ESTIMATING WEALTH EFFECTS WITHOUT EXPENDITURE DATA—OR TEARS: AN APPLICATION TO EDUCATIONAL ENROLLMENTS IN STATES OF INDIA*

DEON FILMER AND LANT H. PRITCHETT

Using data from India, we estimate the relationship between household wealth and children's school enrollment. We proxy wealth by constructing a linear index from asset ownership indicators, using principal-components analysis to derive weights. In Indian data this index is robust to the assets included, and produces internally coherent results. State-level results correspond well to independent data on per capita output and poverty. To validate the method and to show that the asset index predicts enrollments as accurately as expenditures, or more so, we use data sets from Indonesia, Pakistan, and Nepal that contain information on both expenditures and assets. The results show large, variable wealth gaps in children's enrollment across Indian states. On average a "rich" child is 31 percentage points more likely to be enrolled than a "poor" child, but this gap varies from only 4.6 percentage points in Kerala to 38.2 in Uttar Pradesh and 42.6 in Bihar.

This paper has both an empirical and a methodological goal. The empirical goal is to investigate the effect of household economic status on children's educational attainment across the states of India. To accomplish this aim, we propose and defend a method for estimating the effect of household wealth on educational outcomes even in the absence of survey questions on income or expenditures. We use data on asset ownership (e.g., owning a bicycle or radio) and housing characteristics (e.g., number of rooms, type of toilet facilities), henceforth called "asset indicators" or "asset variables," to construct an "asset index." We handle the vexing problem of choosing the appropriate weights by using the statistical procedure of principal components. We demonstrate the empirical validity of this approach for India by comparing the state-level averages of the asset index with data on poverty rates and gross state product per capita. Going further, we use data sets from Indonesia, Nepal, and Pakistan, which contain both expenditures and asset variables for the same households, to show a reasonable correspondence between the classification of households based on the asset index and a classification based on consumption expenditures. The econometric evidence suggests that the asset index, as a proxy of economic status for use in predicting enrollments, is at least as reliable as conventionally measured consumption expenditures, and sometimes more so.

This straightforward and pragmatic method of constructing a proxy for household economic status produces results that are reassuringly consistent with other approaches (see Montgomery et al. 2000 and references therein) and potentially is broadly applicable. Demographic and Health Surveys (DHS) have been conducted with nearly identical survey instruments in more than 40 countries. (In India this survey is called the National Family Health Survey, or NFHS.) These surveys include assets and housing conditions but not consumption expenditures (except for an experimental consumption module included in Indonesia in 1994, a feature that we exploit later in this paper).

Our proposed method for estimating household wealth in the DHS/NFHS allows estimates of the association of wealth with education across households and permits a comparison of wealth gaps across countries (and in India across states). In separate papers we use the asset index to examine wealth and gender gaps in enrollment and educational attainment in more than 35 countries (Filmer 2000; Filmer and Pritchett 1999a).

We limit ourselves here to demonstrating the validity and usefulness of an asset index in analyzing education outcomes. The same method might be equally useful in examining wealth differences in other socioeconomic outcomes in the DHS data, such as mortality, morbidity, utilization of health facilities, fertility, and contraceptive use (Bonilla-Chacin and Hammer 1999; Gwatkin et al. 2000; Stecklov, Bommier, and Boerma 1999). Sahn and Stifel (1999) have taken a similar "index" approach to making comparisons of relative poverty over time and across countries, using DHS data for nine countries in Africa.

A proxy for wealth not only is useful in examining effects of wealth, but also is needed as a "control" variable in estimating effects of variables potentially correlated with household wealth, such as maternal education. The method proposed here provides a simple technique for creating a wealth proxy in the absence of either income or expenditure data (an issue discussed further in Montgomery et al. 2000).

OVERCOMING THE PROBLEM OF RANKING HOUSEHOLDS IN THE ABSENCE OF CONSUMPTION OR EXPENDITURE DATA

The DHS and NFHS data present both an opportunity and a challenge. The opportunity is a rich set of large, representative

*Deon Filmer, Development Research Group, The World Bank, 1818 H Street NW, Washington, DC 20433; E-mail: dfilmer@worldbank.org. Lant H. Pritchett, John F. Kennedy School of Government and The World Bank. This research was funded in part through World Bank research support grant (RPO 682-11). We would like to thank Harold Alderman, Zoubida Allaoua, Gunnar Eskeland, Jeffrey Hammer, Keith Hinchliffe, Valerie Kozel, Alan Krueger, Peter Lanjouw, Marlaine Lockheed, Berk Ozler, and Martin Ravallion for valuable discussions and comments, including some on an earlier version of this paper (Filmer and Pritchett 1998). The findings, interpretations, and conclusions expressed here are entirely those of the authors. They do not necessarily represent the views of the World Bank, its executive directors, or the countries they represent.

surveys with nearly identical questionnaires covering a large number of countries or Indian states. The challenge is that the DHS/NFHS surveys contain no data on income or on household consumption expenditures. The latter is unfortunate because a large literature establishes the theoretical underpinnings of consumption expenditures as a measure of current and long-run household (and implicitly individual) welfare. (For example, see Deaton 1997; Deaton and Muellbauer 1980; Deaton and Zaidi 1999.) As a result, expenditures are routinely used in measuring poverty. We overcome the absence of expenditure data by using the information collected on assets owned by household members and on housing characteristics. These data are used to generate an asset index that proxies for wealth and hence for long-run economic status.

Three clarifications will avoid confusion at the outset. First, we are not proposing this asset index as a measure either of *current* welfare or of poverty.[1] We use the asset index as a determinant of current enrollment, which depends on long-run as well as current expenditures: households will smooth schooling expenditures over time and are unlikely to respond to temporary shocks by withdrawing children from school.

Second, unlike Montgomery et al. (2000), we are not creating an asset index as a proxy for current consumption expenditures. We view the asset index and expenditures as a proxy for something unobserved: a household's long-run economic status. Therefore, although it is reassuring that these two proxies are related, discrepancies in the classification of households cannot be assumed to be "mistakes" of the asset index. The two, as proxies, have conceptually and empirically distinct limitations; therefore discrepancies could just as easily indicate limitations of current consumption expenditures. The major problem with the asset index is that the weights on individual indicators are not grounded theoretically. The major problem with current expenditures as a proxy for long-run wealth is the presence of short-term fluctuations.[2]

Third, we emphasize that the principal-components approach is a pragmatic response to a data constraint problem. In this paper we set the modest goal of establishing the validity of this particular approach in this application; we do not establish that our approach is "optimal" because there may be other methods that possess superior statistical properties.

1. The limitation in poverty analysis is twofold. First, the conventional notion of poverty is based on the *flow* of consumption relative to some predetermined poverty threshold, whereas we, by aggregating assets, are establishing only a measure of a stock. Second, the categorization used is based on a relative measure (that is, the household's ranking within the distribution), whereas poverty thresholds typically are based on the expenditures necessary for the consumption of a determined bundle of goods.

2. Current expenditures would be a perfect measure only under the unrealistically strong assumptions of perfect foresight and perfect capital markets. Current expenditures are a popular proxy for long-run wealth for two main reasons: the theoretical justification that expenditures are superior to current income as a proxy for long-run income because of consumption smoothing, and, perhaps even more important, the pragmatic justification that expenditures are easier to measure than income in most rural settings. Current expenditures, however, are not preferred unambiguously to asset ownership on either of these dimensions: asset ownership also reflects smoothing and is (if anything) easier to measure than either income or expenditures.

Construction of an Asset Index

How does one aggregate various asset ownership indicators into one variable to proxy for household "wealth"? Even if the question is simplified by limiting the aggregation to a linear index, how should the weights be chosen? Equal weights have the appeal of simplicity and apparent objectivity, but these qualities only mask the fact that the imposition of numeric equality is completely arbitrary.

A second option would be to estimate the current value of household assets using explicit and implicit "prices" as weights. If the data did not include current values, then the purchase price, the date of purchase, and a suitable depreciation rate could be used to estimate them. The DHS and NFHS data, however, are typical in that they include binary indicators on asset ownership and housing characteristics but not current value, purchase price, or vintage of assets.

A third possible solution is not to build an index at all but simply to enter all of the asset variables separately in a linear multivariate regression equation. This procedure implicitly creates weights on the variables. Such an approach handles the problem of controlling for wealth in estimating the impact of other, nonwealth, variables. Yet the linear index of the assets using regression weights does not estimate the wealth effect because many assets exert both a direct and an indirect effect on outcomes. For instance, the household's use of electricity for lighting may serve as a proxy for wealth, but also may make study easier and hence may lower the opportunity costs of schooling. The availability of piped water not only may indicate greater wealth but also may reduce the time needed for water collection and thus may reduce the opportunity costs of schooling. This argument is even clearer with health outcomes because water and sanitation variables have strong independent effects on children's health (Bonilla-Chacin and Hammer 1999). Therefore, although linear regression coefficients implicitly produce weights for the linear index of the asset variables that predicts the dependent variable most closely, there is no way to infer from these unconstrained coefficients the impact of an increase in wealth.

Using Principal Components

We implement a different approach: we use the statistical procedure of principal components to determine the weights for an index of the asset variables. Principal components is a technique for extracting from a set of variables those few orthogonal linear combinations of the variables that capture the common information most successfully. Intuitively the first principal component of a set of variables is the linear index of all the variables that captures the largest amount of information that is common to all of the variables. (For readable and intuitive descriptions of principal components, see Lindeman, Merenda, and Gold 1980; StataCorp 1999.)

Suppose we have a set of N variables, a^*_{1j} to a^*_{Nj}, representing the ownership of N assets by each household j. Principal components starts by specifying each variable normalized by its mean and standard deviation: for example, $a_{1j} = (a^*_{1j} - a^*_1) / (s^*_1)$, where a^*_1 is the mean of a^*_{1j} across households and s^*_1 is its standard deviation. These selected variables are expressed as linear combinations of a set of underlying components for each household j:

$$\begin{aligned} a_{1j} &= v_{11} \times A_{1j} + v_{12} \times A_{2j} + \ldots + v_{1N} \times A_{Nj} \\ &\ldots \qquad\qquad\qquad\qquad\qquad\qquad j = 1,\ldots J \\ a_{Nj} &= v_{N1} \times A_{1j} + v_{N2} \times A_{2j} + \ldots + v_{NN} \times A_{Nj}\,, \end{aligned} \quad (1)$$

where the As are the components and the vs are the coefficients on each component for each variable (and do not vary across households). Because only the left-hand side of each line is observed, the solution to the problem is indeterminate.

Principal components overcomes this indeterminacy by finding the linear combination of the variables with maximum variance—the first principal component A_{1j}— and then finding a second linear combination of the variables, orthogonal to the first, with maximal remaining variance, and so on. Technically the procedure solves the equations $(\mathbf{R} - \lambda_n \mathbf{I})\mathbf{v}_n = 0$ for λ_n and $\mathbf{v}_n$, where $\mathbf{R}$ is the matrix of correlations between the scaled variables (the as) and $\mathbf{v}_n$ is the vector of coefficients on the nth component for each variable. Solving the equation yields the characteristic roots of $\mathbf{R}$, λ_n (also known as eigenvalues) and their associated eigenvectors, $\mathbf{v}_n$. The final set of estimates is produced by scaling the $\mathbf{v}_n$s so the sum of their squares sums to the total variance, another restriction imposed to achieve determinacy of the problem.

The "scoring factors" from the model are recovered by inverting the system implied by Eq. (1), and yield a set of estimates for each of the N principal components:

$$\begin{aligned} A_{1j} &= f_{11} \times a_{1j} + f_{12} \times a_{2j} + \ldots + f_{1N} \times a_{Nj} \\ &\ldots \qquad\qquad\qquad\qquad\qquad\qquad j = 1,\ldots J \\ A_{Nj} &= f_{N1} \times a_{1j} + f_{N2} \times a_{2j} + \ldots + f_{NN} \times a_{Nj}. \end{aligned} \quad (2)$$

The first principal component, expressed in terms of the original (unnormalized) variables, is therefore an index for each household based on the expression

$$A_{1j} = f_{11} \times (a^*_{1j} - a^*_1)/(s^*_1) + \ldots + f_{1N} \times (a^*_{Nj} - a^*_N) / (s^*_N). \quad (3)$$

The crucial assumption for our analysis (and it is just that—an assumption) is that household long-run wealth explains the maximum variance (and covariance) in the asset variables. There is no way to test this assumption directly, but in the next two sections we provide evidence that this method produces reasonable results. Below we illustrate the method using the states of India, and show the internal and external coherence of the asset index. Then we compare this the method with the use of consumption expenditures in three countries; we find that this simple asset index has reasonable agreement with consumption expenditures and performs as well, or better, in predicting education outcomes in two of the countries.

CONSTRUCTION AND INTERNAL VALIDITY OF THE ASSET INDEX IN INDIA

Construction of the Asset Index Using the NFHS Data

The NFHS survey covers about 88,000 households and about one-half million individuals. The number of households surveyed in each Indian state varies from 9,963 in Uttar Pradesh to around 1,000 in the small northeastern states. The survey includes data on 21 asset indicators that can be grouped into three types: household ownership of consumer durables, with eight questions (clock/watch, bicycle, radio, television, bicycle, sewing machine, refrigerator, car); characteristics of the household's dwelling, with 12 indicators (three about toilet facilities, three about the source of drinking water, two about rooms in the dwelling, two about the building materials used, and one each about the main source of lighting and cooking); and household landownership.

Table 1 reports the scoring factors from the principal-components analysis of the 21 variables (see Eq. (3)). The mean value of the index is 0; the standard deviation is 2.3. Because all the asset variables (except "number of rooms") take only the values 0 or 1, the weights have an easy interpretation: a move from 0 to 1 changes the index by f_{1i} / s^*_i (reported in column 4). A household that owns a clock has an asset index higher by 0.54 than one that does not; owning a car raises a household's asset index by 1.21 units; using biomass as the main cooking fuel lowers the asset index by 0.67.

We sort individuals by the asset index and establish cutoff values for percentiles of the population. We then assign households to a group on the basis of their value on the index. For expository convenience, we refer to the bottom 40% as "poor," the next 40% as "middle," and the top 20% as "rich," asking the reader to remember that this classification does not follow any of the usual definitions of poverty.

The difference in the average index between the poorest and the middle group is 2.07 units. One example of a combination of assets that would produce this difference is owning a radio (0.54), having a kitchen as a separate room (0.37), having electricity for lighting (0.57), and not having a dwelling of all low-quality materials (0.55). The average asset index is 3.78 units higher for the richest than for the middle group. This difference is equivalent to owning a motor scooter (0.91) and a television (0.83), having a flush toilet (0.75) and a house of all high-quality materials (0.73), and not using biomass as the main cooking fuel (0.67).

The Reliability of the Asset Index

For India the asset index performs well on three dimensions. First, it is internally coherent because average asset ownership differs markedly across the poor, middle, and rich households for each asset; second, it is robust to the assets included; third, it produces reasonable comparisons with measures of state-level poverty and output. The index has drawbacks, however, especially possible problems with urban/rural comparisons.

Internal coherence. The last three columns of Table 1 compare the average ownership of each asset across the poor, middle, and rich households. We find large differences across groups for almost all assets. Clock ownership is 16% for the poor versus 98% for the rich. Also, the poor cook with biomass fuel almost exclusively (96%), whereas only

TABLE 1. SCORING FACTORS AND SUMMARY STATISTICS FOR VARIABLES ENTERING THE COMPUTATION OF THE FIRST PRINCIPAL COMPONENT

	All India				Means		
	Scoring Factors	Mean	*SD*	Scoring Factor × *SD*	Poorest 40%	Middle 40%	Richest 20%
Own Clock/Watch	0.270	0.533	0.499	0.54	0.164	0.739	0.985
Own Bicycle	0.130	0.423	0.494	0.26	0.264	0.510	0.621
Own Radio	0.248	0.396	0.489	0.51	0.101	0.522	0.838
Own Television	0.339	0.209	0.407	0.83	0.000	0.127	0.866
Own Sewing Machine	0.253	0.182	0.385	0.66	0.015	0.179	0.580
Own Motorcycle/Scooter	0.249	0.082	0.274	0.91	0.001	0.031	0.375
Own Refrigerator	0.261	0.068	0.252	1.04	0.000	0.006	0.353
Own Car	0.129	0.012	0.107	1.21	0.000	0.001	0.059
Drinking Water From Pump/Well	–0.192	0.609	0.488	–0.39	0.800	0.569	0.242
Drinking Water From Open Source	–0.041	0.040	0.195	–0.21	0.057	0.036	0.005
Drinking Water From Other Source	–0.002	0.019	0.138	–0.01	0.016	0.027	0.012
Flush Toilet	0.308	0.217	0.412	0.75	0.005	0.175	0.797
Pit Toilet/Latrine	0.040	0.086	0.280	0.14	0.040	0.127	0.111
None/Other Toilet	0.001	0.001	0.029	0.03	0.001	0.001	0.001
Main Source of Lighting Electric	0.284	0.510	0.500	0.57	0.143	0.700	0.989
Number of Rooms in Dwelling	0.159	2.676	1.957	0.08	1.975	2.965	3.739
Kitchen a Separate Room	0.183	0.536	0.499	0.37	0.312	0.643	0.848
Main Cooking Fuel Biomass (Wood/Dung/Coal)	–0.281	0.776	0.417	–0.67	0.956	0.841	0.224
Dwelling All High-Quality Materials	0.309	0.237	0.425	0.73	0.005	0.218	0.821
Dwelling All Low-Quality Materials	–0.273	0.483	0.500	–0.55	0.832	0.308	0.017
Own > 6 Acres Land	0.031	0.115	0.319	0.10	0.075	0.155	0.126
Economic Status Index		0.000	2.32		–2.00	0.071	3.857

Notes: Each variable beside number of rooms takes the value 1 if true, 0 otherwise. Scoring factor is the "weight" assigned to each variable (normalized by its mean and standard deviation) in the linear combination of the variables that constitute the first principal component. The percentage of the covariance explained by the first principal component is 26%. The first eigenvalue is 5.37; the second eigenvalue is 1.66.

Source: Authors' calculation from NFHS 1992–1993.

22% of the rich do so. One might ask whether the asset index tends too much to reflect community variables (especially locally available infrastructure such as electricity for lighting or piped water) rather than household-specific variables. On this score, we are reassured by the clear separation between poor and rich on housing variables not related to infrastructure, such as "all high-quality materials in the dwelling" (less than 1% of the poor versus 82% of the rich) and "having a kitchen as a separate room" (31% of the poor versus 85% of the rich).

TABLE 2. CLASSIFICATION DIFFERENCES OF THE POOREST 40%: ALL INDIA

	Base Case: All Variables	All Variables Except Drinking Water and Toilet Facilities	Asset Ownership, Housing, and Land Ownership	Only 8 Asset Ownership Variables
Poorest 40%	100.0	95.1	87.7	80.2
Middle 40%	0.0	4.9	12.3	19.7
Richest 20%	0.0	0.0	0.0	0.1
Total	100.0	100.0	100.0	100.0
Spearman Rank Correlation Coefficient, Ranking of Households	1.0	.94	.87	.79

Source: Authors' calculation from NFHS 1992–1993.

Robustness. The asset index produces very similar classifications when different subsets of variables are used in its construction. Table 2 reports the percentage of households classified in the poorest 40% when all assets are used, compared with indices based on (1) all the variables except those related to drinking water and toilet facilities, (2) ownership of consumer durables, housing quality, number of rooms, and land ownership, and (3) only the ownership of consumer durables (e.g., watch, radio). Almost no households classified in the poorest group by the index using all variables would be classified as rich by any of the more limited measures. The robustness of the classification is similar for the middle and the rich groups.

A more general measure of the differences in rankings can be derived from the rank correlation coefficient, which compares the degree to which two methods produce the same ranking of households. Even when the index is constructed with only consumer durables, the correlation with the index that uses all the variables is .79 (all correlation coefficients in Table 2 are significant; $p < .001$, $N = 87{,}175$). Adding more variables in constructing the index only increases the similarity of the rankings.

An additional check for robustness is made by using a different methodology for deriving the weights. Although the theoretical underpinnings and the algorithms used in factor analysis are close to those for principal components, the two methodologies differ sufficiently to make factor analysis a possible alternative approach. The first factor derived from a model analogous to that described above yields a household ranking that has a .988 Spearman rank correlation with a ranking derived from principal components. Clearly, the results drawn from the asset index approach are robust to whether one picks one or the other of these methods.

Comparisons across states. Because the poor, middle, and rich groups are defined on an "all-India" basis, states differ as to the percentage of households in each group. Thus we can compare state-by-state rankings with more conventional measures. The national poverty rate based on consumption expenditures is 36%, roughly comparable to defining the "asset poor" as the poorest 40% nationally by the asset index. The rank correlation of the poverty rate with the proportion asset poor across states is .794 ($p < .001$, $N = 16$). In a comparison of the first and the second columns of Table 3, both classifications show that Punjab, Haryana, and Kerala have better than average economic status and that Bihar, Orissa, and Uttar Pradesh are below average. Differences exist, however: Maharashtra appears richer by the asset index, while Andhra Pradesh looks poorer. The rank correlation of the proportion of a state's population asset poor with SDP per capita is –.864 ($p < .001$, $N = 18$). (By comparison, the rank correlation between the poverty rate and per capita state domestic product is –.729, $p = .002$, $N = 15$). Again, although the rankings agree overall, certain states look different according to the two rankings. For example, Kerala appears richer by the index, with only 15% asset poor and a per capita SDP of Rs 5,065, whereas Assam looks poorer, with 58% asset poor but a per capita SDP of Rs 5,056.

TABLE 3. DISTRIBUTION OF INDIVIDUALS ACROSS GROUPS, STATE-LEVEL POVERTY, AND NET DOMESTIC PRODUCT

	Percentage in Nationwide Poorest 40%	State Poverty Rate (Headcount Index)	Per Capita Net State Domestic Product
Delhi	1.3		
Goa	5.6		10,128
Himachal Pradesh	6.8	28.58	
Punjab	8.4	11.46	10,857
Haryana	10.5	25.22	9,609
Jammu	14.5		
Kerala	15.1	25.12	5,065
Mizoram	18.1		
Nagaland	20.3		
Gujarat	26.8	24.15	7,586
Maharashtra	26.9	36.82	9,270
Karnataka	27.6	32.91	6,313
Manipur	27.6		
Tamil Nadu	32.5	35.40	6,205
Meghalaya	37.9		5,769
Arunachal Pradesh	38.1		6,359
Andhra Pradesh	39.0	21.87	5,802
Rajasthan	39.7	27.46	5,035
Tripura	41.8		
West Bengal	44.3	36.94	5,901
Uttar Pradesh	48.6	41.55	4,280
Madhya Pradesh	49.4	42.46	4,725
Orissa	54.4	48.64	3,963
Assam	58.3	41.09	5,056
Bihar	61.5	55.15	3,280
All India	40.0	36.16	6,380

Notes: The rank correlation coefficient between the percentage asset poor and the poverty rate is .794 ($p < .001$, $N = 16$); the rank correlation between the percentage asset poor and per capita state product is –.864 ($p < .001$, $N = 18$).

Sources: Authors' calculation from NFHS 1992–1993, World Bank (1998), and Agrawal and Varma (1996). Data on the headcount index are for 1993–1994.

Alternative interpretations. The first principal component explains 25.6% of the variation in the 21 asset variables; this percentage is substantial but not overwhelming. Although the first principal component might well serve as a reasonable overall index, it is uncertain whether this component alone contains all of the relevant information. The factor scores derived for the second principal component are more difficult to interpret. The procedure produces a pattern of factor scores, which appears to assign positive weights to assets that one might think richer rural households would own, and negative weights to assets that one might expect other types of households to own. (See Appendix A, available from the authors.) This is somewhat worrisome because the asset index rankings show rural households to be less "wealthy" than do conventional expenditure measures.

One explanation for this discrepancy is that many of the asset variables depend on the availability of infrastructure (electricity, piped water, sewerage); therefore urban households are more likely than rural households to appear well-off. On the other hand, this point may imply that standard expenditure measures underestimate the difference between rural and urban households' welfare by not adequately adjusting real incomes for the implicit price differences for the services provided by infrastructure. For the analysis of enrollment decisions, however, we want an index that captures the dimensions of wealth relevant to education, which is not necessarily net welfare.

Finally, in this particular application, nothing in our analyses depends on urban/rural comparisons. The bivariate analysis uses the pooled urban/rural data sets only to analyze the stability of rankings within the entire samples. The subsequent multivariate analysis either uses only the rural data or, when data are pooled, controls for rural/urban status. Thus the analysis should not be affected by any level difference due to systematic over- or understatement of the differences. (We refrain from interpreting the "urban" coefficient because it will combine the "true" effect of urban residence with any effect of the construction of the index.)

Even though the procedure produces higher-order components, they are, by construction, orthogonal to the first and therefore will not create direct "excluded variable" bias in bivariate comparisons of wealth and education. In the multivariate analysis discussed below, the results are robust to including higher-order components as linear indices in the specification.

ASSET INDEX VERSUS CONSUMPTION EXPENDITURES AS PROXIES FOR LONG-RUN WEALTH

Comparisons of Consumption Expenditures With Asset Index Classifications

The Nepal Living Standards Survey (NLSS) conducted in 1996 and the Pakistan Integrated Household Survey (PIHS) conducted in 1991 are "standard" surveys from the Living Standards Measurement Study (LSMS) (Grosh and Glewwe 1998). The Indonesian DHS (IDHS) conducted in 1994 included an experimental module on consumption expenditures (based closely on Indonesia's National Socio-Economic Survey—SUSENAS) for about half of the households.

In these three countries we compare an asset index with total consumption expenditures (C) adjusted for household size (N), C/N^{α}, where the economies of scale parameter α is set to 0.6 (For discussions of the choice of this parameter, see Dreze and Srinivasan 1997; Lanjouw and Ravallion 1995). The asset index is not adjusted for household size because the benefits of many of the assets, such as quality of housing materials, source of fuel, or lighting, are present at the household level. The Spearman rank correlations across households are .64 for Nepal ($p < .001$, $N = 3{,}372$), .56 for Indonesia ($p < .001$, $N = 16{,}242$), and .43 for Pakistan ($p < .001$, $N = 1{,}192$). Clearly, the degree of agreement among the different rankings varies across the countries. Generally, the smaller the α, the better the fit between assets and expenditure classifications; thus the asset index classification fits total household expenditures better than is reported and fits per capita expenditures worse than is reported. At the suggestion of a referee, we repeated the analysis using C/A^{α}: that is, adjusting for the number of adults in the household. The results differ neither qualitatively nor, to any substantial degree, quantitatively.[3]

We assigned households to the poorest 40, middle 40, and richest 20 percentiles, using either the asset index or the expenditure measure. Table 4 shows the results of comparing the two classifications. The results in Indonesia and Nepal are quite similar: roughly two-thirds of those classified into the poorest 40% by expenditures are also classified into the poorest 40% by assets, and only 5% of those in the poorest 40% by expenditures appear in the richest 20% by assets. The classification of the richest 20% shows less agreement: only 49 to 56% of households rich by expenditures are also in the richest 20% by assets. Reassuringly, however, only 10 to 13% of those ranked in the richest 20% by expenditures are in the poorest 40% by assets.

The results for Pakistan show less coherence between the two rankings. Although it is still the case that only 4% of those who are poor by expenditures are rich by assets, only 60% of the expenditure poor are also asset poor. Moreover, only 43% of those in the richest 20% by expenditures are also in the richest 20% by assets, and 22% of the *richest* 20% of households by expenditures are in the *poorest* 40% by assets.

Comparison of Enrollment Rate Regressions Using the Two Measures

Table 5 reports the results of probit regressions of "currently enrolled" and "ever attended" for children age 6 to 14, and "completed grade 5" for those age 15 to 19. The regressions control for child characteristics, residence, and characteristics of the household head—a specification that mimics the one used in the India analysis conducted below. The marginal wealth effect reported is the increase in the probability that the education variable will equal 1 for a child in the rich-

3. We compare the asset index with consumption expenditures and not with *predicted* consumption expenditures, where assets and other household variables are used as instruments. Although some of the results would have appeared more similar if we had taken this "best practice" approach (recently used and explored in Behrman and Knowles 1999), the conventional approach—particularly for bivariate/tabular analysis—is to use actual, not predicted, expenditures. Therefore we use this as our baseline for comparison.

TABLE 4. CLASSIFICATION DIFFERENCES: NEPAL, INDONESIA, AND PAKISTAN, URBAN AND RURAL AREAS

Groups Based on Asset Index	Groups Based on Household Consumption Expenditures per Adjusted Size[a]: Poorest 40%	Middle 40%	Richest 20%
Nepal			
Poorest 40%	65.2	37.8	12.6
Middle 40%	29.9	46.8	31.4
Richest 20%	4.9	15.4	56.0
Total	100.0	100.0	100.0
Indonesia			
Poorest 40%	63.9	35.3	10.5
Middle 40%	31.7	49.1	41.0
Richest 20%	4.4	15.6	48.5
Total	100.0	100.0	100.0
Pakistan			
Poorest 40%	61.2	40.0	20.2
Middle 40%	34.9	42.5	37.1
Richest 20%	3.9	17.5	42.7
Total	100.0	100.0	100.0

Sources: Authors' calculations from NLSS 1996, IDHS 1994, and PIHS 1991.

[a]Adjusted household size is equal to household size to the power 0.6.

est versus the poorest quintile when all other variables are at their mean.

The results for current attendance in Nepal are almost identical for the two proxies: the marginal effect of being in the richest relative to the poorest quintile by assets is 34.0 percentage points, versus 33.8 percentage points by expenditures. In contrast, the marginal effect in Pakistan is 36.7 percentage points when quintiles are defined according to assets, but only 27.5 percentage points when the groups are defined according to expenditures. The Indonesian results fall between the other two: the wealth gap in current enrollment is 10.5 percentage points when quintiles are based on assets and 8.7 percentage points when quintiles are based on expenditures.

In Indonesia and Pakistan the asset index produces a larger predicted gap between the richest and the poorest quintiles than do expenditures for all three education outcomes. In Nepal the differences are very close to 0 for current enrollment and ever attended. For completion of grade 5, they even indicate a slightly smaller gap when the asset index is used.

As discussed earlier, higher-order principal components are orthogonal to the asset index by construction. Nevertheless, the results may be affected by introducing the higher-order terms into a nonlinear multivariate model. When we include up to the fifth principal component in the set of regressors in the models underlying Table 5, the changes in the estimated marginal wealth effects are very small for all three

TABLE 5. MARGINAL EFFECTS OF BEING IN THE RICHEST QUINTILE (RELATIVE TO THE POOREST QUINTILE): NEPAL, INDONESIA, AND PAKISTAN, URBAN AND RURAL AREAS

	Based on Asset Index	Based on Household Consumption Expenditures (Adjusted for Size)	Difference in "Wealth Gap" Between Asset Index and Expenditures Classification
Nepal			
Currently attending school (ages 6 to 14)	34.0	33.8	0.2
Ever went to school (ages 6 to 14)	31.9	30.7	1.2
Completed at least grade 5 (ages 15 to 19)	44.9	48.5	–3.6
Indonesia			
Currently attending school (ages 6 to 14)	10.5	8.7	1.8
Ever went to school (ages 6 to 14)	3.5	3.3	0.2
Completed at least grade 6 (ages 15 to 19)	8.5	4.3	4.2
Pakistan			
Currently attending school (ages 6 to 14)	36.7	27.5	9.2
Ever went to school (ages 6 to 14)	36.5	28.3	8.2
Completed at least grade 5 (ages 15 to 19)	51.0	27.4	23.6

Notes: Values are derived from a probit regression of the outcome on a set of four dummy variables for quintile (the poorest category is the reference category) and controls for gender, urban residence, age, gender of the head of the household, age of the head of the household, and education of the head of the household. (Control variables are available from the authors.) Marginal effects are evaluated at the means of all the variables: that is, a specification that mimics that used in the subsequent analysis for India.

Sources: Authors' calculations from IDHS 1994, NLSS 1996, PIHS 1991.

outcome variables.[4] For example, the marginal effect of belonging to the richest versus the poorest quintile for current attendance declines from 34.0 to 32.2 percentage points in Nepal, remains at 10.5 points for Indonesia, and increases from 36.7 to 39.2 points in Pakistan.

Because the NFHS does not include both assets and consumption expenditures, we cannot conduct the same exercise for India as for Nepal, Indonesia, and Pakistan. Yet we can compare wealth gaps in enrollment across quintiles calculated with the asset index in the NFHS to wealth gaps based on per capita consumption expenditures ($\alpha = 1$) from an analysis of Indian National Sample Survey (NSS) data (Haque, Lanjouw, and Ravallion 1998). We use a simple bivariate analysis because we do not have the multivariate predicted probabilities for the NSS data. (In the case of Nepal, Indonesia, and Pakistan, the bivariate results produce exactly the same pattern as that from the multivariate specification.)

The first row of Table 6 shows the difference in the enrollment rate of rural children age 6 to 14 between the richest and the poorest quintiles, based on either the asset index (first column) or on consumption expenditures (second column) for the national sample. The enrollment of the poorest asset quintile is 7 percentage points lower than for the poorest expenditure quintile (42 versus 49); the enrollment of the richest asset quintile is 12 percentage points higher than for the richest expenditure quintile (94 versus 82). As a result, the wealth gap in enrollment rates between the richest and the poorest quintiles is only 33 percentage points when consumption expenditures are used; when we use the asset index, however, the gap is 52 percentage points, or 19 percentage points higher.

The subsequent rows of Table 6 show that the wealth-education profile is "flatter" (that is, the rich-poor differential is smaller) when rankings are based on consumption expenditures for every state of India. One possible interpretation of this regularity is attenuation bias due to greater "measurement error" in consumption expenditures than in the asset index. "Error" here is defined in relation to its use as a proxy for the relevant indicator of economic status in the analysis of education outcomes.

We explore this issue of relative measurement error further, using two multivariate regression approaches: "pseudo"-instrumental variables (IV) and reverse regression.

Pseudo-IV. Under the hypotheses that (1) expenditures and the asset index are both proxies for long-run economic status and (2) the measurement errors in each are not correlated perfectly, each proxy can be used as an instrument for the other to mitigate the attenuation bias due to measurement error. This is true even if the measurement errors in expenditures and in assets are correlated and hence neither of the resulting IV estimates are consistent. (This is a different application of the method described in Ashenfelter and Krueger 1994.) We refer to these as "pseudo"-IV estimates because we are not making the (strong) assumption that measurement errors are uncorrelated. We do not make such an assumption because some of the asset variables may be used both in the asset index and in estimating expenditures. The resulting instruments therefore are not "valid" in the sense of producing consistent parameter estimates.

Nevertheless, the ratio of OLS to pseudo-IV estimates is an estimate of the relative signal to signal-plus-noise ratio for the two variables. Because the degree of inconsistency in the pseudo-IV estimates depends only on the measurement error common to both measures, the pseudo-IV estimates for expenditures and assets will converge in probability to the same (inconsistent) estimate. In contrast, the degree of inconsistency in the two OLS estimates depends on both the common measurement error and the error specific to either the asset index or consumption expenditures. Hence the ratio of the ratios of OLS to pseudo-IV for each measure is a valid indicator of the relative degree of measurement error in the two proxies. In the multivariate context, the inconsistency in both the OLS and the pseudo-IV estimate will depend in addition on the correlation with the other variables in the regression. Because the measurement error is assumed to be

TABLE 6. DIFFERENCE BETWEEN THE RICHEST QUINTILE AND THE POOREST QUINTILE, AVERAGE ENROLLMENT RATES OF RURAL CHILDREN AGE 6 TO 14

	Quintiles Constructed by		
	Asset Index	Per Capita Consumption Expenditures	Difference
All India	52	33	19
Andhra Pradesh	55	37	19
Assam	36	21	15
Bihar	67	43	25
Gujarat	46	27	19
Haryana	49	39	10
Karnataka	51	38	13
Kerala	12	3	9
Madhya Pradesh	55	33	22
Maharashtra	34	25	9
Orissa	47	38	10
Punjab	56	52	5
Rajasthan	52	41	11
Tamil Nadu	25	15	9
Uttar Pradesh	52	30	21
West Bengal	51	40	11

Sources: Authors' calculation from NFHS 1992–1993. Enrollment for consumption quintiles from Haque, Lanjouw, and Ravallion (1998).

4. To maintain consistency with the rest of the paper, we enter the first principal component into the model as dummy variables for each quintile. We enter the higher-order principal components into the model simply as linear indices as given by the factor scores from the analysis.

TABLE 7. PROBABILITY LIMITS USING TWO ALTERNATIVE NOISY MEASURES OF *X*

	OLS	Pseudo-IV	Ratio of OLS to Pseudo-IV
X^*	$\beta\times(\sigma_x^2/(\sigma_x^2+\sigma_\lambda^2+\sigma_{\upsilon*}^2))$	$\beta\times(\sigma_x^2/(\sigma_x^2+\sigma_\lambda^2))$	$(\sigma_x^2+\sigma_\lambda^2)/(\sigma_x^2+\sigma_\lambda^2+\sigma_{\upsilon*}^2)$
X^+	$\beta\times(\sigma_x^2/(\sigma_x^2+\sigma_\lambda^2+\sigma_{\upsilon+}^2))$	$\beta\times(\sigma_x^2/(\sigma_x^2+\sigma_\lambda^2)$	$(\sigma_x^2+\sigma_\lambda^2)/(\sigma_x^2+\sigma_\lambda^2+\sigma_{\upsilon+}^2)$

Notes: σ_y^2 is the variance of variable y. Additional discussion is available from the authors.

TABLE 8. SCHOOL ENROLLMENT AS A FUNCTION OF THE ASSET INDEX OR CONSUMPTION EXPENDITURES: ALTERNATIVE ESTIMATES OF RELATIVE MEASUREMENT ERROR OF EXPENDITURES VERSUS ASSET INDEX

	Pseudo-IV Method			Reverse Regression Method		
	OLS to IV Ratio: Asset Index[a]	OLS to IV Ratio: Cons. Expend.[b]	Ratio of Ratios	Reverse to Direct Ratio: Asset Index[c]	Reverse to Direct Ratio: Cons. Expend.[d]	Ratio of Ratios
Nepal	0.37	0.71	0.52	27.1	14.2	0.52
Indonesia	0.52	0.29	1.79	60.2	108.8	1.81
Pakistan	0.65	0.35	1.86	20.8	38.5	1.85

Notes: Where β is the coefficient on the asset index (*) or consumption (+) in a regression of enrollment on the asset index or consumption (and other variables), β_{IV} is the IV estimate of β; β_r is the coefficient on enrollment in the (reverse) regression of the asset index or consumption on enrollment (and other variables). Further discussion and full regression results are available from the authors.

Sources: Authors' calculations from IDHS 1994, NLSS 1996, and PIHS 1991.

[a] (β^* / β^*_{IV})

[b] (β^+ / β^+_{IV})

[c] $(\beta_r^{*-1} / \beta^*)$

[d] $(\beta_r^{+-1} / \beta^+)$

uncorrelated with those variables, only the correlation between the error-free underlying variable and other variables in the regression will matter. Consequently an additional term will be introduced into the probability limits, but this will not qualitatively affect the analysis because it will be the same for the two variables.

In a simple case where there are two noisy measures of a variable X, namely X^* and X^+, the common measurement error component is λ, and the measure-specific errors are υ^* and υ^+, we summarize the probability limits and their ratios (see Table 7). The ratio of the OLS to pseudo-IV for the two proxies will depend only on the relative measurement error. The discussion here revolves around probability limits, which are valid asymptotically but may not necessarily hold in a given sample.

In the first two columns of Table 8 we report the ratios of OLS to pseudo-IV estimates for Nepal, Indonesia, and Pakistan. (The full multivariate regressions are reported in Appendix B, which is available from the authors.) To keep the setup simple for this exploratory work, we use the index itself here rather than quintiles based on the index. We compare this with the log of consumption expenditures per adjusted household size. In an alternative specification (not shown here) we used the level of consumption expenditures per adjusted household size; the results were qualitatively similar, and even more striking than those shown here.

In Nepal, when we regress current enrollment on the asset index using OLS and with consumption as an instrument, the ratio of the OLS to IV estimates is 0.37 (0.046/0.124). When we regress enrollment on the consumption measure using OLS and with the asset index as an instrument, the ratio of the OLS to IV estimates is 0.71 (0.211/0.298), yielding a ratio of the two ratios of 0.52 (0.37/0.71). In Nepal it appears that there is more measurement error in the asset index than in consumption expenditures.

With data from Indonesia and Pakistan, the finding is the opposite. When current enrollment is regressed on the asset index, the ratio of the OLS to IV estimate is 0.52 in Indonesia and 0.65 in Pakistan. When we regress enrollment on the consumption measure using the asset index as an instrument, the ratio of the OLS to IV estimates is 0.29 in Indonesia and 0.35 in Pakistan.[5] This yields respective ratios

5. Behrman and Knowles (1999) find that their estimates of the elasticities of various education outcome measures in Vietnam with respect to income, or consumption expenditures, increase by 50 to 60% when they use household assets and other household characteristics as instruments for consumption. In the case of India we do not have expenditures, but we have 21 assets. We divided the asset variables into two groups and constructed an asset index from each set to form repeat measurements on long-run

of the ratios equaling 1.79 and 1.86. In these two countries the results indicate much larger measurement error in consumption expenditures as a proxy in predicting enrollments than in the asset index. For instance, if the true signal-plus-measurement error common to both (which is unobservable) is roughly equal to the idiosyncratic error in assets, then a ratio of ratios equaling about 1.8 would imply a measurement error three times larger in consumption expenditures than in the asset index. (For this derivation, see Appendix A, available from the authors.)

Reverse regression. An alternative approach to measurement error is to use reverse regression: that is, to regress enrollment on the wealth measure and estimate the coefficient on wealth (β), and then to regress the wealth measure on enrollment and estimate the coefficient on enrollment (β_r). If enrollment and wealth are measured with error, then, under certain conditions, the true regression parameter is bounded by β and $1/\beta_r$ (see, for example, Maddala 1988). Using reverse regression in a multivariate context, as we do, introduces several complications that are absent in the bivariate case, but under reasonable conditions they will not substantially affect our conclusions. (For further discussion see Appendix A, which is available from the authors.)

A comparison of the ratio of β to $1/\beta_r$, using the asset index versus using expenditures, indicates the relative measurement error in the two variables (as whatever measurement error in enrollment rates is the same for the two analyses). In Nepal the reverse regression yields an estimate ($1/\beta_r$), which is 27.1 times higher than the direct regression when the asset index is used and 14.2 times higher when expenditures are used, yielding a ratio of 0.52 (Table 8). In Indonesia the reverse regression estimate is 60.2 times higher for the asset index and 108.8 times higher for consumption; for Pakistan the numbers are 20.8 and 38.5, yielding respective ratios of 1.81 and 1.85. For all three countries these figures are strikingly consistent with the pseudo-IV results.

To make more precise statements about the relative magnitude of measurement error, one would need to make more assumptions. If one is willing to assume that the measurement error is uncorrelated with the included control variables and that the true variability and noise for assets are roughly equal (as in the illustration of pseudo-IV above), then one can use the ratio of ratios to work out the relationship between the measurement error variance in expenditures and in the asset index. Under these assumptions, if the ratio is 1.8, approximately the ratio in Indonesia and Pakistan, then the measurement error variance in consumption expenditures is approximately 2.5 times larger than in the asset index.

wealth. Although both of these will be imperfect proxies for long-run wealth, the measurement errors will not be correlated perfectly; hence each can be used as an instrument for the other. In this case the ratio of OLS to IV estimates, approximately one-half, is an estimate of the ratio of the "true" variance to the total variance. The wealth index appears to include a sizable measurement error component.

TABLE 9. CLASSIFICATION DIFFERENCES USING ASSET INDEX DERIVED FROM TWO SAMPLES (WITH OVERLAP) IN MOROCCO

Quintiles Based on 1992 Ranking	Quintiles Based on 1995 Ranking					
	Poorest	2	3	4	Richest	Total
Poorest	78.4	20.3	1.1	0.2	0.0	100.0
2	26.6	53.9	19.5	0.0	0.0	100.0
3	2.8	24.5	54.7	13.6	4.5	100.0
4	0.0	2.5	15.7	58.4	23.4	100.0
Richest	0.0	0.0	3.1	32.6	64.2	99.9[a]

Source: Authors' calculations from Morocco DHS, 1992 and 1995.

[a]Row does not sum to 100.0 because of rounding.

Stability of Household Rankings Over Time

These results from Indonesia and Pakistan are consistent with an asset index that is less sensitive to transitory fluctuations than are consumption expenditures. Therefore one explanation of the "superior" performance of the asset index is that household rankings based on the asset index are more stable than those based on a consumption expenditure measure.[6]

A panel survey of households in Morocco conducted in 1992 and 1995 provides the basis for exploring this issue. A DHS survey conducted in 1992 covered 6,407 households; a 1995 survey covered 2,751, of which 2,489 can be matched across surveys. Table 9 presents the classification differences across the two periods for the subsample of overlapping households. (Households are classified in each period according to their position with respect to the entire sample in that period, not merely the subsample that can be matched over time.) For example, 78.4% of the households that are classified as belonging to the poorest quintile in 1992 are also in the poorest quintile in 1995, and virtually none (1.3%) move out of the poorest 40%.

In a recent survey, Fields (1998) reports a similar analysis of stability of classifications based on expenditures (or incomes) in a secondary analysis of results from four countries (China, Peru, Malaysia, and Chile). In addition, Skoufias (1999) calculates similar numbers for Indonesia between 1997 and 1998 based on a panel survey of about 12,000 households. Table 10 summarizes the results on changes in household rankings from these studies. Particu-

6. On the basis of a recent six-year panel of households in China, Jalan and Ravallion (1998) find that annual consumption expenditures are highly variable. In particular, they find that the average standard deviation of consumption per person across households is 384 yuan (the mean is 342 yuan per person per year at 1985 prices over the period 1985–1990) and that the mean of the intertemporal standard deviation for any given household, over the entire period, is 189 yuan. Thus the standard deviation of a household's measured expenditures over time is about half that in the cross-section.

TABLE 10. STABILITY OVER TIME IN RANKINGS: COMPARISON FROM PANEL DATA SETS

Country	Start Year	End Year	Difference	Variable Used to Rank Households (Individuals)	% Staying in Same Quintile: Poorest Quintile	% Staying in Same Quintile: Richest Quintile
Indonesia	1997	1998	1	Per capita household expenditures	54	51
Morocco	1992	1995	3	Household asset index	78	64
China (Rural)	1978	1983	5	Household income	54	61
Lima, Peru	1985	1990	5	Per capita household expenditures	40	50
China (Rural)	1983	1989	6	Household income	41	49
Malaysia	1967	1976	9	Income of males	55	62
Chile	1968	1986	18	Per capita household income	8	58

Sources: Adapted from Fields (1998); Skoufias (1999) for Indonesia; authors' calculations for Morocco (DHS 1992, 1995).

larly for the poorest quintile, these findings clearly show more variability over time for the classifications based on income or consumption expenditure than do the results for Morocco, which use the measure based on the asset index. This is true even for the consumption-based data from Indonesia, which trace changes over only one year. Although these results may not surprise some readers, they may not be obvious to others, who argue that consumption expenditures are smoothed relative to income. As Table 10 makes clear, consumption expenditures vary substantially across time (possibly because of measurement error, transitory shocks, or unexpected shocks to permanent income).

Methodological Summary

The conventional wisdom is that survey-based household consumption expenditures are not only the best estimates of current expenditures but also the most reliable proxy for a household's long-run wealth. This view implies that surveys lacking consumption expenditures have limited value because they cannot control for, or estimate, wealth effects. Current consumption expenditures are more firmly grounded theoretically and have a much wider range of uses (e.g., estimation of demand functions, absolute poverty analysis, estimates of current welfare); yet there is no a priori argument explaining why current consumption expenditures are a more reliable proxy for long-run household economic status than an index of assets. Because the two have different sources of potential "error," this is an open empirical question.

Our results suggest a methodologically simple solution to the vexing problem of creating a weighted asset index. In the absence of data on expenditures, the straightforward solution of applying the principal-components technique to a collection of asset indicators works quite well. The asset index is not difficult to defend empirically; it appears to be an internally and externally coherent and stable measure. Although each of the methods we use to assess measurement error has its limitations, together they tell a consistent story. In two out of three countries studied (Indonesia and Pakistan), the results are consistent with the asset index containing less measurement error than traditional consumption expenditures as a proxy for long-run economic status in predicting enrollment differences. In Nepal the results are ambiguous.

Ultimately the question in each case is empirical and depends on many factors, such as the quality of underlying data and the degree of expenditure variability. In cases where the data contain both expenditures and assets, we would not use these results to argue that an asset index is the most appropriate variable to employ. In such cases, using assets as instruments for household per capita expenditures is most likely the more effective way of extracting the maximum amount of information from the data while reducing the impact of measurement error. Yet the apparent success of this simple asset index in addressing the problem at hand introduces the possibility of applying the DHS and NFHS data on household wealth to a broader array of socioeconomic outcomes. Here we have limited ourselves to investigating and discussing its validity and usefulness in the study of education enrollments in states of India.

MULTIVARIATE ANALYSIS OF WEALTH GAPS IN EDUCATIONAL ENROLLMENT IN INDIAN STATES

Armed with data on educational outcomes, on the one hand, and the asset index, on the other, we now examine how children's school enrollments differ within Indian states according to the household's economic status.

In India only 68% of children age 6 to 10 and 66% of those age 11 to 14 are reported as being in school. Enrollment rates vary dramatically across the states: the percentage of 6- to 10-year-olds in school ranges from only 50% in Bihar to 96% in Kerala, and the percentage of those age 11 to 14 in school ranges from 54% in Bihar to 94% in Kerala and Mizoram (see Appendix Table A1).

To disentangle the determinants of school enrollment, we estimate probit regressions with the school enrollment of the *i*th child age 6 to 14 as the (latent) dependent variable:

$$E_i^* = \Sigma_{q=2,5}\, \beta_q \times Q_{iq} + \alpha \times \mathbf{X}_i + \Sigma_{k=2,25}\, \delta_k \times \lambda_{ik} + \varepsilon_i. \quad (4)$$

We use a probit model because we observe only whether a child is in school (corresponding to $E_i = 1$ if $E_i^* \geq 0$) or not in school (corresponding to $E_i = 0$ if $E_i^* < 0$). Wealth is specified by including the Q_{iq}s; these are dummy variables equal to 1 if the child's household is in quintile q. (The poorest group is the reference quintile.)

In all of the samples the variables included in addition to wealth quintiles are the *child* variables, namely a dummy variable for gender and the child's age and age squared, and the *household* variables: age of the head of the household, whether the household head ever attended school, the highest grade completed by the household head, whether the household is Hindu, and whether the household belongs to a scheduled caste or tribe (Patrinos 1997). (If the information on the education of the household head is missing, we set the "head ever attended school" and "head's highest grade" variables equal to 0 and set an indicator dummy variable equal to 1 in the regression.) Finally, the specification includes a set of *state* dummy variables λ_{ik} equal to 1 if child i lives in state k.

The other variables used (in $\mathbf{X}_i$) depend on the sample because data on school availability and other village characteristics are limited to rural areas. In the pooled urban/rural samples, we include an urban dummy variable. In the rural samples, the variables include three dummy variables for the presence of (1) a primary school, (2) a primary and a "middle" school, and (3) a primary, "middle," and secondary school. In addition, we include a number of variables capturing village infrastructure and "development" (e.g., post office, bank, cinema).

Table 11 reports the estimation of Eq. (4) for all of India with both urban and rural samples, and with only the rural samples. (Summary statistics are displayed in Appendix Tables A2 and A3.) The first rows of Table 11 report the marginal effect, or (for a dummy variable) the change in a variable from 0 to 1, on the probability of enrollment when all of the other variables in the regression are set to their sample means. The probability of enrollment rises sharply with household wealth. All else being equal, a child from a household in the highest quintile is about 30.7 percentage points more likely to be in school than a child from the poorest quintile. Moreover, the effects are ordered strictly across the quintiles: belonging to the second quintile is associated with a 10-percentage-point increase in the probability of being in school, and each subsequent quintile is associated with an increase of roughly 7 percentage points in the probability of enrollment (10.3 to 16.9 to 24.1 to 30.7).[7]

The wealth effects for rural areas only, where we can include numerous additional village-level factors in the model, are very similar. This is important because the rural sample includes information on school availability; therefore these effects represent the relationship, controlling for the fact that the poor are more likely to live in villages without schools. Even with these additional controls, the marginal effects associated with the wealth quintiles are nearly identical to those in the all-India sample (11.1, 18.5, 26.9, and 31.5).

We estimate the same regressions separately for India as a whole (Table 11) and for each state. (The all-India regressions include state dummy variables.) Table 12 presents the marginal effects of higher wealth on the probability that a child age 6 to 14 is in school when the effects are estimated in state-by-state regressions. The state-level models include all the same control variables as the national models; here, however, we report only the wealth effects. In a separate paper we delve more deeply into the interpretation of the other variables in the Indian context, including an examination of gender effects and the state-specific effects (Filmer and Pritchett 1999b).

Although the effects are large, on average, the states vary substantially in the magnitude of the wealth effects. For example, a child from the highest quintile in Kerala is only 4.6 percentage points more likely to be in school than a child from the poorest quintile in that state, whereas in Bihar the difference is 42.6 percentage points, nearly 10 times larger. The differences are exacerbated in rural areas: the rich-poor difference is 4.2 percentage points in Kerala and 52.6 in Bihar. These enormous differences in the wealth gap, even within the same country, certainly deserve further analysis.

The magnitudes of the wealth gaps found here are consistent with those from other studies, both in the all-India findings and in the state-by-state results. For example, NCAER (1994) finds that among children age 6 to 14 who had ever attended school, there was an average difference of 25 percentage points, over 14 major states, between children from households with per capita incomes of less than Rs 3,000 and children from households with per capita incomes of more than Rs 10,000. The difference ranged from virtually none in Kerala to 55 percentage points in Punjab. Haque et al. (1998), in bivariate tabulations, find similar differences across the quintiles in the raw enrollment rates (see the discussion of Table 6).

Behrman and Knowles (1999) review estimates on the income elasticity of educational attainment from many different countries. These are not strictly comparable with our results because they report elasticities of attainment rather than enrollment probabilities. Even so, the evidence in the

7. The results shown in Table 11 are robust to the inclusion of higher-order principal components in the specification. For example, including indices up to the fifth component changes the estimated percentage-point effect on quintiles to 8.4, 13.5, 20.5, and 29.4 in the all-India model and to 8.0, 12.5, 20.3, and 29.2 in the rural model. Although these point estimates are slightly different, the largest difference is no more than 3.6 percentage points in the all-India model and 6.6 points in the rural-only sample. The gradients implied by the two specifications are also very similar. In all specifications, the children from the richest quintile are 30 percentage points more likely to be enrolled than those in the poorest quintile. As above, to maintain simplicity in this robustness check, we include higher-order components as indices, whereas the first principal component enters in "quintile" form. (The results are not shown here but are available from the authors.)

TABLE 11. MARGINAL EFFECTS ON THE PROBABILITY OF BEING "IN SCHOOL," AGES 6–14 (PROBIT REGRESSION RESULTS)

		All India (Urban and Rural)		Rural Only	
	Zero / One Variable	Marginal Effect	*T*-Ratio	Marginal Effect	*T*-Ratio
Quintile 2[a]	*	0.103	12.32	0.111	9.87
Quintile 3	*	0.169	16.94	0.185	17.92
Quintile 4	*	0.241	22.55	0.269	20.77
Quintile 5	*	0.307	23.53	0.315	18.69
Male	*			0.237	8.42
Rural Male[b]	*	0.070	3.85		
Urban Female	*	–0.107	–6.19		
Rural Female	*	–0.149	–6.70		
Scheduled Caste/Tribe	*	–0.047	–3.87	–0.053	–4.37
Age		0.206	13.37	0.232	13.20
Age Squared		–0.011	–16.89	–0.012	–16.47
Head Is Male	*	–0.092	–5.64	–0.119	–5.90
Head's Age		0.001	4.29	0.002	5.41
Head Ever Attended School	*	0.072	6.73	0.071	6.88
Head's Highest Grade Completed		0.019	16.27	0.023	19.31
Head Information Missing	*	0.094	4.42	0.112	4.75
Hindu	*	0.109	5.11	0.119	5.38
Primary School in Village	*			0.037	2.10
Primary and Middle School in Village	*			0.073	3.05
Primary, Middle, and Secondary in Village	*			0.083	6.43
Nearest Town Within 5 km	*			0.018	1.31
Nearest Railroad Within 5 km	*			–0.001	–0.11
Nearest Bus Within 5 km	*			0.014	1.71
Paved Road in Village	*			0.006	0.42
Electricity in Village	*			0.019	1.10
PHC Clinic in Village	*			–0.006	–0.27
Health Subcenter in Village	*			–0.011	–1.09
Hospital in Village	*			–0.015	–1.00
Dispensary in Village	*			0.001	0.11
Health Guide in Village	*			0.001	0.05
Bank in Village	*			0.009	0.92
Co-op in Village	*			0.007	0.55
Post Office in Village	*			–0.009	–0.60
Market in Village	*			–0.021	–2.95
Cinema in Village	*			0.003	0.31
Pharmacy in Village	*			0.016	1.15
Mahila Mandal (Women's Group) in Village	*			–0.022	–1.01
Flood Within Last 2 Years	*			–0.003	–0.22
Drought in Last 2 Years	*			–0.007	–0.56

Notes: The marginal effect for a zero/one variable is the effect of a change in the variable from 0 to 1 on the probability of a child being in school, evaluated at the means of the other variables. The specification includes dummy variables for each state. *T*-ratios refer to the underlying probit coefficient.

Source: Authors' calculation from NFHS 1992–1993.

[a]Reference group is Quintile 1 (poorest).

[b]Reference group is urban male.

TABLE 12. MARGINAL EFFECTS OF WEALTH ON THE PROBABILITY OF BEING IN SCHOOL, AGES 6 TO 14, URBAN AND RURAL: PROBIT REGRESSION RESULTS FOR SELECTED VARIABLES

	Pooled Urban and Rural Samples				Rural Sample Only			
	Quint. 2	Quint. 3	Quint. 4	Quint. 5	Quint. 2	Quint. 3	Quint. 4	Quint. 5
Mizoram	0.030	0.073	0.112	0.083	–0.012[a]	0.026[a]	0.018[a]	–0.096[a]
Himachal Pradesh	–0.035[a]	0.031[a]	0.045[a]	0.062[a]	–0.086[a]	0.005[a]	0.013[a]	0.026[a]
Kerala	0.017[a]	0.038	0.059	0.046	0.014[a]	0.037	0.058	0.042
Goa	0.019[a]	0.042	0.064	0.098	0.024[a]	0.038	0.063	0.054
Nagaland	–0.004[a]	0.027[a]	0.017[a]	0.064	0.001[a]	0.037[a]	0.007[a]	0.065[a]
Manipur	0.032[a]	0.055	0.085	0.073	0.037	0.049	0.095	0.095
Jammu	0.039[a]	0.079	0.146	0.160	0.028[a]	0.066	0.118	0.119
Tamil Nadu	0.006[a]	0.061	0.106	0.143	–0.001[a]	0.078	0.119	0.142
Tripura	0.080	0.115	0.138	0.079[a]	0.066	0.136	0.137	0.155
Delhi	0.055[a]	0.072[a]	0.115	0.446			0.087	0.160
Maharashtra	0.048	0.084	0.124	0.199	0.049	0.093	0.163	0.164
Assam	0.131	0.202	0.212	0.133	0.139	0.212	0.187	0.172
Haryana	0.072	0.093	0.186	0.234	0.084[a]	0.107	0.229	0.196
Arunachal Pradesh	0.137	0.215	0.239	0.242	0.121	0.217	0.226	0.212
Orissa	0.082	0.206	0.231	0.263	0.095	0.229	0.250	0.251
Meghalaya	0.011[a]	0.081	0.188	0.197	0.011[a]	0.083[a]	0.209	0.257
Gujarat	0.057	0.106	0.179	0.294	0.066	0.145	0.210	0.273
West Bengal	0.152	0.242	0.290	0.271	0.124	0.226	0.287	0.284
Punjab	0.035[a]	0.104	0.207	0.336	0.022[a]	0.110	0.246	0.286
Karnataka	0.088	0.185	0.253	0.296	0.074	0.191	0.267	0.303
Madhya Pradesh	0.121	0.198	0.268	0.348	0.135	0.220	0.297	0.371
Uttar Pradesh	0.135	0.188	0.271	0.382	0.152	0.196	0.282	0.372
Andhra Pradesh	0.077	0.151	0.261	0.322	0.083	0.126	0.270	0.387
Rajasthan	0.082	0.158	0.296	0.388	0.065	0.180	0.339	0.406
Bihar	0.150	0.248	0.400	0.426	0.167	0.255	0.425	0.526
All India	0.103	0.169	0.241	0.307	0.111	0.185	0.269	0.315

Notes: States are sorted by the "wealth gap" as measured by the Quintile 5 coefficient in the rural sample. Marginal effects are evaluated at the means of the other variables. In addition to the displayed variables, the probit regression includes age, age squared; gender, age, and schooling of the head of the household; and a dummy variable for Hindu. The regression for the rural sample includes dummy variables for village infrastructure (for example, for the presence of a paved road, a PHC clinic, a post office, and a market shop). The all-India regression includes dummy variables for state.

Source: Authors' calculation from NFHS 1992–1993.

[a]Not significant at .05 level. For all other coefficients, $p < .05$.

poorer countries is consistent with an income elasticity that would produce wealth gaps similar to those we estimate here.

CONCLUSION

In this paper we show that the relationship between wealth and enrollment can be estimated without income or expenditure data, and largely without apologies or tears, by using household asset variables. Principal-components analysis provides plausible and defensible weights for an index of assets to serve as a proxy for wealth. In the four countries examined—India, Indonesia, Nepal, and Pakistan—this approach produces reasonable results. The results from India, Indonesia, and Pakistan are consistent with the presence of less measurement error in the asset index than in consumption expenditures as a proxy for long-run wealth in predicting educational outcomes.

When the asset index is applied to the Indian data, the results show large school enrollment differences by wealth that vary widely across states of India. On average a rich Indian child is 31 percentage points more likely to be enrolled than a poor child, but the wealth gap varies from only 4.6 percentage points in Kerala to 38.2 in Uttar Pradesh and 42.6 in Bihar.

Many research possibilities are suggested by the ability to generate a proxy for household wealth from DHS-like data. For example, Filmer and Pritchett (1999a) and Filmer (2000) explore how educational attainment profiles differ by wealth and gender in more than 35 countries with a recent DHS. Similar country-specific (or comparative) analyses potentially could be conducted for a wide array of socioeconomic indicators included in the DHS/NFHS, such as health outcomes (mortality, morbidity, immunization, utilization of health facilities), fertility, and use of family planning.

APPENDIX. AVERAGE ENROLLMENT IN INDIAN STATES AND ALL-INDIA REGRESSIONS

The following additional appendix material is available from the authors: description of a simple model of the structural relationship between individual assets and the index created, and the output from the principal-components procedure (factor scores for higher-order components, eigenvalues, and proportion of the variance captured by each component); discussion and illustration of the correspondence between the OLS-to-IV ratios and differential measurement error; discussion of the complications arising from the multivariate case of reverse regression and demonstration that under reasonable conditions they will not substantially affect our conclusions; discussion of additional assumptions required and the implications of relaxing those assumptions, in the derivation of relative magnitudes of measurement error in reverse regression; and full multivariate regressions results for OLS, pseudo-IV, and reverse regressions.

APPENDIX TABLE A1. EDUCATION STATUS, BY WEALTH GROUP

	Proportion of 6- to 14-Year-Olds Currently in School				Proportion of 15- to 19-Year-Olds Who Have Completed at Least Grade 8			
State	All	Poorest 40%	Richest 20%	Wealth Gap (Rich – Poor)	All	Poorest 40%	Richest 20%	Wealth Gap (Rich – Poor)
Kerala	0.949	0.887	0.975	0.088	0.749	0.531	0.923	0.392
Goa	0.937	0.774	0.973	0.200	0.703	0.344	0.848	0.504
Himachal Pradesh	0.908	0.724	0.970	0.246	0.565	0.233	0.818	0.585
Mizoram	0.907	0.768	0.974	0.205	0.567	0.190	0.844	0.654
Manipur	0.902	0.804	0.991	0.186	0.610	0.359	0.927	0.568
Nagaland	0.896	0.824	0.980	0.157	0.572	0.354	0.865	0.511
Delhi	0.872	0.477	0.924	0.448	0.685	N/A	0.766	N/A
Jammu	0.857	0.666	0.979	0.313	0.541	0.195	0.833	0.638
Tamil Nadu	0.825	0.717	0.950	0.232	0.518	0.269	0.838	0.570
Maharashtra	0.820	0.671	0.962	0.290	0.579	0.279	0.832	0.554
Haryana	0.813	0.605	0.957	0.352	0.480	0.189	0.728	0.539
Punjab	0.808	0.427	0.957	0.531	0.571	0.153	0.777	0.624
Tripura	0.795	0.710	0.873	0.163	0.395	0.187	0.789	0.603
Gujarat	0.757	0.552	0.962	0.410	0.504	0.212	0.845	0.633
Meghalaya	0.749	0.601	0.959	0.358	0.326	0.150	0.667	0.516
Arunachal Pradesh	0.711	0.585	0.865	0.279	0.340	0.184	0.585	0.400
Karnataka	0.708	0.507	0.943	0.437	0.447	0.205	0.816	0.611
Assam	0.703	0.615	0.846	0.231	0.422	0.229	0.866	0.637
Orissa	0.697	0.552	0.969	0.416	0.395	0.189	0.908	0.719
West Bengal	0.678	0.527	0.902	0.375	0.338	0.137	0.734	0.597
Andhra Pradesh	0.639	0.457	0.917	0.460	0.419	0.160	0.859	0.698
Madhya Pradesh	0.626	0.461	0.937	0.476	0.367	0.172	0.832	0.661
Uttar Pradesh	0.614	0.484	0.939	0.455	0.424	0.239	0.836	0.598
Rajasthan	0.593	0.414	0.910	0.496	0.345	0.141	0.773	0.632
Bihar	0.514	0.378	0.942	0.564	0.381	0.183	0.864	0.681
All India	0.677	0.500	0.942	0.442	0.447	0.204	0.824	0.620

Source: Authors' calculation from NFHS 1992–1993.

APPENDIX TABLE A2. ESTIMATES AND SUMMARY STATISTICS OF MULTIVARIATE MODELS FOR INDIA (SEE TABLE 11)

	Urban and Rural				Rural Only			
	Coefficient	*T*-Statistic	Mean	Standard Deviation	Coefficient	*T*-Statistic	Mean	Standard Deviation
1 = Quintile 2	0.335	12.32	0.187	0.390	0.322	9.87	0.239	0.426
1 = Quintile 3	0.588	16.94	0.203	0.402	0.559	17.92	0.241	0.428
1 = Quintile 4	0.930	22.55	0.218	0.413	0.912	20.77	0.204	0.403
1 = Quintile 5 (Richest)	1.385	23.53	0.211	0.408	1.385	18.69	0.077	0.267
1 = Male					0.663	8.42	0.520	0.500
1 = Rural Male	0.218	3.85	0.367	0.482				
1 = Urban Female	–0.307	–6.19	0.141	0.349				
1 = Rural Female	–0.442	–6.70	0.339	0.473				
1 = Scheduled Caste/Tribe	–0.140	–3.87	0.254	0.435	–0.144	–4.37	0.291	0.454
Age	0.629	13.37	9.841	2.553	0.642	13.20	9.797	2.548
Age Squared	–0.033	–16.89	103.369	51.017	–0.034	–16.47	102.464	50.783
1 = Head is Male	–0.282	–5.64	0.795	0.404	–0.330	–5.90	0.784	0.412
Head's Age	0.004	4.29	36.386	16.793	0.006	5.41	36.005	17.331
1 = Head Ever Went to School	0.222	6.73	0.523	0.499	0.199	6.88	0.453	0.498
Highest Grade Head Completed	0.059	16.27	4.091	4.802	0.065	19.31	3.125	4.143
1 = Head's Information Missing	0.312	4.43	0.137	0.343	0.331	4.75	0.148	0.355
1 = Hindu	0.315	5.11	0.761	0.426	0.317	5.38	0.781	0.414
Andhra Pradesh	0.204	27.15	0.041	0.198	0.198	4.00	0.042	0.201
Arunachal Pradesh	0.724	12.42	0.012	0.110	0.859	13.65	0.015	0.122
Assam	0.599	31.37	0.039	0.194	0.706	25.27	0.038	0.192
Delhi	0.190	4.86	0.035	0.183	0.113	1.57	0.004	0.061
Goa	1.082	43.87	0.031	0.173	1.213	20.40	0.024	0.153
Gujarat	0.403	31.47	0.039	0.194	0.462	14.36	0.038	0.190
Himachal Pradesh	0.925	40.84	0.031	0.172	1.010	27.99	0.033	0.178
Haryana	0.462	25.83	0.034	0.182	0.488	11.03	0.033	0.178
Jammu	0.860	39.16	0.033	0.179	0.926	20.04	0.034	0.180
Karnataka	0.309	24.39	0.049	0.216	0.335	8.38	0.050	0.218
Kerala	1.405	41.29	0.040	0.195	1.478	16.39	0.043	0.203
Meghalaya	0.872	13.00	0.013	0.114	0.930	11.19	0.015	0.122
Maharashtra	0.661	43.42	0.043	0.202	0.735	16.52	0.038	0.190
Manipur	1.221	30.56	0.014	0.116	1.254	20.61	0.010	0.100
Madhya Pradesh	0.179	21.42	0.072	0.259	0.204	4.92	0.079	0.270
Mizoram	1.301	16.09	0.013	0.114	1.295	11.57	0.010	0.099
Nagaland	1.359	17.60	0.013	0.114	1.387	17.21	0.015	0.122
Orissa	0.473	83.46	0.047	0.211	0.493	16.27	0.049	0.217
Punjab	0.570	17.02	0.036	0.187	0.612	11.98	0.037	0.190
Rajasthan	0.057	5.52	0.068	0.251	0.130	2.99	0.069	0.253
Tamil Nadu	0.728	50.32	0.035	0.184	0.834	16.76	0.033	0.179
Tripura	0.734	59.16	0.013	0.113	0.762	20.10	0.014	0.118
Uttar Pradesh	0.166	29.01	0.136	0.343	0.243	5.22	0.155	0.362
West Bengal	0.417	29.52	0.046	0.210	0.517	11.49	0.046	0.210
Constant	–3.494	–11.29			–4.245	–12.42		

Source: Authors' calculation from NFHS 1992–1993.

APPENDIX TABLE A3. ESTIMATES AND SUMMARY STATISTICS OF MULTIVARIATE MODELS FOR INDIA, RURAL AREAS ONLY (SEE TABLE 11)

	Coefficient	*T*-Statistic	Mean	Standard Deviation
Primary School in Village	0.104	2.10	0.372	0.483
Primary and Middle School in Village	0.208	3.05	0.244	0.429
Primary, Middle, and Secondary School in Village	0.236	6.44	0.284	0.451
Nearest Town Within 5 km	0.050	1.31	0.194	0.395
Nearest Railroad Within 5 km	–0.004	–0.11	0.196	0.397
Nearest Bus Within 5 km	0.037	1.71	0.663	0.473
Paved Road in Village	0.017	0.42	0.506	0.500
Electricity in Village	0.052	1.10	0.783	0.413
PHC Clinic in Village	–0.016	–0.27	0.116	0.321
Health Subcenter in Village	–0.032	–1.09	0.309	0.462
Hospital in Village	–0.041	–1.01	0.146	0.353
Dispensary in Village	0.004	0.11	0.357	0.479
Health Guide in Village	0.002	0.05	0.401	0.490
Bank in Village	0.024	0.92	0.237	0.425
Co-op in Village	0.018	0.55	0.386	0.487
Post Office in Village	–0.026	–0.60	0.459	0.498
Market in Village	–0.059	–2.95	0.478	0.500
Cinema in Village	0.008	0.31	0.147	0.354
Pharmacy in Village	0.044	1.16	0.261	0.439
Mahila Mandal (Women's Group)	–0.062	–1.01	0.335	0.472
Flood in Last 2 Years	–0.008	–0.22	0.152	0.359
Drought in Last 2 Years	–0.020	–0.56	0.199	0.399

Source: Authors' calculation from NFHS 1992–1993.

REFERENCES

Agrawal, A.N., H.O. Varma, and R.C. Gupta. 1996. *India Economic Information Year-Book*. New Delhi: National.

Ashenfelter, O. and A. Krueger. 1994. "Estimates of the Economic Returns to Schooling From a New Sample of Twins." *American Economic Review* 84:1157–73.

Behrman, J.R. and J.C. Knowles. 1999. "Household Income and Child Schooling in Vietnam." *World Bank Economic Review* 13(2):211–56.

Bonilla-Chacin, M. and J.S. Hammer. 1999. "Life and Death Among the Poorest." Johns Hopkins University and Development Economics Research Group, World Bank, Washington, DC. Photocopied document.

Deaton, A. 1997. *The Analysis of Household Surveys*. Washington, DC: Johns Hopkins University Press.

Deaton, A. and J. Muellbauer. 1980. *Economics and Consumer Behavior*. Cambridge, UK: Cambridge University Press.

Deaton, A. and S. Zaidi. 1999. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis." Working Paper 192, Princeton University Research Program in Development Studies, Princeton, NJ.

Dreze, J. and P.V. Srinivasan. 1997. "Widowhood and Poverty in Rural India: Some Inferences From Household Survey Data." *Journal of Development Economics* 54:217–34.

Fields, G.S. 1998. "Income Mobility: Meaning, Measurement, and Some Evidence for the Developing World." Department of Industrial and Labor Relations, Cornell University. Photocopied document.

Filmer, D. 2000. "The Structure of Social Disparities in Education: Gender and Wealth." World Bank Policy Research Working Paper 2268, Development Economics Research Group, World Bank, Washington, DC.

Filmer, D. and L. Pritchett. 1998. "Estimating Wealth Effects Without Expenditure Data—Or Tears: With an Application to Educational Enrollments in States of India." World Bank Policy Research Working Paper 1994, Development Economics Research Group, World Bank, Washington, DC.

———. 1999a. "The Effect of Household Wealth on Educational Attainment: Evidence From 35 Countries." *Population and Development Review* 25(1):85–120.

———. 1999b. "Determinants of Education Enrollment in India: Child, Household, Village and State Effects." *Journal of Educational Planning and Administration* 13(2):135–64.

Grosh, M. and P. Glewwe. 1998. "The World Bank's Living Standards Measurement Study Household Surveys." *Journal of Economic Perspectives* 12:187–96.

Gwatkin, D.R., S. Rutstein, K. Johnson, R. Pande, and A. Wagstaff. 2000. "Socio-Economic Differences in Health, Nutrition, and Population." HNP/Poverty Thematic Group, World Bank, Washington, DC.

Haque, T., P. Lanjouw, and M. Ravallion. 1998. “A Poverty Profile for India 1993–94.” Development Research Group, The World Bank, Washington, DC. Photocopied document.

Jalan, J. and M. Ravallion. 1998. “Transient Poverty in Postreform Rural China.” *Journal of Comparative Economics* 26:338–57.

———. 1995. “Poverty and Household Size.” *Economic Journal* 105(433):1415–34.

Lindeman, R.H., P.F. Merenda, and R.Z. Gold. 1980. *Introduction to Bivariate and Multivariate Analysis*. New York: Scott, Foresman.

Maddala, G.S. 1988. *Introduction to Econometrics*. New York: Macmillan.

Montgomery, M.R., M. Gragnolati, K.A. Burke, and E. Paredes. 2000. “Measuring Living Standards With Proxy Variables.” *Demography* 27:155–74.

National Council of Applied Economic Research (NCAER). 1994. “Non-Enrollment, Drop-Out, and Private Expenditures on Elementary Education: A Comparison Across States and Population Groups.” New Delhi. Photocopied document.

Patrinos, H.A. 1997. “Differences in Education and Earnings Across Ethnic Groups in Guatemala.” *Quarterly Review of Economics and Finance* 37:809–21.

Sahn, D. and D.C. Stifel. 1999. “Poverty Comparisons Over Time and Across Countries in Africa.” Cornell Food and Nutrition Policy Program, Cornell University. Photocopied document.

Skoufias, E. 1999. “Krismon and Its Impact on Household Welfare: Preliminary Evidence From Household Panel Data From the 100 Village Survey in Indonesia.” International Food Policy Research Institute, Washington, DC. Photocopied document

StataCorp. 1999. *Stata Statistical Software: Release 6.0*. College Station, TX: Stata Corporation.

Stecklov, G., A. Bommier, and T. Boerma. 1999. “Trends in Equity in Child Survival in Developing Countries: A Illustrative Example Using Ugandan Data.” MEASURE Evaluation Project/Carolina Population Center, INED, Chapel Hill, NC. Photocopied document.

World Bank. 1998. “Reducing Poverty in India: Options for More Effective Public Services.” Report 17881-IN, The World Bank, Washington, DC.