Births, Infants and Children: an Econometric Portrait of Women and Children in India

Borooah, Vani

University of Ulster

2003

Births, Infants and Children: an Econometric Portrait of Women and Children in India

Vani K. Borooah[*]
University of Ulster

Abstract

This paper undertakes an econometric analysis of the constellation of factors that serve to determine some outcomes with respect to demography and to schooling in India. These are: the numbers of pregnancies, live births and infant survivals to women and the chances of children being enrolled at school and, if enrolled, of continuing in school. The econometric estimates are based on *unit record data* from a survey - carried out by the National Council of Applied Economic Research (NCAER), New Delhi - of 33,000 *rural* households - encompassing 195,000 individuals - spread over 1,765 villages, in 195 districts, in 16 states of India. The study concludes that a broad spectrum of factors affect these outcomes. The literacy of women is important but so is the literacy of men. Infrastructure, in the form of safe drinking water and easy access to medical facilities, is important for infant survivals and, in the shape of easy access to schools, is important for school enrolment. Parental occupation matters for both infant survivals and schooling: children born to women who work as labourers are disadvantaged, relative to other children, in terms of their chances both of surviving infancy and, if they do survive, of receiving schooling. The number of siblings that a child has affects his/her schooling outcomes and gender, religion and region play an important role.

---

[*] School of Economics and Politics, University of Ulster,
Newtownabbey BT37 0QB, Northern Ireland, UK. (email: vk.borooah@ulst.ac.uk).
This study is based on unit record data from the 1993-94 Human Development Survey carried out by the National Council of Applied Economic Research (NCAER), New Delhi. I am grateful to the NCAER for providing me with the data and to Abusaleh Shariff and Vibha Singh for help and advice with the data. Thanks are also due to two referees whose comments have greatly improved the paper. However, needless to say, I alone am responsible for the results reported in the paper, for their interpretation and, indeed for any of its deficiencies.

1. Introduction

There can be little doubt that, in terms of a number of indicators relating to the welfare of women and children, India's record, relative to that of other countries in Asia, has been woefully inadequate. In the context of some of the indicators which are the focus of this study, India's total fertility rate, over 1995-2000, of 3.1 contrasted poorly, for example, with that of: Thailand's 1.7; China's 1.8; Sri Lanka's 2.1; and Indonesia's 2.6. Its infant mortality rate in 1998 of 69, and under-five mortality rate of 105, per 1,000 live births, was high relative to: Sri Lanka's 17 and 19; Thailand's 30 and 37; China's 38 and 47; and Indonesia's 40 and 56. The 77% of its children – in the primary age group – enrolled at school was low compared to the 99.9% enrolment rates in China and Sri Lanka; the 99.2% rate in Indonesia; and the 88% enrolment rate in Thailand.

Furthermore, India's adult literacy rate of 56% in 1998 needs to be set against: Thailand's 95%; Sri Lanka's 91%; Indonesia's 86%; and China's 83%. Underlying India's low adult literacy rate, is the considerable disparity between the literacy rates of its men (67%) and of its women (44%). Once again, low rates of adult female literacy in India compare unfavourably with corresponding rates in other countries: 93% in Thailand ; 88% in Sri Lanka; 81% in Indonesia; and 75% in China. If there is any solace to be found in this litany of failure it is that, on all the above indicators, India outperformed its unfriendly neighbour - Pakistan! (The United Nations Development Programme, 2000).

Against this background, the purpose of this paper is to undertake an econometric analysis of the constellation of factors that serve to determine some of the outcomes, noted above, with respect to demography and to schooling in India. The econometric estimates are based on *unit record data* from a survey of 33,000 *rural* households - encompassing 195,000 individuals - which were spread over 1,765 villages, in 195 districts, in 16 states of India. This survey - commissioned by the Indian Planning Commission and funded by a consortium of United Nations agencies - was carried out by the National Council of Applied Economic Research (NCAER) over January-June 1994 and most of the data from the survey pertains to the year prior to the survey, that is to 1993-94. Details of the survey - hereafter referred to as the NCAER Survey - are to be found in Shariff (1999), though some of the salient features of data from the NCAER Survey, insofar as they are relevant to this study, are described in this paper.

A number of empirical studies have examined demographic outcomes in India and in other countries, particularly with respect to fertility and infant mortality rates (*inter alia* Caldwell, 1979 and 1986; Subbarao and Rainey, 1992; Murthi *et. al.*, 1995; Borooah, 2000) and they have broadly pointed to the important role that the literacy of mothers plays in reducing the incidence of child mortality. However, a weakness of these studies is that while they purported to examine the behaviour of individuals, they were, in fact, based on data pertaining to geographical units. For example, the studies of Murthi *et. al.*, (1995) and of Borooah (2000) were both based on district-level data. The dangers of inferring individual behaviour from an analysis of aggregate data were recognised, nearly half a century ago, by Theil (1954): "when models of individual behaviour are estimated from variation in average behaviour and average conditioning variables for large aggregates …[then] the properties of the estimates depend upon many tenuous aggregation assumptions". But, given the paucity of large sets of data relating to individuals, researchers sought exculpation in the fact that there was no alternative.

More recently, Parikh and Gupta (2001) have enquired into the effectiveness of female literacy in reducing fertility in the two Indian states of Andhra Pradesh and Uttar Pradesh, using unit record data for 'ever married' women from the National Family Household Survey's 1992-93 data set. In so doing they noted that, to the best of their knowledge, these data had not been used for the multiple regression analysis of the relationship between literacy and fertility.

These observations then point to a general problem that vitiates empirical studies of demographic outcomes India: when they are cast in a multiple regression mould, their results are derived from aggregate data; on the other hand, when they are based on unit record data they do little more than present bi-variate cross-tabulations[1]. This paper, like that of Parikh and Gupta (2001), addresses this general problem by marrying data on individuals to the methods of econometrics. But, in so doing, its scope is much broader than investigating the determinants of fertility. Given the richness of the data set used – described in section 4 – the paper encompasses analysis of both demographic and schooling outcomes.

In contrast to other studies of demographic behaviour, which analyses variations in the total fertility rate, this study begins with an analysis of the *number of pregnancies* to currently married women, up to their date of interview by the NCAER Survey. From this, the analysis moves to the *number of live births* to these women and, from there, to their *number of infant survivals*. These are the three components of demographic outcomes considered in this paper. The schooling equations focus on the *school enrolment* experiences of children between the ages of 6 and 14 and on the *school continuation* experiences of children between the ages of 10 and 14. The demographic and the schooling equations, considered collectively, paint – as the title of the paper suggests - an econometric portrait of women and children in India. The econometric methods are, as it were, a 'camera lucida', filtering the external subject through the prism of a regression equation, to provide an image that – notwithstanding some distortion, which is a necessary concomitant of the device - is reasonably life-like[2].

[1] See, for example, the papers in Jeffery and Basu (1996). See also Bose (2001) on this point.
[2] See Hockney (2001) for a description of the way the use of the 'camera lucida' and the 'camera obscura', made European painting, from the 13th century onwards, more 'life-like'.

2. Econometric Methods

Schultz (2001) has observed that, in the study of demographic outcomes, there is a fundamental methodological divide between those want to employ a 'structural' model to explore the *interactions* between a variety of choices made at the household level – fertility, age at marriage, labour force participation – and those who are content to work with a 'reduced form' model in which say, fertility decisions are analysed *conditional* upon the constraints embodied in the prior outcomes of other family choices and, also, upon the environmentally fixed constraints exogenous to the family.

The former approach has the advantage of appreciating that family choices involve a more complex mechanism than a simple causal effect from one variable to another. However, its disadvantage is that strong identification restrictions are needed to derive the parameters of such a model, a process which imbues the conclusions drawn from such models with controversy (Manski, 1995). The second method, while glossing over the aspect of inter-choice interaction and, thereby, not clarifying the *process* through which fertility changes occur, has the advantage that it provides an (unbiased) estimate of the effect of the constraint upon fertility outcome; it is precisely information of this sort that is of central interest to the policy maker.

Against the backdrop of this methodological divide, this study falls into the 'reduced form' camp: it examines outcomes relating to demography and schooling *at a point in time*, conditional upon outcomes arrived at earlier in the life-history of the respondents and conditional upon the 'environment' – both in terms of the family/household and in terms of the village – within which the respondent was placed. With this clarificatory preamble, the econometric model used in this paper consisted of *five* equations whose dependent variables were:

1. The number of pregnancies to a woman (up to the interview date[3]): PRG=0,1, 2,...

2. The number of live births to a woman: LVB=0, 1, 2...

3. The number of infant survivals[4] to a woman: INS=0,1, 2...

4. Whether a child, between the ages of 6-14 inclusive, was (ENR=1) or was not (ENR=0) enrolled at school

5. Whether, after enrolment, a child, between the ages of 9-14 inclusive, was (CON=1) or was not (CON=0) continuing at school

[3] The qualification that the numbers relate to the period up to the date of interview applies to all the variables below.
[4] Defined as live births less infant deaths

The first three dependent variables related to the experience of women with respect to their pregnancies, their live births and the mortality of their infants. For each woman (indexed i), the number of pregnancies was specified as a linear function of a vector $\mathbf{x_i}=\{x_{ij}\}$ of j=1…J determining variables:

$$PRG_i = \alpha_0 + \sum_{j=1}^{J} \alpha_j x_{ij} \qquad (1)$$

The number of live births was specified as a linear function of the number of pregnancies. However, it was also hypothesised that the translation of pregnancies into live births – as measured by the coefficient on the number of pregnancies - would *ceteris paribus* depend upon the spacing of pregnancies, as measured by the average number of years between pregnancies: the longer the spacing the greater the proportion of pregnancies that would result in live births.

This idea was encapsulated in the equations:

$$\begin{aligned} &LVB_i = \beta_0 + \beta_1 PRG_i \text{ where: } \beta_1 = \beta_2 + \beta_3 SPC_i,\ \beta_1, \beta_3 > 0 \\ &\Rightarrow LVB_i = \beta_0 + \beta_2 PRG_i + \beta_3 (PRG_i * SPC_i) \end{aligned} \qquad (2)$$

where: $SPC_i= YRM_i/PRG_I$ is the spacing of pregnancies with YRM_i being the number of ('cohabitation') years that a woman had been married.

The number of infant survivals to a woman was hypothesised to depend on her number of live births as well as upon a vector $\mathbf{z_i}=\{z_{ik}\}$ k=1…K of determining variables pertaining to her. The equation was estimated with separate coefficients for the number of female (FLB_i) and male (MLB_i) births to allow for the possibility that birth gender might influence infant survival:

$$INS_i = \gamma_0 + \gamma_1 FLB_i + \gamma_2 MLB_i + \sum_{k=3}^{K+3} \gamma_k z_{ik} \qquad (3)$$

Equations (1), (2) and (3) – for respectively, the number of pregnancies, live births and infant survivals - were estimated as a system of *'seemingly unrelated regression equations'* (SURE) in order to allow for any correlation between the error terms of the three equations.

The alternative to SURE estimation would have been to estimate equations (1)-(3) individually but using methods (like the Poisson regression model) which explicitly took account of the 'count' nature of the data (Greene, 2000, p. 880). This, however, posed two problems: first, the cross-equation error correlation, which *prima facie* would appear to be important, would be ignored; second, the Poisson distribution embodied the unattractive restriction of setting the mean of the distribution equal to the variance. On balance, therefore, the estimation method adopted was that of SURE.

The schooling equations related to the experience of children between the ages of 6 and 14 (inclusive). The likelihood of children being enrolled in school, and of continuing in school after enrolment, was estimated using logit methods whereby:

$$\log\left(\frac{\Pr(ENR_i=1)}{1-\Pr(ENR_i=1)}\right)=\rho_0+\sum_{l=1}^{L}\rho_l v_{il}=R_i \qquad (4)$$

and

$$\log\left(\frac{\Pr(CON_i=1)}{1-\Pr(CON_i=1)}\right)=\sigma_0+\sum_{m=1}^{M}\sigma_m w_{im}=S_i \qquad (5)$$

where: v_{il} (l=1..L) and w_{im} (m=1…M) are values, for child i, of the determining variables of, respectively, school enrolment and school continuation.

The probabilities of being enrolled at school and of continuing in school, after enrolment, are respectively:

$$\Pr(ENR_i=1)=\frac{\exp(R_i)}{1+\exp(R_i)} \text{ and } \Pr(CON_i=1)=\frac{\exp(S_i)}{1+\exp(S_i)} \qquad (6)$$

where R_i and S_i are as defined in equations (4) and (5), above.

The quantities $\frac{\Pr(ENR_i=1)}{1-\Pr(ENR_i=1)}=\exp(R_i)$ and $\frac{\Pr(CON_i=1)}{1-\Pr(CON_i=1)}=\exp(S_i)$ are the 'odds-ratio' of, respectively, being enrolled in school and of continuing in school, after enrolment. The change in the odds-ratio, for the i^{th} child, of being enrolled at school, in the face of a unit change in v_{il}, the value of the i^{th} determining variable, is $\exp(\rho_l)$, l=1…L; and the change in the odds-ratio, for the ith child, of continuing in school, in the face of a unit change in w_{im}, the value of the m^{th} determining variable, is $\exp(\sigma_m)$, m=1…M.

The explanatory power of the logit equations are shown in terms of the 'Pseudo-R^2'. The 'Pseudo-R^2' is a popular measure of the model's performance in binary models and compares the maximised log-likelihood value of the full model (log L) to that obtained when all the coefficients, expect the intercept term, are set to zero (log L_0) and is defined as: 1-(log L/log L_0). The measure has an intuitive appeal in that it is bounded by 0 (all the slope coefficients are zero) and 1 (perfect fit). Unfortunately, there is no natural interpretation to the numbers between 0 and 1 (Greene, 2000).

Another way of assessing the predictive ability of a model with a binary dependent variable is by constructing a 2x2 table of the 'hits' and 'misses' emanating from a prediction rule such that a child is classified as being enrolled (ENR_i=1) or as continuing (CON_i=1) if the *estimated*[5] probability of the child being enrolled at, or continuing in, school > p*. Given a cut-off point, p*, the

[5] Estimated using equation (6) from the estimates of ρ_l and σ_m

'sensitivity' and the 'specificity' of an equation are, respectively, the proportions of positive and negative cases that are correctly classified.

One can, further, plot the graph of sensitivity versus 1-specificity as the cut-off point p* is varied. The curve starts at (0,0) corresponding to p*=1: no positive case is correctly classified (sensitivity=0) and every case is classified negative (specificity =1 or 1-specificity=0); it ends at (1,1) corresponding to p*=0: every positive case is correctly classified (sensitivity=1) and no case is classified as negative (specificity =0 or 1-specificity=1). A model with no predictive power would be the 45^0 line connecting the two extreme points (0,0) and (1,1). The more bowed the curve, the greater the predictive power. Hence the area under the curve – known as the 'receiver operating characteristic' (ROC) curve - is a measure of the model's predictive power: a model with no predictive power has an area of 0.5, while perfect predictive power implies an area of 1 (StataCorp, 2001).

3. Economic Issues

There are several routes to arriving at the specification of the vector of determining variables in the demographic (equations (1), (2) and (3), above) and schooling outcome equations (equations (4) and (5), above) but a convenient path is offered by Becker's (1991) observation that the *quantity* and *quality* of children are substitutes. If the utility to parents of having children depends both upon their number and upon the expenditure on each child then the marginal rate of substitution between quantity and quality is the number of children parents are prepared to give up in order to gain an additional unit of quality, utility remaining unchanged.

The structure of preferences with respect to children may change with economic and social development: literate parents may be more aware of the importance of the quality of children, and thus have a higher marginal rate of substitution, than illiterate parents[6]. This, in turn, would lead them to have fewer children and to invest more in their children's future. Such investment could take the form of: better diet; preventing illness through vaccination and immunisation; seeking medical help promptly in the event of illness. These investments in children's health would raise infant and child survivals and would, therefore, go some way towards blunting the effects of a reduced number of births[7]. Investment could also go towards building the human capital of children by enrolling them at school and ensuring that, after enrolment, they continued to remain in school.

[6] For example, literate (and, more pertinently, educated) parents may be more knowledgeable than illiterate parents of the high returns associated with schooling.
[7] Conversely, bringing up a child in a dirty or unhygienic environment - for example, mothers who are laborers and who take their infants with them to their place of work – would reduce the number of infant survivals.

Preferences may also be determined by cultural forces which shape attitudes towards family size. Religion may play a big role in determining family size. For example, the use of contraception methods, including the preference for certain types of contraceptive methods over others, may be heavily influenced by religious beliefs. McQuillan (1999), in a study of the Alsace region between 1750 and 1870, demonstrated the importance of religious identity on marriage, child-bearing and mortality.

In the context of fertility, another dimension of cultural mores, explored in this paper, is the 'preference for sons' that many families in India (and, indeed, in East Asia) display. Thus, while the increased chances of births surviving past infancy (infant survivals) serve to blunt the appetite for births, it is really the presence of male infants and male children that sets a cap on a family's demand for children.

Cultural preferences towards family size carry implications for the size of the investments undertaken in children. If the quantity and quality of children are, as assumed by Becker (1991), substitutes then one would expect that communities characterised by large families would have a lower proportion of their (eligible) children in school than communities in which family sizes were smaller. These cultural effects would be compounded if groups with a preference for large families had ancillary disadvantages such as relatively low literacy rates and incomes. In this context, Patrinos and Pscharopoulous (1997), in the context of Peru, analysed the effect of being indigenous and of the number of siblings on child schooling progess.

Impinging upon these preferences are a set of constraints. One set of constraints concerns the 'price' of investment in quality. If villages have poor infrastructure in the form of poor quality drinking water and an absence of schools and medical facilities then the price of investment will be high. Unsafe drinking water or the absence of preventive and curative medical facilities will reduce infant and child survivals notwithstanding the best efforts of parents. If children have to travel long distances to school then the journey time – particularly when it is lengthened by an absence of good transport facilities – could add appreciably to the costs of schooling. On the other hand, villages which have 'mother and child' centres - providing pre-school education for children and raising awareness among mothers of infants and toddlers of the importance of investing in the health and education of their children - should harvest the benefit of such centres (for example, *anganwadis* in India, discussed below) in the form of higher school enrolment.

Another set of constraints relates to the opportunity cost of children. If, say, because of the poverty of their families, children are viewed as an economic resource, supplementing the income of the family, then the opportunity cost of schooling investment will be high. For example, a critical assumption underpinning the Basu and Van (1998) model of child labour is that 'a family will

send the child to the labour market if, and only if, the family's income from non-child-labour sources drops very low'. This assumption, which they term the 'luxury assumption', is supported by a number of pieces of empirical evidence – cited in their paper – but, for the purposes of this study, the most pertinent fact is that the children of the non-poor rarely work, even in very poor countries. Conversely, Jensen and Nielsen (1997), in the context of Zambia, find support for the hypothesis that poverty forces households to keep their children away from school.
From the perspective of the econometric model of this paper, the opportunity cost of children's investments has a number of implications. First, one would expect a positive relationship between household income and the likelihood of children from a household to be enrolled at school and, after enrolment, to continue in school. Second, one would expect that the larger the number siblings to a child, the lower the likelihood of that child being enrolled at, or continuing in, school; a large number of siblings suggests that parents have made the 'quantity-quality decision' in favour of quantity. Third, 'education outcomes' for girls - by virtue of the fact that their parents would reap lower returns on their education than on the education of their brothers – would not be as good as that for boys[8]. Fourth, in the cultural setting of rural India, where, broadly speaking, women are in paid work only if the needs of the family so demand, children whose mothers worked would *ceteris paribus* have a lower likelihood of being enrolled at school than children whose mothers were 'unoccupied'. Fifth, given that the degree of economic prosperity varied across the regions of India, it might be expected that children would be more likely to be seen as economic resources - serving to boost current family income - in the poorer, as compared to the richer, regions of India; on this expectation, the poorer regions would have a larger number of pregnancies per woman and also a lower likelihood of children being in school.

4. The Data

The data used for estimating the five econometric equations, whose dependent variables were described above, was obtained from the NCAER survey, referred to earlier. The salient features of this data are set out in this section. The data from the NCAER survey are organised as a number of 'reference' files, with each file focusing on specific subgroups of individuals. However, the fact that in every file an individual is identified by a household number and, then, by an identity number within the household, means that the 'reference' files can be joined – as will be described below – to form larger files.

So, for example, the schooling equations were estimated on data from the 'individual' file. This file, as the name suggests, gave information on the 194,473 individuals in the sample with particular reference to their educational

[8] Girls, after marriage, leave home and, in a traditional Indian context, are 'lost' to their parents. Needless to say, culture would also play a role – perhaps a bigger role than economic calculation – in the educational deprivation of women.

attainments[9]. From this file, data on the school enrolments and continuations of each child aged 6-14 were extracted (the variables ENR and CON) and associated with this information was data on: the educational attainments and occupation of the child's father and/or mother; the income and size of the household to which the child belonged; the state, district and village in which it lived; its caste/tribe (scheduled or non-scheduled only); its religion; the number of its siblings etc.

Another file – the 'village file' – contained data relating to the existence of infrastructure in, and around, each of the 1,765 villages over which the survey was conducted. This file gave information as to whether *inter alia* a village:

(i) was supplied with safe drinking water[10] and how far its households had to travel to obtain water;
(ii) had *anganwadis*[11], primary schools, middle schools and high schools and, if it did not, what was the nature of access to such institutions
(iii) had a pharmacy; dispensary; hospital (or a sub-centre of a hospital) and, if it did not, what was the nature of access to such institutions
(iv) the quality of the approach roads to the village and its access to public transport
(v) the presence of utilities like electricity, telephones, radio and television

The village file could be joined to the individual file so that for each individual (say, child between 6-14) there was information not just on the child's schooling outcome and its family and household circumstances but also on the quality of the educational facilities – and general infrastructure - in the village in which the child lived.

The pregnancy / live births /infant survival equations were estimated on data from the 'eligible women' file. This file contained data on women who were in the reproductive group in the year prior to the commencement of the survey. Some of these women had never been married; others were currently married; and yet others had once been married but were now widowed, divorced or separated. In particular, this file contained details of the experiences of women with respect to: marriage, pregnancies, abortions, miscarriages, still births, live births and infant deaths. In addition to such reproductive data, this file also contained information about the occupation and the educational attainments of the women (and, where relevant, of their husbands) and about the households in which they lived (area of residence, income, size, caste, religion). The 'eligible women' file

[9] Needless to say, the file also contained other information on the individuals.
[10] Protected wells; tanker truck; hand pump; piped water.
[11] *Anganwadis* are village-based early childhood development centres. They were devised in the early 1970s as a baseline village health centre, their role being to: provide state government-funded food supplements to pregnant women and children under five; to work as an immunization outreach agent; to provide information about nutrition and balanced feeding, and to provide vitamin supplements; to run adolescents girls' and women's groups; and to monitor the growth, and promote the educational development of, children in a village.

too could be joined by the 'village file' to provide additional information about the village-level facilities (described above) that were available to the women and their families.

The caste and religion variables in the Survey deserve comment. The respondents to the Survey were distinguished along caste lines as: Schedule Caste; Schedule Tribe; 'other'. They were *separately* distinguished by religion as: Hindu; Muslim; Christian; 'other'. Consequently, membership of the two categories, caste and religion, could overlap: persons of the Schedule Castes/Tribes could be Hindu, Muslim or Christian and, say, Hindus could either be, or not be, from the Schedule Caste/Tribes. In this study, the two categories of caste and religion were rendered mutually exclusive by defining Hindus, Muslims, Christians (and persons of 'other' religions) as those persons professing the relevant faith *and not belonging to Schedule Castes or Tribes*. No distinction was made by religion within the Schedule Caste/Tribe category: hereafter, 'SCT' is used as the collective abbreviation for persons belonging to the 'Schedule Castes' or 'Schedule Tribes'. Because of the small number of Christians and persons of 'other' religions[12] in the Survey, the analysis reported in this paper was confined to Hindus, Muslims and SCTs.

The Survey contained information for each of sixteen states. In this study, the states were aggregated to form five regions: the *Central* region consisting of Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh; the *South* consisting of Andhra Pradesh, Karnataka, Kerala and Tamilnadu; the *West* consisting of Maharashtra and Gujarat; the *East* consisting of Assam, Bengal and Orissa; and the *North* consisting of Haryana, Himachal Pradesh and Punjab.

4.1 Pregnancies, Births and Infant Survivals

The equations relating to the number of pregnancies, live births and infant survivals – respectively, equations (1), (2) and (3) above - were estimated on data for *currently married women*. After dropping cases which:

(a) offered evidence of possible 'data irregularities' from women who reported: more births than pregnancies [9 cases], more deaths than births [0 cases]; a negative value for the number of years they had been married [90 cases]; more than one pregnancy in less than one year of marriage [30 cases]
(b) were 'extreme' cases: women who reported more than 15 pregnancies [238 cases];
(c) 'unlikely' cases for reproduction: women who married after the age of 45 [63 cases]

there remained a sample of 29,088 currently married women. Table 1 shows some of the salient features of the data relating to these women. The average

[12] That is, non schedule caste/tribe persons.

number of pregnancies per woman (after, an average of, 14 years of marriage) was 3.48 and this was converted into 3.20 live births (a conversion factor of 92%) which were further converted into 2.93 infant survivals (a survival factor of 84% over pregnancies). There was some inter-regional variation in these numbers: the lowest number of average pregnancies per women were recorded in the South and the highest in the Central region. However, the largest variation in pregnancy was not by region but by religion: the average number of pregnancies for Muslim women was 4.0 while for Hindu women it was only 3.34.

The average all-India household income of the currently married women was Rupees (Rs.) 30,868 and this showed considerable variation by region: the average household income in the West (Rs. 39,890) was over one-third greater than that in the Central region (Rs. 29,161). The average literacy rate of 31% for the currently married women in the Survey also masked considerable inter-regional variation, from the abysmal 16% for the Central region to the more respectable 40%+ for the South, the West and the East.

4.2 School Enrolment and Continuation

The equation relating to school enrolment (equation (4), above) was estimated on data from the NCAER Survey's 'Individual' file', described above, for children between the ages of 6-14 (inclusive): this yielded a total of 40,697 observations. The equation relating to school continuation (equation (5) above) was estimated on data from the same file but for children between the ages of 10-14 (inclusive) *who had been enrolled at school*. This yielded a total of 16,565 observations. The salient features of the data relating to school enrolments and school continuations are shown in Tables 2 and 3, respectively.

Table 2 shows that, in terms of educational infrastructure, only 11% of the children in the sample lived in villages which *did not* have a primary school, though 50% lived in villages without *anganwadis* and 30% lived in villages without a middle school within a distance of 2 kilometres. This table also shows that, of children aged 6-14 (inclusive) in the overall sample, 71% were enrolled in school, the percentage for boys (77%) being considerably higher than for girls (64%). However, underlying the aggregate figures, there was considerable variation in enrolment rates by: region; community; and the literacy status of the children's parents.

Of the regions, the Central region had the lowest rate of enrolment (60%) and nearly half the girls in this region were not enrolled at school. By contrast, enrolment rates in the South and in the West were in excess of 80% with well over three-fourths of girls being enrolled.

Of the communities, the enrolment rate of Hindu children (77%) was considerably higher than that for Muslim (63%) and SCT children (62%) and the difference in enrolment rates was particularly marked for girls: 73% of Hindu girls, as compared to only 58% of Muslim girls and 54% of SCT girls were enrolled in schools.

The enrolment rates of children also differed widely according to the literacy status of their parents. When both parents were literate, the enrolment of children was 94% with little difference between rates for boys and girls. In contrast, enrolment rates were abysmally low when both parents were illiterate: just over half of children - and only 43% of girls - with illiterate parents were enrolled at school. An interesting aspect of the data is that when only one parent was literate, it was the literacy of the mother, rather than that of the father, which provided the greater boost to literacy rates and – perhaps even more importantly – reduced gender differences in enrolment rates.

Once a child had been enrolled at school, the likelihood of it continuing in school (as Table 3 shows) was high. Nine out of ten children, between the ages of 10-14 (inclusive), were still at school. Although there continued to be a gender disparity in continuation rates, the magnitude of this disparity was not as great as it was for enrolment rates. Moreover, continuation rates were high - for boys and for girls - for children with illiterate parents. This would suggest that the major obstacle to the educational prospects of children in India was one of being enrolled at school, rather than one of being allowed to continue in school after enrolment.

5. Econometric Specification and Results

In the light of the discussion in section 3, the determining variables used to specify the equations for the five dependent variables were grouped as:

1. *Caste and religion* variables (discussed in section 4).

2. *Regional variables* (discussed in section 4).

3. The *educational attainments* of the women and of the men. These were classed as:

(i) illiterate
(ii) low, if the person was literate but had not completed primary school
(iii) medium, if the person was educated to primary level or above, but below that of matric
(iv) high, if the person was educated to matric level or above

In the schooling equations, the 'women' were the mothers, and the 'men' were the fathers, of the children whose schooling experience was being studied[13]. In

[13] In the schooling equations, the educational attainments of the father and mother were defined only if the relevant parent was in the household. If the relevant parent was not in the household (say, due to bereavement, divorce or separation), then it was undefined. An identical remark applies to the occupation of the father and mother in these equations.

the pregnancy/live births/infant deaths equations, the ‘women’ were currently married women and the ‘men’ were the husbands.

4. The *occupations* of the men and of the women. For men, the mutually exclusive and collectively exhaustive categories were:
(i) cultivator: if the man was (primarily) engaged in cultivation or allied agricultural activities
(ii) labourer: if the man was (primarily) a (agricultural or non-agricultural) labourer, cattle tender or domestic servant
(iii) non-manual worker

For women, in addition to (i)-(iii) above there was a fourth (residual) category: ‘unoccupied’.

5. *Personal* and *household* variables. These were, for the pregnancy/live births/infant deaths equations: years of marriage after child-bearing age[14]; household income; gender of the live birth. For the schooling equations these were: household income; household size; number of siblings[15]; years of schooling[16] (school continuations equation only); gender.

6. *Village level* variables (discussed above)

In addition to these variables, it was important to incorporate into the pregnancy equations variables that would capture the ‘preference for sons’ that is commonly acknowledged as being endemic among Indian households (Murthi *et. al*., 1995). Two variables were used to proxy this preference.

First, for each woman, the proportions of her male and female infant survivals to the number of her pregnancies was constructed. These proportions were included in the pregnancy equation with, say, coefficients α and β, respectively. The expectation was that $\alpha, \beta<0$, so that the greater the proportion of pregnancies that were converted into survivals, the smaller would be the number of pregnancies. However, by ‘son preference’, the expectation was also that $|\alpha| >|\beta|$, so that a given proportion of male survivals would imply a smaller number of pregnancies than the *same* proportion of female survivals.

The second proxy for ‘son preference’ was provided by the number of abortions that a woman had had. Sen (2001) has observed that, with modern techniques to determine the sex of a foetus, sex-selective abortions allow parents to realise their desire for sons and that such abortions, though mainly popular in East Asia, are beginning to emerge as a statistically significant phenomenon in India. The expectation was that the coefficient on the number of abortions would be positive

[14]Age of woman-age at marriage (AGM) if AGM>15; otherwise defined as age of woman-16.
[15] The number of persons with the same mother; if the mother was not present in the household, then the number of persons with the same father.
[16] Age of child-age at enrolment.

implying that the greater the number of abortions to a woman (in pursuit of a male child) the larger the number of pregnancies she would have.

The results from estimating the pregnancy/live births/infant mortality equations (equations (1), (2) and (3), respectively), using SURE, are shown in Table 4A and 4B and the logit results from the schooling equations (equations (4) and (5)) are shown in Table 5A, 5B and 5C. Before discussing the estimates, it is worth making three general points.

The first point is that variables whose associated coefficients were 'insignificant' were dropped from the equation, where 'insignificance' was defined as a z-score less than 1: the exclusion of such variables, as is well known, adds to the explanatory power - defined in terms of the adjusted R^2 - of regression equations.

The second point is that the logit coefficient estimates in Table 5A are shown in terms of 'odds-ratios', discussed earlier[17]. They report by how much the relevant odds-ratio would change, given a unit change in the associated determining variable. A positive/negative sign (before a coefficient estimate) in Table 5A implies that the relevant odds-ratio would increase/decrease - or equivalently that the probability of the event occurring would rise/fall - for a *ceteris paribus* unit increase in the associated variable.

The third point is that the number of cases over which the equations were estimated as shown in Table 5A (37,566 cases for the enrolment equation and 15,179 cases for the continuation equation) are fewer than the number of cases reported in Tables 2 (40,697 children between 6-14) and 3 (16,565 children between 10-14, who had been enrolled in school). Because the educational attainments (and the occupations) of fathers and of mothers entered as determining variables into the enrolment and the continuation equations, the two schooling equations were estimated over the sub-sample of children who had *both* parents living in the household[18]: these are the 37,566 and 15,179 children of Table 5A.

Against the background of the above three general points, the 'pregnancy' equation explained 43% of the variation in the numbers of pregnancies - while the 'live births' and the 'infant survival' equations explained, respectively, 90% and 87% of the variation in the numbers of live births and infant survivals - across the 29,088 currently married women in the estimation sample (Table 4B). The high explanatory success of the latter two equations was, undoubtedly, due

[17] That is, the coefficient estimates of Table 5A refer to $\exp(\rho_l)$ and $\exp(\sigma_m)$ *not* to the ρ_l and σ_m of equations (4) and (5).

[18] As observed in an earlier footnote, the value of the educational variable for a 'missing' (from the household) parent would be recorded as a 'missing value'. Consequently, a case involving a child with a 'missing' parent would be dropped when the estimated equation required the educational variables of both parents to be included.

to the fact that were anchored by, respectively, the number of pregnancies and the number of live births. The use of SURE estimation methods was justified by the value of the Breusch-Pagan test which rejected the null hypothesis that the cross-equation errors were independently distributed.

As observed earlier, the 'pseudo-R^2' values for binary models often need to be supplemented by other indicators of goodness of fit. Table 5C shows the 2x2 table of 'hits' and 'misses' when p*=0.5. This shows that the probabilities predicted from the school enrolment equation correctly classified 76% of the 37,566 children studied and that the probabilities predicted from the school continuation equation correctly classified 90.5% of the 15,179 children studied. When the cut-off probability was varied from 1 to 0, the area under the ROC curve (discussed earlier) was 79% for the enrolment equation and 76% for the continuation equation. On all indications, therefore, the 'fit' of the logit equations for school enrolment and school continuation was satisfactory.

The discussion of the coefficient estimates associated with the determining variables for the five econometric equations (Tables 4A and 5A) are cast in terms of the effects of changes in the values of the determining variables on the average (mean) number of pregnancies, live births and infant survivals and on the average probability of being enrolled at school and, after enrolment, on the average probability of continuing in school. The determining variables used in the equations can be grouped as those which take a continuum of values and those which take discrete (usually binary) values. Household income, household size, the number of siblings, years of marriage, the spacing of pregnancies are examples of 'continuous' variables and region, religion/caste, educational attainments, occupation are examples of binary variables.

The effects of changes in the values of the 'continuous' variables, on the values of the dependent variables, were traced by constructing *ceteris paribus* either a percentage change (income increases by x%) or an absolute change (years of marriage increase by y) – as thought appropriate – in the value of the determining variable. The effect of changes in the 'binary' variables were traced by comparing the outcomes that resulted when the binary variable took one value with the outcomes associated with it taking the other value, the values of the other variables remaining unchanged between the two comparisons[19].

5.1 Household-level variables: household income and size, the number of siblings, years of marriage

Table 4A shows that an increase in the household income of currently married women would reduce their average number of pregnancies, and increase their average number of infant survivals. In addition, as Table 5A shows, an

[19] See Borooah (2001) for a detailed analysis of computing marginal effects in discrete choice models.

increase in household income would raise the probabilities of both being enrolled at, and of continuing in, school.

On the other hand, household size was positively associated with a higher number of pregnancies (Table 4A) and also with a higher probability of continuing in school[20] (Table 5A). The greater the number of siblings that a child had, the lower would be its likelihood of being enrolled at school and, if enrolled, of continuing in school (Table 5A).

The numbers of years that a woman had been married would have a (strong) positive effect on the number of her pregnancies and (though weaker) on the number of her infant survivals. By affecting the spacing of her pregnancies, it would also affect the translation of pregnancies into live births.

Table 6 shows the consequences of the following changes (all of which were made under *ceteris paribus* assumptions):

(i) All incomes were raised by 20%
(ii) All household sizes were increased by 1 person
(iii) The number of siblings of each child was increased by 1
(iv) The number of years that a woman had been married was increased by 1 year

Although, as Table 4A and 5A show, a household's income significantly affected a number of its demographic and schooling outcomes, the income effects were generally very weak: even a 20% increase in all incomes barely shifted the predicted values from their sample averages. Increasing household sizes by 1 person increased the number of pregnancies from the sample average of 3.48 to a predicted 3.58 and it increased the probability of continuing at school from 0.90 to 0.91. While, therefore, this study – like that of Jensen and Nielsen (1997) for Zambia - finds some support for the hypothesis that low household income kept children away from school, the non-economic environment of the household was much more important in determining schooling outcomes[21].

On the other hand, increasing the number of siblings[22] by 1 lowered both the likelihood of being enrolled at school (from the sample average of 0.71 to the predicted 0.69) and of continuing at school (from the sample average of 0.91 to the predicted 0.89). In this, the results agree with those of Patrinos and Pscharopoulos (1997) who showed, using unit record data from the 1991 Peru Living Standards Survey, that family size was important in determining child

[20] Though not with a higher probability of being enrolled at school.
[21] Needless to say, as Strauss and Thomas (1995) point out, there may a strong interaction between income and other non-economic characteristics.
[22] The average number of siblings for children aged 5-14 was 2.6.

schooling progress[23]. Raising the number of years married by 1 raised the number of pregnancies from 3.48 to 3.63 and the number of infant survivals from 2.93 to 2.95[24].

5.2 Caste and Religion

Caste and religion exerted a strong influence on the pregnancy equation and on the school enrolment and continuation equations[25]. In the pregnancy equation, the number of pregnancies were significantly higher for Muslim and SCT women relative to Hindu women (Table 4A); in the schooling equations, the likelihood of being enrolled at school and, after enrolment, of continuing in school was significantly lower for Muslim and SCT children than for Hindu children. The number of live births and infant survivals to a woman was not affected by her religion or caste[26].

Amin et. al. (1997), have discussed the relationship between contraception use and religiosity in contemporary Bangladesh and Moulasha and Rao (1999) have argued, on the basis of evidence from National Family Health Survey data for India, that although contraceptive use was as prevalent among Muslims as among Hindus, their preferences, between the different contraception methods, were different: Hindus favored 'permanent' measures, like sterilisation, while Muslims had a preference for less permanent measures. Lastly, on the theme of religion and fertility, Iyer (2001) contains a comprehensive discussion of those aspects of Muslim and Hindu religious beliefs which have a bearing upon the fertility rates of adherents to the two faiths.

Three scenarios were constructed in order to quantify the effects of religion/caste on the number of pregnancies, enrolments and continuations. In the first, 'all-Hindu', scenario *all* the 29,088 women in the pregnancy equation, *all* the 37,566 children in the enrolment equation and *all* the 15,179 children in the continuation equation, were assumed to be Hindu. In the second, 'all-Muslim', scenario they were all assumed to be Muslim and in the third, 'all-SCT' scenario, they were all assumed to be SCT. If prg_i^R, p_i^R and q_i^R (R = H [Hindu]; M [Muslim); S [SCT]) represent, respectively, the (estimated) number of pregnancies, the (estimated) probability of being enrolled at school and the (estimated) probability of continuing in school under each of the three scenarios then, for any woman i (i=1…29,088) the difference between say prg_i^H and prg_i^M is *entirely* due to the effect of religion since nothing was changed between the 'all-Hindu' and the 'all-

[23] They also showed that, in addition to the number of siblings, the activities of siblings were also important in determining a child's educational outcome.
[24] Note these are 'static' calculations. In a 'dynamic' context, the increase in pregnancies consequent upon an increase in the number of years married, would affect the number of live births and, thereby, the number of infant survivals.
[25] Note that in section 3 it was pointed out, in the context of the 'quantity-quality trade-off', that poor schooling outcomes could be a corollary of high fertility rates
[26] Except, insofar, as the spacing of pregnancies was shorter for Muslim and SCT women, compared to Hindus: see the live births equation (2) and its estimates in Table 4A.

Muslim' scenarios *except* the religion of the respondents. The computation of prg_i^M is the consequence of the 'Muslim coefficient' in Table 4A being 'switched on' and the computation of prg_i^H is the consequence of the Muslim and the SCT coefficients in Table 4A being 'switched off'[27].

The average (mean) values of prg_i^R, p_i^R and q_i^R (R=H,M,S) under the three scenarios were computed and are shown in Table 7. These show that the 'effect of religion/caste' is to yield a lower average number of pregnancies for Hindu, compared to Muslim or SCT, women[28] and a higher average probability of being enrolled at school, and of continuing in school, for Hindu, compared to Muslim or SCT children. The term 'effect of religion/caste' is used in the sense that while Hindu, Muslim and SCT women and children might differ in respect of several attributes (household income, region of residence etc.) these differences are abstracted from in arriving at the calculations reported in Table 7.

5.3 Educational Attainment

The educational attainment of the men and women in this study affected outcomes with respect to several of the dependent variables. The literacy of women affected the number of pregnancies, with literate women having fewer pregnancies *ceteris paribus* than illiterate women[29]. The literacy of both men (as husbands) and of women (as mothers of the infants) affected the number of infant survivals: the number of survivals were highest when both husband and wife were literate and lowest when both husband and wife were illiterate.

Since Caldwell's (1979) pioneering work in Nigeria, the positive relation between maternal education and infant survivals has been confirmed for many countries (Strauss and Thomas, 1995). The results reported here add another stone to this intellectual cairn but, in so doing, point also to another, more general, result. This is that it was the literacy of *both* parents, rather than just the literacy of the mother, which was important in influencing infant survivals. Caldwell (1979) argued that more educated women were more able to understand evidence from the media as well as from medical and public health personnel. One could equally make this argument for educated men. It would appear that, in the context of infant mortality, there are powerful economies of scale to be gained when both parents are literate: one parent's understanding of 'health information' could corroborate and reinforce that of the other[30].

[27] Note that Hindu is the 'residual' category.
[28] See also Jeffery and Jeffery (1997) and Iyer (2001) for a discussion of the relationship between religion and fertility.
[29] In the case of pregnancies it was only the woman's literacy was significant and pregnancy-live births conversion was unaffected by the educational attainment of either the woman or her husband.
[30] See Basu and Foster (1998) for a discussion of the externalities that literate members of a household confer upon those members who are illiterate.

While educational attainments higher than basic literacy (of the women and their husbands) did not exert a significant effect on either the number of pregnancies or on the number of infant survivals, the post-literacy levels of educational attainment of parents were a significant influence on both the probability of school enrolment and the probability of school continuation: in general, both probabilities rose with the parental level of education.

It is important to re-emphasise the point that the educational status of *both* the man and the woman were important in determining infant survival and schooling outcomes. The correlation between the literacy of currently married women and of their husbands was 0.44; the correlation between fathers and mothers with 'low' educational attainments was 0.16; the correlation between fathers and mothers with 'medium' educational attainments was 0.09; and the correlation between fathers and mothers with 'high' educational attainments was 0.35[31]. The values of these correlation coefficients suggested that multicollinearity was unlikely to be a problem when the educational attainments of men and of women were included as regressors[32].

In order to assess the importance of literacy in determining the number of pregnancies and infant survivals, three scenarios were constructed: in the first ('all-illiterate') scenario all the 29,088 women and their husbands were taken to be illiterate; in the second ('all-women literate') scenario, all the women were taken to be literate (that is, the literacy coefficients for women were 'activated' in the pregnancy and the infant survivals equation) but their husbands were assumed to be illiterate; in the third ('all-person literate') scenario all the women and their husbands were assumed to be literate (that is, now, in addition, the literacy coefficient for men was 'activated' in the infant survivals equation). Since literacy was the *only* factor that was altered between the scenarios, the difference in outcomes between the scenarios, with respect to the number of pregnancies and infant survivals, was *entirely* the result of differences in literacy.

Table 8 shows that the universal literacy of women, relative to their universal illiteracy, would lower the average number of pregnancies (per woman) from 3.53 to 3.36 and would raise the average number of infant survivals (per woman) from 2.91 to 2.93: female literacy would, therefore, raise infant survivals as a proportion of pregnancies from 82% to 87%. When, in addition to the universal literacy of currently married women, it was assumed that all the husbands were also literate the number of infant survivals went up to 2.96 (the number of pregnancies was unchanged) and survivals as a proportion of pregnancies rose to 88%.

[31] 'low' is literate, but educational attainment primary or less; 'medium' is educational attainment more than primary but less than matric; 'high' is educational attainment of matric or higher.
[32] This is the result of using unit-record data. On more aggregated data there is likely to be a high degree of correlation between the educational attainments of men and women. For example, districts which have a relatively high rate of male literacy are also likely to have a relatively high rate of female literacy (Borooah, 2000).

Graff and March (1979) have commented that “the higher the level of aggregation and the further the data are removed from the level of individuals, families or households, the higher the degree of association” between literacy and fertility. The results reported here support Parikh and Gupta’s (2001) findings that the evidence, based on unit record data, is that the relationship between literacy and fertility (or, in this case, the number of pregnancies) is surprisingly weak; consequently, it would be a mistake, as they put it, “to regard it [female literacy] as the magic bullet for population control”.

In order to assess the influence of educational attainment on the likelihood of school enrolment and of continuation after enrolment, four scenarios were constructed: in the first (‘all-illiterate’) scenario, both parents of all the children were assumed to be illiterate; in the second (‘all-low’) scenario, both parents of all the children were assumed to be literate but with a ‘low’ level of educational attainment; in the third (‘all-medium’) scenario, both parents of all the children were assumed to have a ‘medium’ level of educational attainment; in the fourth (‘all-high’) scenario, both parents of all the children were assumed to have a ‘high’ level of educational attainment.

The probabilities of enrolment and continuation under these scenarios are shown in Table 9. These show that, in terms of determining the educational future of children, the crucial divide was between children with parents who were both illiterate and children with literate parents[33]. When both parents were illiterate, the average chance of a child being enrolled at school was only 57% and, if enrolled, that of continuing was 87%. When both parents were literate, albeit with a ‘low’ level of educational attainment, the two probabilities rose, respectively, to 90% and 92%. Thereafter, further improvements in the educational attainment of parents raised the probabilities of children being enrolled at school and, after enrolment, of continuing in school but these increases were much less impressive than those effected by the transition from parental illiteracy to parental literacy.

5.4 Regional Variables

Table 4A shows that even after controlling for *inter alia* household level variables (income, household size, years of marriage); caste and religion; educational attainment; and infrastructure availability, the region in which a woman lived had a significant effect on the number of her pregnancies; similarly, Table 5A shows that even after controlling for similar non-regional factors in the schooling

[33] As Dreze and Sen (1996, p.109) have observed: ”literacy is a basic tool of self-defence in a society where social interaction often involves the written media...an illiterate person is that much less equipped to defend herself in court, to obtain a bank loan, to enforce her inheritance rights, to take advantage of new technology, to compete for secure employment, to get on the right bus, to take part in political activity, in short to participate successfully in the modern economy and society".

equations, the region in which a child lived had a significant effect on the likelihood of it being enrolled at school and, after enrolment, of continuing in school. In that sense, one could plausibly refer to a 'regional factor' affecting demographic and educational outcomes. In order to assess the strength of this factor, five scenarios were constructed: in the first, *all* the women and the children, under study, were assumed to live in the Central region; in the second, third, fourth and fifth scenarios they were *all* assumed to live in, respectively, the South, the West, the East and the North[34].

The differences between the scenarios in the number of pregnancies and in the probabilities of school enrolment and school continuation (Table 10) could then be ascribed to 'the regional factor' since the values of all the non-regional variables were the same between the scenarios. Table 10 shows that the number of pregnancies was highest when it was assumed that all the currently married women lived in the North (an average of 3.68 pregnancies per woman) and lowest when they were all assumed to live in the South (3.25).

Interestingly, for the South, the West, the East and the North the sample average of the number of pregnancies (Table 1) was lower than the 'synthetic' average shown in Table 10. Since differences in the sample averages incorporate the effects of inter-regional differences in the values of the non-regional variables, while the synthetic averages abstract from them, the inference is that in these regions the non-regional factors worked towards *reducing* the number of pregnancies. On the other hand, for the Central region, the sample average of 3.78 was higher than the synthetic average of 3.53: in this region the non-regional factors worked towards *raising* the number of pregnancies.

When the effects of non-regional factors were abstracted from, Table 10 shows that, with respect to the number of pregnancies, there were two 'outliers': the South had a low 'synthetic' average number of pregnancies (3.25) while the corresponding number was high in the North (3.68); there was little variation in these synthetic averages between the other regions.

A similar story can be told with respect to school enrolments. The non-regional factors determining the likelihood of school enrolment served to lower this probability in the Central region (the average 'synthetic' probability of 0.64 reported in Table 10 being greater than the 60% of children enrolled at children in the Central region, shown in Table 2). On the other hand, these non-regional factors served to raise the probability of school enrolment in the other regions as evidenced by the fact that their average 'synthetic' probabilities (Table 10) were lower than the corresponding sample proportions of children enrolled at school in these regions (Table 2).

With respect to school continuation, the only significant outlier – judged on a purely regional basis - was the North in which the average 'synthetic' probability

[34] See an earlier section for the definition of the regions in terms of their constituent states.

of continuing in school, after enrolment, was 3 percentage points higher than in the other regions. Since, in all the five regions, the sample proportion of children continuing in school (Table 2) was within 1 percentage point of the corresponding average synthetic probability, the inference was that inter-regional differences in the values of the non-regional variables did not significantly influence differences between them in their respective probabilities of school continuation.

5.5 Occupational Variables

Table 4A shows that while the occupations of currently married women and their husbands did not affect the number of their pregnancies, or the translation of their pregnancies into live births, the fact of the father and the mother working as labourers served *ceteris paribus* to reduce the number of infant survivals.

Furthermore, as Table 5A shows, children with fathers working as cultivators or in non-manual occupations had a higher likelihood of being enrolled at school than children with fathers who were labourers[35]. However, in terms of the likelihood of continuing at school, children with fathers in non-manual occupations had a higher likelihood of continuing than children with cultivator fathers or with fathers who were labourers; however, between children with fathers in the latter two occupations, there was no significant difference in their respective probabilities of continuing in school.

Table 5A also shows that children with mothers who worked as labourers or as non-manuals were less likely to be enrolled at school than children whose mothers worked as cultivators or were 'unoccupied'. In terms of the likelihood of continuing in school, this likelihood was reduced by the very fact of the mother working (relative to the mother being unoccupied), the reduction being greatest for cultivator mothers and smallest for mothers in non-manual occupations.

The effects of the occupation of husbands/fathers on the number of infant survivals to their wives and on the probability of their children being enrolled at school and of continuing in school were traced through three scenarios. In the first of these scenarios, the husbands of *all* the currently married women, and the fathers of *all* the children studied in the schooling equations, were assumed to work as cultivators. In the second scenario, they were all assumed to work as labourers and, in the third scenario, they were all assumed to be in non-manual occupations.

The effect of the occupation of currently married women on the number of their infant survivals, and of mothers on the probability of their children being enrolled at school and of continuing in school, were traced through four scenarios: in the

[35] Note that, for fathers and husbands, the occupational categories 'cultivator', 'labourer and 'non-manual were collectively exhaustive.

first three, *all* the currently married women, and the mothers of *all* the children studied in the schooling equations, were first assumed to work as cultivators; then as labourers; and then in non-manual occupations; in the fourth scenario, they were all assumed to be unoccupied.

The upper panel of Table 11 shows the results from the three scenarios relating to the occupations of the husbands/fathers while the lower panel of Table 11 shows the results from the four scenarios relating to the occupations of the currently married women in the infant survivals equation, and of the mothers of the children in the schooling equations. Since, the values of the other variables were unchanged between these the scenarios, differences between the results for the three scenarios shown in Table 11 (upper panel) are purely the consequence of differences in occupations of husbands and of fathers. Similarly, differences between the results for the four scenarios, shown in Table 11 (lower panel) are purely the consequence of differences in occupations of the currently married women (in the infant survivals equation), and of the mothers of the children (in the schooling equations).

Table 11 shows that *ceteris paribus* the number of infant survivals fell when the husband was a labourer (compared to when the husband was a cultivator or a non-manual worker), from 2.94 to 2.92, but the fall in the number of survivals was particularly pronounced when the woman was a labourer (compared to when she was a cultivator, a non-manual worker or unoccupied), from 2.94 to 2.90. The likelihood of children with non-manual fathers being enrolled in school (75%) was 3 percentage points higher than that for children with cultivator parents and 8 percentage points higher than that for children whose fathers were labourers. Table 11 also shows that the likelihood of continuing at school was higher for children with unoccupied mothers (92%) than for children whose mothers worked (89%, 87% and 84%, respectively, for children with cultivator, labourer and non-manual mothers). This would suggest that a major reason for the currently married women in the sample to work outside the home was economic necessity and that the children of such mothers had a relatively greater chance of being taken out of school in order to contribute to their family's finances.

5.6 The Role of Gender

Gender issues enter the model in two ways. First, in the infant survivals equation, the proportion of live births that survived their first year was significantly higher for female (79%), than for male (76%), live births (Table 4A). This is consistent with the well documented fact that *ceteris paribus* women tend to have lower age-specific mortality rates than men (Sen, 2001). Secondly, *ceteris paribus* the probability of being enrolled at school and, then, of continuing in school was lower for girls than it was for boys (Table 5A).

If the sample of children studied was assumed to consist entirely of boys (that is, the 'female' coefficient in the enrolment and the continuation equations was 'switched off' for all the children) then the likelihood of being enrolled at school would be 77% and the likelihood of continuing in school, after enrolment, would be 93%. If the sample of children studied was assumed to consist entirely of girls (that is, the 'female' coefficient in the enrolment and the continuation equations was 'switched on' for all the children) then the corresponding probabilities would be 64% and 87%. In respect of school enrolment, therefore, girls suffered a 'gender penalty' of 13% and, in terms of school continuation, the gender penalty was 7%[36].

5.7 'Son Preference'

Table 4A suggests that 'son preference' played a significant role in influencing the number of pregnancies to currently married women. Although the number of pregnancies was predicted to fall, as the proportion of male and female infant survivals in the total number of pregnancies rose, the fall was predicted to be significantly greater for a rise the proportion of male survivals, relative to the same rise in the proportion of female survivals.

The average proportion of infant survivals (male and female) to pregnancies for the 29,088 currently married women was 85%. If *all* the live births to *each* of the 29,088 women (and, therefore, all the infant survivals) were assumed to be female (so that the 'male infant survivals' coefficient in Table 4A was 'switched off') then the average number of pregnancies was predicted to be 3.79. On the other hand, if *all* the live births to *each* of the 29,088 women (and, therefore, all the infant survivals) were assumed to be male (so that the 'female infant survivals' coefficient in Table 4A was 'switched off') then the average number of predicted pregnancies was 3.22. 'Son preference', as measured by the different responses to male and female infant survivals, increased, on average, the number of pregnancies per woman by 0.57.

If it is assumed that the abortions recorded in the survey were carried out for sex selection purposes, then, to this increase must be added the further effect of abortions. The coefficient on the 'number of abortions' variable suggests that every additional abortion leads to another 0.83 pregnancies[37]. This difference of 0.83 may be decomposed in terms of the following two scenarios: if none of the women in the sample had had an abortion, then the number of predicted pregnancies would be 3.46; on the other hand, if all the women had had exactly one abortion each, the predicted number of pregnancies would be 4.29.

5.8 Infrastructure

[36] Computed as 0.13=1-(0.64/.77) and 0.07=1-(0.87/0.93).
[37] Abortions, however, have implications for women's health (Mishra, 2001).

The existence, and the quality of, infrastructure available in, and around, the villages in which the women and the children of this study lived affected their outcomes with respect to demography and schooling. First, the absence of a pharmacy within 2 kilometres of a village – implying difficulty of access to both birth-control methods and to medicines - served to both increase the number of pregnancies and to reduce the number of infant survivals[38]; the absence of safe drinking water in the village and the absence of a dispensary[39], with a trained doctor, within 5 kilometres of the village both served to reduce the number of infant survivals[40].

Table 12 shows that, in the universal absence of 'good' access to pharmacies, the average number of pregnancies per currently married woman would be predicted to be 3.51, while the universal presence of 'good' access to pharmacies would reduce this to 3.47. The absence of pharmacies in the neighbourhood of a village, therefore, raised the average number of pregnancies by 0.04. The universal presence of safe drinking water in a village, and of 'good' access to a dispensary by the inhabitants of a village, were predicted to yield, respectively, average numbers of infant survivals of 2.95 and 2.93 per currently married woman; in the universal absence of these facilities, these numbers would be 2.91 and 2.93, respectively. Safe drinking water in a village would, therefore, raise the average number of infant survivals by 0.04; on the other hand, easy access to dispensaries would raise the number of survivals by only 0.01.

The absence of *anganwadis* in a village, and the absence of middle schools within 1 km of a village, reduced the probability of school enrolment. Interestingly, the probability of school enrolment was not affected by the absence/presence of a primary school. This is because only 11% of the children in the study lived in a village *without* a primary school: whatever the factors depressing the enrolment rate of children, it was not – in terms of the data from the NCAER sample – an absence of primary schools.

Equally interestingly, the absence of a primary school in a child's village raised the probability of a child continuing in school after enrolment. Most likely, this is because the probability of continuation is a conditional probability, the condition being that a child was enrolled in school; if parents were sufficiently motivated to enrol their child at school even though the school was in another village, they would be motivated enough to have the child continue in school. Apart from this, there was no evidence of any other component of school infrastructure (access to middle schools or to high schools) affecting school continuation.

---

[38] Though, it should be noted, the latter effect was not significantly different from zero; consequently, its effect is not reported in Table 12.
[39] Though the absence of good hospital access did not significantly affect the number of infant survivals. The correlation between hospital and dispensary access was 0.41.
[40] Though the effect of dispensaries on the number of infant survivals was only significant at the 10% level of significance.

6. Conclusions

The fact that there may be a diversity of factors which impinge upon and influence fertility outcomes has been noted by *inter alia* Davis and Blake (1956), Gillis et. al. (1992) and Basu (1997). This is not so say that, at different times, different persons (or organisations) have not – do not – seize upon a particular factor, or set of factors, and attempt to promote it to the exclusion of all else. A case in point is women's literacy which, today, is regarded, by both academic researchers and institutions like the World Bank, as being the key that will unlock the door leading to fertility reductions.

As this study points out much of this optimism about the effects of literacy on fertility is based on studies based on aggregated data. The contribution of this paper has been to take an encompassing view of demographic and schooling outcomes using unit record data. This view has lent support to Basu's (1997) contention that there cannot be "one grand explanation of fertility behaviour". Undoubtedly, the literacy of mothers is important in reducing the number of their pregnancies but their religion, caste, region, household income and size and degree of 'preference for sons' also matter. Also important are supply-side factors relating to the easy availability of contraceptive advice and methods. In this welter of factors, it is not even clear that female literacy is the most important factor.

The situation becomes more complicated when the discussion embraces a broader set of outcomes relating to infant survivals and children's schooling. Here the paper makes clear that is the literacy of the father is almost as important as that of the mother. But, in this broader context, a wider set of factors come into play. Infrastructure, in the form of safe drinking water and easy access to medical facilities, is important for infant survivals and, in the shape of easy access to schools, is important for school enrolment. Parental occupation matters for both infant survivals and schooling: children born to women who work as labourers are disadvantaged, relative to other children, in terms of their chances both of surviving infancy and, if they do survive, of receiving schooling. The number of siblings that a child has affects his/her schooling outcomes and gender, religion and region play an important role.

However, in one respect, literacy is important. Our study shows that the illiteracy / literacy of parents establishes a clear divide between children whose educational chances are poor and children whose educational chances are good. The further upgrading of the educational levels of parents, beyond literacy, delivers benefits of a much smaller order of magnitude compared to the transition of literacy to literacy. This would suggest that the effects of literacy may be more important in an inter-generational context (literate parents ensure the literacy of their children) than in the context – as in this paper - of a single generation.

References

Amin, S., Diamond, I., and Steele, F. (1997), "Contraception and Religiosity in Bangladesh", in G.W. Jones, R.M. Douglas, J.C. Caldwell and R.M. D'Souza (eds), The Continuing Demographic Transition, Oxford: Clarendon Press.

Basu, A. M. (1997), "The 'Politicization' of Fertility to Achieve Non-Demographic Objectives", *Population Studies*, vol. 51, pp. 5-18.

Basu, K. and Foster, J.E. (1998), "On Measuring Literacy", *Economic Journal*, vol. 108, pp. 1733-49.

Basu, K. and Van, P.H. (1998), "The Economics of Child Labor", *American Economic Review*, vol. 88, pp. 412-27.

Becker, G.S. (1991), *The Economics of the Family*, Cambridge MA: Harvard University Press.

Borooah, V.K. (2000), "The Welfare of Children in Central India: Econometric Analysis and Policy Simulation", *Oxford Development Studies*, vol. 28, pp. 263-87.

Borooah, V.K. (2001), *Logit and Probit: Ordered and Multinomial Models*, Thousand Oaks, CA: Sage Publications Inc.

Bose, A. (2001), "Demographic Data: Overflow and Non-Ultilisation", *Economic and Political Weekly*, vol. 36 (44), pp. 4176-79 (November 3-9).

Caldwell, J.C. (1979), "Education as a Factor in Mortality Decline: an Examination of Nigerian Data", *Population Studies*, vol. 33.

Caldwell, J.C. (1986), "Routes to Low Mortality in Poor Countries", *Population and Development Review*, vol. 12, pp. 171-220.

Davis, K. and Blake, J. (1956), "Social Structure and Fertility: An Analytical Framework", *Economic Development and Cultural Change*, vol. 4, pp. 211-35.

Dreze, J. and Sen, A.K. (1996), *India: Economic Development and Social Opportunity*, New Delhi: Oxford India Paperbacks.

Gillis, J.R., Tilly, L.A., and Levine, D. (1992), "The quiet revolution" in J.R. Gillis, L.A. Tilly and D. Levine (eds), *The European Explanation of Declining Fertility*, Cambridge: Blackwell.

Graff, H.J. and A.C. March (1979), "Literacy, Education and Fertility, past and Present: A Critical Review", *Population and Development Review*, vol. 5, pp. 105-40.

Greene, W.H. (2000), *Econometric Analysis*, Englewood Cliffs: Prentice Hall.

Hockney, D. (2001), *Secret Knowledge: Recovering the lost technique of the Old Masters*, London: Thames and Hudson.

Iyer, S. (2001), *Religion and Demographic Behaviour*, Delhi: Oxford University Press.

Jeffery, R. and Basu, A.M. (eds.) (1996), *Girls' Schooling, Women's Autonomy and Fertility Change in South Asia*, New Delhi: Sage Publications India.

Jeffery, R. and Jeffery, P. (1997), *Population, Gender and Politics*, Cambridge: Cambridge University Press.

Jensen, P. and Nielsen, H.S. (1997), "Child Labour or School Attendance? Evidence from Zambia", *Journal of Population Economics*, vol. 10, pp. 407-24.

Manski, C.F. (1995), *Identification Problems in the Social Sciences*, Cambridge MA: Harvard University Press.

McQuillan, K. (1999), *Culture, Religion, and Demographic Behaviour*, Liverpool: Liverpool University Press.

Mishra, Y. (2001), "Unsafe Abortions and Women's Health", *Economic and Political Weekly*, vol. 36 (40), pp. 3814-17 (October 6-12).

Moulasha, K. and Rao, G.R. (1999), "Religion-Specific Differentials in Fertility and Family Planning", *Economic and Political Weekly*, vol. 34 (42), pp. 3047-51 (October 16-23).

Murthi, M., Guio, A-C, and Dreze, J. (1995), "Mortality, Fertility and Gender Bias in India", *Population and Development Review*, vol. 34, pp. 745-782.

Patrinos , H.A. and Pscharopoulous, G. (1997), "Family Size, Schooling and Child Labor in Peru – An Empirical Analysis", *Journal of Population Economics*, vol. 10, pp. 387-405.

Parikh, K. and Gupta ,C. (2001), "How Effective is Female Literacy in Reducing Fertility?", *Economic and Political Weekly*, vol. XXXVI, pp. 3391-98 (September 1-7).

Schultz, T.P. (1997), “Demand for Children in Low Income Countries”, in M.R. Rosenzweig and O. Stark (eds), Handbook of Population and Family Economics, vol. 1A, chapter 8, Amsterdam: Elsevier Science.

Schultz, T.P. (2001), *The Fertility Transition: Economic Explanations*, Discussion Paper 833, Economic Growth Centre, New Haven CT: Yale University.

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

Sen, A.K. (2001), “The Many Faces of Gender Inequality”, *The New Republic*, vol. 225, pp. 35-40 (September 17).

StataCorp (2001), *Stata Statistical Software: Release 7.0*, College Station, TX: Stata Corporation.

Strauss, J. and Thomas, D. (1995), “Human Resources: Empirical Modelling of Household and Family Decisions”, in J. Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics, volume III*, Amsterdam: Elsevier Science.

Subbarao, K. and Rainey, L. (1992), *Social Gains from Female Education: a Cross-National Study*, Policy Research Working Papers WPS 1045, Population and Human Resources Department, Washington, DC: World Bank.

United Nations Development Programme (2000), *Human Development Report 2000*, New York: Oxford University Press.

Table 1
Selected Data for Currently Married Women

	All-India (29,088)	Central (11,575)	South (6,265)	West (3,895))	East (4,218))	North (3,135)
Pregnancies	3.48	3.78	3.05	3.38	3.35	3.55
Births	3.20	3.51	2.80	3.09	3.08	3.23
Survivals	2.93	3.14	2.63	2.87	2.80	3.01
Age	31	32	31	32	30	31
Age at Marriage	17	16	17	17	17	17
Spacing of Pregnancies	4.6	4.7	4.9	4.6	4.3	4.1
Hindu	15,630	5,869	3,757	2,428	1,707	1,869
Muslim	2,758	1,162	621	161	681	133
SCT	10,700	4,544	1,887	1,306	1,830	1,133
Hhld Income (Rupees)	30,868	29,161	29,945	39,890	22,891	38,543
Hhld Size (Persons)	6.9	7.4	6.3	6.8	6.6	7.3
% literate	31	16	41	40	44	31
% with literate husbands	61	54	61	68	66	68
Husband's occupation						
% Cultivator:	45	49	42	52	42	37
% Labourer:	27	25	35	28	24	26
% living in villages without safe drinking water*	51	72	26	22	69	34
% living in villages without pharmacy within 1 km	27	22	42	14	35	23
% living in villages without hospital within 5 km**	32	44	25	22	20	28
% living in villages without dispensary (with doctor) within 5 km	51	59	41	49	53	40

* from: protected wells; tanker truck; hand pump; piped water.
** hospital sub-centre not in village and hospital > 5 km from village
SCT = Schedule Caste/Tribe
Source : NCAER Survey

Table 2
Selected Data for School Enrolments:
Children Aged 6-14

	All Children (40,697)	Boys (21,455)	Girls (19,242)
% enrolled	71	77	64
% enrolled: Hindu	77	84	73
% enrolled: Muslim	63	68	58
% enrolled: SCT	62	69	54
% enrolled: Central	60	69	50
% enrolled: South	81	85	77
% enrolled: West	84	88	79
% enrolled: East	71	75	66
% enrolled: North	82	87	77
% enrolled with both parents literate	94	95	93
% enrolled with literate father & illiterate mother	78	85	70
% enrolled with literate mother & illiterate father	86	86	86
% enrolled with illiterate parents	53	61	43
% living in village *without* anganwadi	50		
% living in village *without* primary school	11		
% living in village *without* middle school within 2 km	30		
% living in village *without* high school within 5 km	25		

Source: NCAER Survey

Table 3
Selected Data for School Continuations after Enrolment: Children Aged 10-14

	All Children (16,565)	*Boys (9,263)*	*Girls (6,942)*
% continuing	90	92	88
% continuing: Hindu	92	94	89
% continuing: Muslim	87	89	85
% continuing: SCT	89	91	85
% continuing: Central	92	92	93
% continuing: South	87	90	85
% continuing: West	88	91	85
% continuing: East	90	91	88
% continuing: North	94	95	93
% continuing with both parents literate	95	97	94
% continuing with literate father & illiterate mother	90	93	87
% continuing with literate mother & illiterate father	86	94	79
% continuing with illiterate parents	87	89	84

Source: NCAER Survey

Table 4A
SURE estimation results of the equations for the number of: pregnancies, live births and infant survivals

Equation→ Determining Variables↓	Number of Pregnancies (29,088 cases)	Number of Live Births (29,088 cases)	Number of Infant Survivals (29,088 cases)
Central	-0.1542523 (5.19)		
South	-0.4316747 (13.32)		
West	-0.1689665 (4.80)		
East	-0.1729036 (4.93)		
SCT	0.1859089 (9.73)		
Muslim	0.6415909 (20.97)		
Years Married	0.1453043 (135.39)		0.0152758 (27.04)
Household Size	0.0917523 (34.17)		
Household Income	-0.0037475 (16.08)		0.003927 (4.23)
Woman Literate	-0.1780996 (8.94)		0.0256519 (2.81)
Husband Literate	-		0.0311537 (3.65)
Male infant survivals	-1.862419 (49.79)		
Female infant survivals	-1.190382 (30.74)		
Number of abortions	0.8296623 (16.94)		
Absence of Pharmacy*	0.045633 (2.31)		-0.116614 (1.37)
Absence of 'safe' water*			-0.0353449 (4.80)
Absence of Dispensary*			-0.0138997 (1.82)
Number of Pregnancies		0.8309316 (369.54)	
Pregnancy Spacing		0.0249295 (47.21)	
Number of female Live births			0.7944025 (260.96)
Number of male live births			0.7641552 (222.38)
Husband Labourer			-0.0225011 (2.39)
Woman Labourer			-0.0419583 (3.67)
Intercept	2.286332 (49.75)	-0.0428352 (5.48)	0.2169783 (16.80)

Figures in parentheses are z-values; * 'absence' as defined in Table 1

Table 4B
Numbers of Pregnancies, Live Births and Infant Survivals: Equation Statistics

	Number of Pregnancies	Number of Live Births	Number of Infant Survivals
Observations	29,088	29,088	29,088
Parameters	14	2	11
Adjusted-R^2	0.4312	0.8971	0.8669
Breusch-Pagan Test**		$\chi^2(3)=428.4$	

* This statistic tests the null hypothesis that the errors across the equations are independent.

Table 5A
Logit Estimates of the School Enrolment and Continuation Equations

Equation→ Determining Variables↓	*School Enrolments (37,566 cases)*	*School Continuations (15,179 cases)*
Central	-0.3693563 (30.99)	-0.7683893 (2.44)
South	-	-0.6552453 (3.68)
West	-	-0.6181977 (3.95)
East	-0.4734663 (17.20)	-0.5384413 (5.00)
SCT	-0.7229936 (11.33)	-0.8996944 (1.60)
Muslim	-0.5971857 (12.32)	-0.6042515 (5.35)
Household Income	1.002821 (5.66)	1.00189 (1.80)
Father educated: low*	2.519997 (28.52)	-
Mother educated: low*	3.282561 (21.47)	1.701709 (6.28)
Father educated: medium***	2.987703 (23.77)	1.686987 (5.47)
Mother educated: medium**	3.171892 (11.02)	3.878081 (5.95)
Father educated: high******	4.429366 (27.07)	2.455744 (7.47)
Mother educated: high***	2.483103 (7.03)	5.511184 (4.59)
Father cultivator	1.314194 (8.32)	1.353905 (1.73)
Father labourer	-	1.400528 (1.85)
Father non-manual	1.599029 (11.63)	1.890149 (3.48)
Mother Cultivator	-	-0.7541737 (2.80)
Mother labourer	-0.8795938 (3.44)	-0.5789348 (6.25)
Mother non-manual	-0.7997272 (1.92)	-0.4416161 (4.14)
No anganwadi in village	-0.8739616 (4.79)	
No primary school in village	-	1.251527 (2.20)
No middle school within 2 km	-0.8510045 (5.81)	
Number of Siblings	-0.9069286 (10.51)	-0.913069 (3.99)
Household size	-	1.025234 (2.17)
Number of years at school		-0.6746576 (20.42)
Female	-0.4417544 (31.26)	-0.5294515 (10.80)

Figures in parentheses are z-values and coefficients are shown in terms of 'odds-ratios'
*literate, but educational attainment primary or less; **educational attainment more than primary but less than matric;
***educational attainment matric or higher

Table 5B
School Enrolments and Continuations: Equation Statistics

	School Enrolments	*School Continuations*
Observations	37,566	15,179
Psuedo-R2	0.1925	0.1173
Test of 'intercept' only model	$\chi^2(23)=1116.14$	$\chi^2(19)=8727.66$
LR test of zero restrictions	$\chi^2(6)=2.08$	$\chi^2(4)=4.14$

Table 5C
School Enrolments and Continuations: 'Hits' and 'Misses'

	School Enrolments		*School Continuations*	
Actually→ Classified as↓	Enrolled	Not-Enrolled	Continuing	Not-continuing
Enrolled	23,884 (sensitivity=89.7)	6,286		
Not-Enrolled	2,732	4,664 (specificity=42.6)		
Total	26,616	10,950		
Continuing			13,726 (sensitivity=99.9)	1,431
Not-continuing			15	7 (specificity=0.49)
Total			13,741	1,438

Table 6
Effects of Increases in Household Income, Household Size, Years of Marriage and Number of Siblings on Demographic and Schooling Outcomes

	Number of Pregnancies	*Number of Infant Survivals*	*Probability of School Enrolment*	*Probability of School Continuation*
Sample averages	3.48	2.93	0.71	0.90
All incomes raised by 20%	3.46	2.93	0.71	0.91
All household sizes increased by 1 person	3.58	na	na	0.91
Number of siblings increased by 1 for all children	na	na	0.69	0.89
Years of marriage increased by 1 year	3.63	2.95	na	na

na = not applicable

Table 7
The Effects of Religion/Caste on the Number of Pregnancies of Currently Married Women and on the Probabilities of School Enrolment and Continuation

Variable→ Scenario↓	*Number of Pregnancies*	*Probability of School Enrolment*	*Probability of School Continuation*
All-Hindu	3.35	0.74	0.91
All-Muslim	3.99	0.66	0.87
All-SCT	3.54	0.69	0.90

Table 8
The Effects of Literacy on
the Number of Pregnancies and Infant Survivals
of Currently Married Women

Variable→ Scenario↓	*Number of Pregnancies*	*Number of Infant Survivals*
All currently married women and their husbands are illiterate	3.53	2.91
All currently married women are literate, but their husbands are illiterate	3.36	2.93
All currently married women and their husbands are literate	3.36	2.96

Table 9
The Effects of Parents Educational Attainment on
the Probability of School Enrolments and Continuations

Variable→ Scenario↓	*School Enrolment*	*School Continuation*
Both parents are illiterate	0.57	0.87
Both parents have 'low' level of educational attainment	0.90	0.92
Both parents have 'medium' level of educational attainment	0.91	0.98
Both parents have 'high' level of educational attainment	0.92	0.99

'low' is literate, but educational attainment primary or less;
'medium' is educational attainment more than primary but less than matric;
'high' is educational attainment of matric or higher;

Table 10
The Effects of Region on
the Number of Pregnancies of Currently Married Women
and on the Probabilities of School Enrolment and Continuation

Variable→ Scenario↓	*Number of Pregnancies*	*Probability of School Enrolment*	*Probability of School Continuation*
All-Central	3.53	0.64	0.90
All-South	3.25	0.80	0.89
All-West	3.51	0.80	0.89
All-East	3.51	0.69	0.88
All-North	3.68	0.80	0.93

Table 11
The Effects of Occupation on
the Number of Infant Survivals to Currently Married Women
and on the Probabilities of School Enrolment and Continuation

Variable→ Scenario↓	*Number of Infant Survivals*	*Probability of School Enrolment*	*Probability of School Continuation*
All husbands/fathers cultivators	2.94	0.72	0.90
All husbands/fathers labourers	2.92	0.67	0.90
All husbands/fathers non-manual	2.94	0.75	0.92
All women/mothers cultivators	2.94	0.72	0.89
All women/mothers labourers	2.90	0.69	0.87
All women/mothers non-manual	2.94	0.68	0.84
All women/mothers unoccupied	2.94	0.72	0.92

Table 12
The Effects of Village-Level Infrastructure on the Number of Pregnancies and Infant Survivals to Currently Married Women and on the Probabilities of School Enrolment and Continuation

Variable→ Scenario: All persons live in villages with↓	Number of Pregnancies	Number of Infant Survivals	Probability of School Enrolment	Probability of School Continuation
A pharmacy within 1km	3.47			
No pharmacy within 1km	3.51			
'Safe' Drinking Water*		2.95		
'Unsafe' Drinking Water		2.91		
A dispensary within 5 km		2.94		
No dispensary within 5 km		2.93		
Anganwadi in village			0.72	
No anganwadi in village			0.70	
Primary school in village				0.90
No primary school in village				0.92
Middle school within 2 km			0.72	
No middle school within 2 km			0.69	

* from: protected wells; tanker truck; hand pump; piped water.