Where Has All the Bias Gone? Detecting Gender Bias in the Intrahousehold Allocation of Educational Expenditure

GEETA GANDHI KINGDON
Oxford University

I. Introduction

Two approaches have been used in the literature to detect gender bias in the intrahousehold allocation of consumption or expenditure: the direct comparison of expenditure on males and females where data are available at the level of the individual and the indirect household expenditure methodology, commonly referred to as the Engel curve approach. Since information on the consumption of or expenditure on each individual member of a household is typically not available in household surveys (where generally only total household expenditure on specific items is available), it is usually not possible to directly observe gender bias in the allocation of expenditure within the household. A researcher must therefore use an indirect method. The Engel curve method seeks to detect differential treatment within the household indirectly by examining how household expenditure on a particular good changes with household gender composition.

However, the reliability of the Engel curve methodology as a way of detecting gender bias has been called into question because it has generally failed to confirm discrimination even where it is known to exist (Deaton 1997, 239–41).[1] Deaton notes that "it is a puzzle that expenditure patterns so con-

I would like to thank Angus Deaton, Jean Drèze, Marcel Fafchamps, Måns Söderbom, Shankar Subramaniam, Francis Teal, and participants at the Northeast Universities Development Consortium conference at Yale University for helpful discussions on this article. I am also grateful for the generous help of Måns Söderbom with STATA programming and of Abusaleh Shariff with the data set. Any errors are mine. The research was supported by Wellcome Trust grant no. 053660 and by an Economic and Social Research Council grant under the Global Poverty Research Group. The acquisition of the data was funded by the Department for International Development, India.

[1] For example, the use of the Engel curve method failed to detect significant differential treatment in the intrahousehold distribution of food consumption in Maharashtra (Subramanian and Deaton 1990) and also in Thailand and Cote d'Ivoire (Deaton 1989). It might be thought that much better laboratories to test the Engel curve techniques are provided by Indian states such as Rajasthan, Haryana, and Punjab, with their very skewed sex ratios, or from Bangladesh and Pakistan, two

 0013-0079/2005/5302-0006$10.00

sistently fail to show strong gender effects even when measures of outcomes show differences between girls and boys" (240). Case and Deaton (2003, 11) say "it is not clear whether there really is no discrimination or whether, for some reason that is unclear, the method simply does not work." Ahmad and Morduch (2002, 17) say, "coupled with evidence on [significant gender differences in] mortality and health outcomes, the results on household expenditures pose a challenge in understanding consumer behaviour."

This article tests two potential reasons for this puzzle. First, there are two possible channels through which pro-male bias may occur in expenditure on any particular commodity: (1) via zero purchases for daughters and positive purchases for sons and (2) conditional on positive purchase for both daughters and sons, via lower expenditure on daughters than on sons. If gender bias operates through only one of these mechanisms, then averaging across the two mechanisms may lead to the conclusion of no significant gender bias. Second, there is the issue of the effect of aggregation (of expenditure data across individuals within the household) on the ability to detect gender bias in household expenditures. It may be that somehow aggregation mutes gender effects.

On the first issue, suppose that bias against girls in education takes the form mainly of zero expenditure on girls' education (nonenrollment of girls) but that, conditional on enrollment, expenditure on girls' education is similar to that on boys or even somewhat exceeds that on boys—for, say, sample selectivity reasons or because certain components of expenditure on girls' education are higher than those on boys (e.g., expenditure on school transport and clothing). Then, averaging across these two mechanisms, there may not be significant gender bias but, via the nonenrollment mechanism only, there may be strong bias. Thus, one would be interested in asking whether significant bias occurs via either of the two mechanisms separately and whether it is the averaging across the two mechanisms that leads to the conclusion of nonbias. One would be interested not only in the average unconditional expenditure on girls and boys but also in the distribution of the expenditure.[2]

countries from which comes much of the other evidence on differential treatment by gender. However, a study by Subramanian (1995) failed to find evidence of gender bias in these three Indian states. Similarly, Ahmad and Morduch (2002) found no evidence in favor of boys in Bangladesh, even though the survey they use itself shows that there is an excess of boys over girls of 11%. A similar finding of roughly identical treatment of boys and girls is confirmed for Pakistan (Deaton 1997, 240; Bhalotra and Attfield 1998) and with 1999–2000 NSS data for India (Case and Deaton 2003).

[2] Another reason why the conventional application of the Engel curve method might fail to pick up discrimination against girls, even when in fact it exists, is that the distributional assumption about the dependent variable and thus the specification of the budget share equation could be

Second, the failure of the conventional approach to detect gender discrimination may have to do with the aggregate nature of the data employed in the method. Even expenditure on an individually assignable good such as education is at best typically available only at the household level, though it is, in principle, more readily measurable on an individual basis than food expenditure. It could be that somehow household-level analysis mutes gender effects. It could also be that the way in which household gender-age composition variables are defined makes it difficult to pick up discrimination.

Much of the work using Engel curve methods has focused on detecting gender bias in the allocation of food. Our focus here is on detecting gender bias in the allocation of education. Previous work on India on the allocation of education expenditure using Engel curve methods has generally failed to find consistent evidence of gender bias. For example, Subramanian and Deaton (1991) find that, in National Sample Survey (NSS) data from rural Maharashtra, there is no evidence of pro-boy gender bias in educational expenditure in the 5–9 and 15–54 age groups, though there is weak evidence of bias in the 10–14 age group. Using similar NSS data from a decade later, Lancaster, Maitra, and Ray (2003) do find significant gender difference in educational expenditure in rural Bihar and rural Maharashtra in the 10–16 age group but not in urban areas and not in the primary school 6–9 age group. In his study of five Indian states, Subramanian (1995, 17) wondered "how [to] explain the finding of discrimination against females under [age] 14 in only two states, when school enrolment data suggest discrimination is pervasive?"

Ahmad and Morduch (2002) provide some possible frameworks to explain the lack of evidence of gender bias in household consumption expenditures in Bangladesh. One of their explanations is two-stage budgeting, namely, that parents' choices about aggregate expenditures are separable from their choices about how those expenditures are allocated. That is, parents may not change buying habits (budget share on a commodity might remain unchanged with a change in gender composition of the household), but they might allot different portions of a commodity to sons than to daughters. This will not show up in investigations of aggregate expenditures, but it will show up in investigations of individual outcomes. Another explanation for the failure to find the expected gender bias in resource allocation is provided by differential mortality selection (Rose 1999). Rose finds that excess mortality of girls in

---

wrong. For example, if the education budget share for households with positive education spending is distributed log normally but, because the budget share equation is fitted on all (zero and positive education budget share) households, one is obliged to use absolute budget share rather than the log of budget share as the dependent variable. This would lead to incorrect standard errors. However, this is not a particularly important worry in large samples such as ours.

rural India is related to households' inability to smooth consumption. If households sacrifice some daughters' survival in order to cope with adverse shocks, there may be no gender bias in resource allocations among surviving children but this masks prior gender bias in mortality selection. Jensen (2002) suggests a plausible explanation why gender inequality of outcomes need not arise from inferior parental allocations to daughters than sons. If couples have a strong desire for male offspring, they will continue childbearing until they have at least one (or some given desired number of) sons. This kind of fertility behavior will lead to girls having more siblings and larger household size than boys. Because of fewer resources for each child in larger households, girls will be worse off than boys, even in the absence of any differential treatment by parents. This type of behavior means that household size will be endogenous. Of course, with individual-level expenditure data, it is possible to test these explanations of the apparent lack of gender bias, and I do so in this article.

The 1994 National Council of Applied Economic Research (NCAER) rural household survey of 16 major states in India collected data on individual educational outcomes, that is, on school enrollment, years of schooling, and education expenditure data, on each household member 35 years old or younger. Thus, it is possible, using these data, to investigate gender bias in the allocation of educational expenditure, both by direct examination of educational spending on boys and girls and by the indirect Engel curve method. In other words, it is possible to test whether the indirect, aggregate-data method confirms gender bias in states where the direct individual-data method shows bias. A vindication of the indirect methodology for detecting bias should be of considerable practical interest beyond this study and beyond India since most data sets only permit the use of the indirect method.

Schooling has costs in India. Even apparently "free" government schooling has substantial costs, such as expenditure on books, stationery, travel, and school uniforms.[3] Some studies have also shown that girls are less likely to be sent to fee-charging private schools that are costlier (Drèze and Sen 1995, 133; Kingdon 1996a, 1996b). Our data show that the overwhelming majority (98%) of enrolled 5–19-year-olds have positive expenditure incurred on their education.

In this article, we find that the Engel curve method does fail to find evidence of discrimination, even when significant boy-girl differences are manifest in individual-level expenditure data. The research tests the two explanations for

[3] Household survey data on educational spending show that even so-called fee-free schooling has substantial costs in India. For instance, the PROBE report (Probe Team 1999, 16) found that, in rural north India, parents spend about 318 rupees per year on each child who attends government (i.e., tuition-free) school, so that an agricultural laborer in Bihar with three such children would have to work for about 40 days of the year just to send the children to primary school.

the failure that were outlined above. The first explanation is tested by separating out the two mechanisms through which bias can occur to "unpack" the total gender bias into its two components. The second potential explanation, namely, that aggregation is responsible for the failure to find significant gender bias, is tested by examining whether the effects of gender variables in an education expenditure equation at the household level are similar to those in an equation (with as similar a specification as possible) at the individual-child level. Section II discusses the methodology, including both the Engel curve method and the hurdle model. Section III discusses data and estimation issues. The results are discussed in Section IV, and the final section concludes.

II. Indirect Methodology for Detecting Discrimination

The Engel curve method utilizes the fact that household composition is a variable that exerts an effect on household consumption patterns. The needs that arise with additional household members act in such a way as to increase expenditure on items of consumption associated with the additional member. The approach examines whether budget share of a good consumed by, say, children (such as education), rises as much when an additional girl is added to the household as when an additional boy is added, in a given age range.

The approach is to estimate an Engel curve for the commodity being examined, education in the present case. While there are many possible functional forms for the Engel curve linking expenditure on a good to total expenditure, the Working and Leser specification has the theoretical advantage of being consistent with a utility function and its postulation of a linear relationship between budget share of a good and the log of total expenditure conforms to the data in a wide range of circumstances (Deaton 1997). We use the Working and Leser specification but—so as not to prejudge the issue—later relax it to allow for nonlinearity in the shape of the Engel curve. Working's Engel curve can be extended to include household demographic composition by writing

$$s_i = \alpha + \beta \ln (x_i/n_i) + \gamma \ln n_i + \left[\sum_{j=1}^{J-1} \theta_j (n_{ji}/n_i)\right] + \eta z_i + u_i, \tag{1}$$

where x_i is total expenditure of household i; s_i is the budget share of education eduexp/x_1; n_i is household size; z_i is a vector of other household characteristics such as religion, caste, and household head's education and occupation; and u_i is the error term. The term $\ln n_i$ allows for an independent scale effect for household size, while $j = 1, \dots, J$ refers to the Jth age-gender class within the household and n_{ji}/n_i is the fraction of household members in the jth age-gender class. Since these fractions add up to unity, one of them is omitted from the regression. In this article, there are 14 age-sex categories. These are males and females in age groups 0–4 years, 5–9 years, 10–14 years, 15–19 years, 20–24 years, 25–60 years, and age 61 and above. The fraction of women age 61 years or older in the household is the omitted category. The variables of most interest pertain to persons of school-going age, that is, they are fractions of males and females aged 5–9, 10–14, and 15–19 in the household. These variables are named M5to9, F5to9, M10to14, F10to14, M15to19, and F15to19, respectively. The θ_j coefficients represent the effects (on budget share) of changing household composition while holding household size constant, for example, by replacing a child in a younger age group with one in an older age group or replacing a man by a woman in a given age category. Testing for gender differences simply involves testing the hypothesis that $\theta_{jm} = \theta_{jf}$, where the subscripts m and f are the gender groups male and female and the subscript j refers to the age group. Thus, testing for gender difference in educational expenditure in the 5–9 age group will involve testing whether the coefficient on M5to9 is significantly different from the coefficient on F5to9.

The above method has been used to fit the budget share equations for a wide range of commodities, including food items, clothing, and medical and educational expenses. Conventionally, the model has been fitted on the sample of all households, irrespective of whether they incurred zero or positive expenditure on the particular commodity. Much of the extant Engel curve literature has not conditioned on zero values, that is, it includes both zero and positive values of the dependent variable, the budget share. For example, Subramanian (1995) and Subramanian and Deaton (1990) fit OLS Engel curves on the sample of all households, despite the preponderance of households with zero education budget share (89% and 70% of households had zero education budget shares in these studies, respectively).[4]

Given censoring of the dependent variable (education budget share) at zero for a large percentage of the sample households, an important estimation issue is the choice of the appropriate statistical model. While the extant literature has used OLS, in much of the applied econometrics literature, there is a well-

[4] Some studies have used flexible-form or semi/nonparametric regression, e.g., Bhalotra and Attfield (1998). In Subramanian and Deaton (1990), only 11% of rural Maharashtran households reported positive educational expenditures. In Subramanian's (1995) study, using 1987–88 data, only 30% of rural Maharashtran households had positive spending on education. In the current NCAER data, 56% of rural Maharashtran households incurred some education spending. In Subramanian (1995), in Andhra Pradesh, Haryana, Punjab, and Rajasthan, 21%, 56%, 51%, and 23% of households, respectively, reported positive education spending. In the current NCAER data, the corresponding figures are 49%, 64%, 58%, and 55%, respectively. That is, between 1988 and 1994, the proportion of rural households incurring positive spending on education rose quite sharply.

justified reluctance to include both zero and positive values in an OLS regression because of the biased estimates that result. A standard solution often suggested is the use of a Tobit model. However, apart from the potentially severe problem of heteroskedasticity (Deaton 1997), an important limitation of the Tobit (as well as of the suggested alternative, namely, a partially nonparametric censored least absolute deviation or CLAD estimator) is that it assumes that a single mechanism determines the choice between $s = 0$ versus $s > 0$. In particular, $\partial P(s > 0|x)/\partial x_j$ and $\partial E(s|x, s > 0)/\partial x_j$ are constrained to have the same sign.

The alternatives to censored Tobit that allow the initial decision of $s = 0$ versus $s > 0$ to be separate from the decision of how much s is, given that $s > 0$, are called hurdle models (Wooldridge 2002, 536). These models allow the effect of a variable to differently affect the decision to incur any expenditure ($s = 0$ vs. $s > 0$) and how much to spend ($s|s > 0$). The hurdle or first tier is whether or not to choose positive s. In addition to estimating the conventional Engel curve equation, I propose to use hurdle model estimation to allow the decision of whether to incur any education expenditure to be modeled separately from the decision of how much to spend on education, conditional on spending anything. The simple hurdle model used is

$$P(s = 0|x) = 1 - \Phi(x\gamma), \tag{2}$$

$$\log(s)|(x, s > 0) \sim \text{Normal}(x\beta, \sigma^2), \tag{3}$$

where s is the (total household, not per capita) budget share of education, x is a vector of explanatory variables, γ and β are parameters to be estimated, and σ is the standard deviation of s. Equation (2) stipulates the probability that s is zero or positive, and equation (3) states that, conditional on $s > 0$, $s|x$ follows a log normal distribution. An examination of the distribution of s shows that, conditional on positive education spending, s is log normally rather than normally distributed.

The maximum likelihood estimate (MLE) of γ is simply the probit estimator using $s = 0$ versus $s > 0$ binary response. The MLE of β is just the OLS estimator from the regression of $\log(s)$ on x, using those observations for which $s > 0$. A consistent estimator of $\hat{\sigma}$ is the usual standard error from the latter regression. Estimation is straightforward because we assume that, conditional on $s > 0$, $\log(s)$ follows a classical linear model. The conditional expectation of $E(s|x, s > 0)$ and the unconditional expectation of $E(s|x)$ are easy to obtain using properties of the log normal distribution:

$$E(s|x, s > 0) = \exp(x\beta + \sigma^2/2), \tag{4}$$

$$E(s|x) = \Phi(x\gamma)\exp(x\beta + \sigma^2/2), \tag{5}$$

and these are easily estimated, given $\hat{\beta}$, $\hat{\sigma}$, and $\hat{\gamma}$. The marginal effect of x on s can be obtained by transforming the marginal effect of x on $\log(s)$, using the exponent. Thus, the marginal effect of x on s in the OLS regression of $\log(s)$ conditional on $s > 0$ is obtained by taking the derivative of the conditional expectation of s with respect to x:

$$\frac{\partial E(s|x, s > 0)}{\partial x} = \beta \cdot \exp(x\beta + \sigma^2/2). \tag{6}$$

The marginal effect of a variable x on s, taking into account the effect of x on both the probability that $s > 0$ and on the size of s conditional on $s > 0$, is obtained by taking the derivative of the unconditional expectation of s with respect to x. Differentiating (5), using the product rule, we have

$$\begin{aligned}\frac{\partial E(s|x)}{\partial x} &= \gamma\phi(x\gamma)\exp(x\beta + \sigma^2/2) + \Phi(x\gamma)\beta \cdot \exp(x\beta + \sigma^2/2) \\ &= [\gamma\phi(x\gamma) + \Phi(x\gamma)\beta] \cdot \exp(x\beta + \sigma^2/2), \end{aligned} \tag{7}$$

where $\phi(.)$ is the standard normal density function and $\Phi(.)$ is the cumulative normal distribution function.

It is possible that β in the conditional OLS equation of $\log(s)$ will suffer from sample selectivity bias. We are particularly concerned to see whether the coefficients on the male and female demographic variables, such as proportion of males ages 5–9 in the household (M5to9), proportion of females ages 5–9 (F5to9), and so forth, suffer from selectivity bias, as that would have implications for our measure of gender bias in educational spending. If both male and female demographic variables are equally affected by selectivity bias, then there will be no underestimation or overestimation in the measurement of gender bias. However, if unobserved characteristics, such as child ability, child motivation, and parental attitudes, have a greater influence in enrollment decisions about daughters than sons, then sample selectivity bias in the coefficients of the female demographic variables will be greater than for males and this will lead to an overestimation of pro-male gender bias.

This can be shown by focusing on any one pair of demographic variables, for example, M5to9 and F5to9. Suppose that a girl's ability is an important unobserved trait that determines both whether positive expenditure is incurred on her schooling and how much is spent on her schooling, conditional on positive education spending. Suppose that, for boys, ability does not matter (or matters less) to those two decisions. Thus, girls' ability is an element of the error term both in the probit equation of positive education spending and the OLS equation of conditional education spending for girls. Suppose that the effect of F5to9 is positive in both probit and conditional OLS equations, that is, the greater the proportion of 5–9-year-old females in the household, the greater is the likelihood of the household incurring positive education expenditure and the higher the conditional education expenditure (or education budget share). Now, if the observed F5to9 variable is very large, the household will be almost certain to incur positive education spending. But suppose that, on the basis of the size of the observed variable F5to9, the household is equally likely to have positive education spending as to have zero education spending; then the ability of girls in the household (unobserved to us but observed to parents) will determine whether the household has positive or zero education spending. If the girls in a household have high ability, that household will be observed to have positive education spending, and if they have low ability, the household will not incur positive education spending. Thus, at high values of F5to9, there is no correlation between ability and F5to9, but at low levels of F5to9, there is a negative correlation between ability and F5to9, that is, $(\mathrm{Corr}(x, u) < 0)$. Averaging over all households, the correlation between the explanatory variable (F5to9) and the error term is not equal to zero $(\mathrm{Corr}(x, u) \neq 0)$ and, in fact, the correlation is negative; this implies a violation of the basic assumptions of the classical linear regression model, and there will be endogenous sample selection bias. Due to this negative correlation, the coefficient of F5to9 in the conditional OLS equation of education expenditure will be biased downward. If the coefficient on the corresponding male demographic variable (M5to9) does not suffer from selectivity bias or suffers from it less than the female variable (as is likely), then any pro-male bias will be overestimated.

An important issue is the potential endogeneity of fertility and thus of household size (Browning 1992). Use of the Engel curve method requires that the number of children and household size be exogenous, but household size is itself likely to be determined by education budget-share decisions: for instance, couples with higher taste for schooling may choose to have smaller families. Household size may also be endogenous if son preference that affects budget-share decisions also results in differences in household size by the sex-composition of early-born children (Jensen 2002). Alluding to such difficulties in modeling the effects of children on various economic decisions, Browning (1992, 1435) testifies to "how difficult it is to draw any robust and credible inferences in this area of economics." However, use of family fixed effects in this article provides a powerful way to control for unobserved factors such as parental tastes and preferences that cause household size to be endogenous. I examine whether endogeneity of household size alters the results on gender bias by estimating individual-level education expenditure equations with and without family fixed effects.

Another estimation issue is the potential endogeneity of household per capita expenditure in the education budget-share equation 1.[5] In my specification, endogeneity remains possible since household expenditure reflects labor, savings, and other consumption decisions made at the same time as the choice of education budget share: parents who have higher aspirations for their children's education may work harder to generate income to pay for school fees, and so forth. Endogeneity of per capita expenditure is sought to be addressed with two-stage least squares (2SLS) estimation, using land ownership and nonearned income as identifying instruments after testing for their validity. In the final analysis, however, OLS is used because the Hausman-Wu test failed to reject exogeneity of per capita expenditure in the majority of states.

The analysis will proceed as follows. I will estimate the marginal effect of the male and female demographic variables in the conventional OLS model of the budget share of education in order to compare my results with extant studies. I will also estimate the marginal effects of the demographic variables in a hurdle model, that is, in each of its two tiers—the binary probit of whether the household incurs positive education expenditure and an OLS of household education spending, conditional on spending a positive amount. The marginal effects will be computed using STATA. The main object of interest is to see whether the difference in the marginal effects of the male and female demographic variables is statistically significant in each age group.

III. Data and Estimation Issues

This study uses household survey data collected by the National Council of Applied Economic Research (NCAER), New Delhi. This 1994 survey covered 33,230 households across 16 major states in India. Sampling information and other details about the data set are available in Shariff (1999).

The major advantage of this data set is its detailed information on education of each person aged 35 or less in the household, including educational expenditure information. However, an important drawback is that it did not collect comprehensive information on total household expenditure. Only house-

[5] Deolalikar and Rose (1998) and Rose (2000) show that, in rural South India, men and women's labor supply and household savings respond to child gender. We therefore expect household consumption expenditure to also respond to child gender. However, since we control for sex directly, this correlation is not the source of endogeneity bias here.

hold expenditures on food, health, and education were collected. This implies that the denominator in the budget share expression is not household total expenditure but a (large) subset of it, namely, food, health, and education (FHE) expenditure.[6] The missing component of household total expenditure is the non-FHE expenditure. This would include expenditure on items such as fuel/energy, transport, housing, and entertainment. Given that we have data only on FHE, differential treatment depends upon two components,

$$s = \frac{\text{Eduexp}}{\text{Totalexp}} = \frac{\text{Eduexp}}{\text{FHEexp}} \times \frac{\text{FHEexp}}{\text{Totalexp}},$$

that is, it depends on (1) how Eduexp/FHEexp changes with more girls in the household and (2) how FHEexp/Totalexp changes with more girls in the household. We are able to model only the first component, that is, the share of education expenditure in FHE expenditure. However, the combined FHE share in total expenditure (i.e., the second component) is unlikely to rise with the proportion of girls in the household. If it is the case that, with a greater proportion of girls in the household, education expenditure falls but this reduction is compensated for by an increase in food expenditure (which is the overwhelming part of FHE expenditure), then one could doubt the evidence from a test of component 1 only. However, there is little reason to suppose that FHE expenditure as a proportion of total expenditure rises with proportion of girls in the household. In fact, the contrary has often been suggested in the literature, that is, it has been hypothesized that additional girls in the household decrease the share of food expenditure in total expenditure. If the latter is true, then the evidence here based on component 1 only would underestimate gender bias. We believe that additional girls in the household are unlikely to increase or decrease the share of food (or of FHE) expenditure in total household expenditure; it is most likely that the effect is neutral.[7] In other words, modeling how the share of education in FHE expenditure changes with household gender composition should neither under- nor overestimate gender bias in the allocation of education expenditure. Thus, although we use the budget subshare of education in this article for simplicity, we refer to it simply as the budget share of education.

The analysis here is limited to households that have children of school-going age, that is, those with children aged 5–19. This yields a sample of

[6] We know from Subramanian's (1995) study that, in 1987–88 in five Indian states, food, health, and education expenditure together accounted for about 63% of total household expenditure.
[7] None of the several extant studies provides any convincing evidence of systematic gender bias in food allocation within Indian households.

25,954 households. In this sample, the mean budget share of education is 4.40% and the percentage of households with zero education spending is 31%.[8]

IV. Discussion of Results

We present the results in three subsections. The first explores gender bias by means of descriptive statistics using individual-level data. The second examines whether incorrect functional form is responsible for the failure of the conventional Engel curve approach to detect gender bias. The third subsection asks whether aggregation of data at the household level is to blame for the failure of the Engel curve approach to detect gender bias.

A. Descriptive Statistics

The second column of table 1 shows the sex ratio in the 0–14 age group in sample households. It shows that the proportion of girls is only 46.4% in rural India, but it also shows considerable variation across states, with Bihar, Gujarat, Haryana, Rajasthan, Uttar Pradesh, and Assam having lower proportions of girls than the all-India average.[9] This gives us our prior belief that gender difference in the intrahousehold allocation of educational expenditure is likely to be strongest in these states.

In the remaining columns of table 1, we divide all households with children less than 15 years of age into two groups—all-girl households, where all the

[8] The mean budget share of education percentage is considerably higher than the budget share of education in previous studies on India. For example, the average budget share of education for the five Indian states studied in Subramanian (1995) was 1.34%. In our data, it is 3.69% for those same five states. However, the data used in the two studies are not comparable because, first, the earlier studies do not restrict the sample to only households with children in school-going age range. Second, as stated above, our denominator is not total household expenditure (as in Subramanian) but rather a subset of it, consisting only of food, medical, and educational expenditure. In Subramanian's NSS data on five states, these three expenditure items together constitute 63% of total expenditure, so it is possible to "adjust" our education budget share by deflating it appropriately ((3.69 × 63)/100). This yields a budget share of 2.32% for education, which, though considerably higher than the 1.34% figure in Subramanian for the year 1987–88, is closer to the 2.87% figure for rural India in the Micro Impact of Macro and Adjustment Policies Survey of the mid-1990s (Pradhan and Subramanian 2000, 27). The main explanation for the fact that the budget share of education (s) in our data (2.32%) is greater than that in Subramanian's study (1.34%) is that the education budget share has increased between 1987–88 and 1994, the reference dates of the data in the two studies. This is plausible because of (i) reductions in poverty over time (Drèze and Srinivasan 1996, 4–5; Datt and Ravallion 1998, 30; Dubey and Gangopadhyay 1998) and (ii) increased demand for and more widespread supply of education. That demand for education increased may be gleaned by examining changes over time in the percentage of households that incurred any positive educational expenditure. Figures available for rural Maharashtra at three points in time—1983, 1988, and 1994—show that the percentage of households incurring positive educational expenditures rose from 11% in 1983 to 30% in 1988 and further to 55% in 1994.

[9] The figure for Assam seems implausibly low.

TABLE 1
DESCRIPTIVE STATISTICS BY STATE

State (1)	Proportion of Girls in All Children (Ages 0–14) (2)	Proportion of All-Girl Households in All Households (3)	% of At-Least-One-Boy Households That Incurred Positive Education Expenditure (4)	% of All-Girl Households That Incurred Positive Education Expenditure (5)	Percentage Point Difference (4 − 5) (6)	*t*-Value of the Difference in 4 and 5 (7)
Andhra	48.1	25.6	62.8	48.9	13.9	4.8
Bihar	44.7[a]	17.1[a]	56.3	43.2	13.1	4.3
Gujarat	45.9[a]	17.8[a]	62.9	42.6	20.3[a]	5.5
Haryana	46.3[a]	15.5[a]	72.4	52.2	20.2[a]	5.7
Himachal	46.8	19.1	85.4	69.6	15.8	4.3
Karnatak	47.6	20.6	72.1	59.0	13.1	4.9
Kerala	50.2	28.9	72.1	58.7	13.4	4.2
Maharashtra	46.3[a]	19.2	69.2	48.3	20.9[a]	7.8
Madhya	46.4	18.5[a]	61.1	42.3	18.8[a]	8.6
Orissa	48.4	21.1	64.7	44.0	20.7[a]	6.7
Punjab	46.4	17.7[a]	70.5	46.5	24.0[a]	6.0
Rajasthan	45.0[a]	15.0[a]	67.9	32.3	35.6[a]	11.1
Tamilnadu	48.7	28.5	60.1	39.5	20.6[a]	6.2
Uttar	44.8[a]	15.0[a]	66.9	44.1	22.8[a]	9.6
West Bengal	49.2	20.4	60.2	42.8	17.4	5.1
Assam	39.6	12.2	62.4	55.6	6.8	1.5
All India	46.4	19.0	66.0	47.3	18.7	24.4

Note. The figures for Assam in the first two columns are implausibly low. The states with the greatest expected gender bias are Bihar, Gujarat, Haryana, Maharashtra, Madhya, Orissa, Punjab, Rajasthan, and Uttar Pradesh.
[a] Value is above or below the national average.

TABLE 2
CURRENT ENROLLMENT RATE OF CHILDREN BY AGE GROUP AND GENDER

	Ages 5–9			Ages 10–14			Ages 15–19		
State	Female	Male	Gap	Female	Male	Gap	Female	Male	Gap
Andhra	65	77	12*	57	70	13*	15	40	25*
Bihar	35	46	11*	50	64	14*	24	46	22*
Gujarat	57	65	8*	68	82	14*	24	44	20*
Haryana	55	60	5	70	85	15*	21	49	28*
Himachal	79	83	4	89	94	5	46	74	28*
Karnatak	60	64	4	64	74	10*	27	44	23*
Kerala	81	85	4	98	96	−2	54	55	1
Maharashtra	69	70	1	71	85	14*	26	56	30*
Madhya	40	47	7*	52	69	17*	15	42	27*
Orissa	51	58	7*	56	76	20*	18	42	24*
Punjab	71	76	5	73	83	10*	26	45	19*
Rajasthan	32	58	26*	36	79	43*	9	46	37*
Tamilnadu	61	74	13*	67	80	13*	23	38	15*
Uttar	40	56	16*	49	72	23*	19	47	28*
West Bengal	47	48	1	62	66	4	25	40	15*
Assam	52	60	8*	77	86	9*	49	59	10*
All India	51	60	9*	60	76	16*	24	47	23*

* Gender gap is statistically significant at the 5% level.

children below the age of 15 are girls, and at-least-one-boy households, where there are one or more boys in the household. Table 1 shows quite a dramatic difference in the percentage of households incurring positive educational spending, depending on whether it is an all-girl or at-least-one-boy household. It shows that, in rural India, the percentage of all-girl households reporting positive education spending is only 47.3%, whereas the corresponding percentage for at-least-one-boy households is 66.0%. In other words, all-girl households are nearly 19 percentage points more likely to report zero educational spending than at-least-one-boy households. This large and statistically very significant difference indicates an important correlation between the gender composition of the household child population and the household's decision to incur positive educational spending.

Table 2 shows that, in the 10–14 and 15–19 age groups, girls have a significantly and substantially lower current enrollment rate (than boys), that is, a higher probability of reporting zero educational spending due to nonenrollment, in nearly all of the 16 sample states (exceptions are Kerala and West Bengal). However, this is not so in the 5–9 age group, where the gender gap in enrollment rate is significant only in about half the states. Using PROBE (Public Report on Basic Education) data, Drèze and Kingdon (2001) find that, in the 5–12 age group in northern India, boys are 13.6 percentage points more likely to be currently enrolled than girls, which is similar to the implied raw gender difference in table 2 in the 5–14 age group.

TABLE 3
EDUCATIONAL EXPENDITURE ON ENROLLED CHILDREN BY AGE GROUP AND GENDER

	Ages 5–9			Ages 10–14			Ages 15–19		
State	Female	Male	Gap	Female	Male	Gap	Female	Male	Gap
Andhra	258	219	−1.0	305	330	.5	864	885	.1
Bihar	249	309	2.1*	378	431	1.4	651	652	.0
Gujarat	258	247	−.3	313	350	1.0	912	1171	1.5
Haryana	633	634	.0	721	859	2.3*	1,115	1,434	2.6*
Himachal	671	707	.6	974	1,049	1.2	1,686	1,966	1.9
Karnatak	285	337	1.3	446	455	.2	751	918	1.8
Kerala	490	611	2.7*	677	745	1.3	1,269	1,373	.8
Maharashtra	210	222	1.1	359	397	1.7	688	786	1.7
Madhya	218	242	1.6	301	289	−.7	651	582	−1.1
Orissa	222	188	−1.6	295	289	−.3	852	831	−.2
Punjab	498	651	2.3*	674	793	2.0*	1,712	1,365	−2.0*
Rajasthan	324	348	1.0	496	520	.8	1,109	1,164	.4
Tamilnadu	333	331	−.0	386	418	.6	1069	910	−.8
Uttar	343	316	−1.0	375	411	1.8	710	780	1.0
West Bengal	200	204	.1	382	379	−.1	863	945	.8
Assam	357	353	−.1	352	449	2.2*	905	1,007	.6
All India	331	345	1.5	455	477	2.2*	981	994	.4

Note. Expenditure is measured in rupees.
* Gender gap is statistically significant at the 5% level.

Table 3 shows average educational expenditure, conditional on enrollment. It is clear that, once enrolled in school, girls and boys are not treated differently in terms of educational spending in most states in any of the three age groups. Thus, the main form of differential treatment is via the differential current enrollment rates of girls and boys. Table 4 includes zero education-expenditure (i.e., nonenrolled) children, and it shows that, in the 5–9 age group, the states with the greatest gender gap in unconditional educational expenditure are Bihar, Madhya Pradesh, Punjab, Rajasthan, and Uttar Pradesh.[10] In the 10–14 age group, the gender difference in unconditional education expenditure is significant in 12 of the 16 states, and in the 15–19 age group it is so in 14 of the 16 states. Thus, there is fairly strong evidence of gender bias in the raw data, and the bias is stronger in the older age groups. The gender gap in educational expenditure occurs mainly via girls' significantly higher probability of nonenrollment (i.e., via zero education expenditures) and only rarely via lower expenditures once enrolled.

[10] While Kerala appears to have a significant gender gap in the 5–9 age range, this seems implausible. Moreover, this gap becomes insignificant after controlling for household characteristics, as will be demonstrated later.

TABLE 4
EDUCATIONAL EXPENDITURE ON ALL (ENROLLED AND NONENROLLED) CHILDREN BY AGE GROUP AND GENDER

	Ages 5–9			Ages 10–14			Ages 15–19		
State	Female	Male	Gap	Female	Male	Gap	Female	Male	Gap
Andhra	168	168	.0	174	232	1.7	130	355	4.9*
Bihar	88	142	4.1*	191	275	3.4*	153	302	5.0*
Gujarat	147	160	.5	212	288	2.6*	215	514	4.6*
Haryana	348	378	.9	503	731	4.5*	236	703	8.7*
Himachal	528	586	1.2	872	989	2.0*	780	1,458	6.3*
Karnatak	171	218	1.8	284	339	1.7	199	406	5.2*
Kerala	399	520	2.9*	662	718	1.1	679	758	.9
Maharashtra	144	154	1.2	254	339	4.6*	180	438	8.4*
Madhya	86	113	3.5*	156	200	4.0*	99	247	8.6*
Orissa	112	109	.3	165	219	3.1*	155	351	5.0*
Punjab	352	491	2.7*	495	660	3.2*	449	611	2.2*
Rajasthan	104	202	7.6*	176	410	11.4*	95	540	9.8*
Tamilnadu	204	244	1.0	259	336	1.9	248	348	1.4
Uttar	137	176	2.9*	182	297	8.6*	136	368	9.4*
West Bengal	95	99	.3	235	249	.6	212	376	3.8*
Assam	186	210	.9	271	387	3.0*	444	593	1.5
All India	170	206	6.6*	274	364	12.1*	234	468	19.5*

* Gender gap is statistically significant at the 5% level.

B. *Is Incorrect Functional Form the Reason for the Engel Curve Method's Failure to Detect Gender Bias?*

The conventional Engel curve equation is fitted using least squares regression on the absolute value of the household's unconditional budget share of education. Thus, the functional form used for the dependent variable is linear, and the analysis models both zero and positive education budget shares in a single equation. As stated earlier, this is problematic. We unpack the unconditional education budget share into its two components: the probability of positive budget share and, conditional on positive budget share, the size of budget share. Using household-level data, we estimate three equations for each of the 16 states: (*a*) the conventional Engel curve equation; (*b*) a binary probit of whether the household's education budget share is positive or zero; and (*c*) the OLS of the natural log of education budget share, conditional on positive education budget share. The resulting 48 equations are presented in table A1 in the appendix.

The first column under each state presents the conventional Engel curve of education expenditure share (or ESHARE) fitted on all zero and positive education expenditure households. This is the unconditional OLS of ESHARE.

The budget share of education varies from 2.7% in Andhra Pradesh to 8.7% in Himachal Pradesh. The goodness of fit of the conventional Engel curves varies substantially by state. The shape of the education Engel curve was nonlinear in several states when I allowed for a quadratic term in expenditure, confirming that, at low levels of log of per capita expenditure (LNPCE), education is a luxury but that it becomes a necessity at higher levels of expenditure.

Note that LNPCE has a significant positive relationship with budget share of education and that the total expenditure elasticity is close to or above unity in all states. This suggests that education is treated as a luxury. The elasticities are mostly lower than those found in Subramanian and Deaton (1990) and Subramanian (1995), suggesting that education has come to be treated as less of a luxury than in the mid-1980s (the date of data in previous studies).[11]

Since parents with higher educational aspirations for their children's may work harder to generate income, education budget share and household per capita expenditure may be jointly determined. To allow for this potential endogeneity of LNPCE, acres of land owned, its square, and household's nonearned income (from rent, interest, and dividends) were used as instruments for household per capita expenditure.[12] However, the Hausman-Wu test failed to reject exogeneity of LNPCE in the majority of states, and IV estimates were typically not much different from the OLS results.[13] In no state and age group did the coefficients of the age-gender variables change significantly from the OLS specification, and P-values of the F-tests (that the coefficients on male and female variables were equal) changed little, never altering a result from statistical insignificance to significance or vice versa. This suggests that endogeneity of per capita expenditure is not driving the results.

The coefficient on household size is positive and significant in every state.[14]

[11] In Subramanian's study, the total expenditure elasticities for Andhra Pradesh, Haryana, Maharashtra, Punjab, and Rajasthan were 2.14, 1.13, 1.79, 1.58, and 1.75, respectively. When we repeat our analysis to resemble Subramanian's, i.e., this time including households without children of school-going age, our estimated elasticities for the five states are 1.49, 1.41, 1.19, 1.17, and 1.08, respectively. That is, except for Haryana, the elasticities for the other four states are very considerably lower than in Subramanian (1995).

[12] Land is an illiquid asset and unlikely to be sold in response to short-term requirements such as schooling costs (land transactions are infrequent in rural India). Similarly, rental income (generated from illiquid assets such as property) is unlikely to change in the short run. A Sargan test of overidentifying restrictions showed that the instruments were accepted as valid in 12 out of the 16 states.

[13] A Hausman-Wu endogeneity test showed that LNPCE was accepted as exogenous in 11 of the 16 states, and a Hausman specification test for all coefficients (not just LNPCE coefficient) showed that OLS is a consistent estimator in the majority of (i.e., in 11) states.

[14] This is in line with theoretical considerations that suggest that, at any given level of per capita resources, larger households will be better off due to economies of scale that accrue from shared household public goods. The evidence of scale economies here is of interest given its usual elusiveness (Deaton and Paxson 1998). However, as mentioned in the methodology section, household size is potentially endogenous, since couples with a higher taste for schooling may choose to have both smaller families and higher education budget share. As we do not have data on households at two or more points in time, we cannot introduce household fixed effects to address the endogeneity of household size here. However, Sec. IV.D reports results of individual-level education expenditure

Household head's schooling (HEDYRS) increases the budget share of education very significantly across all sample states, indicating a higher "taste"/demand for child schooling among more educated households. The effects of caste and occupation are generally not significant or consistent across states. However, religion matters. Even after controls for household per capita expenditure and head's education, Muslim households have significantly lower education budget shares than Hindus and Sikhs (the omitted category) in Andhra Pradesh, Himachal Pradesh, Karnataka, Orissa, Uttar Pradesh, West Bengal, and Assam. The parameters of the gender- and age-composition variables (M5to9, F5to9, M10to14, etc.) show that education budget share generally increases with proportion of male and female children of school-going age within the household.

What does the fitted conventional Engel curve in each state tell us about gender bias in the within-household allocation of educational expenditure? The *P*-values of the *F*-test of the null hypothesis that the coefficients on the male and female demographic variables are equal are presented in the last three rows of table A1. The row for the *P*-value for age 5–9 of the first columns under each state shows that, in the 5–9 age group, the hypothesis that the coefficient on M5to9 (the male demographic variable for the 5–9 age group) is the same as the coefficient on F5to9 (the female demographic variable for the 5–9 age group) is rejected at the 5% significance level only for Madhya Pradesh, Rajasthan, and Uttar Pradesh. This lack of evidence of significant gender bias in all but three states shows that the conventional Engel curve technique is not good at picking up gender-differentiated treatment in educational expenditure within households, given that enrollment data show significant gender differences in nine out of the 16 states in the 5–9 age group (table 2).

Next, in attempting to examine why the Engel curve method fails to detect gender bias, I unpack total household education budget share into its two underlying components, using the hurdle model outlined earlier. The second and third columns under each state in table A1 present equations, respectively, for (*a*) the probability that the household budget share of education is positive (the probit equation of ANYEDEXP) and (*b*) the natural log of education budget share, conditional on positive spending (the conditional OLS equation). In the conditional budget share equation, sample selection could be a problem.

equations and confirms that the findings of gender bias survive the introduction of household fixed effects.

The likely direction of selectivity bias was discussed in the methodology section.[15]

In table A1, it is conspicuous in the second and third columns under each state that some variables have opposing effects on the two outcomes. For example, the effect of log of household per capita expenditure (LNPCE) is invariably positive and highly significant in the probit of ANYEDEXP in all states, but it is invariably negative and often highly significant in the conditional OLS of budget share. As per Engel's law, this is as expected. While the household size variable (LNHHSIZE) has a large positive and significant effect on the probability of spending a positive sum on education, its effect on the conditional budget share is small and typically insignificant.[16]

Of most interest, from the point of view of the central question about gender bias, is the impact on the two outcomes of the demographic variables M5to9 and F5to9 (household's proportion of males and females aged 5–9), M10to14 and F10to14 (proportion of males and females aged 10–14), and M15to19 and F15to19 (proportion of males and females aged 15–19). To investigate this impact, we compute the marginal effects of the male and female demographic variables in each equation and then take the difference between the male and female marginal effects. For example, in any given equation, the marginal effect of the variable M5to9 minus the marginal effect of the variable F5to9 is the difference in marginal effect (DME) of the gender variables in the 5–9 age group.

Table 5 presents the difference in marginal effects (DME) of the demographic variables for the 5–9, 10–14, and 15–19 age groups, respectively, calculated from the results in table A1. The figures in parentheses below each DME are the *P*-values of the *F*-test that the DME is equal to zero. The *P*-values of statistically significant DMEs (at the 5% level or better) are identified with an asterisk. The meaning of the DME is best illustrated with an example. For instance, in the probit of ANYEDEXP in Bihar in table A1, the marginal effect of the variable M5to9 was 0.6408 and the marginal effect of F5to9 was 0.4437.

[15] As suggested by a referee, we controlled for sample selectivity by using caste and religion variables as identifying exclusion restrictions. We used these where it was empirically justifiable, i.e., in states where caste and/or religion mattered to the choice of positive versus zero education expenditure but did not matter to the conditional education expenditure. However, even in the few cases where the selectivity term was statistically significant, the correction did not change any of our inferences: the results of the *F*-tests (in the conditional LNESHARE column under each state) in the last three rows of table A1 never changed from significant to insignificant or vice versa in these states.

[16] The marginal effects of the demographic variables are sometimes above one because these variables take values from zero to one rather than from one to 100. Redefining them to be bounded by one and 100 simply leads to the reported marginal effects being divided by 100.

TABLE 5

DIFFERENCE IN MARGINAL EFFECT (DME) × 100 OF GENDER VARIABLES BY AGE GROUP AND *P*-VALUES OF THE ASSOCIATED *T*-TESTS (HOUSEHOLD-LEVEL RESULTS)

	Males Ages 5–9 and Females Ages 5–9				Males Ages 10–14 and Females Ages 10–14				Males Ages 15–19 and Females Ages 15–19			
State	Probit (1)	Conditional OLS (2)	Combined Probit + OLS (3) = *f*(1, 2)	Unconditional OLS (Conventional Engel Curve) (4)	Probit (1)	Conditional OLS (2)	Combined Probit + OLS = *f* (3) = *f*(1, 2)	Unconditional OLS (Conventional Engel Curve) (4)	Probit (1)	Conditional OLS (2)	Combined Probit + OLS = *f* (3) = *f*(1, 2)	Unconditional OLS (Conventional Engel Curve) (4)
AP	66.23	−.17	2.31	.48	35.28	1.22	2.17	2.36	73.52	1.16	3.53	2.99
	(.00)*	(.90)	(.05)*	(.70)	(.06)	(.38)	(.06)	(.07)	(.00)*	(.55)	(.05)*	(.03)*
BIH	19.71	−.51	.65	.56	46.00	−2.24	.86	−.52	80.52	−.18	3.86	4.21
	(.24)	(.76)	(.64)	(.65)	(.01)*	(.17)	(.54)	(.69)	(.00)*	(.94)	(.04)*	(.01)*
GUJ	48.30	−.14	1.47	.43	65.55	−3.69	−.69	−.98	39.06	4.01	4.35	4.56
	(.04)*	(.93)	(.35)	(.79)	(.01)*	(.01)*	(.63)	(.55)	(.09)	(.02)*	(.00)*	(.01)*
HAR	24.07	2.63	4.04	3.54	21.01	.47	1.96	−1.58	39.13	3.28	5.71	5.97
	(.11)	(.25)	(.09)	(.08)	(.22)	(.82)	(.37)	(.43)	(.01)*	(.25)	(.03)*	(.01)*
HIM	13.65	−2.97	−1.66	.55	12.93	−.43	.70	.47	19.17	8.82	10.08	10.86
	(.29)	(.27)	(.58)	(.83)	(.31)	(.86)	(.78)	(.84)	(.02)*	(.00)*	(.00)*	(.00)*
KAR	−3.47	−.80	−.82	−.13	23.73	.80	1.80	1.48	38.72	3.81	4.97	5.89
	(.80)	(.59)	(.56)	(.93)	(.09)	(.57)	(.15)	(.29)	(.01)*	(.03)*	(.00)*	(.00)*
KER	6.36	1.29	2.02	1.59	−8.98	−1.41	−2.45	−1.08	.86	−1.99	−1.80	.00
	(.33)	(.57)	(.48)	(.50)	(.45)	(.51)	(N.A.)	(.63)	(.88)	(.41)	(.52)	(.99)
MAH	16.39	−1.34	−.07	.89	52.03	−.68	2.62	1.18	79.64	2.16	6.59	7.27
	(.26)	(.37)	(.96)	(.50)	(.00)*	(.63)	(.09)	(.36)	(.00)*	(.20)	(.00)*	(.00)*
MP	27.59	.05	1.05	1.66	65.66	.54	2.80	.82	88.06	4.58	6.55	5.00
	(.02)*	(.96)	(.18)	(.03)*	(.00)*	(.57)	(.00)*	(.31)	(.00)*	(.00)*	(.00)*	(.00)*

ORI	77.58	−2.00	1.46	−.26	72.55	.15	2.82	−.55	70.45	5.43	6.53	4.99
	(.00)*	(.11)	(.31)	(.83)	(.00)*	(.91)	(.04)*	(.69)	(.00)*	(.00)*	(.00)*	(.00)*
PUN	38.36	3.27	4.89	2.98	9.99	1.99	2.22	2.81	22.70	3.63	4.29	2.48
	(.06)	(.33)	(.08)	(.18)	(.66)	(.55)	(.47)	(.22)	(.22)	(.33)	(.12)	(.26)
RAJ	55.88	2.09	3.88	4.02	130.48	3.64	8.12	6.54	105.99	7.72	10.15	8.82
	(.00)*	(.10)	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*
TN	52.10	.44	2.68	2.55	43.35	1.61	3.21	2.23	17.23	5.04	4.76	1.67
	(.02)*	(.83)	(.15)	(.17)	(.09)	(.44)	(.12)	(.26)	(.44)	(.05)*	(.06)	(.39)
UP	45.77	.26	2.35	2.30	78.53	−.40	3.41	1.26	59.92	6.70	7.77	7.21
	(.00)*	(.82)	(.01)*	(.01)*	(.00)*	(.72)	(.00)*	(.19)	(.00)*	(.00)*	(.00)*	(.00)*
WB	16.01	−1.77	−.59	−.13	−6.14	.70	.24	−1.17	34.73	3.74	3.73	2.02
	(.41)	(.29)	(.67)	(.92)	(.78)	(.67)	(.86)	(.44)	(.18)	(.08)	(.04)*	(.24)
ASS	−3.82	.96	.69	−.67	18.93	1.30	1.81	1.97	−16.08	3.59	2.42	3.95
	(.84)	(.55)	(.67)	(.67)	(.55)	(.51)	(.34)	(.35)	(.59)	(.17)	(.38)	(.11)

Note. In the conditional OLS equation fitted only for households with positive education spending, the dependent variable is the natural log of the household education budget share. Thus, the coefficients of the gender dummy variables were transformed so that the marginal effects reported in col. 2 are comparable to those in col. 4, where the dependent variable is in absolute rather than log terms. Col. 4 pertains to the unconditional OLS of absolute household education budget share, fitted on all households, including those with zero education budget shares. The table displays 100 times the difference in marginal effects (DME) of the variables "proportion of males aged 5–9" and "proportion of females aged 5–9," etc. The figures in parentheses are *P*-values of the *t*-test of the DME, where standard errors for the *t*-test in each cell of col. 3 were obtained by bootstrapping with 500 replications.

* Statistically significant at the 5% level.

Thus the gender DME in the 5–9 age group was 0.1971. Table 5 shows this difference multiplied by 100, that is, as 19.71. The *P*-value of the *F*-test that this difference is equal to zero was 0.24, that is, this gender difference in marginal effect is statistically insignificant. In table 5, the probit results in column 1 refer to male-female DME from the probit of whether the household had a positive education budget share. Column 2 refers to the male-female DME in the conditional OLS of the log of education budget share (LNESHARE). Since the dependent variable here is in logs, the marginal effects of the male and female demographic variables were transformed before taking differences, so that the DMEs reported in column 2 are comparable to those in column 4, where the dependent variable was absolute ESHARE.[17] Column 3 shows the DME of the combined marginal effects from the probit and conditional OLS equations, the combined marginal effect having been derived in the way shown in equation (7). Column 4 pertains to the unconditional OLS results, that is, the OLS of the absolute budget share of education fitted on all (including zero education expenditure) households—the commonly reported Engel curve equation.

Table 5 demonstrates two interesting facts. First, the DME is almost always positive in the probit. That is, in most cases, having an extra boy in the household has a greater positive impact on the probability of having ANYEDEXP than having an extra girl in the household. Second, the gender DME is often negative in the conditional OLS in the 5–9 and 10–14 age groups (though not in the 15–19 group). Thus, in the basic education age group (ages 5–14), in many states, there is slight pro-female bias in conditional education budget share: having an extra girl in the household increases the conditional household budget share of education more than having an extra boy in the household. This could be because certain costs of girls' education are somewhat greater than those for boys.[18]

[17] For example, the coefficient on the variable M5to9 in the conditional OLS of LNESHARE for Gujarat is −0.87 and the coefficient on F5to9 is −0.83. The log transforms of these are obtained by using the property of the log normal distribution that the conditional expectation of $E(s \mid x, s > 0)$ equals $\exp(x\beta + \sigma^2/2)$. For the Gujarat conditional log expenditure equation, exp(.) is equal to 0.03458. Thus the marginal effect of M5to9 is $b \times \exp(.)$, i.e., it is $-0.87 \times 0.03458 = -0.0301$; the marginal effect of F5to9 is $-0.83 \times 0.03458 = -0.0287$. The gender difference in marginal effect for the 5–9 age group in Gujarat in the conditional OLS of budget share (as opposed to the log of budget share) is thus $(-0.0301) - (-0.0287) = -0.0014$. In table 5, all (differences in) marginal effects are multiplied by 100, so this appears as −0.14.

[18] For instance, girls' school clothes may cost more since girls should be well covered. However, there is no consistent evidence of systematically greater expenditure on girls than boys in particular education expenditure categories. In the questionnaire, tuition fee and school uniform are lumped together in one category, so we cannot check if more is spent on a girl's school uniform than a boy's. In the 5–9, 10–14, and 15–19 age groups, mean transport costs are higher for girls than boys in only three, five, and six of the 16 states, respectively.

In the 5–9 age group of table 5, the gender DME in the probit is positive for all states except two (Karnataka and Assam) and is statistically significant in seven states. In eight out of the 16 states, the gender DME in the conditional OLS of LNESHARE is negative (albeit insignificant), and in no state is there significant pro-male gender bias in conditional education expenditure. The inference from the "conventional" Engel curve results, presented in column 4, is that there is no significant gender bias in education expenditure in the 5–9 age group in any state other than Madhya Pradesh, Rajasthan, and Uttar Pradesh. However, such an inference masks the fact that, in four states other than these, there is significant gender bias in the decision whether to enroll a child in school. To overlook the difference is to miss an important discriminatory process.

In the 10–14 age group of table 5, the gender DME in the probit is positive for all states except Kerala and West Bengal, and it is significant in seven states. But the DME from the conditional OLS is insignificant in all but two states. The conventional Engel curve result in column 4 would lead to the inference of no significant gender bias in any state other than Rajasthan. As in the 5–9 age group, such an inference would neglect the fact that in six states other than Rajasthan, there is significant bias in the enrollment decision. Age 10–14 results of table 5 also show that, using the hurdle model approach (col. 3), four states have significant gender bias in unconditional education expenditure. In other words, when the decision to incur positive education expenditure is modeled separately from the decision of how much to spend conditional on positive expenditure (using appropriate functional forms), we are more successful in "picking up" gender bias in education spending than with the conventional Engel approach that imposes linear regression of unconditional education expenditure.

In the 15–19 age group, both the DME in the probit and the DME in the conditional OLS are typically positive (significant only in 10 states in the probit and in eight states in the conditional OLS). Thus, unlike in the case of the 5–9 and 10–14 age groups, here both the probit and conditional OLS results mostly work in the same direction, that is, they reinforce each other.

It is not clear what explains the lack of significant gender difference in conditional education expenditure in the primary and junior high school age groups but its presence in the secondary school age group. One possibility might be that gender-differentiated treatment in conditional education expenditure only begins at the secondary school stage because at that stage children are closer to further education courses and to employment. However, at the secondary school stage, there may be supply-side reasons for not interpreting lower conditional educational expenditure on girls necessarily as evidence of parental discrimination. By the early 1990s, state-provided elementary education was tuition free for both genders, but this was not so for secondary education. However, certain states operated an affirmative action policy for girls at the secondary school stage by providing tuition-free secondary schooling to girls.[19] Thus, in these states, lower conditional education expenditure on girls cannot be taken as evidence of parental bias against girls. Moreover, the dearth of (and distance to) single-sex girls' secondary schools may deter parents from sending girls to school for safety and social reasons, rather than for reasons of discrimination; thus it is difficult to know what part of girls' observed inferior enrollment outcomes in the ages 15–19 range is due to parental discrimination and what is due to supply-side factors.

To sum up, the discussion so far suggests two conclusions. First, the Engel curve approach does not pick up gender bias partly because it uses the wrong functional form. It estimates a single budget share equation to encompass two different decisions: the binary decision of whether to make a purchase and the decision, conditional on purchase, of how much to spend on the good. If the correct functional form for the binary decision is nonlinear and the correct distribution of conditional expenditure is log normal rather than normal, then a hurdle model seems better able to capture gender biases in unconditional expenditure. Second, the discussion shows the importance of "unpacking" the total gender difference in expenditure into its two constituent parts—the difference due to a greater incidence of zero purchases for girls than boys and the difference due to lower conditional expenditures on girls than boys—so as to avoid lumping together two different (often divergent) processes. Averaging over the two dilutes the effect of the former difference, which is clearly the main discriminatory process. While averaging may lead to the conclusion of no pro-male bias, there is evidence of significant pro-male bias in one of the processes, and policy makers may be as concerned with the distribution of educational expenditure for girls and boys as with its average. Indeed, it is possible that, for children's long-term life chances, being in school is more important than expenditure on schooling once enrolled.

C. Is Aggregation the Reason for the Engel Curve Method's Failure to Detect Gender Bias?

We turn next to examine whether aggregation of data at the household level makes it more difficult to detect gender differences in educational expenditure

[19] Bihar, Haryana, Himachal, Maharashtra, Orissa, Rajasthan, and Assam provided free access to secondary education for girls.

than when using individual child-level data. Individual-level expenditure provides the most reliable way of detecting gender bias. As we have educational expenditure information at the level of the individual child and also, by aggregation, at the level of the household, it is possible to compare household-level Engel curve results with individual-level analysis. In the individual-level analysis, the dependent variable is education expenditure on the individual child (rather than household budget share of education). Moreover, instead of demographic variables such as household proportion of males aged 5–9 and household proportion of females aged 5–9, and so forth, the gender variable of interest is simply the dummy variable MALE, which is one for males and zero for females. The remainder of the explanatory variables in the individual-level equations are identical to those in the household equations of table A1, that is, they are household-level variables. The three age groups of interest, as before, are ages 5–9, ages 10–14, and ages 15–19, corresponding roughly with primary, junior high, and secondary education.

At the individual child level, we estimated 144 separate equations (16 states × 3 age groups × 3 equations). We do not display all 144 equations, but the marginal effects on the gender variable MALE from these equations are presented in table 6 for the three age groups.

The marginal effects on MALE in table 6 are not comparable with the difference in marginal effects of the household demographic variables in table 5. This is because the household demographic variables in a household-level regression are not identical to the dummy variable MALE in the individual-level regression. It is also because the dependent variable in the conditional and unconditional OLS equations in table 6 is education expenditure on the individual child but in table 5 the corresponding dependent variable is household education budget share. Thus, the scaling of the coefficients and marginal effects will be different in the two tables. However, we are interested mainly in whether any statistically significant gender differences in the individual-level table 6 are also significant in the household-level table 5.

The individual-level results of table 6 confirm what we saw earlier, namely, that, in each of the three age groups, much of the gender-differentiated treatment occurs at the stage of the decision whether to even incur positive education expenditure (enroll a child in school) and not in the decision of how much to spend, conditional on school enrollment. In some instances, the marginal effect of MALE in the conditional expenditure equation is negative, that is, girls have somewhat higher education expenditure, conditional on being in school, though this pro-female bias is never statistically significant.

Since MALE is a discrete variable, the marginal effect of MALE in the combined hurdle model (col. 3) is estimated by calculating the expected values

TABLE 6

MARGINAL EFFECT OF THE GENDER DUMMY VARIABLE MALE AND *P*-VALUE OF THE ASSOCIATED *T*-TEST (INDIVIDUAL-LEVEL DATA FOR THE THREE AGE GROUPS)

	Ages 5–9				Ages 10–14				Ages 15–19			
State	Probit (1)	Conditional OLS (2)	Combined Probit + OLS (3) = *f*(1, 2)	Unconditional OLS (4)	Probit (1)	Conditional OLS (2)	Combined Probit + OLS (3) = *f*(1, 2)	Unconditional OLS (4)	Probit (1)	Conditional OLS (2)	Combined Probit + OLS (3) = *f*(1, 2)	Unconditional OLS (4)
AP	.148	9.8	29.5	31.4	.214	13.0	49.7	61.0	.345	56.6	198.3	177.9
	(.00)*	(.30)	(.00)*	(.05)*	(.00)*	(.34)	(.00)*	(.00)*	(.00)*	(.44)	(.00)*	(.00)*
BIH	.126	15.9	33.8	38.3	.237	16.1	78.6	74.7	.337	−23.2	140.5	122.0
	(.00)*	(.24)	(.00)*	(.00)*	(.00)*	(.28)	(.00)*	(.00)*	(.00)*	(.51)	(.00)*	(.00)*
GUJ	.045	1.6	7.1	−1.5	.200	2.6	7.4	36.0	.252	212.2	203.6	203.0
	(.24)	(.86)	(.29)	(.93)	(.00)*	(.82)	(.21)	(.08)	(.00)*	(.01)*	(.00)*	(.00)*
HAR	.061	64.3	66.1	57.3	.210	68.3	177.3	165.7	.337	314.7	454.2	426.4
	(.04)*	(.02)*	(.00)*	(.02)*	(.00)*	(.02)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*
HIM	.052	30.5	34.1	40.0	.123	59.4	152.5	96.3	.340	149.8	588.5	548.5
	(.15)	(.26)	(.12)	(.28)	(.00)*	(.11)	(.00)*	(.06)	(.00)*	(.05)*	(.00)*	(.00)*
KAR	.060	10.0	18.3	23.4	.160	13.0	40.5	44.3	.206	32.4	148.6	141.2
	(.02)*	(.36)	(.04)*	(.15)	(.00)*	(.31)	(.00)*	(.01)*	(.00)*	(.50)	(.00)*	(.00)*
KER	.008	19.6	19.3	45.5	−.116	−17.5	−92.2	−40.8	.049	−41.7	27.7	−5.7
	(.84)	(.41)	(.21)	(.24)	(.01)	(.44)	(.15)	(.19)	(.33)	(.50)	(.82)	(.92)
MAH	−.001	−1.7	−1.2	5.3	.176	10.9	75.4	53.7	.391	59.9	207.3	210.0
	(.97)	(.87)	(.86)	(.53)	(.00)*	(.31)	(.00)*	(.00)*	(.00)*	(.15)	(.00)*	(.00)*
MP	.090	13.4	19.7	25.9	.255	23.8	67.6	46.3	.369	44.0	178.3	145.4
	(.00)*	(.04)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.18)	(.00)*	(.00)*

ORI	.062	−16.7	.95	−9.4	.283	2.9	64.8	45.7	.303	49.9	143.9	162.5
	(.05)*	(.07)	(.31)	(.46)	(.00)*	(.79)	(.00)*	(.00)*	(.00)*	(.29)	(.00)*	(.00)*
PUN	.095	100.3	114.6	105.1	.104	66.5	115.5	114.8	.287	51.7	345.3	264.8
	(.01)*	(.04)*	(.01)*	(.03)*	(.01)*	(.15)	(.00)*	(.02)*	(.00)*	(.62)	(.00)*	(.00)*
RAJ	.305	31.6	94.5	87.5	.617	60.5	289.6	243.8	.425	57.3	240.7	390.6
	(.00)*	(.02)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.61)	(.00)*	(.00)*
TN	.156	11.8	40.8	16.1	.183	18.5	66.4	56.0	.200	−43.6	100.0	62.9
	(.00)*	(.52)	(.10)	(.61)	(.00)*	(.38)	(.05)*	(.08)	(.00)*	(.59)	(.30)	(.16)
UP	.196	−5.7	50.5	41.8	.341	32.7	132.6	111.8	.358	86.0	223.7	234.2
	(.00)*	(.60)	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.00)*	(.02)*	(.00)*	(.00)*
WB	.012	−3.6	.12	1.2	.064	30.3	37.6	26.8	.160	11.6	88.9	154.6
	(.71)	(.75)	(.68)	(.93)	(.07)	(.11)	(.02)*	(.16)	(.00)*	(.84)	(.00)*	(.00)*
ASS	.076	.44	17.1	10.4	.096	9.9	27.5	6.9	.185	75.2	152.2	105.9
	(.04)*	(.97)	(.10)	(.62)	(.12)	(.59)	(.87)	(.77)	(.01)*	(.29)	(.09)	(.12)

Note. In the conditional OLS equation fitted only for children with positive education spending, the dependent variable is the natural log of education expenditure. Thus, the coefficients of the gender dummy variables were transformed so that the marginal effects reported in col. 2 are comparable to those in col. 4, where the dependent variable is in absolute rather than log terms. Col. 4 pertains to the unconditional OLS of absolute education expenditure, fitted on all children, including those with zero education expenditure. The table shows the marginal effect on the gender dummy variable MALE. The figures in parentheses are *P*-values of the *t*-test of the marginal effect of MALE, where standard errors for the *t*-test in each cell of col. 3 are obtained by bootstrapping with 500 replications.

* Statistically significant at the 5% level.

of unconditional expenditure in equation (5) with MALE = 1 and with MALE = 0 and then taking the difference, rather than by taking derivatives, as in equation (7).[20] Column 4 presents the marginal effect of the variable MALE in the unconditional expenditure equation, that is, the single OLS equation estimated including zero education expenditures.

While a comparison of columns 3 and 4 shows quite good correspondence between the two, the hurdle model is still more effective at picking up gender bias than the conventional unconditional OLS model. For example, in table 6, for the 5–9 age group, the hurdle model detects overall gender bias in Karnataka where the unconditional OLS fails to pick it up. The same is true for Himachal, Tamil Nadu, and West Bengal for the 10–14 age group in table 6.

The most noteworthy fact to emerge from a comparison of tables 5 and 6 is that the gender difference in education expenditure is statistically significant in many more states when individual-level data are used (table 6) than when household aggregated data are used (table 5). This may be taken to suggest that there is something in the aggregation that makes it more difficult to pick up gender differences in expenditure.

D. Is There Gender Inequality in Education Expenditure without There Being Gender Bias?

Jensen (2002) suggests that gender inequality in outcomes (such as education expenditure) could arise even in the absence of any parental bias against daughters. If parents' fertility behavior displays differential stopping rules after the birth of sons and daughters, then, even if parents treat all children equally, girls will have worse outcomes in the population as a whole simply because of across-household differences in household size, that is, because girls have more siblings and thus live in larger households.[21] If this is true, the substantial male-female differences in education expenditure observed so far may not represent parental bias per se. Since household size—an outcome of parental preferences—is endogenous in this situation, controlling for household size (as we have done so far) will not control adequately for this effect. However, the inclusion of family fixed effects in education outcome equations provides a powerful control for household unobserved factors such as parental preferences.

[20] However, when we estimated the marginal effects of the continuous gender variables M5to9 and F5to9, etc., in the Engel curve equation, using household-level data earlier in this article, we used derivatives as set out in eqq. (6) and (7).

[21] Households where early births were male offspring will have smaller household size and be better-off economically, implying higher expenditures on children of both sexes. Households in which early births were female children continue childbearing until a boy (or desired number of boys) is born. Such households will be larger due to there being more girls.

Table 7 presents the coefficients and t-values of the gender dummy variable MALE in three individual-level equations for each age group: (i) the probit equation of ANYEDEXP (whether any positive expenditure was incurred on a child's education), (ii) the equation of the natural log of education expenditure (LNEDEXP) conditional on positive education expenditure, and (iii) the equation of the unconditional education expenditure (EETOTAL). The equations were fitted on the subset of households that had at least one child of each gender in the relevant age group. Table 7 shows significant within-household gender difference in education expenditures in the 5–9, 10–14, and 15–19 age groups in most of the states where there are significant gender differences in the raw descriptive statistics of tables 2–4. Thus, most of the observed gender differences can indeed be interpreted as differential treatment of sons and daughters by parents within the home, rather than as arising from across household differences in household size.

V. Conclusion

The individual-level data on educational expenditures confirm that (i) in Indian states with the most skewed sex ratios, educational outcomes, such as school enrollment rates for girls, are significantly worse than those for boys and that (ii) in those Indian states where there is evidence of significantly worse educational outcomes for girls than boys, household expenditure on girls' education is indeed significantly lower than that on boys', that is, lower educational inputs are an important mechanism by which girls' educational outcomes turn out to be inferior than boys'. The data show that the most important way in which gender bias in educational resource allocation manifests itself in rural Indian households is via nonenrollment of girls, which implies zero educational spending. There is little gender bias in educational expenditure among enrolled children.

The existing explanations for why the household expenditure methodology fails to find gender bias include the suggestion that there may not, in fact, be any within-household bias against girls in education expenditure, as discussed earlier (Rose 1999; Jensen 2002). However, since individual-level data and household fixed-effects estimations here show pervasive within-household gender differences in education expenditure in rural India, this article has sought to test alternative potential explanations for the Engel curve method's inability to detect bias in intrahousehold allocation.

The analysis here shows a low degree of correspondence between results in individual-level and household-level data. Particularly in the 5–9 and 10–14 age groups, the household expenditure method fails to find significant discrimination. Tests suggest that the failure is partly because the Engel curve method as conventionally applied suffers from an incorrect functional form and the

TABLE 7

COEFFICIENT OF THE MALE DUMMY IN INDIVIDUAL-LEVEL EQUATIONS WITH FAMILY FIXED EFFECTS

	Ages 5–9			Ages 10–14			Ages 15–19		
State	ANYEDEXP	LNEDEXP	EETOTAL	ANYEDEXP	LNEDEXP	EETOTAL	ANYEDEXP	LNEDEXP	EETOTAL
Andhra	.048	.091	.2	.165	.080	46.0	.184	−.053	153.8
	(1.4)	(1.3)	(.0)	(4.3)*	(.9)	(2.4)*	(5.2)*	(.3)	(3.2)*
Bihar	.102	.087	63.0	.184	.171	81.2	.257	.190	159.9
	(3.9)*	(1.3)	(3.6)*	(6.9)*	(2.6)*	(5.1)*	(6.9)*	(1.8)	(5.1)*
Gujarat	.074	.008	20.9	.155	.053	62.2	.163	.221	238.6
	(2.0)*	(.2)	(1.7)	(4.6)*	(.7)	(1.7)	(4.0)*	(1.2)	(2.5)*
Haryana	.071	.255	90.4	.154	.104	212.1	.312	.178	478.8
	(2.5)*	(3.5)*	(3.6)*	(4.7)*	(2.1)*	(3.8)*	(8.1)*	(1.0)	(6.0)*
Himachal	.091	.059	77.1	.065	.098	153.5	.265	.047	548.5
	(2.4)*	(1.0)	(1.9)	(2.5)*	(1.9)	(1.7)	(5.8)*	(.4)	(4.4)*
Karnatak	−.014	.007	−20.6	.114	.017	26.1	.184	.047	167.7
	(.5)	(.1)	(1.0)	(4.4)*	(.3)	(.9)	(4.8)*	(.7)	(3.9)*
Kerala	−.003	.090	52.2	−.048	.019	−33.4	−.062	−.210	−70.5
	(.1)	(1.2)	(1.2)	(1.6)	(.3)	(.7)	(1.2)	(1.5)	(.9)
Maharashtra	.037	.030	19.3	.115	.089	58.4	.281	−.006	241.2
	(1.2)	(.5)	(1.9)	(4.3)*	(2.0)*	(3.5)*	(6.5)*	(.0)	(4.2)*
Madhya	.081	.138	29.7	.178	.046	56.7	.251	.061	161.6
	(3.9)*	(2.7)*	(4.2)*	(8.2)*	(1.1)	(6.4)*	(10.4)*	(.6)	(7.3)*

Orissa	.001	−.087	−1.5	.203	.106	51.4	.139	.005	87.3
	(.0)	(1.1)	(.1)	(5.8)*	(1.9)	(3.3)*	(3.5)*	(.0)	(2.4)*
Punjab	.051	.122	82.2	.099	.072	69.9	.241	−.055	284.1
	(1.2)	(1.2)	(2.5)*	(2.6)*	(1.4)	(1.5)	(5.9)*	(.8)	(3.4)*
Rajasthan	.258	.120	90.5	.524	.094	256.6	.286	.084	328.9
	(9.5)*	(2.3)*	(4.6)*	(17.3)*	(1.4)	(9.4)*	(8.0)*	(.6)	(5.4)*
Tamilnadu	.138	.132	34.4	.112	.077	44.8	.225	−.161	212.9
	(2.2)*	(1.4)	(.9)	(2.3)*	(1.1)	(1.8)	(4.0)*	(.7)	(2.8)*
Uddar	.146	.065	45.6	.294	.111	135.6	.269	.139	210.6
	(7.1)*	(1.3)	(3.0)*	(13.8)*	(3.0)*	(10.8)*	(10.5)*	(1.3)	(6.9)*
West Bengal	−.050	.030	−13.2	.111	−.004	13.6	.121	.045	104.9
	(1.5)	(.4)	(.6)	(3.2)*	(.1)	(.6)	(2.7)*	(.3)	(2.4)*
Assam	.021	.066	2.9	.030	.054	12.7	.025	.009	27.9
	(.6)	(1.2)	(.1)	(1.1)	(.7)	(.7)	(.5)	(.1)	(.4)

Note. The three individual-level equations for each age group are (i) the probit equation of ANYEDEXP (whether any positive expenditure was incurred for the index child's education), (ii) the equation of the natural log of education expenditure (LNEDEXP) conditional on ANYEDEXP being positive, and (iii) the unconditional education expenditure equation (EETOTAL). The right-hand-side variables are age and the gender dummy MALE. The equations are fitted only on that subset of households that have at least one child of each gender in the relevant age group. The average number of households per state in the fixed effects estimation of EETOTAL and ANYEDEXP is 233, 203, and 153 in the 5–9, 10–14, and 15–19 age groups, respectively. In the conditional LNEDEXP equation, the average number of households per state in the fixed effects estimation in these three age groups is 173, 168, and 77, respectively. The figures in parentheses are *t*-statistics.

* Statistically significant at the 5% level.

limitation that the effects of the household gender composition variables on both (*a*) the decision to enroll in school and (*b*) the decision of how much to spend—conditional on enrolling—are constrained to be in the same direction. Our data suggest that the effects are in divergent directions in a substantial number of cases in the primary and junior high school age groups. However, in the 15–19 age group, these two effects work in the same direction and tend to reinforce each other. Thus, it is only in this group that results from the Engel curve method correspond well with the results from the direct inspection of individual-level expenditure. Given that the two processes of discrimination often diverge, neither the unconditional OLS nor the Tobit are appropriate modeling strategies. The hurdle model has greater power to detect discrimination.

The results also suggest that aggregation of data at the household level makes it more difficult to pick up gender differences. Even when individual- and household-level variables and equations are made as similar as possible, household-level equations consistently fail to capture the full extent of the gender bias. This suggests that aggregation of data does prevent the household expenditure method from detecting gender bias and that this is not due to measurement error in the household expenditure variable. We are left with the conclusion that, for those concerned with reliably measuring the extent of gender discrimination in household expenditure allocation, household-level data are a poor substitute for individual-level expenditure data. Household expenditure data are of some use providing that one models the hurdle, but it still understates the extent of the problem of gender discrimination.

The results here highlight that there are two distinct processes by which gender bias occurs in the within-household allocation of educational expenditure. Thus, a method that integrates/jointly models these two processes dilutes the powerful gender differentiation that exists in many states in the main discriminatory mechanism, namely, the nonenrollment of girls. This insight may be generalizable to some other goods. For instance, it is possible that this is also the reason why no significant or consistent evidence of gender bias has been detected in medical expenditures in India (Subramanian and Deaton 1990; Subramanian 1995). It is fairly plausible to imagine scenarios whereby parents delay seeking medical care for girls as compared with boys in the same state of illness but, conditional on seeking medical advice, the expenditure on girls is the same as that on boys.[22] Policy makers may be as,

[22] The results suggest that, where the conventional Engel method shows significant bias, this is indicative of the existence of strong gender differences in education expenditure. The findings also have implications for the adult good method of detecting gender bias. Any parental response to additional boys and girls could (i) modify their purchasing behavior (e.g., whether or not they purchase an adult good such as alcohol, cigarette, or pan) and/or (ii) modify their conditional

or even more, concerned with the former source of bias, since it may be more important for children's longer-term life chances.

Our discussion also points out the need to consider the supply side when investigating household expenditures on particular commodities. If certain facilities and institutions (such as schools or health clinics) are not locally available and there are social taboos or difficulties about girls' use of nonlocal facilities, or if there are affirmative action policies in place for girls' health or their participation in certain levels of education, household expenditures on girls may be lower not due to parental discrimination per se but rather due to these supply-side conditions.

While our data show very significantly lower educational allocations to girls than boys in rural India, explanations underlying these differential allocations are not explored here. Gender-differentiated treatment could be due to son preference or due to an investment motive. The investment motive attributes unequal allocations to the differential returns of girls and boys, or differential returns accruing to parents. Differential returns may arise from dowry, different labor returns of males and females, or patrilocal family structure (Rose 2000). Foster and Rosenzweig (2000) find that, where there are economic returns to women's human capital, parents do invest in girls' education. Estimates for urban India suggest that women face lower economic returns to education than men (Kingdon 1998).[23] Further evidence on returns to men and women's education in the rural Indian labor market would be useful in analyzing whether gender bias in intrahousehold educational resource allocation in rural India is attributable to gender differentials in the returns to education.

amount expended on the adult good. That is, biased parents could adjust adult good consumption (via either or both of the above two mechanisms) more for an additional boy than for an additional girl. If additional children affect adult good consumption mainly via mechanism i and not via ii, then, as here, averaging across the two mechanisms could potentially dilute the capacity of the adult good method to detect gender bias. In any case, discovering that parents are more willing to adjust purchasing behavior for boys than girls would be indicative of bias and thus be of interest in its own right.

[23] Indian estimates in Kingdon (1998) do not conform to the worldwide pattern (noted in Schultz 2002) that returns to women's education are generally comparable to or higher than those to men's. Duraisamy (2002) and Kingdon and Unni (2001) find mixed evidence on returns to men and women's education in India, though neither study controlled for omitted family background bias, which, in Kingdon (1998), substantially reduces women's returns but not men's.

Appendix

TABLE A1

OLS REGRESSION OF BUDGET SHARE OF EDUCATION; BINARY PROBIT OF ANY EDUCATION EXPENDITURE; AND OLS REGRESSION OF NATURAL LOG OF BUDGET SHARE OF EDUCATION, CONDITIONAL ON POSITIVE EDUCATION (ALL EQUATIONS WITH VILLAGE FIXED EFFECTS)

	Andhra Pradesh						Bihar					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	.58	1.8	.22	4.4	−.39	−3.8	1.15	4.1	.32	7.9	−.35	−4.5
LNHHSIZE	1.86	5.9	.47	9.2	−.05	−.5	1.65	5.0	.49	9.8	−.01	−.1
M0TO4	−5.24	−2.3	−.98	−3.0	−1.36	−1.8	−2.34	−.8	−.58	−1.5	−.41	−.5
M5TO9	3.36	1.5	1.58	4.7	.19	.3	3.37	1.2	.64	1.8	.31	.4
M10TO14	4.08	1.8	.81	2.5	.83	1.2	7.48	2.7	.94	2.5	1.27	1.6
M15TO19	5.52	2.4	−.07	−.2	2.32	3.2	7.19	2.6	.39	1.0	1.90	2.4
M20TO24	−2.69	−1.1	−.47	−1.4	.52	.6	3.88	1.3	−.42	−1.0	1.33	1.5
M25TO60	−1.31	−.6	−.43	−1.3	−.42	−.6	1.23	.4	−.43	−1.1	.24	.3
M61MORE	−3.07	−1.1	−.69	−1.7	.55	.6	4.47	1.2	−.31	−.6	.79	.7
F0TO4	−4.33	−1.9	−.41	−1.3	−1.62	−2.2	−2.55	−.9	−.59	−1.6	−.70	−.9
F5TO9	2.88	1.3	.91	2.8	.24	.3	3.93	1.4	.44	1.2	.41	.5
F10TO14	1.73	.8	.45	1.4	.50	.7	8.01	2.9	.48	1.3	1.73	2.1
F15TO19	2.53	1.1	−.81	−2.5	2.00	2.7	2.98	1.0	−.41	−1.0	1.93	2.3
F20TO24	.94	.4	−.12	−.3	−.14	−.2	−.08	.0	−.65	−1.6	.59	.7
F25TO60	2.67	1.2	.60	1.9	.57	.8	.12	.0	−.05	−.1	.01	.0
HEDYRS	.41	9.9	.06	8.3	.08	6.7	.36	10.2	.06	10.9	.05	5.5
SC	−.52	−2.1	.01	.3	−.22	−2.8	−.79	−2.8	−.07	−1.9	−.17	−2.2
ST	−.40	−.5	−.03	−.3	.00	.0	−.37	−.8	−.11	−1.7	−.02	−.2
MUSLIM	−.90	−1.8	.00	.1	−.31	−2.1	.36	1.0	−.11	−2.3	.13	1.3
CHRISTN	−.03	.0	.07	.7	−.27	−1.2	1.34	1.0	.29	1.9	−.01	.0
INTERCEPT	−7.68	−2.1			−1.26	−1.1	−13.24	−3.6			−1.85	−1.8
Adjusted R^2	.2056		.3661		.3935		.2979		.3294		.3422	
N	1,571		1,548		1,001		1,787		1,768		1,042	
Dependent variable mean	.0269		.6415		−3.7624		.0346		.5865		−3.2549	
Expenditure elasticity	1.22						1.33					
P-value:												
Age 5–9	.70		.00		.90		.65		.24		.76	
Age 10–14	.07		.06		.38		.69		.01		.17	
Age 15–19	.03		.00		.55		.01		.00		.94	

	Gujarat						Haryana					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	1.36	3.9	.18	3.8	−.08	−.8	2.19	4.9	.24	7.2	−.07	−1.0
LNHHSIZE	1.40	3.6	.51	8.2	−.10	−.8	2.84	6.2	.36	9.8	.06	.8
M0TO4	−2.68	−.9	−.94	−2.3	−.92	−1.1	−6.88	−2.0	−.30	−1.3	−1.46	−2.5
M5TO9	−.35	−.1	1.06	2.7	−.87	−1.1	9.06	2.7	.81	3.4	.39	.7
M10TO14	3.20	1.2	1.40	3.6	.05	.1	10.88	3.2	1.03	4.0	.97	1.7
M15TO19	4.50	1.7	.07	.2	1.87	2.5	11.36	3.4	.28	1.2	1.38	2.4
M20TO24	−1.98	−.7	−.97	−2.4	.26	.3	.37	.1	−.21	−.9	.39	.6
M25TO60	−3.51	−1.2	−.16	−.4	−.62	−.8	.12	.0	−.09	−.4	−.20	−.3
M61MORE	−4.59	−1.2	−1.22	−2.1	−.48	−.4	−2.21	−.5	−.21	−.6	.00	.0
F0TO4	−4.34	−1.6	−.75	−1.9	−1.62	−2.1	−4.82	−1.4	−.17	−.7	−1.46	−2.4
F5TO9	−.78	−.3	.58	1.5	−.83	−1.1	5.52	1.6	.57	2.4	.03	.1
F10TO14	4.18	1.5	.74	1.8	1.12	1.5	12.46	3.8	.82	3.5	.91	1.6
F15TO19	−.06	.0	−.32	−.8	.71	.9	5.39	1.5	−.12	−.5	.93	1.6
F20TO24	−2.29	−.8	−.91	−2.1	−1.23	−1.3	−1.95	−.5	−.32	−1.3	−.12	−.2
F25TO60	−.07	.0	−.20	−.5	−.18	−.2	2.57	.8	.26	1.2	.07	.1
HEDYRS	.29	5.2	.05	5.7	.04	2.5	.35	6.6	.02	3.9	.05	6.1
SC	−.85	−2.1	−.05	−.7	−.18	−1.6	−.44	−1.2	−.07	−2.5	−.06	−1.0
ST	−.25	−.5	−.12	−1.7	−.16	−1.2	1.37	.6	…	…	.39	1.2
MUSLIM	1.03	1.6	−.09	−.9	.16	.8	−.89	−.8	−.07	−.9	.00	.0
CHRISTN	3.77	1.9	…	…	.42	.9	−4.37	−.7	…	…	…	…
INTERCEPT	−10.38	−2.6			−2.77	−2.5	−19.06	−3.9			−2.15	−2.5
Adjusted R^2	.3286		.4252		.4515		.2560		.3708		.2793	
N	1,182		1,083		776		1,409		1,361		1,074	
Dependent variable mean	.0317		.6463		−3.6331		.0614		.7546		−2.8315	
Expenditure elasticity	1.43						1.36					
P-value:												
Age 5–9	.79		.04		.93		.08		.11		.25	
Age 10–14	.55		.01		.01		.43		.22		.82	
Age 15–19	.01		.09		.02		.01		.01		.25	

TABLE A1 (*Continued*)

Variable	Himachal Pradesh: Unconditional OLS (ESHARE) Coefficient × 100	t-Value	Himachal Pradesh: Probit (ANYEDEXP) Marginal Effect	t-Value	Himachal Pradesh: Conditional OLS (LNESHARE) Coefficient	t-Value	Karnataka: Unconditional OLS (ESHARE) Coefficient × 100	t-Value	Karnataka: Probit (ANYEDEXP) Marginal Effect	t-Value	Karnataka: Conditional OLS (LNESHARE) Coefficient	t-Value
LNPCE	−.30	−.6	.07	3.0	−.40	−6.1	.48	1.4	.19	5.6	−.29	−3.9
LNHHSIZE	3.48	6.2	.15	5.8	.13	1.8	1.75	5.1	.36	10.4	.01	.1
M0TO4	−11.02	−3.0	−.32	−2.1	−2.14	−4.8	−2.86	−1.1	−.38	−1.7	−1.48	−2.6
M5TO9	5.05	1.5	.39	2.3	−.04	−.1	4.48	1.8	.84	3.8	.09	.2
M10TO14	13.83	4.3	.29	1.9	1.34	3.5	10.37	4.5	1.11	5.1	1.40	2.7
M15TO19	16.11	5.0	−.03	−.2	1.93	4.9	9.09	3.8	.24	1.1	1.95	3.7
M20TO24	−4.02	−1.1	−.33	−2.3	.07	.2	5.08	2.0	−.33	−1.5	1.56	2.7
M25TO60	−3.59	−1.1	−.24	−1.8	−.54	−1.4	1.68	.7	−.25	−1.1	.55	1.0
M61MORE	−7.90	−2.0	−.25	−1.4	−.61	−1.2	.00	.0	−.27	−.9	.36	.5
F0TO4	−7.91	−2.1	−.24	−1.5	−1.70	−3.6	−3.49	−1.3	−.41	−1.8	−.92	−1.6
F5TO9	4.50	1.3	.25	1.7	.30	.7	4.61	1.9	.87	3.9	.25	.5
F10TO14	13.36	4.0	.16	1.1	1.39	3.5	8.89	3.8	.88	4.0	1.24	2.4
F15TO19	5.25	1.6	−.22	−1.6	.92	2.3	3.20	1.3	−.15	−.7	1.19	2.2
F20TO24	−.17	−.1	−.14	−.9	.13	.3	2.64	1.0	−.28	−1.2	.47	.8
F25TO60	−5.79	−1.9	.00	.0	−.63	−1.7	3.21	1.4	.06	.3	.25	.5
HEDYRS	.32	4.6	.01	2.2	.05	5.5	.22	5.1	.02	3.7	.04	4.4
SC	−.95	−1.9	−.04	−1.6	−.03	−.6	−.10	−.3	−.08	−2.5	−.04	−.5
ST	−1.52	−1.3	−.15	−1.9	−.16	−1.1	−.99	−2.4	−.04	−1.0	−.26	−2.9
MUSLIM	−4.17	−3.0	−.45	−3.4	−.14	−.8	−.73	−1.8	−.10	−2.5	−.10	−1.2
CHRISTN	…	…	…	…	…	…	.21	.2	−.18	−1.1	.16	.5
INTERCEP	6.64	1.1			.94	1.3	−1.55	−.4			−.92	−1.1
Adjusted R^2	.3631		.4136		.4386		.2402		.2983		.4275	
N	949		725		838		1,979		1,850		1,435	
Dependent variable mean	.0868		.8469		−2.5700		.0427		.7216		−3.3266	
Expenditure elasticity	.97						1.11					
P-value:												
Age 5–9	.83		.29		.27		.93		.80		.59	
Age 10–14	.84		.31		.86		.29		.09		.57	
Age 15–19	.00		.02		.00		.00		.01		.03	

	Kerala						Maharashtra					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	1.42	2.4	.05	3.1	−.07	−1.0	−.67	−2.3	.20	6.3	−.49	−8.2
LNHHSIZE	2.52	4.1	.13	7.5	−.08	−1.0	1.63	5.3	.37	10.1	.01	.1
M0TO4	−5.52	−1.6	−.09	−1.2	−1.00	−2.2	−5.76	−2.7	−.42	−1.9	−2.20	−4.7
M5TO9	10.65	3.2	.48	5.2	.81	2.0	5.65	2.8	1.18	5.4	−.11	−.3
M10TO14	17.74	5.7	.61	6.1	1.47	3.8	10.12	4.9	1.22	5.5	1.11	2.6
M15TO19	14.49	4.4	.25	3.0	1.50	3.7	11.48	5.6	.66	3.2	1.65	3.9
M20TO24	1.53	.5	.06	.8	.21	.5	3.51	1.6	−.07	−.3	.34	.8
M25TO60	−.96	−.3	.10	1.5	−.35	−.9	.89	.4	−.13	−.6	−.12	−.3
M61MORE	−4.85	−1.1	−.03	−.3	−.55	−1.0	−2.23	−.8	−.21	−.7	−.51	−.8
F0TO4	−4.11	−1.2	.05	.6	−1.15	−2.6	−4.88	−2.3	−.48	−2.2	−1.92	−4.1
F5TO9	9.07	2.7	.42	4.6	.66	1.5	4.76	2.3	1.02	4.6	.12	.3
F10TO14	18.82	6.2	.70	6.0	1.64	4.3	8.93	4.3	.70	3.2	1.24	2.9
F15TO19	14.49	4.6	.24	3.1	1.74	4.5	4.21	2.0	−.13	−.6	1.27	2.8
F20TO24	3.72	1.1	−.05	−.7	.66	1.4	−1.91	−.8	−.25	−1.1	−.04	−.1
F25TO60	4.13	1.4	.21	3.0	−.15	−.4	.15	.1	.13	.6	.12	.3
HEDYRS	.38	4.4	.01	3.0	.03	2.6	.27	6.1	.02	4.5	.03	4.0
SC	.44	.7	.00	.2	−.09	−1.0	−.09	−.3	.01	.2	.00	.0
ST	1.36	.5	…	…	.03	.1	−1.18	−2.8	−.09	−1.9	−.32	−3.6
MUSLIM	.21	.3	−.01	−.3	.05	.6	−.29	−.4	−.04	−.5	−.04	−.3
CHRISTN	1.08	2.1	.00	−.2	.07	1.1	1.86	.9	…	…	.31	.8
INTERCEP	−19.99	−3.3			−2.94	−3.8	1.45	.5			−.09	−.1
Adjusted R^2	.3416		.4642		.3623		.2855		.4027		.3518	
N	948		772		809		2,039		1,874		1,507	
Dependent variable mean	.0788		.8199		−2.6175		.0474		.7161		−3.0901	
Expenditure elasticity	1.18						.86					
P-value:												
Age 5–9	.50		.33		.57		.50		.26		.37	
Age 10–14	.63		.45		.51		.36		.00		.63	
Age 15–19	.99		.88		.41		.00		.00		.20	

TABLE A1 (*Continued*)

	Madhya Pradesh						Orissa					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	1.34	6.2	.35	9.0	−.03	−.5	1.34	4.1	.27	5.0	−.08	−.8
LNHHSIZE	1.87	10.4	.57	17.0	−.04	−.7	1.65	5.6	.47	9.2	.09	1.0
M0TO4	−.17	−.1	−.35	−1.4	−.48	−.9	1.63	.7	.18	.5	−.18	−.2
M5TO9	6.55	4.5	1.03	4.3	.84	1.8	4.73	2.0	1.58	4.4	.27	.4
M10TO14	9.57	6.6	1.50	6.1	1.82	3.8	7.82	3.3	1.72	4.7	1.50	2.1
M15TO19	9.00	6.1	.66	2.7	2.60	5.3	8.64	3.7	.63	1.8	2.57	3.5
M20TO24	3.57	2.2	−.18	−.7	1.17	2.1	2.67	1.1	−.04	−.1	1.44	1.9
M25TO60	1.19	.8	.06	.2	.25	.5	1.82	.8	.28	.8	−.29	−.4
M61MORE	2.26	1.1	−.01	.0	.60	.8	.48	.2	.20	.5	−.39	−.4
F0TO4	−.27	−.2	−.23	−.9	−.56	−1.1	1.27	.5	.39	1.1	−.34	−.5
F5TO9	4.89	3.3	.76	3.2	.83	1.7	4.99	2.2	.81	2.3	.82	1.2
F10TO14	8.75	6.0	.84	3.5	1.68	3.5	8.37	3.6	.99	2.8	1.46	2.1
F15TO19	4.00	2.6	−.22	−.9	1.43	2.8	3.65	1.6	−.07	−.2	1.08	1.5
F20TO24	.88	.5	−.38	−1.4	.34	.6	2.60	1.1	.01	.0	.62	.8
F25TO60	2.59	1.8	.19	.8	.48	1.0	4.40	1.9	.92	2.6	1.02	1.4
HEDYRS	.31	11.7	.05	10.2	.05	6.8	.28	6.4	.04	6.3	.04	3.5
SC	−.34	−1.8	−.09	−2.9	−.08	−1.4	−1.15	−3.9	−.04	−.9	−.38	−4.5
ST	−.58	−3.0	−.10	−3.1	−.16	−2.6	−1.43	−3.9	−.18	−3.2	−.32	−2.9
MUSLIM	−.14	−.3	−.11	−1.3	−.11	−.8	−2.47	−2.3	−.11	−.6	−.62	−2.3
CHRISTN	1.79	1.4	.00	.0	−.01	.0	2.14	2.0	.20	1.5	.30	1.1
INTERCEP	−15.62	−6.0			−4.21	−5.3	−10.72	−2.9			−3.02	−2.8
Adjusted R^2	.3443		.3636		.4579		.3031		.4009		.3165	
N	3,305		3,221		2,036		1,522		1,442		979	
Dependent variable mean	.0297		.6129		−3.4757		.0305		.6526		−3.5070	
Expenditure elasticity	1.45						1.44					
P-value:												
Age 5–9	.03		.02		.96		.83		.00		.11	
Age 10–14	.31		.00		.57		.69		.00		.91	
Age 15–19	.00		.00		.00		.00		.00		.00	

Variable	Punjab: Unconditional OLS (ESHARE) Coefficient × 100	t-Value	Punjab: Probit (ANYEDEXP) Marginal Effect	t-Value	Punjab: Conditional OLS (LNESHARE) Coefficient	t-Value	Rajasthan: Unconditional OLS (ESHARE) Coefficient × 100	t-Value	Rajasthan: Probit (ANYEDEXP) Marginal Effect	t-Value	Rajasthan: Conditional OLS (LNESHARE) Coefficient	t-Value
LNPCE	1.21	2.5	.26	5.8	−.12	−.9	.10	.4	.26	7.2	−.40	−6.3
LNHHSIZE	2.01	4.2	.32	6.6	.20	1.5	2.39	7.8	.51	11.2	.19	2.3
M0TO4	−4.34	−1.2	−.06	−.2	−2.33	−2.3	−2.85	−1.2	−.14	−.5	−.80	−1.3
M5TO9	6.53	1.8	1.21	3.6	−.07	−.1	6.73	2.9	1.02	3.4	1.71	3.0
M10TO14	6.96	2.0	1.16	3.6	−.04	.0	11.59	4.8	1.95	6.2	2.49	4.2
M15TO19	7.40	2.1	.20	.6	1.28	1.3	10.16	4.2	.69	2.2	2.89	4.8
M20TO24	−4.39	−1.2	.00	.0	−1.03	−1.0	4.56	1.7	−.46	−1.4	1.91	2.9
M25TO60	−1.56	−.4	.04	.1	−.50	−.5	4.14	1.6	.24	.7	1.42	2.1
M61MORE	−.15	.0	.06	.1	−.57	−.4	4.32	1.3	.09	.2	2.46	2.8
F0TO4	−7.14	−2.0	.05	.2	−3.46	−3.3	−1.87	−.8	−.22	−.7	.17	.3
F5TO9	3.55	1.0	.83	2.5	−.63	−.6	2.71	1.1	.46	1.5	1.21	2.0
F10TO14	4.15	1.2	1.06	3.3	−.38	−.4	5.05	2.1	.65	2.1	1.61	2.8
F15TO19	4.93	1.3	−.03	−.1	.66	.7	1.34	.5	−.37	−1.1	1.03	1.6
F20TO24	−7.79	−1.9	−.40	−1.1	−1.05	−.9	2.16	.7	−.45	−1.2	1.82	2.4
F25TO60	−2.37	−.7	.17	.5	−.72	−.7	2.36	1.0	−.26	−.9	1.77	2.8
HEDYRS	.42	6.8	.02	3.5	.06	3.7	.33	7.6	.02	4.0	.06	5.9
SC	−.49	−1.3	−.01	−.3	−.30	−2.8	−.63	−2.4	−.06	−1.7	−.12	−2.0
ST	−2.62	−.7	−.11	−1.1	−.16	−.2	−1.15	−2.6	−.14	−2.3	−.21	−1.9
MUSLIM	.04	.0	−.13	−1.1	−.11	−.3	−.58	−.9	−.03	−.4	−.12	−.7
CHRISTN	.10	.1	…	…	−.41	−1.2	…	…	…	…	…	…
INTERCEP	−9.43	−1.8			−2.52	−1.7	−7.48	−2.3			−2.83	−3.5
Adjusted R^2	.2898		.3208		.2731		.2861		.3627		.3594	
N	964		916		720		1,599		1,592		1,063	
Dependent variable mean	.0456		.7336		−3.3441		.0353		.6677		−3.3213	
Expenditure elasticity	1.27						1.03					
P-value:												
Age 5–9	.18		.06		.33		.00		.00		.10	
Age 10–14	.22		.66		.55		.00		.00		.00	
Age 15–19	.26		.22		.33		.00		.00		.00	

TABLE A1 (*Continued*)

	Tamil Nadu						Uttar Pradesh					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	1.78	4.0	.28	5.0	−.07	−.6	.37	1.9	.20	7.7	−.41	−8.2
LNHHSIZE	2.19	4.3	.55	8.0	−.01	−.1	1.97	9.2	.44	14.5	.05	.8
M0TO4	−2.08	−.6	−.64	−1.7	−1.16	−1.3	−2.71	−1.5	−.41	−1.8	−.58	−1.3
M5TO9	9.19	2.8	2.06	5.2	.15	.2	4.16	2.4	.61	2.8	.66	1.5
M10TO14	11.73	3.6	2.03	4.9	.93	1.1	5.69	3.3	.95	4.3	1.18	2.7
M15TO19	6.55	2.0	.34	.9	1.46	1.6	7.15	4.1	.12	.5	2.36	5.3
M20TO24	2.65	.8	−.38	−1.0	.12	.1	5.32	2.8	−.13	−.5	1.52	3.0
M25TO60	−1.15	−.3	−.24	−.7	−.81	−.9	2.27	1.2	−.10	−.4	.83	1.7
M61MORE	−.77	−.2	−.30	−.6	−.25	−.2	1.79	.8	−.26	−.9	1.19	1.9
F0TO4	−1.31	−.4	−.53	−1.4	−1.68	−1.9	−2.23	−1.2	−.38	−1.7	−.31	−.7
F5TO9	6.63	2.0	1.54	4.0	.06	.1	1.85	1.1	.15	.7	.60	1.4
F10TO14	9.50	3.0	1.60	4.4	.58	.7	4.43	2.5	.16	.7	1.26	2.9
F15TO19	4.88	1.5	.17	.4	.35	.4	−.06	.0	−.48	−2.1	.97	2.1
F20TO24	4.01	1.0	−.80	−1.9	1.66	1.5	−.82	−.4	−.50	−2.0	.24	.5
F25TO60	3.55	1.1	−.08	−.2	.10	.1	1.60	.9	.21	1.0	.30	.7
HEDYRS	.27	4.5	.02	2.8	.05	3.3	.36	13.3	.04	10.3	.07	10.4
SC	−.70	−1.8	−.01	−.1	−.19	−2.0	−.89	−4.6	−.09	−3.7	−.21	−4.2
ST	−1.49	−.7	.09	.4	.07	.1	−.57	.9	−.15	−2.0	.03	.2
MUSLIM	−.52	−.4	−.15	−.9	−.23	−.7	−1.22	−4.7	−.15	−4.5	−.12	−1.9
CHRISTN	.36	.4	−.15	−1.1	−.02	−.1	6.78	1.5	…	…	.92	1.1
INTERCEP	−19.00	−4.1			−3.57	−2.8	−7.00	−2.8			−1.63	−2.6
Adjusted R^2	.2136		.4236		.3096		.2550		.3171		.2738	
N	916		872		624		3,337		3,190		2,229	
Dependent variable mean	.0350		.6651		−3.4513		.0363		.6602		−3.3172	
Expenditure elasticity	1.51						1.10					
P-value:												
Age 5–9	.17		.02		.83		.01		.00		.82	
Age 10–14	.26		.09		.44		.19		.00		.72	
Age 15–19	.39		.44		.05		.00		.00		.00	

	West Bengal						Assam					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
Variable	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value	Coefficient × 100	*t*-Value	Marginal Effect	*t*-Value	Coefficient	*t*-Value
LNPCE	1.03	3.1	.29	5.4	−.24	−2.2	1.16	2.6	.30	4.9	−.23	−2.0
LNHHSIZE	1.36	4.2	.61	11.1	−.14	−1.2	2.16	3.9	.51	7.0	.05	.4
M0TO4	−5.38	−1.9	−.68	−1.5	−2.08	−2.3	1.31	.4	−.52	−1.3	−.18	−.2
M5TO9	−1.44	−.5	.59	1.3	−1.69	−1.9	6.86	1.9	.83	2.0	.63	.7
M10TO14	3.42	1.2	.65	1.5	.37	.4	11.93	3.3	1.13	2.6	2.11	2.3
M15TO19	3.93	1.4	−.12	−.3	1.14	1.3	19.64	5.1	.43	1.0	3.44	3.6
M20TO24	−.27	−.1	−.69	−1.5	.01	.0	8.03	2.0	−.62	−1.4	3.90	3.8
M25TO60	−3.90	−1.4	−.46	−1.0	−1.26	−1.3	−.34	−.1	−.62	−1.5	.76	.8
M61MORE	−6.07	−1.7	−.75	−1.4	−.51	−.4	−.96	−.2	−.73	−1.4	1.46	1.3
F0TO4	−5.49	−2.0	−1.20	−2.7	−2.15	−2.4	2.92	.8	−.56	−1.3	.11	.1
F5TO9	−1.31	−.5	.43	1.0	−1.20	−1.4	7.53	2.1	.87	2.1	.41	.5
F10TO14	4.59	1.7	.71	1.6	.17	.2	9.96	2.7	.94	2.1	1.80	1.9
F15TO19	1.92	.7	−.46	−1.0	.12	.1	15.70	4.1	.59	1.4	2.60	2.7
F20TO24	.48	.2	−.32	−.7	−.99	−1.0	14.99	3.6	−.35	−.8	2.77	2.6
F25TO60	.51	.2	.28	.6	−.92	−1.0	8.65	2.3	−.22	−.5	.91	.9
HEDYRS	.40	8.6	.04	6.1	.11	7.4	.32	5.6	.04	5.1	.07	4.5
SC	−1.08	−3.4	−.12	−2.5	−.13	−1.3	−1.19	−1.9	−.09	−1.0	−.21	−1.4
ST	−.64	−.8	−.13	−1.2	.16	.6	−.11	−.2	−.10	−1.2	.06	.3
MUSLIM	−1.23	−3.0	−.17	−2.8	−.14	−1.1	−1.84	−2.1	−.19	−1.7	−.07	−.3
CHRISTN	−1.08	−1.0	.04	.3	−.26	−.8	.49	.4	−.03	−.2	.55	1.6
INTERCEP	−6.93	−1.8			−1.08	−.8	−16.58	−3.2			−3.46	−2.6
Adjusted R^2	.3066		.3246		.3794		.4244		.3963		.4677	
N	1,243		1,235		768		941		786		707	
Dependent variable mean	.0292		.6154		−3.6473		.0428		.7023		−3.3651	
Expenditure elasticity	1.35						1.27					
P-value:												
Age 5–9	.92		.41		.29		.67		.84		.55	
Age 10–14	.44		.78		.67		.35		.55		.51	
Age 15–19	.24		.18		.08		.11		.59		.17	

Note. All equations include village fixed effects. For the unconditional OLS, the dependent variable is ESHARE or the budget share of education and coefficients have been multiplied by 100. For the conditional OLS, i.e., that fitted only on households with positive ESHARE, the dependent variable is the natural log of ESHARE, or LNESHARE. The dependent variable in the probit is ANYEDEXP, i.e., whether household had any positive education expenditure in past year, as opposed to zero education spending. Where a variable predicts success perfectly, that is indicated with ellipses (. . .). For example, where all Christian households have ANYEDEXPp = 1, then the marginal effect of that variable is not identified and it is denoted with ellipses (. . .). Similarly, if there are no Christians in the rural part of a state in the sample, this is denoted with ellipses (. . .). The last three rows present the *P*-values of *F*-test that, in a given age group, the coefficients of male and female demographic variables in that model/column are equal. The number of observations on which the probit equation is fitted is usually somewhat smaller than the number of observations on which the unconditional ESHARE equation is fitted; this is because we have included village fixed effects, so any villages in which all sample households take the value of 0 or of 1 (for ANYEDEXP) perfectly predict failure/success and drop out.

References

Ahmad, Asif, and Jonathan Morduch. 2002. "Identifying Sex Bias in the Allocation of Household Resources: Evidence from Linked Household Surveys from Bangladesh." Unpublished manuscript, Department of Economics, New York University.

Bhalotra, Sonia, and Chris Attfield. 1998. "Intrahousehold Resource Allocation in Rural Pakistan: A Semiparametric Analysis." *Journal of Applied Econometrics* 13, no. 5:463–80.

Browning, Martin. 1992. "Children and Household Economic Behavior." *Journal of Economic Literature* 30, no. 3:1434–75.

Case, Anne, and Angus Deaton. 2003. "Consumption, Health, Gender, and Poverty." Working Paper no. 3020, World Bank Poverty Reduction and Economic Management Network, Washington, DC.

Datt, Guarav, and Martin Ravallion. 1998. "Why Have Some Indian States Done Better than Others at Reducing Rural Poverty?" *Economica* 65, no. 257:17–38.

Deaton, Angus. 1989. "Looking for Boy-Girl Discrimination in Household Expenditure Data." *World Bank Economic Review* 3, no. 1:1–15.

———. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Washington, DC: Johns Hopkins University Press for the World Bank.

Deaton, Angus, and Christina Paxson. 1998. "Economies of Scale, Household Size, and the Demand for Food." *Journal of Political Economy* 106, no. 5:897–930.

Deolalikar, Anil, and Elaina Rose. 1998. "Gender and Savings in Rural India." *Journal of Population Economics* 11, no. 4:453–70.

Drèze, Jean, and Geeta G. Kingdon. 2001. "Schooling Participation in Rural India." *Review of Development Economics* 5, no.1:1–26.

Drèze, Jean, and Amartya Sen. 1995. *India: Social Opportunity and Economic Development.* Oxford: Oxford University Press.

Drèze, Jean, and P. V. Srinivasan. 1996. "Poverty in India: Regional Estimates, 1987–88." STICERD Development Economics Discussion Paper no. 70, London School of Economics.

Dubey, A., and S. Gangopadhyay. 1998. "Counting the Poor: Where Are the Poor in India? " Sarvekshana Analytical Report no. 1, Department of Statistics, Government of India, New Delhi.

Duraisamy, P. 2002. "Changes in Returns to Education in India, 1983–94: By Gender, Age-Cohort, and Location." *Economics of Education Review* 21, no. 6:609–22.

Foster, Andrew, and Mark Rosenzweig. 2000. "Missing Women, the Marriage Market, and Economic Growth." Unpublished manuscript, Department of Economics, Brown University.

Jensen, Robert. 2002. "Equal Treatment, Unequal Outcomes? Generating Gender Inequality through Fertility Behavior." Unpublished manuscript, John F. Kennedy School of Government, Harvard University.

Kingdon, Geeta Gandhi. 1996a. "Private Schooling in India: Size, Nature, and Equity Effects." *Economic and Political Weekly* 31, no. 51:3306–14.

———. 1996b. "The Quality and Efficiency of Private and Public Education: A Case Study of Urban India." *Oxford Bulletin of Economics and Statistics* 58, no. 1: 57–81.

———. 1998. "Does the Labour Market Explain Lower Female Schooling in India?" *Journal of Development Studies* 35, no. 1:39–65.

Kingdon, Geeta Gandhi, and Jeemol Unni. 2001. "Education and Women's Labour Market Outcomes in India." *Education Economics* 9, no. 2:173–95.

Lancaster, Geoffrey, Pushkar Maitra, and Ranjan Ray. 2003. "Endogenous Power, Household Expenditure Patterns, and New Tests of Gender Bias: Evidence from India." Technical Report Discussion Paper no. 20/03, Department of Economics, Monash University.

Pradhan, Basant, and Arjunan Subramanian. 2000. "Education, Openness, and the Poor: Analysis of an All-India Sample of Households." Discussion Paper no. 14, National Council of Applied Economic Research, New Delhi.

PROBE Team. 1999. *Public Report on Basic Education in India*. Delhi: Oxford University Press.

Rose, Elaina. 1999. "Consumption Smoothing and Excess Female Mortality in Rural India." *Review of Economics and Statistics* 81, no. 1:41–49.

———. 2000. "Gender Bias, Credit Constraints, and Time Allocation in Rural India." *Economic Journal* 110, no. 465:738–58.

Schultz, T. Paul. 2002. "Why Governments Should Invest More to Educate Girls." *World Development* 30, no. 2:207–25.

Shariff, Abusaleh. 1999. *India Human Development Report*. New Delhi: Oxford University Press.

Subramanian, Shankar. 1995. "Gender Discrimination in Intra-Household Allocation in India." Unpublished manuscript, Department of Economics, Cornell University.

Subramanian, Shankar, and Angus Deaton. 1990. "Gender Effects in Indian Consumption Patterns." Discussion Paper no. 147, Research Program in Development Studies, Woodrow Wilson School, Princeton University.

Wooldridge, Jeffery. 2002. *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, MA: MIT Press.