CSAE WPS/2003-13

Where has all the bias gone?
Detecting gender-bias in the household allocation of educational expenditure

by
Geeta Gandhi Kingdon

Department of Economics
University of Oxford

October 2003

Abstract

The reliability of the household consumption based (Engel curve) methodology in detecting gender bias has been called into question because it has generally failed to confirm bias even where it exists. This paper seeks to find explanations for this failure by exploiting a dataset that has educational expenditure information at the individual level and also, by aggregation, at the household level. We find that in the basic education age groups, the discriminatory mechanism in education is via differential enrolment rates for boys and girls. Education expenditure *conditional* on enrolment is equal for boys and girls. The Engel curve method fails for two reasons. Firstly, it models a single equation for the two stage process. Second, even when we make individual and household level expenditure equations as similar as possible, the household level equation still fails to 'pick up' gender bias in about one third of the cases where the individual-level equation shows significant bias. The paper concludes that only individual based data can accurately capture the full extent of gender bias.

Correspondence: Geeta G. Kingdon, Department of Economics, University of Oxford, Oxford OX1 3UQ, United Kingdom; tel: 00 44 1865 271065; email: geeta.kingdon@economics.ox.ac.uk

Acknowledgements: I would like to thank Angus Deaton, Jean Drèze, Marcel Fafchamps, Måns Söderbom, Shankar Subramaniam, Francis Teal and participants at the NEUDC conference at Yale University (especially Nistha Sinha) for helpful comments on this paper. I am also grateful for the generous help of Abusaleh Shariff with the dataset and of Måns Söderbom with STATA programming. Any errors are mine. The research was supported by a Wellcome Trust grant No. 053660 and by an ESRC grant under the Global Poverty Research Group. The acquisition of the data was funded by DFID, India.

Where has all the bias gone?
Detecting gender-bias in the household allocation of educational expenditure

1. Introduction

Two approaches have been used in the literature to detect gender bias in the intra-household allocation of consumption or expenditure: the direct comparison of expenditure on males and females where data is 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/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 perforce 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, p239-41)[1]. Deaton notes: "it is a puzzle that expenditure patterns so consistently fail to show strong gender effects even when measures of outcomes show differences between girls and boys". Case and Deaton (2003) 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) 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 paper 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: one, via zero purchases for daughters and positive purchases for sons and two, 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. Secondly, 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.

[1] For example, the use of the Engel curve method failed to detect significant differential treatment in the intra-household distribution of food consumption in Maharashtra (Deaton and Subramanian, 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 very skewed sex-ratios, or from Bangladesh and Pakistan, two 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 favour 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, p240; Bhalotra and Attfield, 1998) and with 1999-2000 NSS data for India (Case and Deaton, 2003).

On the first issue, suppose that bias against girls in education takes the form mainly of zero expenditure on girls' education (non-enrolment of girls), but that *conditional* on enrolment, 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. on school transport and clothing). Then, averaging across these two mechanisms there may not be significant gender bias but, via the non-enrolment 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 non-bias. One would be interested not only in the average unconditional expenditure on girls and boys but also in the distribution of the expenditure[2].

Secondly, the failure of the conventional approach to detect gender discrimination may be 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 NSS data from rural Maharashtra, there is no evidence of pro-boy gender bias in educational expenditure in the age groups 5 to 9 and 15-54, 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 age group 10-16 but not in urban areas and not in the primary school age-group 6-9. In his study of five Indian states Subramanian (1995) 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?"

[2] Another reason why the conventional application of the Engel curve method may fail to pick up discrimination against girls even where it exists may be because the distributional assumption about the dependant variable and thus the specification of the budget-share equation are 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 dependant variable. This would lead to incorrect standard errors. However, this is not a particularly important worry in large samples, such as ours.

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 is 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 daughters. This will not show up in investigations of aggregate expenditures but it will show up in investigations of individual outcomes[3].

The 1994 NCAER rural household survey of 16 major states in India collected data on individual educational outcomes, i.e. on school enrolment, years of schooling, and education expenditure data on each household member aged <=35 years old. Thus, it is possible - using this data - to investigate gender bias in the allocation of educational expenditure both by direct examination of educational spending on boys and girls, and also 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 datasets 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 uniform[4]. 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, p133; Kingdon, 1996a and 1996b). Our data show that the overwhelming majority (98%) of enrolled 5-19 year olds have positive expenditure incurred on their education.

In this paper 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. It tests two explanations for this failure 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

[3] Some of Ahmad and Morduch's explanations are ex post rationalisations of gender differences in mortality and morbidity in the supposed absence of gender bias in expenditure allocation within the household. For instance, they consider sex-bias in fertility (i.e. the fact that girls are in larger households due to parents' going on having births till they get a boy, and thus having lower per capita expenditure) as an explanation for the fact that there are significant gender differences in outcomes such as mortality and morbidity even though there may not be any gender discrimination within the family in the allocation of food and medical expenditure. In other words, they ask: if we believe what we find in the household expenditure methodology i.e. that there is no significant gender difference in consumption expenditure, then how can we explain that individual outcomes differ for girls and boys. We are asking the question the other way round. Given we know that educational outcomes differ for girls and boys, how can we explain that the household expenditure patterns do not pick this up.

[4] 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, p16) found that in rural north India, parents spend about Rupees 318 per year on each child who attends government (i.e. tuition-free) school, so that an agricultural labourer in Bihar with 3 such children would have to work for about 40 days in the year just to send them to primary school.

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 2 discusses the methodology, including both the Engel curve method and the hurdle model. Section 3 discusses data and estimation issues. The results are discussed in Section 4 and the final section concludes.

2. Indirect methodology for detecting discrimination

The Engel curve method utilises 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 it does 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-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-Leser specification but - so as not to pre-judge the issue - later relax it to allow for non-linearity in the shape of the Engel curve. Working's Engel curve can be extended to include household demographic composition by writing:

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

where x_i is total expenditure of household *i*, s_i is the budget share of education $\left\{ \frac{edu\exp}{x_i} \right\}$, n_i is household size, and z_i is a vector of other household characteristics such as religion, caste, and household head's education and occupation. u_i is the error term. The term $\ln n_i$ allows for an independent scale effect for household size. j=1,…,J refers to the Jth age-gender class within the household. 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 paper, there are 14 age-sex categories. These are males and females in age groups 0-4, 5-9, 10-14, 15-19, 20-24, 25-60, and 61 and above. The fraction of women aged >=61 years old in the household is the omitted category. The variables of most interest pertain to persons of school-going age, i.e. they are males and females aged 5-9, 10-14, and 15-19. These variables are named M5to9, F5to9, M10to14, F10to14, M15to19 and F15to19. 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 to 9 age group will involve testing whether the coefficient on M5to9 is significantly different to 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, *i.e.* it includes both zero and positive values of the dependant variable - the budget share. For example, Subramanian and Deaton (1990) and Subramanian (1995) 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)[5].

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-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 non-parametric 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, $\partial P(s > 0 \mid x) / \partial x_j$ and $\partial E(s \mid x, s > 0) / \partial x_j$ 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 ($s = 0$ versus $s > 0$) and how much to spend ($s \mid s > 0$). 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

[5] Some studies have used flexible-form or semi/non-parametric regression, for example, 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 per cent of households respectively reported positive education spending. In the current NCAER data, the corresponding figures are 49, 64, 58 and 55 per cent respectively. That is, between 1988 and 1994, the proportion of rural households incurring positive spending on education rose quite sharply.

model estimation to allow the decision of whether to incur any education expenditure to be modelled separately from the decision of how much to spend on education, conditional on spending anything.

A simple hurdle model can be written down as:

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

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

where *s* is the 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 $s>0$, $s \mid x$ follows a lognormal distribution. An examination of the distribution of *s* in Figure 1 suggests that conditional on positive education spending, *s* is more lognormally than normally distributed.

The maximum likelihood estimate 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 $E(s \mid x, s>0)$ and the unconditional expectation of $E(s \mid x)$ are easy to obtain using properties of the lognormal distribution:

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

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

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*:

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

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:

$$\frac{\partial E(s \mid x)}{\partial x} \quad = \quad \gamma\phi(x\gamma)\exp(x\beta+\sigma^2/2) + \Phi(x\gamma)\beta\exp(x\beta+\sigma^2/2)$$

$$= \quad \{\gamma\phi(x\gamma) + \Phi(x\gamma)\beta\}.\exp(x\beta+\sigma^2/2) \tag{7}$$

$\phi(.)$ 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 aged 5to9 in the household (M5to9), proportion of females aged 5to9 (F5to9), etc. 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 under- or over-estimation in the measurement of gender bias. However, if unobserved characteristics such as child ability, child motivation, and parental attitudes have a greater influence in enrolment 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 over-estimation of pro-male gender bias.

This can be shown by focusing on any one pair of demographic variables, e.g. 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, i.e. the greater the proportion of 5 to 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, i.e. [$Corr(x,u)<0$]. Averaging over all households, the correlation between the explanatory variable (F5to9) and the error term is not equal to zero [$Corr(x,u) \neq 0$] 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 less from it than the female variable (as is likely), then any pro-male bias will be over-estimated.

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, i.e. 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.

3. Data and estimation issues

This study utilises 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 dataset are available in Shariff (1999).

The major advantage of this dataset is its detailed information on education of each person aged <=35 years in the household, including educational expenditure information. However, an important drawback is that it did not collect comprehensive information on *total* household expenditure. Only household expenditure on food, health, and education was 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, entertainment, *etc*. Given that we have data only on food, health, and education (*FHE*), differential treatment depends upon two components:

$$s = \frac{Edu\exp}{Total\exp} = \frac{Edu\exp}{FHE\exp} \times \frac{FHE\exp}{Total\exp}$$

that is, it depends on:

(i) how $\frac{Edu\exp}{FHE\exp}$ changes with more girls in the household, and

(ii) how $\frac{FHE\exp}{Total\exp}$ changes with more girls in the household

We are able to model only the first component, *i.e.* 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

[6] We know from Subramanian's (1995) study that in 1987-88 in five Indian states, food, health and education expenditure together account for about 63% of total household expenditure.

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 (i) 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, *i.e.* it has been hypothesised 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 (i) 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 - most likely the effect is neutral[7]. In other words, modelling how the share of education in *FHE* expenditure changes with household gender composition should neither under- nor over-estimate gender bias in the allocation of education expenditure. Thus, although we use the budget sub share of education in this paper, for simplicity, we refer to it simply as the budget share of education.

The analysis here is limited to households which have children of school-going age, i.e. children aged 5 to 19 years old. This yields a sample of 25954 households. In this sample, the mean budget share of education is 4.40%[8] and the percentage of households with zero education spending is 31%.

4. Discussion of results

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

---

[7] None of the several extant studies provides any convincing evidence of systematic gender bias in food allocation within Indian households.

[8] This is considerably higher than the budget share of education in previous studies on India. For example, the average budget share of education for the 5 Indian states studied in Subramanian (1995) was 1.34%. In our data, it is 3.69% for those same 5 states. However, the data used in the two studies are not comparable because firstly, the earlier studies do not restrict the sample to only households with children in school-going age range. Secondly, 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 5 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 MIMAP survey of the mid-1990s (Pradhan and Subramaniam, 2000, p27). 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, p4-5; Datt and Ravallion, 1998, p30; 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.

4.1 *Descriptive statistics*

The first column of Table 1 shows the sex-ratio in the 0-14 year age group in sample households. It shows that the proportion of girls is only 46.4% in rural India but 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 intra-household 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 aged <15 years into two groups - 'all-girl' households, where all the children below age 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 very large 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 age groups 10-14 and 15-19 years, girls have a significantly and substantially lower current enrolment rate (than boys), *i.e.* a higher probability of reporting zero educational spending due to non-enrolment, in virtually every one of the 16 sample states (except Kerala and West Bengal). However, this is not so in the age group 5-9 where the gender gap in enrolment rate is significant only in about half the states.

Table 3 shows average educational expenditure, conditional on enrolment. 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 differential *current enrolment* rates of girls and boys. Table 4 includes zero education-expenditure (i.e. non-enrolled) 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 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 non-enrolment (i.e. via zero education expenditures) and only rarely via lower expenditures once enrolled.

[9] The figure for Assam seems implausibly low.
[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 seen later.

4.2 Incorrect functional form as the reason for 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 dependant 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 state (a) the conventional Engel curve equation; (b) a binary probit of whether the household's education budget share is positive or zero; and (c) OLS of the natural log of education budget share, conditional on positive education budget share. The resulting 48 equations are presented in Appendix Table 1.

The first column under each state in Appendix Table 1 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 non-linear in several states when I allowed for a quadratic term in LNPCE, confirming that at low levels of per capita expenditure, education is a luxury but that it becomes a necessity at higher levels of expenditure[11].

Per capita expenditure has a significant positive impact on budget share of education, and the total expenditure elasticity is close to or above unity in all states, suggesting 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 (date of data in previous studies) [12].

[11] I report the specification using log of per capita expenditure (LNPCE) on the right hand side but I also tried two variations: one was to include LNPCI (log of per capita income) instead. The other was an instrumental variable estimation, using LNPCI to instrument LNPCE. Using *LNPCI* as an instrument for *LNPCE* is justified because the two are highly correlated and because income will not be correlated with budget shares independently of its correlation with *LNPCE*. The coefficients of the household demographic variables (M5to9, F5to9, M10to14, etc.) and the F-tests of the significance of the gender gaps in educational expenditure were very robust to these alternative formulations.

[12] In Subramanian's study the total expenditure elasticities for AP, 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 5 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).

The effect of household size is positive and significant in every state. This is in line with theoretical considerations which 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 finding of a positive and consistent effect of household size is of particular interest given the failure to find this effect in the seven high and low income countries studied in Deaton and Paxson (1998).

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 sub-shares than Hindus and Sikhs (the omitted category) in Haryana, HP, Karnataka, Kerala, Orissa, UP, WB, 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? 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 Appendix Table 1. The row for 'p-value: age 5 to 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 age group 5 to 9) is the same as the coefficient on F5to9 (the female demographic variable for age group 5 to 9) is rejected at the 5% significance level <u>only</u> for Rajasthan. This lack of evidence of significant gender bias in all but one state shows that the conventional Engel curve technique is not good at picking up gender-differentiated treatment in educational expenditure within households, given that enrolment data show significant gender differences in 9 out of the 16 states (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 Appendix Table 1 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 (conditional OLS equation). In the conditional budget share equation, sample selection could be a problem. However, as explained in the methodology section, the conditional OLS equation will tend to *over-estimate* any pro-male bias. We attempted to control for sample selectivity but its effects were largely insignificant[13].

[13] We allowed for sample selectivity using 'index of productive assets owned by household' as the exclusion restriction for identifying the constructed variable Lambda. Index of productive assets (INDEXPA) seemed a good exclusion restriction in that, for each age group (5-9, 10-14, 15-19), INDEXPA was significant in the probit of current enrolment and insignificant in the educational expenditure function. One would expect this *a priori* since the presence of productive assets would be likely to raise the opportunity cost of school attendance by increasing the returns to

In Appendix Table 1, 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 almost invariably negative and 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[14].

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, F5to9 (household's proportion of males and females aged 5 to 9); M10to14, F10to14 (proportion of males and females aged 10 to 14); and M15to19 and F15to19 (proportion of males and females aged 15 to 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 age group 5-9.

Tables 5a, 5b and 5c present 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 Appendix Table 1. The figures in parentheses below each DME are the p-values of the F-test that the DME is equal to zero. P-values of statistically significant DME's (at the 5% level or better) are shaded. The meaning of the DME is best illustrated with an example. For instance, in the probit of ANYEDEXP in Gujarat in Appendix Table 1, the marginal effect of the variable M5to9 was 0.4867 and the marginal effect of F5to9 was 0.0712. Thus the gender DME in the 5-9 age group there was 0.4155. Table 5a shows this difference multiplied by 100, i.e. as 41.55. The p-value of the F-test that this difference is equal to zero was 0.04, i.e. this gender difference in marginal effect is statistically significant at the 4% level. In Tables 5a, 5b and 5c, the probit results in column (a) refer to male-female DME from the probit of whether the household had a positive education budget share. Column (b) refers to the male-female DME in the conditional OLS of the log of

---

child labour. However, once a child is in school, productive assets should not matter to how much the household spends on the education of the child since we control for household per capita expenditure. The selectivity variable Lambda was significant at the 5% level (p value 0.043, t=2.02) only in the 10-14 age group but even there, there was no significant difference in the coefficients of the OLS and selectivity corrected equations. In the age-groups 5-9, 10-14, and 15-19, the t-values on lambda were -0.44, 2.02, and -1.69 respectively. INDEXPA turned out not to be a good identifying exclusion restriction when doing the regressions by state, since it was frequently insignificant in the current enrolment probit and occasionally significant in the educational expenditure function. It is possible that with better identifying exclusion restrictions, we could achieve better identification of lambda and so the OLS coefficient on MALE should be taken as the lower bound on the effect of MALE. We also tried CLAD estimates (available from author) but these were not significantly different to OLS estimates.

[14] The marginal effects of the demographic variables are sometimes above 1 because these variables take values from 0 to 1 rather than from 1 to 100. Redefining them to be bounded by 1 and 100 simply leads to the reported marginal effects being divided by 100.

education budget share (ESHARE). 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 (b) are comparable to those in column (d), where the dependent variable was absolute ESHARE[15]. Column (c) 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 (d) pertains to the unconditional OLS results, i.e. the OLS of the absolute budget share of education fitted on all (including zero education expenditure) households – the commonly reported Engel curve equation.

Tables 5a, 5b and 5c demonstrate two interesting facts. Firstly that 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. Secondly, it shows that 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 (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[16].

In the 5 to 9 age group, the gender DME in the probit is positive for all states except one (Karnataka), and is statistically significant in six states. In 10 out of 16 states the gender DME in the conditional OLS of LNESHARE is negative (albeit insignificant), and in only one of the 16 states is it positive and statistically significant, i.e. in the vast majority of states there is no pro-male gender bias in conditional education expenditure. The inference from the 'conventional' Engel curve results in column (d) is that there is no significant gender bias in education expenditure in the 5-9 age group in any state other than Rajasthan. However, such an inference masks the fact that in 5 states other than Rajasthan, there is significant gender bias in the decision whether to enrol a child in school. To overlook the difference is to miss an important discriminatory process.

In the 10-14 age group, the gender DME in the probit is positive for all states except Kerala and West Bengal, and it is significant in 7 states. But the DME from the conditional OLS is insignificant in all but one state. Here too, as in the 5-9 age group, the conventional Engel curve result in column (d) would

[15] For example, the coefficient on the variable M5to9 in the conditional OLS of LNESHARE for Gujarat is –1.58 and the coefficient on F5to9 is –1.08. The log transforms of these are obtained by using the property of the lognormal distribution that the conditional expectation of $E(s \mid x, s>0)$ equals $\exp(x\beta + \sigma^2/2)$. For the Exp(.) is equal to 0.04836. Thus the marginal effect of M5to9 is b*exp(.), i.e. it is –1.58*0.04836 = -0.0766 ; the marginal effect of F5to9 is –1.08*0.04836 = -0.0524. Thus, 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 –0.0766 – -0.0524 = -0.0242. In Table 5 all (differences in) marginal effects are multiplied by 100, so this appears as -2.42.

[16] 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 girls' school uniform than on boys'. In the 5-9, 10-14 and 15-19 age groups, mean transport costs are higher for girls than boys in only 3, 5 and 6 of the 16 states respectively.

lead to the inference of no significant gender bias in any state other than Rajasthan. Again, such an inference would neglect the fact that in 6 states other than Rajasthan, there is significant bias in the enrolment decision. Table 5b also shows that using the hurdle model approach (column c), 5 states have significant gender bias in unconditional education expenditure. In other words, when the decision to incur positive education expenditure is modelled separately from the decision 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 which 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 6 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, i.e. they reinforce each other.

In order to show graphically that the two processes of gender differentiation are different, I present a scatter plot of the DMEs separately for the three age groups in Figure 2. If the two processes were the same, we would expect all the points to fall on the diagonal 45 degree line through the origin. It is clear in Figure 2 that for youngest two age groups (denoted age=1 and age=2 in Figure 2), there is little suggestion that the states are on the upward diagonal. Indeed if anything, the points appear to lie on the downward diagonal. However, for the 15-19 year olds (denoted age=3), except for a few states such as Assam, Bihar and Himachal Pradesh (denoted by as, bi, and hp respectively), most of the other states lie roughly on the positive diagonal. In other words, below the age of 15, the two processes oppose each other but beyond age 15, 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 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 but certain states operated an affirmative action policy for girls in the secondary school stage by providing tuition-free secondary schooling[17]. 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

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

of discrimination; thus it is difficult to know what part of girls' observed inferior enrolment outcomes in the 15-19 age range is due to parental discrimination and what due to supply-side factors.

To sum up, the discussion so far suggests two conclusions. Firstly, that 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 non-linear and the correct distribution of conditional expenditure is lognormal rather than normal, then a hurdle model seems better able to capture gender biases in unconditional expenditure. Secondly, 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.

4.3 Aggregation as the reason for 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 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 dependant 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 to 9' and 'household proportion of females aged 5 to 9', etc., the gender variable of interest is simply the dummy variable MALE which is 1 for males and 0 for females. The rest of the explanatory variables in the individual level equations are identical to those in the household equations of Appendix Table 1, i.e. they are household level variables. The three age groups of interest, as before, are ages 5 to 9, ages 10 to 14 and ages 15 to 19, corresponding roughly with primary, junior and secondary education.

At the individual child level, we estimated 144 separate equations (16 states x 3 age groups x 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 6a (for age group 5 to 9), Table 6b (for age group 10-14) and Table 6c (for age group 15-19).

The marginal effects on MALE in Tables 6a, 6b and 6c are not comparable with the difference in marginal effects of the household demographic variables in Tables 5a, 5b and 5c. 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 dependant variable in the conditional and unconditional OLS equations in Table 6 is education expenditure but in Table 5 the corresponding dependant variable is 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 Tables 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 (enrol a child in school), and not in the decision of how much to spend, conditional on school enrolment. In several instances, the marginal effect of MALE in the conditional expenditure equation is negative, i.e. girls have somewhat higher education expenditure, conditional on being in school, though this pro-female bias is rarely statistically significant.

Since MALE is a discrete variable, the marginal effect of MALE in the combined hurdle model (column c) is estimated by calculating the expected values 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) [18]. Column (d) presents the marginal effect of the variable MALE in the unconditional expenditure equation, i.e. the single OLS equation estimated including zero education expenditures.

While a comparison of columns (c) and (d) 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 6a, the hurdle model detects overall gender bias in Andhra Pradesh and Tamil Nadu where the unconditional OLS fails to pick it up. The same is true for Assam in Tables 6b, 6c.

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 is used (Tables 6) than when household aggregated data is used (Tables 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. However, when comparing Tables 5 and 6 one is not comparing like with like. While we

[18] 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 the paper, we used derivatives as set out in equations (6) and (7).

have ensured that in all other respects the specification of the probit, conditional OLS and unconditional OLS equations are identical in the individual and household level analyses, the dependant variables (except in the probit) as well as the gender variables are different in the individual and household level analyses. In order to compare like with like, one would need to use a similar gender variable in the household level analysis as the MALE dummy variable in the individual level analysis.

Since gender bias manifests itself mainly in the zero versus positive expenditure (ANYEDEXP) decision, we examine whether the effect of gender is similar in the individual and household level probits of ANYEDEXP. Table 7 compares the marginal effects of gender in the individual and household level probit equations that are as alike as we could make them. Column (1) reproduces the marginal effect of MALE in the individual level probits (taken from the first columns of Tables 6a, 6b, and 6c). Columns (2) presents the marginal effects of gender variables in the household level probit, the three gender variables being: 'proportion of males among all 5-9 year olds in the household'; 'proportion of males among all 10-14 year olds in the household'; and 'proportion of males among all 15-19 year olds in the household'. These gender variables at the household level differ from those used so far in that they represent the proportion of males within a given age group (e.g. number of males aged 5 to 9 divided by number of males and females aged 5 to 9 in the family) rather than, for example, proportion of males aged 5 to 9 within the household as a whole. This is the gender concept that comes closest to the gender dummy MALE in the individual level regression. Since both individual and household level gender variables are now bounded between 0 and 1, their marginal effects should be comparable.

Table 7 shows that at conventional levels of significance, household level data fails to detect significant discrimination in about 11 (or about one-third of) cases where individual level data shows significant bias. We can also compare the sizes of the marginal effects of gender across the individual and household probits. Such a comparison shows that even when we have done the best we can to achieve similar explanatory and dependant variables in individual and household level equations, we still fail to capture the full extent of gender bias when we use household level data. The marginal effect of the gender variable is consistently and significantly lower in the household level probit than in the individual level probit in each of the three age groups. The average marginal effect of gender in each age group is presented in the last row of Table 7 and depicted in Figure 3. It shows that the marginal effect of the gender variable increases with age group and, within each age group, is always higher in individual level data than in household level data. This suggests that there *is* something about aggregation that prevents household level data from picking up the full extent of gender bias. It is not that measurement error is greater in household total education expenditure than in individual education expenditure, since in the dataset used for this study, household education expenditure is obtained by aggregating individual education expenditure.

5. 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 enrolment rates for girls are significantly worse than those for boys. They also confirm 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', *i.e,* 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 non-enrolment of girls, which implies zero educational spending. There is little gender bias in educational expenditure among enrolled children.

The analysis shows a low degree of correspondence between results in individual level and household level data; particularly in the younger two age groups the household expenditure method fails to find significant discrimination. Our approach in this paper suggests important explanations for why the conventional Engel method fails to detect gender bias in intra-household allocation. Tests suggest that this failure is partly because the Engel curve method as conventionally applied suffers from incorrect functional form and the limitation that the effects of the household gender composition variables on both (a) the decision to enrol 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 school age groups. However, in the 15-19 year 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 modelling 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 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 is a poor substitute for individual level expenditure data. Household expenditure data is of some use providing 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 non-enrolment of girls. 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 compared with for boys in the same state of illness but, conditional on seeking medical advice, the expenditure on girls is the same as that on boys. Policy makers may be as 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 non-local 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 labour 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)[19]. Further evidence on returns to men and women's education in the *rural* Indian labour market would be useful in analysing whether gender bias in intra-household educational resource allocation in rural India is attributable to gender differentials in the returns to education.

[19] Duraisamy (2002) and Kingdon and Unni (2001) find mixed evidence on returns to men and women's education in India. However, neither study could control for omitted family background bias, which substantially reduces women's returns but not men's in Kingdon (1998). Indian estimates in Kingdon (1998) do not conform to the worldwide pattern that returns to women's education are generally higher than those to men's (Schultz, 1993).

References

Ahmad, Asif and Jonathan Morduch (2002) "Identifying Sex Bias in the Allocation of Household Resources: Evidence from linked Household Surveys from Bangladesh", New York University mimeo.

Behrman, Jere, A. Foster, M. Rosenzweig, and P. Vashishtha (1999) "Women's Schooling, Home Teaching, and Economic Growth", *Journal of Political Economy,* 107(4): 682-714.

Bhalotra, Sonia and Chris Attfield (1998) "Intrahousehold Resource Allocation in Rural Pakistan: A Semiparametric Analysis", *Journal of Applied Econometrics*, 13: 463-480.

Case, Anne and Angus Deaton (2003) "Consumption, health, gender and poverty", Working Paper No: 3020, World Bank Poverty Reduction and Economic Management Network.

Datt, G. and M. Ravallion (1998) "Why have some Indian States Done Better Than Others at Reducing Rural Poverty?", *Economica*; 65(257): 17-38.

Deaton, Angus (1997) "The Analysis of Household Surveys: A Microeconometric Approach to Development Policy", Johns Hopkins Press for the World Bank.

Deaton, Angus, and C. Paxson (1998) "Economies of Scale, Household Size, and the Demand for Food", *Journal of Political Economy*, 106, No. 5: 897-910.

Drèze, Jean and Geeta G. Kingdon (2001) "Schooling Participation in Rural India", *Review of Development Economics*, 5, No. 1, February: 1-26.

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

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 Number 1, *Sarvekshana*, Government of India.

Duraisamy, Palanigounder (2002) "Changes in Returns to Education in India, 1983-94: By Gender, Age-Cohort and Location", *Economics of Education Review*, 21(4), pp. 609-622.

Foster, Andrew and M. Rosenzweig (2000) "Missing Women, the Marriage Market, and Economic Growth", Mimeo, Brown University, August.

Kingdon, Geeta Gandhi (1996a) "The Quality and Efficiency of Private and Public Education: A Case Study of Urban India", *Oxford Bulletin of Economics and Statistics,* 58(1): 57-81.

Kingdon, Geeta Gandhi (1996b) "Private Schooling in India: Size, Nature and Equity Effects", *Economic and Political Weekly*, 31(51): 3306-3314.

Kingdon, Geeta Gandhi (1998) "Does the Labour Market Explain Lower Female Schooling in India?", *Journal of Development Studies*, 35 (1): 39-65.

Kingdon, Geeta Gandhi and Jeemol Unni (2001) "Education and Women's Labour Market Outcomes in India", *Education Economics*, 9 (2):173-195.

Lancaster, G., P. Maitra and R. Ray (2003) "Endogenous Power, Household Expenditure Patterns and New Tests of Gender Bias: Evidence from India", mimeo, Monash University and University of Tasmania.

Pradhan, B. and A. 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*, Oxford University Press, Delhi.

Rose, Elaina "Gender Bias, Credit Constraints and Time Allocation in Rural India", *Economic Journal*, 110 (465): 738-758.

Schultz, T. Paul (1993) "Returns to Women's Education", Chapter 2 in E. King and M. Hill (eds.) *Women's Education in Developing Countries*, Washington D.C., Johns Hopkins Press for the World Bank.

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

Subramanian, S. and Angus Deaton (1990) "Gender Effects in Indian Consumption Patterns", Discussion Paper 147, Research Programme in Development Studies, Woodrow Wilson School, Princeton University, also published in *Sarvekshana*, 14: 1-12, 1991.

Subramanian, Shankar (1995) "Gender Discrimination in Intra-Household Allocation in India", mimeo, Department of Economics, Cornell University.

Wooldridge, Jeffery (2002) "Econometric Analysis of Cross-Section and Panel Data", MIT Press.

Table 1
Descriptive statistics, by State

STATE	Proportion of girls in all children (aged 0-14)	Proportion of 'all-girl' households in all households	% of 'at-least-one-boy' households that incurred positive education expenditure	% of 'all-girl' households that incurred positive education expenditure	Percentage point difference (d) - (e)	t-value of the difference in (d) and (e)
(a)	(b)	(c)	(d)	(e)	(f)	(g)
ANDRHA	48.1	25.6	62.8	48.9	13.9	4.8
BIHAR	44.7	17.1	56.3	43.2	13.1	4.3
GUJARAT	45.9	17.8	62.9	42.6	20.3	5.5
HARYANA	46.3	15.5	72.4	52.2	20.2	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
MAHARASH	46.3	19.2	69.2	48.3	20.9	7.8
MADHYA	46.4	18.5	61.1	42.3	18.8	8.6
ORISSA	48.4	21.1	64.7	44.0	20.7	6.7
PUNJAB	46.4	17.7	70.5	46.5	24.0	6.0
RAJASTHAN	45.0	15.0	67.9	32.3	35.6	11.1
TAMILNADU	48.7	28.5	60.1	39.5	20.6	6.2
UTTAR	44.8	15.0	66.9	44.1	22.8	9.6
W.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
INDIA	46.4	19.0	66.0	47.3	18.7	24.4

Note: Shaded cells represent cells with values above or below the national average. 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.

Table 2
Current enrolment rate of children, by age-group and gender

State	Age 5-9			Age 10-14			Age 15-19		
	female	male	gap	female	male	gap	female	male	gap
ANDRHA	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
MAHARASH	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
W.BENGAL	47	48	1	62	66	4	25	40	15
ASSAM	52	60	8	77	86	9	49	59	10
INDIA	51	60	9	60	76	16	24	47	23

Note: The shaded cells represent statistically significant gender-gaps, at the 5% level.

Table 3
Educational expenditure on ENROLLED children, by age-group and gender

	Age 5 - 9			Age 10-14			Age 15-19		
	girls	boys	t	girls	boys	t	girls	boys	t
ANDRHA	258	219	-1.0	305	330	0.5	864	885	0.1
BIHAR	249	309	2.1	378	431	1.4	651	652	0.0
GUJARAT	258	247	-0.3	313	350	1.0	912	1171	1.5
HARYANA	633	634	0.0	721	859	2.3	1115	1434	2.6
HIMACHAL	671	707	0.6	974	1049	1.2	1686	1966	1.9
KARNATAK	285	337	1.3	446	455	0.2	751	918	1.8
KERALA	490	611	2.7	677	745	1.3	1269	1373	0.8
MAHARASH	210	222	1.1	359	397	1.7	688	786	1.7
MADHYA	218	242	1.6	301	289	-0.7	651	582	-1.1
ORISSA	222	188	-1.6	295	289	-0.3	852	831	-0.2
PUNJAB	498	651	2.3	674	793	2.0	1712	1365	-2.0
RAJASTHAN	324	348	1.0	496	520	0.8	1109	1164	0.4
TAMILNADU	333	331	-0.0	386	418	0.6	1069	910	-0.8
UTTAR	343	316	-1.0	375	411	1.8	710	780	1.0
W.BENGAL	200	204	0.1	382	379	-0.1	863	945	0.8
ASSAM	357	353	-0.1	352	449	2.2	905	1007	0.6
INDIA	331	345	1.5	455	477	2.2	981	994	0.4

Note: The shaded cells represent statistically significant gender-gaps, at the 5% level.

Table 4
Educational expenditure on all (enrolled and non-enrolled) children, by age-group and gender

	Age 5 - 9			Age 10-14			Age 15-19		
	girls	boys	t	girls	boys	t	girls	boys	t
ANDRHA	168	168	0.0	174	232	1.7	130	355	4.9
BIHAR	88	142	4.1	191	275	3.4	153	302	5.0
GUJARAT	147	160	0.5	212	288	2.6	215	514	4.6
HARYANA	348	378	0.9	503	731	4.5	236	703	8.7
HIMACHAL	528	586	1.2	872	989	2.0	780	1458	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	0.9
MAHARASH	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	0.3	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
W.BENGAL	95	99	0.3	235	249	0.6	212	376	3.8
ASSAM	186	210	0.9	271	387	3.0	444	593	1.5
INDIA	170	206	6.6	274	364	12.1	234	468	19.5

Note: The shaded cells represent statistically significant gender-gaps, at the 5% level.

Table 5a
Difference in Marginal Effect (DME) x 100 of gender variables male5-9 and female5-9, and p-value of the associated t-test
(Household-level results)

State	Probit (a)	Conditional OLS (b)	Combined probit+OLS (c) = *f*(a,b)	Unconditional OLS (Conventional Engel curve) (d)
AP	60.09	-1.38	1.67	-0.03
	(.00)	(.41)	(.24)	(.98)
BIH	24.53	-0.73	1.00	0.40
	(.10)	(.73)	(.54)	(.76)
GUJ	41.55	-2.42	0.31	0.30
	(.04)	(.34)	(.87)	(.87)
HAR	10.41	3.78	3.95	3.17
	(.49)	(.11)	(.11)	(.11)
HIM	11.79	-2.71	-1.48	0.93
	(.25)	(.37)	(.62)	(.72)
KAR	-1.94	-2.03	-1.67	-0.31
	(.88)	(.30)	(.32)	(.83)
KER	11.02	2.86	3.71	3.65
	(.10)	(.22)	(.11)	(.12)
MAH	17.24	-2.30	-0.82	0.56
	(.18)	(.17)	(.61)	(.68)
MP	14.50	-0.55	0.31	0.60
	(.19)	(.66)	(.75)	(.45)
ORI	70.57	-2.68	1.00	0.23
	(.00)	(.07)	(.40)	(.86)
PUN	30.46	2.70	4.11	2.85
	(.11)	(.46)	(.24)	(.20)
RAJ	40.86	3.34	4.32	4.10
	(.00)	(.03)	(.00)	(.00)
TN	50.96	1.26	3.50	2.93
	(.01)	(.59)	(.08)	(.11)
UP	32.99	-0.58	1.31	1.55
	(.00)	(.64)	(.19)	(.11)
WB	18.82	-0.18	0.62	0.77
	(.30)	(.92)	(.66)	(.58)
ASS	5.33	2.62	2.46	-0.12
	(.73)	(.24)	(.25)	(.94)

Note: In the conditional OLS equation fitted only for households with positive education spending, the dependant 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 column (b) are comparable to those in column (d), where the dependent variable is in absolute rather than log terms. Column (d) 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 to 9'. The figures in parentheses are p-values of the t-test of the DME, where standard errors for the t-test in column (c) were obtained by bootstrapping.

Table 5b
Difference in Marginal Effect (DME) x 100 of gender variables male10-14 and female10-14, and p-value of the associated t-test
(Household-level results)

State	Probit (a)	Conditional OLS (b)	Combined probit+OLS (c) = *f*(a,b)	Unconditional OLS (Conventional Engel curve) (d)
AP	19.70	-0.11	0.78	1.23
	(.25)	(.95)	(.55)	(.34)
BIH	33.90	-2.35	0.57	-0.54
	(.04)	(.26)	(.70)	(.70)
GUJ	67.30	-3.58	0.75	0.77
	(.00)	(.15)	(.71)	(.67)
HAR	16.63	-0.62	0.73	-1.19
	(.32)	(.78)	(.72)	(.54)
HIM	11.96	0.92	2.01	1.28
	(.25)	(.74)	(.36)	(.60)
KAR	5.92	-0.02	0.34	0.21
	(.65)	(.99)	(.83)	(.88)
KER	-8.86	-0.81	-1.53	-0.55
	(.50)	(.72)	(NA)	(.81)
MAH	43.70	0.50	3.05	2.21
	(.00)	(.75)	(.03)	(.09)
MP	42.99	-0.92	1.40	-0.22
	(.00)	(.45)	(.16)	(.80)
ORI	73.26	0.24	3.16	-0.20
	(.00)	(.88)	(.01)	(.88)
PUN	14.54	4.05	4.20	3.69
	(.49)	(.26)	(.21)	(.11)
RAJ	108.87	3.13	7.35	5.33
	(.00)	(.04)	(.00)	(.00)
TN	42.80	1.79	3.51	2.37
	(.07)	(.45)	(.10)	(.22)
UP	56.12	-0.50	2.56	0.87
	(.00)	(.68)	(.02)	(.38)
WB	-3.53	-1.16	-0.90	-2.20
	(.86)	(.52)	(.50)	(.15)
ASS	16.64	5.10	5.09	3.51
	(.50)	(.08)	(.04)	(.13)

Note: See note in Table 5a. The table displays 100 times the difference in marginal effects (DME) of the variables 'proportion of males aged 10-14' and 'proportion of females aged 10 to 14'.

Table 5c
Difference in Marginal Effect (DME) x 100 of gender variables male15-19 and female15-19, and p-value of the associated t-test
(Household-level results)

State	Probit (a)	Conditional OLS (b)	Combined probit+OLS (c) = *f*(a,b)	Unconditional OLS (Conventional Engel curve) (d)
AP	66.25 (.00)	2.57 (.31)	4.62 (.01)	3.41 (.01)
BIH	60.75 (.00)	0.34 (.91)	3.36 (.10)	4.64 (.01)
GUJ	17.13 (.38)	4.50 (.12)	3.98 (.06)	4.59 (.01)
HAR	50.03 (.00)	4.67 (.11)	7.67 (.00)	6.49 (.01)
HIM	15.71 (.02)	8.27 (.01)	9.37 (.00)	10.14 (.00)
KAR	39.19 (.00)	5.80 (.01)	6.78 (.00)	5.93 (.00)
KER	-0.49 (.94)	-2.88 (.25)	-2.82 (.26)	-0.57 (.81)
MAH	48.13 (.00)	3.48 (.07)	5.72 (.00)	6.60 (.00)
MP	66.29 (.00)	5.32 (.00)	6.57 (.00)	4.92 (.00)
ORI	59.59 (.00)	7.10 (.00)	7.40 (.00)	5.78 (.00)
PUN	8.96 (.60)	2.39 (.55)	2.51 (.40)	1.13 (.61)
RAJ	102.35 (.00)	9.44 (.00)	11.58 (.00)	10.01 (.00)
TN	34.39 (.09)	3.22 (.29)	4.18 (.09)	1.97 (.30)
UP	38.72 (.00)	6.87 (.00)	6.88 (.00)	6.57 (.00)
WB	36.93 (.12)	3.62 (.12)	3.81 (.04)	2.82 (.10)
ASS	-19.24 (.44)	3.79 (.31)	2.12 (.54)	3.90 (.14)

Note: See note in Table 5a. The table displays 100 times the difference in marginal effects (DME) of the variables 'proportion of males aged 15-19' and 'proportion of females aged 15 to 19'.

Table 6a
Marginal effect of the gender dummy variable MALE and p-value of the associated t-test, Individual level data, age group 5-9

State	Probit	Conditional OLS	Combined probit+OLS	Unconditional OLS
	(a)	(b)	(c) = *f*(a,b)	(d)
AP	0.129	0.27	27.3	14.9
	(.00)	(.99)	(.01)	(.39)
BIH	0.105	17.4	32.9	40.5
	(.00)	(.28)	(.00)	(.00)
GUJ	0.056	12.2	18.8	10.0
	(.08)	(.56)	(.09)	(.61)
HAR	0.039	74.9	62.4	58.1
	(.16)	(.01)	(.02)	(.02)
HIM	0.042	12.3	37.5	54.8
	(.13)	(.74)	(.32)	(.16)
KAR	0.046	-5.6	9.3	24.5
	(.05)	(.76)	(.56)	(.15)
KER	0.042	41.8	55.0	92.4
	(.15)	(.10)	(.04)	(.01)
MAH	0.005	-6.1	-2.7	4.8
	(.82)	(.61)	(.25)	(.57)
MP	0.074	21.5	24.4	26.7
	(.00)	(.04)	(.00)	(.00)
ORI	0.077	-31.7	-3.8	-7.2
	(.01)	(.00)	(.22)	(.56)
PUN	0.096	93.5	120.2	116.7
	(.00)	(.12)	(.00)	(.01)
RAJ	0.266	38.3	96.5	94.8
	(.00)	(.02)	(.00)	(.00)
TN	0.132	11.2	39.6	31.9
	(.00)	(.63)	(.00)	(.31)
UP	0.175	-10.2	46.0	40.4
	(.00)	(.41)	(.00)	(.00)
WB	0.015	2.4	3.5	5.6
	(.62)	(.85)	(.85)	(.65)
ASS	0.078	11.1	28.3	17.7
	(.02)	(.62)	(.13)	(.83)

Note: In the conditional OLS equation fitted only for children with positive education spending, the dependant 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 column (b) are comparable to those in column (d), where the dependent variable is in absolute rather than log terms. Column (d) 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 column (c) are obtained by bootstrapping.

Table 6b
Marginal effect of the gender dummy variable MALE and p-value of the associated t-test, Individual level data, age group 10-14

State	Probit (a)	Conditional OLS (b)	Combined probit+OLS (c) = *f*(a,b)	Unconditional OLS (d)
AP	0.140 (.00)	-1.6 (.93)	35.2 (.41)	40.4 (.04)
BIH	0.178 (.00)	12.6 (.45)	64.4 (.00)	72.7 (.00)
GUJ	0.161 (.00)	14.9 (.57)	59.9 (.02)	57.1 (.02)
HAR	0.164 (.00)	63.5 (.04)	157.9 (.00)	153.9 (.00)
HIM	0.037 (.01)	105.0 (.01)	134.5 (.01)	135.0 (.01)
KAR	0.121 (.00)	-4.0 (.83)	40.2 (.23)	26.0 (.13)
KER	-0.012 (.28)	-13.1 (.60)	-20.3 (.44)	-26.6 (.41)
MAH	0.143 (.00)	21.3 (.09)	66.4 (.00)	65.4 (.00)
MP	0.196 (.00)	16.3 (.16)	63.2 (.00)	42.0 (.00)
ORI	0.242 (.00)	4.3 (.75)	60.5 (.00)	42.8 (.00)
PUN	0.085 (.01)	68.5 (.16)	111.8 (.00)	118.0 (.02)
RAJ	0.515 (.00)	59.5 (.01)	262.9 (.00)	230.2 (.00)
TN	0.145 (.00)	10.4 (.68)	54.0 (.00)	52.0 (.12)
UP	0.289 (.00)	26.6 (.04)	120.4 (.00)	106.7 (.00)
WB	0.048 (.13)	3.3 (.89)	17.4 (.08)	3.9 (.84)
ASS	0.048 (.07)	42.7 (.12)	54.1 (.04)	34.9 (.22)

Note: See note in Table 6a.

Table 6c
Marginal effect of the gender dummy variable MALE and p-value of the associated t-test, Individual level data, age group 15-19

State	Probit (a)	Conditional OLS (b)	Combined probit+OLS (c) = *f*(a,b)	Unconditional OLS (d)
AP	0.269 (.00)	0.4 (.99)	152.0 (.00)	166.5 (.00)
BIH	0.248 (.00)	34.7 (.40)	142.0 (.00)	144.9 (.00)
GUJ	0.198 (.00)	234.1 (.03)	212.8 (.00)	211.2 (.00)
HAR	0.311 (.00)	286.7 (.00)	419.6 (.00)	433.4 (.00)
HIM	0.306 (.00)	102.0 (.19)	519.1 (.00)	522.4 (.00)
KAR	0.184 (.00)	48.4 (.30)	135.8 (.00)	157.1 (.00)
KER	0.019 (.66)	-86.5 (.15)	-27.8 (.95)	-23.8 (.66)
MAH	0.300 (.00)	74.5 (.06)	211.1 (.00)	203.3 (.00)
MP	0.300 (.00)	11.6 (.75)	151.8 (.00)	149.1 (.00)
ORI	0.248 (.00)	64.0 (.14)	131.2 (.00)	173.3 (.00)
PUN	0.216 (.00)	7.8 (.94)	265.8 (.00)	202.8 (.00)
RAJ	0.384 (.00)	-57.1 (.65)	362.8 (.00)	404.2 (.00)
TN	0.171 (.00)	-46.8 (.57)	91.7 (.00)	83.5 (.05)
UP	0.312 (.00)	98.5 (.01)	211.0 (.00)	226.1 (.00)
WB	0.189 (.00)	18.1 (.78)	124.4 (.00)	167.9 (.00)
ASS	0.113 (.03)	103.6 (.15)	130.3 (.00)	136.1 (.08)

Note: See note in Table 6a.

Table 7
Marginal effect (x100) of the gender variables in the probit equation of ANYEDEXP and p-value of the associated t-test:
(Gender variables in household level equation redefined)

	Marginal effect of MALE in individual level probit (1)			**Marginal effect of gender variable in household level probit (2)**		
	Age 5-9	Age 10-14	Age 15-19	Age 5-9	Age 10-14	Age 15-19
AP	0.129 (.00)	0.140 (.00)	0.269 (.00)	0,127 (.00)	0.069 0.09	0.203 (.00)
BIH	0.105 (.00)	0.178 (.00)	0.248 (.00)	0.074 (.03)	0.083 (.02)	0.143 (.00)
GUJ	0.056 (.08)	0.161 (.00)	0.198 (.00)	0.094 (.03)	0.132 (.01)	0.065 (.16)
HAR	0.039 (.16)	0.164 (.00)	0.311 (.00)	0.008 (.80)	0.055 (.14)	0.145 (.00)
HIM	0.042 (.13)	0.037 (.01)	0.306 (.00)	0.058 (.02)	0.011 (.65)	0.046 (.01)
KAR	0.046 (.05)	0.121 (.00)	0.184 (.00)	0.013 (.68)	0.010 (.76)	0.108 (.00)
KER	0.042 (.15)	-0.012 (.28)	0.019 (.66)	0.043 (.04)	-0.047 (.26)	0.009 (.67)
MAH	0.005 (.82)	0.143 (.00)	0.300 (.00)	0.039 (.18)	0.121 (.00)	0.143 (.00)
MP	0.074 (.00)	0.196 (.00)	0.300 (.00)	0.042 (.10)	0.136 (.00)	0.152 (.00)
ORI	0.077 (.01)	0.242 (.00)	0.248 (.00)	0.133 (.00)	0.198 (.00)	0.152 (.00)
PUN	0.096 (.00)	0.085 (.01)	0.216 (.00)	0.066 (.11)	0.061 (.24)	0.018 (.66)
RAJ	0.266 (.00)	0.515 (.00)	0.384 (.00)	0.136 (.00)	0.236 (.00)	0.220 (.00)
TN	0.132 (.00)	0.145 (.00)	0.171 (.00)	0.130 (.01)	0.141 (.02)	0.125 (.02)
UP	0.175 (.00)	0.289 (.00)	0.312 (.00)	0.091 (.00)	0.144 (.00)	0.094 (.00)
WB	0.015 (.62)	0.048 (.13)	0.189 (.00)	0.031 (.47)	0.015 (.74)	0.129 (.01)
ASS	0.078 (.02)	0.048 (.07)	0.113 (.03)	0.020 (.59)	0.075 (.17)	0.031 (.58)
Average marginal effect	0.086	0.156	0.236	0.069	0.090	0.111

Note: In the individual level probit, the gender variable is simply the MALE gender dummy. In the household level probit in column (2), there were three gender variables: for each of the three age groups, the 'proportion of males in all children of that age group within the household'. The column (1) figures are reproduced from the first columns of Tables 6a, 6b and 6c.

Appendix Table 1

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 expenditure

	Andhra Pradesh						**Bihar**						**Gujarat**					
	Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS	
Dep. variable	(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)	
	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value
Variable																		
LNPCE	0.56	2.1	0.19	5.0	-0.40	-4.3	0.37	1.4	0.21	6.6	-0.58	-7.7	1.12	3.7	0.15	4.4	0.06	0.6
LNHHSIZE	1.55	5.0	0.40	8.7	-0.15	-1.2	0.57	1.7	0.31	7.3	-0.29	-3.0	1.28	3.0	0.36	7.3	-0.19	-1.3
M0TO4	-6.79	-3.0	-1.01	-3.3	-2.38	-2.8	-0.92	-0.3	-0.29	-0.8	-0.86	-1.0	-4.53	-1.5	-1.13	-3.2	-0.73	-0.7
M5TO9	1.25	0.6	1.30	4.2	-1.12	-1.4	4.82	1.7	0.67	2.0	0.03	0.0	-1.54	-0.5	0.49	1.4	-1.58	-1.7
M10TO14	1.65	0.7	0.56	1.9	-0.39	-0.5	8.89	3.1	0.88	2.6	1.13	1.3	2.96	1.0	0.77	2.3	-0.44	-0.5
M15TO19	3.76	1.7	-0.16	-0.5	1.16	1.4	8.21	2.8	0.41	1.2	1.55	1.8	3.30	1.1	-0.41	-1.3	1.20	1.3
M20TO24	-3.78	-1.6	-0.48	-1.5	0.07	0.1	5.73	1.8	-0.12	-0.3	1.39	1.5	-3.28	-1.0	-1.10	-3.1	-0.54	-0.5
M25TO60	-3.17	-1.4	-0.45	-1.5	-1.38	-1.7	3.86	1.3	-0.08	-0.2	0.56	0.6	-2.36	-0.8	-0.38	-1.1	-0.67	-0.7
M61MORE	-4.68	-1.6	-0.68	-1.8	-0.80	-0.8	6.03	1.5	-0.01	0.0	0.69	0.6	-4.54	-1.1	-1.13	-2.3	-0.60	-0.4
F0TO4	-5.86	-2.6	-0.47	-1.5	-2.97	-3.6	-0.96	-0.3	-0.31	-0.9	-0.79	-0.9	-5.29	-1.8	-0.90	-2.6	-1.59	-1.7
F5TO9	1.28	0.6	0.70	2.4	-0.80	-1.0	4.42	1.5	0.42	1.3	0.15	0.2	-1.84	-0.6	0.07	0.2	-1.08	-1.2
F10TO14	0.42	0.2	0.36	1.2	-0.36	-0.5	9.43	3.2	0.54	1.6	1.53	1.8	2.19	0.7	0.09	0.3	0.30	0.3
F15TO19	0.35	0.2	-0.82	-2.7	0.57	0.7	3.57	1.2	-0.20	-0.5	1.60	1.8	-1.29	-0.4	-0.58	-1.7	0.27	0.3
F20TO24	-2.68	-1.1	-0.20	-0.6	-1.16	-1.3	-0.02	0.0	-0.47	-1.3	0.40	0.4	-1.11	-0.3	-0.77	-2.1	-0.12	-0.1
F25TO60	-0.17	-0.1	0.44	1.5	-0.88	-1.1	2.09	0.7	0.21	0.6	0.06	0.1	3.00	1.0	0.02	0.1	0.91	1.0
HEDYRS	0.37	9.1	0.05	8.5	0.05	3.9	0.37	10.1	0.05	10.6	0.05	5.0	0.43	7.8	0.05	7.9	0.05	3.4
SC	-0.55	-2.3	0.00	0.0	-0.27	-3.3	-0.66	-2.3	-0.06	-1.9	-0.18	-2.3	0.31	0.7	0.00	-0.1	0.22	1.7
ST	-1.17	-2.0	-0.08	-1.1	-0.53	-2.4	-0.77	-2.4	-0.07	-1.8	-0.28	-3.1	-0.16	-0.5	-0.03	-0.8	-0.16	-1.5
MUSLIM	-0.37	-0.8	0.05	0.8	-0.22	-1.4	-0.18	-0.6	-0.11	-3.4	0.04	0.5	-0.10	-0.2	-0.07	-1.1	-0.02	-0.1
CHRISTN	0.87	1.3	0.04	0.4	0.32	1.3	0.97	0.7	0.27	1.8	-0.05	-0.1	1.78	0.9	---	---	0.25	0.5
WAGELAB	-0.56	-2.4	-0.08	-2.5	-0.20	-2.4	-0.77	-2.8	-0.11	-3.6	-0.09	-1.2	-0.25	-0.8	-0.04	-1.1	0.01	0.1
INTERCEP	-3.42	-1.1			0.52	0.5	-4.85	-1.4			1.14	1.1	-8.52	-2.3			-3.72	-3.3
Adjusted R^2	0.1444		0.2755		0.1615		0.1570		0.1960		0.1506		0.1354		0.2250		0.1152	
N	1571		1571		1001		1787		1787		1042		1182		1182		776	
Depvar mean	0.0269		0.6372		-3.7624		0.0346		0.5831		-3.2549		0.0317		0.6548		-3.6331	
Exp. elasticity	1.21						1.11						1.35					
p-value: age5-9	0.98		0.00		0.41		0.76		0.10		0.73		0.87		0.04		0.34	
age 10-14	0.34		0.25		0.95		0.70		0.04		0.26		0.67		0.00		0.15	
age 15-19	0.01		0.00		0.31		0.01		0.00		0.91		0.01		0.38		0.12	

Note: 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 a dash ---. For example, where all Christian households have anyedexp=1, then the marginal effect of that variable is not identified and it is denoted with a dash ---. Similarly, if there are no Christians in the rural part of a state in the sample, this is denoted with a dash ---. The last 3 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.

Appendix Table 1, continued

Dep. variable	Haryana						Himachal Pradesh						Karnataka					
	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value
Variable																		
LNPCE	1.83	4.5	0.19	6.2	-0.01	-0.1	0.17	0.4	0.09	5.4	-0.31	-5.1	0.19	0.6	0.10	3.6	-0.34	-5.0
LNHHSIZE	2.49	5.6	0.31	9.1	0.04	0.6	3.57	6.4	0.14	6.7	0.12	1.6	1.58	4.6	0.29	9.3	-0.08	-1.0
M0TO4	-5.83	-1.8	-0.25	-1.1	-1.46	-2.5	-13.34	-3.6	-0.30	-2.4	-2.01	-4.2	-5.22	-2.0	-0.54	-2.5	-1.37	-2.2
M5TO9	9.23	2.9	0.76	3.3	0.28	0.5	2.57	0.8	0.25	1.8	-0.16	-0.4	2.55	1.0	0.55	2.6	-0.10	-0.2
M10TO14	11.75	3.6	1.02	4.2	0.80	1.4	13.43	4.1	0.21	1.7	1.46	3.6	7.67	3.2	0.75	3.7	1.28	2.3
M15TO19	11.99	3.7	0.36	1.6	1.34	2.3	14.21	4.3	-0.05	-0.4	1.87	4.5	8.13	3.3	0.16	0.8	2.35	4.1
M20TO24	1.68	0.5	-0.13	-0.6	0.31	0.5	-5.44	-1.5	-0.31	-2.6	0.02	0.1	3.34	1.3	-0.40	-1.9	1.17	1.9
M25TO60	1.05	0.3	0.05	0.2	-0.43	-0.7	-7.17	-2.3	-0.25	-2.3	-0.89	-2.2	-0.96	-0.4	-0.35	-1.7	0.60	1.0
M61MORE	-1.37	-0.3	-0.08	-0.2	-0.35	-0.4	-7.58	-1.8	-0.25	-1.8	-0.65	-1.2	-1.94	-0.6	-0.28	-1.0	0.58	0.7
F0TO4	-4.45	-1.3	-0.18	-0.8	-1.52	-2.6	-10.16	-2.6	-0.24	-1.9	-1.55	-3.2	-4.27	-1.6	-0.41	-1.8	-0.69	-1.1
F5TO9	6.06	1.8	0.65	2.7	-0.22	-0.4	1.64	0.5	0.13	1.0	0.13	0.3	2.86	1.1	0.57	2.7	0.24	0.4
F10TO14	12.94	4.1	0.86	3.6	0.89	1.6	12.15	3.6	0.09	0.7	1.36	3.3	7.46	3.1	0.69	3.3	1.28	2.3
F15TO19	5.50	1.6	-0.14	-0.6	0.72	1.2	4.07	1.2	-0.21	-1.8	0.98	2.3	2.20	0.9	-0.23	-1.1	1.37	2.3
F20TO24	-1.59	-0.4	-0.34	-1.4	-0.27	-0.4	-0.40	-0.1	-0.20	-1.7	0.15	0.3	2.02	0.7	-0.33	-1.5	1.17	1.7
F25TO60	3.04	1.0	0.20	1.0	-0.14	-0.3	-5.05	-1.6	-0.04	-0.4	-0.65	-1.6	1.92	0.8	-0.02	-0.1	0.66	1.1
HEDYRS	0.37	7.4	0.02	4.5	0.05	5.9	0.39	5.6	0.01	2.7	0.05	5.3	0.28	6.3	0.02	5.3	0.04	4.3
SC	-0.36	-1.0	-0.04	-1.4	-0.04	-0.7	-0.44	-1.0	-0.02	-1.1	-0.05	-0.8	-0.09	-0.3	-0.08	-2.6	0.03	0.4
ST	1.39	0.6	---	---	0.38	1.1	-1.10	-1.1	-0.10	-2.2	-0.09	-0.7	-0.53	-1.4	-0.02	-0.5	-0.20	-2.3
MUSLIM	-2.48	-3.7	-0.39	-6.5	-0.12	-0.9	-3.51	-2.7	-0.26	-3.3	-0.26	-1.5	-0.63	-1.8	-0.10	-3.2	0.02	0.2
CHRISTN	-4.59	-0.8	---	---	---	---	---	---	---	---	---	---	2.51	1.7	-0.08	-0.6	0.63	1.8
WAGELAB	-0.86	-2.3	-0.06	-2.3	-0.15	-2.3	-0.82	-1.0	-0.05	-1.7	-0.08	-0.7	-0.83	-2.8	-0.08	-3.3	-0.13	-1.9
INTERCEP	-17.49	-3.9			-2.94	-3.7	0.36	0.1			-0.32	-0.5	-2.45	-0.7			-1.32	-1.7
Adjusted R^2	0.2333		0.2824		0.1897		0.2757		0.3163		0.3046		0.1126		0.1865		0.1389	
N	1409		1409		1074		949		949		838		1979		1979		1435	
Depvar mean	0.0614		0.7619		-2.8315		0.0868		0.8830		-2.5700		0.0427		0.7251		-3.3266	
Exp. elasticity	1.00						1.02						1.04					
p-value of F-test																		
Age 5-9	0.11		0.49		0.11		0.72		0.25		0.37		0.83		0.88		0.30	
Age 10-14	0.54		0.32		0.78		0.60		0.25		0.74		0.88		0.65		0.99	
Age 15-19	0.01		0.00		0.11		0.00		0.02		0.01		0.00		0.00		0.01	

Appendix Table 1, continued

	Kerala						**Maharashtra**						**Madhya Pradesh**					
Dep. variable	Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)		Unconditional OLS (ESHARE)		Probit (ANYEDEXP)		Conditional OLS (LNESHARE)	
	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value
Variable																		
LNPCE	0.50	0.9	0.03	2.3	-0.17	-2.4	0.23	0.9	0.15	6.3	-0.17	-3.0	1.25	6.7	0.31	11.7	-0.09	-1.4
LNHHSIZE	1.72	2.8	0.12	6.9	-0.15	-2.0	1.83	6.0	0.30	9.7	0.08	1.1	1.64	8.7	0.50	17.1	-0.18	-2.7
M0TO4	-8.44	-2.4	-0.08	-1.0	-1.51	-3.2	-5.19	-2.4	-0.53	-2.8	-1.58	-3.2	-3.24	-2.0	-0.51	-2.3	-0.92	-1.6
M5TO9	9.01	2.8	0.53	5.6	0.27	0.7	6.18	3.0	0.91	4.7	0.25	0.6	2.48	1.6	0.59	2.8	0.22	0.4
M10TO14	15.88	5.1	0.68	6.3	1.00	2.6	11.55	5.7	0.95	4.9	1.61	3.7	5.78	3.8	1.02	4.7	0.82	1.6
M15TO19	11.88	3.6	0.23	2.7	0.99	2.4	10.92	5.3	0.34	1.8	1.89	4.2	5.98	3.9	0.31	1.4	2.02	3.8
M20TO24	1.23	0.4	0.07	1.0	-0.06	-0.1	3.20	1.5	-0.22	-1.2	0.61	1.3	-0.37	-0.2	-0.37	-1.5	0.44	0.7
M25TO60	-0.78	-0.3	0.12	1.7	-0.42	-1.1	1.01	0.5	-0.16	-0.9	0.06	0.1	-1.63	-1.0	-0.06	-0.3	-0.42	-0.7
M61MORE	-5.33	-1.2	0.00	0.0	-0.81	-1.5	-3.70	-1.3	-0.30	-1.2	-0.58	-0.9	-1.01	-0.5	-0.12	-0.4	-0.02	0.0
F0TO4	-7.23	-2.1	0.00	0.1	-1.77	-4.0	-4.26	-2.0	-0.52	-2.7	-1.42	-2.9	-3.55	-2.3	-0.41	-1.8	-1.38	-2.5
F5TO9	5.36	1.6	0.42	4.5	-0.06	-0.2	5.62	2.7	0.74	3.8	0.63	1.4	1.88	1.2	0.45	2.1	0.34	0.6
F10TO14	16.43	5.4	0.77	6.1	1.09	2.9	9.34	4.5	0.51	2.7	1.53	3.4	6.00	3.9	0.59	2.7	1.02	1.9
F15TO19	12.45	4.0	0.23	3.0	1.32	3.4	4.32	2.0	-0.14	-0.7	1.32	2.8	1.06	0.7	-0.35	-1.5	0.88	1.6
F20TO24	2.17	0.6	0.01	0.1	0.35	0.8	-0.76	-0.3	-0.28	-1.3	0.11	0.2	-1.63	-0.9	-0.32	-1.3	-0.02	0.0
F25TO60	2.75	0.9	0.17	2.5	-0.50	-1.3	0.69	0.3	0.06	0.4	0.23	0.5	0.25	0.2	0.23	1.1	0.16	0.3
HEDYRS	0.46	5.6	0.01	3.1	0.04	4.1	0.23	5.4	0.03	6.1	0.02	2.8	0.38	14.2	0.05	13.0	0.07	7.8
SC	0.45	0.7	0.00	0.3	-0.03	-0.4	0.29	0.9	0.03	1.0	0.12	1.8	-0.32	-1.8	-0.09	-3.5	-0.07	-1.1
ST	-2.17	-0.9	---	---	-0.45	-1.6	-1.74	-5.9	-0.22	-7.4	-0.24	-3.4	-0.03	-0.2	-0.07	-3.2	-0.02	-0.3
MUSLIM	-1.10	-2.1	-0.01	-0.7	-0.14	-2.1	-0.45	-0.7	-0.02	-0.4	-0.04	-0.3	-0.29	-0.7	-0.12	-1.8	-0.05	-0.4
CHRISTN	1.38	3.0	0.00	0.2	0.10	1.8	1.89	0.9	---	---	0.45	1.2	8.44	7.7	0.18	1.2	1.35	4.3
WAGELAB	-1.25	-3.0	-0.03	-2.2	-0.11	-2.2	-0.68	-2.4	-0.04	-1.5	-0.13	-1.9	-0.29	-1.7	-0.05	-2.3	-0.12	-2.0
INTERCEP	-4.47	-0.8			-1.08	-1.5	-3.48	-1.3			-2.43	-4.1	-10.91	-5.3			-2.80	-3.9
Adjusted R^2	0.2700		0.3599		0.2815		0.2097		0.2767		0.1932		0.2031		0.2405		0.1391	
N	948		948		809		2039		2039		1507		3305		3305		2036	
Depvar mean	0.0788		0.8526		-2.6175		0.0474		0.7384		-3.0901		0.0297		0.6160		-3.4757	
Exp. elasticity	1.06						1.05						1.42					
p-value of F-test																		
Age 5-9	0.12		0.10		0.22		0.68		0.18		0.17		0.45		0.19		0.66	
Age 10-14	0.81		0.50		0.72		0.09		0.00		0.75		0.80		0.00		0.45	
Age 15-19	0.81		0.94		0.25		0.00		0.00		0.07		0.00		0.00		0.00	

Appendix Table 1, continued

	Orissa						Punjab						Rajasthan					
	Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS	
Dep. variable	(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)	
	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value
Variable																		
LNPCE	1.69	5.4	0.34	7.5	0.07	0.8	0.38	0.9	0.18	4.7	-0.25	-2.1	0.05	0.2	0.20	6.7	-0.32	-5.2
LNHHSIZE	1.80	6.2	0.46	10.0	0.02	0.2	1.67	3.5	0.29	6.7	0.12	0.9	2.34	7.9	0.47	11.5	0.20	2.4
M0TO4	0.91	0.4	0.02	0.1	0.23	0.3	-4.47	-1.2	-0.10	-0.3	-1.81	-1.7	-3.27	-1.4	-0.49	-1.7	-1.01	-1.6
M5TO9	5.07	2.1	1.41	4.3	0.22	0.3	8.01	2.1	1.10	3.5	0.93	0.9	6.73	2.9	0.73	2.6	1.43	2.4
M10TO14	7.55	3.2	1.64	4.9	1.40	1.9	9.46	2.6	1.15	3.9	1.05	1.1	11.53	4.9	1.54	5.3	2.20	3.7
M15TO19	8.33	3.5	0.47	1.5	2.56	3.5	8.18	2.2	0.21	0.7	2.19	2.2	10.34	4.3	0.43	1.5	2.84	4.6
M20TO24	2.49	1.0	-0.13	-0.4	1.14	1.5	-3.73	-1.0	0.00	0.0	-0.70	-0.7	5.15	2.0	-0.39	-1.2	1.74	2.5
M25TO60	1.31	0.5	0.36	1.1	-0.43	-0.6	-1.04	-0.3	0.18	0.6	-0.29	-0.3	3.69	1.4	-0.02	-0.1	1.54	2.2
M61MORE	2.23	0.8	0.43	1.1	0.19	0.2	1.39	0.3	0.08	0.2	0.22	0.2	4.99	1.5	-0.09	-0.2	1.98	2.2
F0TO4	1.30	0.6	0.40	1.2	-0.34	-0.5	-6.44	-1.7	-0.04	-0.2	-2.62	-2.5	-2.63	-1.1	-0.55	-1.9	-0.04	-0.1
F5TO9	4.84	2.1	0.71	2.2	0.87	1.2	5.16	1.4	0.80	2.6	0.51	0.5	2.63	1.1	0.32	1.1	0.72	1.2
F10TO14	7.75	3.3	0.91	2.8	1.34	1.9	5.77	1.6	1.01	3.4	0.41	0.4	6.20	2.6	0.45	1.6	1.53	2.6
F15TO19	2.55	1.1	-0.12	-0.4	0.82	1.1	7.05	1.9	0.12	0.4	1.82	1.8	0.33	0.1	-0.59	-1.9	0.83	1.2
F20TO24	1.75	0.7	-0.09	-0.3	0.84	1.0	-2.44	-0.6	-0.23	-0.7	0.76	0.7	1.98	0.7	-0.59	-1.7	1.39	1.7
F25TO60	3.83	1.7	0.72	2.3	0.98	1.3	1.82	0.5	0.34	1.2	0.75	0.7	2.27	1.0	-0.29	-1.0	1.29	1.9
HEDYRS	0.38	8.8	0.05	7.7	0.07	5.7	0.56	9.6	0.03	4.8	0.08	5.3	0.44	10.9	0.03	6.5	0.08	8.3
SC	-0.74	-2.6	0.00	0.0	-0.23	-2.6	0.38	0.9	0.01	0.2	0.02	0.2	-0.12	-0.5	-0.05	-1.7	0.01	0.1
ST	-1.33	-5.0	-0.19	-5.2	-0.47	-5.4	0.43	0.1	---	---	0.25	0.3	-1.33	-4.0	-0.16	-3.8	-0.43	-4.6
MUSLIM	-2.04	-1.9	-0.03	-0.2	-0.40	-1.5	0.14	0.1	-0.11	-1.2	0.25	0.7	-1.17	-2.3	-0.15	-2.4	-0.25	-1.9
CHRISTN	0.45	0.7	0.22	2.8	-0.21	-1.2	0.01	0.0	-0.10	-1.1	-0.26	-0.9	---	---	---	---	---	---
WAGELAB	-0.44	-1.7	-0.05	-1.4	-0.14	-1.7	-1.79	-4.1	-0.07	-2.0	-0.54	-4.6	-0.12	-0.4	0.04	1.2	-0.04	-0.5
INTERCEP	-16.68	-5.0			-4.78	-4.6	-4.79	-0.9			-1.91	-1.4	-5.62	-1.9			-2.62	-3.5
Adjusted R^2	0.2313		0.3102		0.1869		0.2121		0.2516		0.1857		0.2240		0.2742		0.1965	
N	1522		1522		979		964		964		720		1599		1599		1063	
Depvar mean	0.0305		0.6432		-3.5070		0.0456		0.7464		-3.3441		0.0353		0.6648		-3.3213	
Exp. elasticity	1.55						1.08						1.01					
p-value of F-test																		
Age 5-9	0.86		0.00		0.07		0.20		0.11		0.46		0.00		0.00		0.03	
Age 10-14	0.88		0.00		0.88		0.11		0.49		0.26		0.00		0.00		0.04	
Age 15-19	0.00		0.00		0.00		0.61		0.60		0.55		0.00		0.00		0.00	

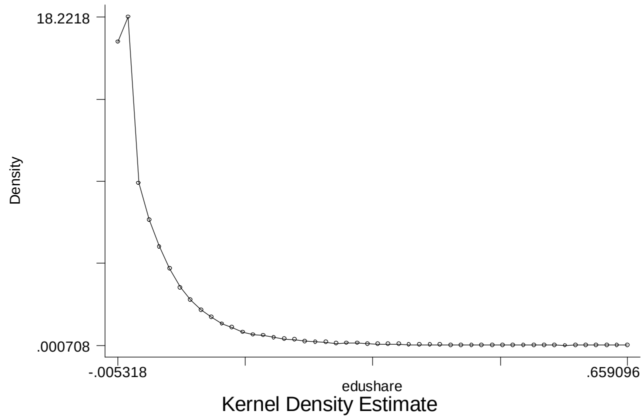
Appendix Table 1, continued

	Tamil Nadu						Uttar Pradesh						West Bengal					
	Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS		Unconditional OLS		Probit		Conditional OLS	
Dep. variable	(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)		(ESHARE)		(ANYEDEXP)		(LNESHARE)	
	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value	coeff x 100	t-value	marginal effect	t-value	coeff	t-value
Variable																		
LNPCE	2.34	6.0	0.22	5.1	0.24	2.2	0.01	0.0	0.16	7.7	-0.46	-9.7	0.90	3.0	0.21	4.8	-0.23	-2.4
LNHHSIZE	2.32	4.7	0.49	8.4	-0.02	-0.1	1.88	8.9	0.42	15.8	-0.08	-1.4	1.14	3.5	0.53	10.8	-0.20	-1.8
M0TO4	-3.51	-1.1	-0.54	-1.7	-1.59	-1.7	-3.73	-2.0	-0.62	-3.0	-0.35	-0.7	-4.23	-1.5	-0.61	-1.5	-2.04	-2.2
M5TO9	8.10	2.6	1.84	5.3	0.01	0.0	3.38	1.9	0.28	1.4	0.96	2.1	0.12	0.0	0.56	1.4	-1.20	-1.3
M10TO14	9.40	3.0	1.80	5.2	0.49	0.5	6.14	3.5	0.66	3.4	1.57	3.5	5.10	1.9	0.58	1.5	1.02	1.2
M15TO19	5.50	1.7	0.43	1.3	0.90	1.0	7.17	4.0	-0.09	-0.5	2.78	6.1	6.11	2.1	-0.05	-0.1	1.84	2.0
M20TO24	0.55	0.2	-0.28	-0.8	-0.23	-0.2	4.29	2.2	-0.31	-1.4	1.83	3.6	2.06	0.7	-0.51	-1.2	0.91	1.0
M25TO60	-3.12	-1.0	-0.24	-0.8	-0.95	-1.0	1.99	1.0	-0.24	-1.1	1.18	2.3	-2.69	-1.0	-0.42	-1.0	-0.86	-0.9
M61MORE	-2.97	-0.7	-0.35	-0.8	-0.39	-0.3	0.96	0.4	-0.33	-1.2	1.25	2.0	-5.55	-1.6	-0.64	-1.3	-0.99	-0.8
F0TO4	-3.83	-1.2	-0.33	-1.0	-2.31	-2.4	-2.07	-1.1	-0.41	-2.0	0.24	0.5	-4.01	-1.4	-1.07	-2.7	-1.50	-1.6
F5TO9	5.17	1.6	1.33	4.0	-0.25	-0.3	1.83	1.0	-0.05	-0.2	1.08	2.4	-0.65	-0.2	0.38	1.0	-1.15	-1.3
F10TO14	7.03	2.3	1.38	4.3	0.12	0.1	5.27	3.0	0.10	0.5	1.67	3.7	7.32	2.7	0.61	1.6	1.32	1.5
F15TO19	3.53	1.1	0.09	0.3	0.25	0.3	0.60	0.3	-0.48	-2.3	1.46	3.0	3.29	1.2	-0.41	-1.0	0.91	1.0
F20TO24	2.03	0.5	-0.48	-1.3	0.70	0.6	-1.13	-0.5	-0.52	-2.2	0.34	0.6	3.92	1.3	-0.19	-0.5	0.48	0.5
F25TO60	1.46	0.5	-0.02	-0.1	-0.50	-0.5	1.64	0.9	0.21	1.0	0.30	0.6	3.02	1.1	0.22	0.6	0.25	0.3
HEDYRS	0.27	4.7	0.02	3.5	0.06	3.8	0.40	14.7	0.04	11.6	0.06	10.1	0.45	10.1	0.04	7.0	0.11	8.0
SC	0.25	0.8	0.05	1.4	0.00	-0.1	-0.60	-3.2	-0.05	-2.2	-0.14	-3.0	-0.49	-1.7	-0.08	-2.0	-0.03	-0.4
ST	0.42	0.2	0.05	0.2	0.41	0.6	-1.32	-2.3	-0.17	-2.5	-0.20	-1.3	-0.82	-1.3	-0.11	-1.4	-0.05	-0.2
MUSLIM	-0.86	-0.7	-0.03	-0.2	-0.30	-0.9	-1.27	-6.2	-0.16	-7.0	-0.18	-3.4	-1.51	-5.0	-0.15	-3.8	-0.36	-3.7
CHRISTN	0.31	0.5	-0.01	-0.1	-0.01	-0.1	5.24	1.2	---	---	0.79	0.9	-0.07	-0.1	0.11	1.0	-0.06	-0.2
WAGELAB	-0.55	-1.7	-0.04	-1.1	0.02	0.2	-0.42	-2.0	-0.04	-1.9	-0.03	-0.5	-0.36	-1.3	-0.09	-2.6	0.06	0.6
INTERCEP	-20.67	-4.8			-5.05	-4.0	-2.65	-1.2			-0.75	-1.3	-7.49	-2.1			-1.62	-1.3
Adjusted R^2	0.1676		0.3388		0.1040		0.1671		0.2009		0.1489		0.2490		0.2532		0.2736	
N	916		916		624		3337		3337		2229		1243		1243		768	
Depvar mean	0.0350		0.6812		-3.4513		0.0363		0.6679		-3.3172		0.0292		0.6179		-3.6473	
Exp. elasticity	1.67						1.00						1.31					
p-value of F-test																		
Age 5-9	0.11		0.01		0.59		0.11		0.00		0.64		0.58		0.30		0.92	
Age 10-14	0.22		0.07		0.45		0.38		0.00		0.68		0.15		0.86		0.52	
Age 15-19	0.30		0.09		0.29		0.00		0.00		0.00		0.10		0.12		0.12	

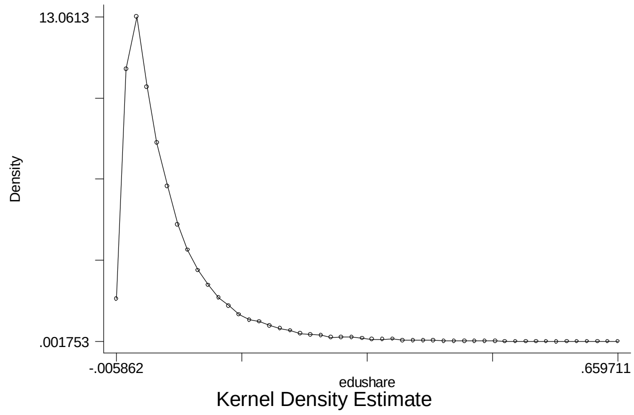
Appendix Table 1, continued

Dep. variable	Assam Unconditional OLS (ESHARE) coeff x 100	t-value	Probit (ANYEDEXP) marginal effect	t-value	Conditional OLS (LNESHARE) coeff	t-value
Variable						
LNPCE	0.98	2.4	0.13	3.4	-0.22	-2.1
LNHHSIZE	2.19	3.7	0.42	7.2	-0.11	-0.7
M0TO4	2.41	0.6	-0.45	-1.3	0.35	0.3
M5TO9	8.87	2.3	0.50	1.4	0.97	0.9
M10TO14	15.24	3.9	0.76	2.1	2.76	2.7
M15TO19	18.83	4.5	0.22	0.6	2.99	2.7
M20TO24	8.22	1.9	-0.71	-1.9	3.18	2.7
M25TO60	2.35	0.6	-0.48	-1.3	0.73	0.7
M61MORE	0.86	0.2	-0.35	-0.8	0.73	0.6
F0TO4	3.64	0.9	-0.39	-1.1	0.71	0.7
F5TO9	8.99	2.3	0.44	1.3	0.47	0.5
F10TO14	11.73	2.9	0.59	1.6	1.79	1.7
F15TO19	14.93	3.6	0.41	1.1	2.27	2.1
F20TO24	13.97	3.1	-0.42	-1.1	1.78	1.4
F25TO60	11.79	2.9	-0.13	-0.4	1.80	1.6
HEDYRS	0.51	8.9	0.03	7.0	0.11	7.0
SC	-2.03	-4.1	-0.04	-0.8	-0.52	-4.2
ST	0.16	0.4	0.02	0.6	0.20	2.0
MUSLIM	-1.88	-4.4	-0.30	-6.5	-0.29	-2.4
CHRISTN	-1.40	-2.8	-0.04	-0.9	-0.18	-1.5
WAGELAB	-0.24	-0.5	-0.11	-2.3	-0.14	-1.0
INTERCEP	-17.03	-3.4			-3.23	-2.4
Adjusted R^2	0.2647		0.3038		0.2051	
N	941		941		707	
Depvar mean	0.0428		0.7513		-3.3651	
Exp. elasticity	1.23					
p-value of F-test						
Age 5-9	0.94		0.73		0.24	
Age 10-14	0.13		0.50		0.08	
Age 15-19	0.14		0.44		0.31	

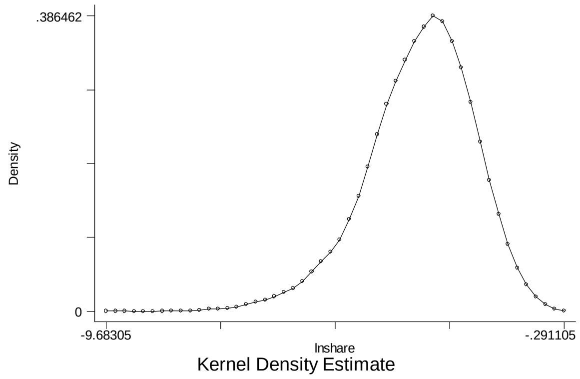
1. kdensity eshare



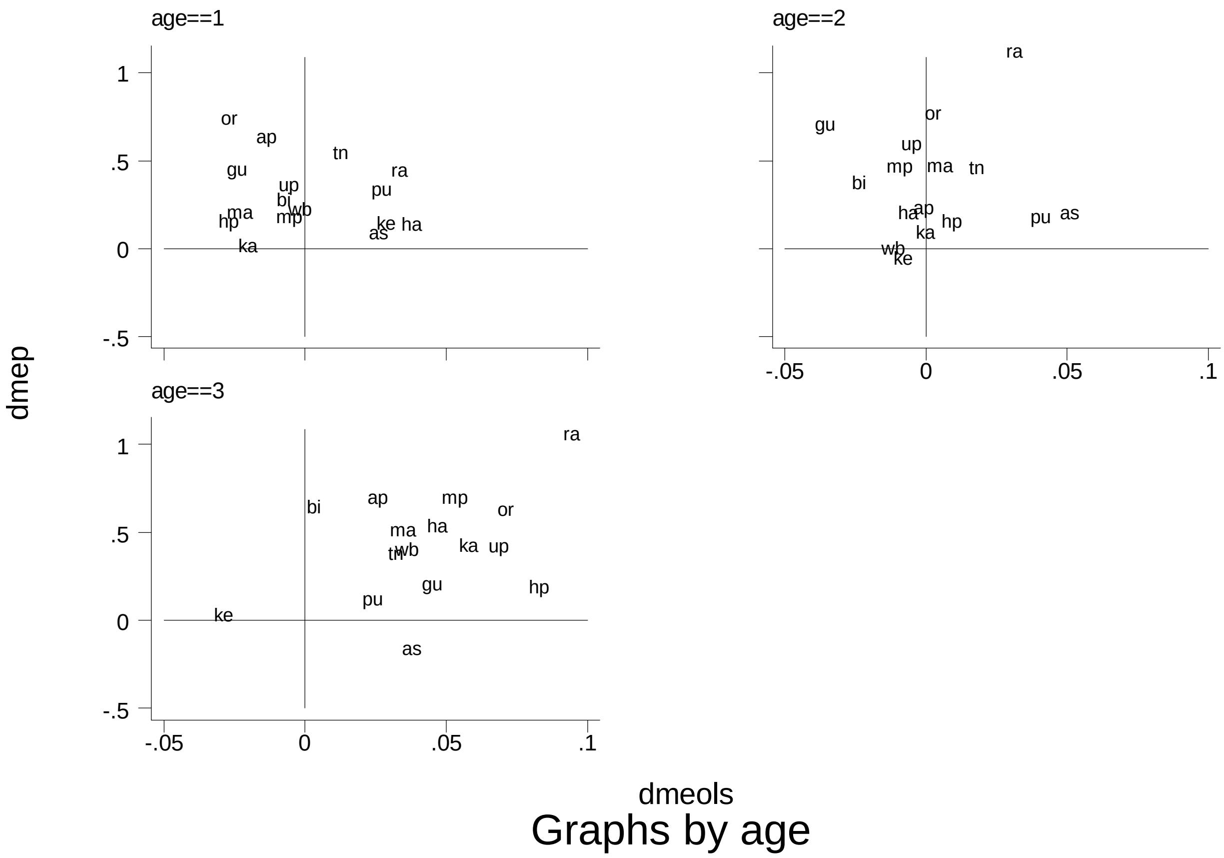
2. kdensity eshare if eshare>0



3. kdensity lneshare if eshare>0



<u>Figure 1</u>



DMEp is the male-female Difference in Marginal Effects in the probit of ANYEDEXP.
DMEols is the male-female Difference in Marginal Effects in the OLS of LNESHARE.

<u>Figure 2</u>

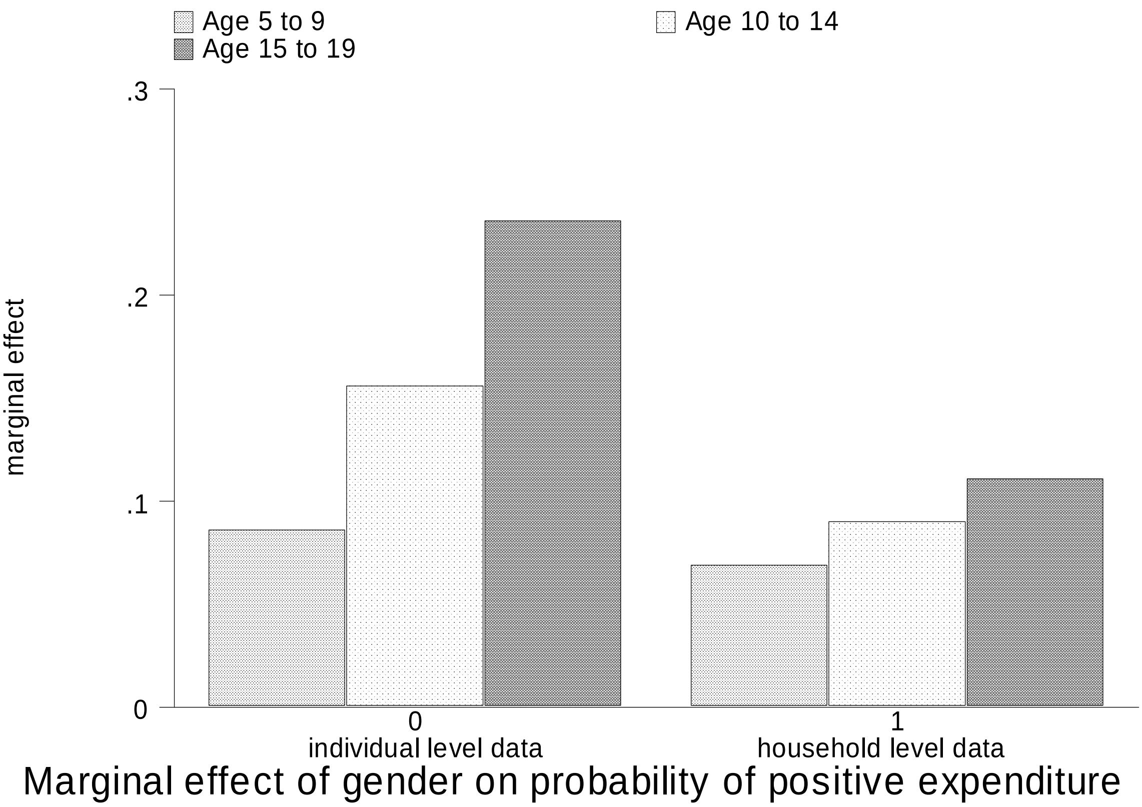

Age 5 to 9
Age 10 to 14
Age 15 to 19
.3
.2
.1
0
marginal effect
0
individual level data
1
household level data
Marginal effect of gender on probability of positive expenditure

Figure 3