expenditure on secondary education

Model		Explanatory Variables					Test Statistics
	Constant	SENROL	LITERACY	SSTRATIO	PRICEDIF	SCSTSEC	LR
S_2R_2	-3.069 (-4.605)*	0.1444 (2.475)*	0.7618 (4.35)*	0.0304 (0.347)	0.9498 (9.418)*	-0.0065 (-0.207)	315.0444

Expenditure head: Higher education
Dependent variable: Per capita expenditure on higher education

Model		Explanatory Variables				Test Statistics
	Constant	HENROL	LITERACY	HSTRATIO	PRICEDIF	LR
S_2R_2	-11.316 (-19.442)*	-0.1921 (-3.936)*	1.471 (18.908)*	0.1453 (2.566)*	0.6052 (6.586)*	61.63

Notes: Figures in parentheses refer to t-ratios.
* Significant at 1 per cent level.
For degrees of freedom 16, the critical values at 0.99 and 0.95 levels are 32.00 and 26.30 respectively.

In contrast to the insignificant relationship observed in the case of primary and secondary education, the student-teacher ratio for higher education bears a significant direct relationship with expenditure per beneficiary group. Such a relationship may suggest an overcrowded situation, where the number of students per teacher is already very high. Thus, any further increase in the number of students would require appointment of new teachers in higher education. The diagnostic LR statistic reported for all the regressions rejects the null hypothesis of heteroskedasticity and cross-group correlation.

To facilitate meaningful comparison, actual expenditure levels are expressed as a proportion of the normative expenditure levels for 15 large Indian states with regard to primary, secondary and higher education. The figures are presented in Table 3. Gujarat and Maharashtra emerge as champions for the cause of primary education since, for these states, the actual expenditure levels are much higher than the normative expenditure levels. On the contrary, Bihar, Haryana, Uttar Pradesh and West Bengal lag far behind the normative expenditure levels in terms of their actual expenditure levels.

In respect of secondary education, Andhra Pradesh, Punjab and West Bengal spent significantly more than the normative levels, whereas Assam, Bihar, Madhya Pradesh, Orissa and Uttar Pradesh spent much less than the normative levels. For higher education, the states that spent significantly more than the normative levels are Andhra Pradesh, Haryana, Kerala and Punjab. Similarly, the states that spent significantly less than the normative levels on higher education are Assam and Uttar Pradesh. Thus, our findings are consistent with the generally maintained hypothesis that rich states spend more and poor states spend less as far as social sectors are concerned. The lone exception being Orissa, where the expenditure on higher education is significantly higher than the normative level.

Policy Implications

Utilising the pooled data for 15 large Indian states over six years, from 1992-93 to 1997-98, this study attempts to estimate the normative expenditure levels for providing a standardised unit of service with regard to three disaggregated expenditure heads, viz, primary education, secondary education and higher education. The findings of this study are consistent with the generally maintained hypothesis that 'rich states spend more and poor states spend less as far as social sectors are concerned'. Gujarat and Maharashtra emerge as champions of the cause of primary education, whereas Andhra Pradesh and Punjab spend substantially more on secondary and higher education. Poor states like Uttar Pradesh and Bihar lag far behind the normative levels for all the three expenditure heads considered in this study.

The findings also suggest existence of unutilised capacity in both primary and higher education. This, to a great extent, can be attributed to the shift in preference from government education to private education. Thus, a policy implication that might emerge here is that any efforts to rationalise the utilisation of existing capacities must involve improvement of the quality of education in the government sector. Furthermore, actual spending on educational services in low-income states is found to be lower than their 'needs'. This implies that the existing fiscal equalisation mechanism is unable to offset the revenue and cost disabilities of poorer states. Hence, there is justification for suitable changes in the existing equalisation mechanism. EPW

Appendix I: Pooled Regression Models

A pooled regression can be represented as:

$$Y_{it} = \alpha + X_{it}\beta + \varepsilon_{it} \qquad i = 1,2,\ldots,N. \quad t = 1,2,\ldots,T.$$

where i denotes the cross-section and t denotes the time dimensions, α is a scalar, β is $K \times 1$ vector of coefficients and X_{it} is the it[th] observation on K[th] explanatory variable, and ε_{it} is classical disturbance term which is assumed to be homoskedastic and non-autocorrelated, that is:

$E[\varepsilon_{it}] = 0$, and $Var[\varepsilon_{it}] = \sigma\varepsilon^2$

Violation of the assumptions of homoskedasticity and autocorrelated errors releases three different possibilities. They are:

(i) Groupwise heteroskedasticity, $E[\varepsilon_{it}^2] = \sigma_{ii}$

(ii) Cross-group correlation, $Cov[\varepsilon_{it}, \varepsilon_{jt}] = \sigma_{ij}$

Appendix II: Description of Variables

Symbol	Variable	Definition	Source
Primary	Per capita expenditure on Primary Education	Expenditure on primary education per child in the age group 5-9.	Finance Accounts of state government, CAG government of India.
Secondary	Per capita expenditure on secondary education	Expenditure on secondary education per child in the age group 10-18.	"
Higher	Per capita expenditure on higher education	Expenditure on higher education per youth in the age group 20-24.	"
PENROL	Enrolment in primary education	umber of students in primary education per 1000 children in the age group 5-9.	Selected Educational Statistics, Ministry of HRD, Department of Education, Government of India.
PDROPOUT	Proportion of dropout in primary education		"
PSTRATIO	Student Teacher ratio in primary education	Number of students per teacher in primary education	"
SENROL	Enrolment in secondary education	Number of students in secondary education per 1000 children in the age group 10-18	"
SSTRATIO	Student teacher ratio in secondary education	Number of students per teacher in secondary education	"
HENROL	Enrolment in higher education	Number of students in higher education per 1000 youth in the age group 20-24	"
HSTRATIO	Student-teacher ratio in higher education	Number of students per teacher in higher education	"
LITERACY	Literacy rate	Proportion of literacy in the states	Office of the Registrar General, Government of India.
PRICEDIF	Price Differential across the states*		

Note: * The Ninth Finance Commission had estimated the index of price differences based on the information on prices of different commodities collected in various urban centres by the Labour Bureau, Simla. Price differences between the states used in this exercise are obtained by applying the changes in prices recorded in different states (averaged for different centres) on the base year differences of 1986-87.

(iii) Within group autocorrelation,

$\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + u_{it}$

With regard to homoskedasticity, three important cases may be considered:

S_0: It refers to the most naïve situation of classical homoskedastic regression, where: $\Sigma = \sigma^2 I$, where I is an identity Matrix and Σ is the error variance-covariance matrix.

S_1:It stands for groupwise heteroskedasticity, and thus

Σ = diag $[\sigma_{11}, \sigma_{22}, \ldots \sigma_{NN}]$

S_2:This is a general case, where, Σ is assumed to be a positive definite matrix signifying, apart from groupwise heteroskedasticity, cross group correlated errors. That is,

Σ = N x N positive definite matrix.

In respect of autocorrelation, a similar set of assumptions may be considered. They are:

R_O : $\rho = 0$, representing a situation where disturbances are non-autocorrelated, ρ being an $N \times 1$ vector of group-specific autocorrelation coefficients.

R_1: $\rho = (\rho, \rho, \ldots, \rho)$, where all the units have common autocorrelation coefficient.

R_2: $\rho = (\rho_1, \rho_2, \ldots, \rho_N)$, where individual units have specific coefficients.

By combining these two sets of restrictions on Σ and ρ, it is possible to generate nine combinations leading to nine distinct models. They may be represented by (S_oR_o), (S_1R_0), (S_2R_0), (S_oR_1), (S_1R_1), (S_2R_1), (S_oR_2), (S_1R_2), (S_2R_2). The first eight models form as special cases of the last one, by imposing appropriate restrictions on S_2R_2.

For non-autocorrelated models, the estimator is a two-step generalised least squares (GLS) and for models with autocorrelation the required estimator is a three-step GLS. For testing the assumptions of homoskedasticity and absence of cross-group error-correlation as restrictions on the most general case of errors being heteroskedastic and cross-group correlated, three diagnostic test statistics are computed, namely, LM, LR and Wald. The test statistics are given by:

$$LM = (T/2) - \Sigma_i \, [S_{ii}/S^2 - 1]^2$$
$$LR = T \, (N \ln S^2 - \Sigma_i \ln S_{ii})$$
$$Wald = (T/2) - \Sigma_i \, [S^2/ S_{ii} - 1]^2$$

where S^2 is the pooled OLS residual variance and S_{ii} is the OLS residual variance of the ith state. All the three test statistics follow the Chi-squared distribution with (N + R −1) degrees of freedom, where N is the number of units in the panel (in our case the number of states), R is the number of restrictions imposed on the estimated models.

In this exercise, for each behavioural specification corresponding to expenditure heads, we have estimated all the nine distinct models ranging from the most restricted case of S_0R_0 to the most general case of S_2R_2. Out of the nine distinct models only one model is chosen. Since statistical significance of the individual coefficients undergo considerable variation over the range of models estimated (from S_0R_0 to S_2R_2), the final model selection criteria rest upon a judicious compromise between the model diagnostic statistics and the desired statistical properties of the individual coefficients.

References

Baltagi, B H (1995): *Econometric Analysis of Panel Data*, John Wiley and Sons.

Dielman, T E (1989): *Pooled Cross-Sectional and Time Series Data Analysis*, Marcel Dekker, New York.

Rao, M G and V Aggarwal (1992): 'Intergovernmental Fiscal Transfers in India: Some Issues of Design and Measurement', *Indian Economic Review*, Vol 26, pp 13-34.

Rao, M G (1993): 'Report of the Ninth Finance Commission: Some Conceptual and Methodological Comments', *Economic and Political Weekly*, Vol XXV, 9, June 9, pp 1275-80.