DETERMINANTS OF FERTILITY IN THE STATES OF BIHAR, MADHYA PRADESH, RAJASTHAN AND UTTAR PRADESH

Dissertation submitted to Jawaharlal Nehru University
in partial fulfilment of the requirements
for the award of the degree of

MASTER OF PHILOSOPHY

SISIR DEBNATH

Centre for Economic Studies and Planning
School of Social Studies
JAWAHARLAL NEHRU UNIVERSITY
New Delhi 110067
2006

CENTRE FOR ECONOMIC STUDIES & PLANNING
SCHOOL OF SOCIAL SCIENCES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI-110067, INDIA

CERTIFICATE

Certified that the dissertation entitled " Determinants of Fertility in the States of Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh" Submitted by Sisir Debnath in partial fulfillment for the award of the degree of Master of Philosophy (M.Phil.) of this University, is his original work and may be placed before the examiners for evaluation. This dissertation has not been submitted for the award of any other degree of this University or of any other University.

Pradipta Chaudhury
Prof. Pradipta Chaudhury
(Chairperson)

Pradipta Chaudhury
for Dr. Sugato Dasgupta
(Supervisor)

Phone : 91-11-26717676, 26717557 Ext. 4421/26704421 Cable : JAYENU Fax : 91-11-26717603, 26717585, 2671758

Maa ar Baba ke

Acknowledgment

I would like to extend my heartfelt gratitude to my advisor Dr. Sugato Dasgupta for his perceptive ideas and persistent encouragement while I was working under him. He taught me the intricacies of writing academic papers and endured the pain to go through the draft of the dissertation rectifying the errors. I consider it a privilege to work under him.

I thank Tirthatanmoy Das for helping me out with the dataset and the software. I am grateful to Mr. Saswata Ghosh of Council for Social Development, New Delhi for his valuable advice. I acknowledge my friends Amarendra, Trinankur, Sachin and Tathagata for their help and support.

My family members have been extremely supportive and inspired me throughout for higher studies. Without their love, inspiration and the values they have embedded in me I could not have reached this stage.

Specially, I am thankful to Piyali for motivating and supporting me throughout. No one can match her patience as she was always there extending her help unasked.

This dissertation is a result of all these contributors, named or not; any remaining faults are solely mine.

Sisir Debnath

Contents

Abstract:

Using data from the second round of the National Family Health Survey, this study examines the determinants of fertility in the states of Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh. In accord with conventional wisdom, I found that the educational attainments and the occupational status of the father and mother affect fertility in these four states. Somewhat surprisingly, socioeconomic factors (e.g., religion) are also important determinants of fertility.

1. Introduction:

For many developing nations high population growth is an impediment to economic development. The goal of population policies in almost every developing country is to reduce fertility. Knowledge of determinants of fertility is an important primary requirement to accomplish that goal. This empirical study estimates the effects of socioeconomic factors on fertility in four Indian states: Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh. Using survey data from National Family Health Survey, 1998-99, and quantitative economic analysis, this study analyses the determinants of fertility measured as number of children ever born to a woman. The study reveals that besides economic factors (education and occupation of the parents) a few other factors that are more social in nature (religion, caste), have considerable influence on family size in these four states.

The main objective of this paper is to determine the major determinants of fertility in the states of Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh. This query appeals since very few studies, as yet, have focused on the determinants of fertility in India. The existing studies on developed countries, even for developing countries, often try to explain fertility on the basis of biological factors, which determine fertility directly. But in case of developing countries there are, other than biological factors, socioeconomic factors, which have considerable impact on fertility decisions, even though they may not seem to be directly relevant in decision-making on fertility. For example, religion and caste exert substantial influence on fertility decision in India. The results of this study may give an insight into the fertility behavior of South-East Asian developing countries. Secondly, the states that have been chosen for analysis in this study are characterized by high population and are economically deprived compared to other states of India. The Total Fertility Rate in these states were 3.49, 3.31, 3.78 and 3.99 per cent respectively in 1999, while the Total Fertility Rate of India was 2.85. Mean age of marriage, level of education, median age at first marriage, and variables of this kind are considerably below the national average for all these four states. Therefore, knowledge of determinants of family size in these states may be of interest from the point of view of policy prescription.

The existing literature on determinants of fertility can be divided into three major categories depending on the methodology and discipline: microeconomics of fertility, general socioeconomic study, and psychological approaches.

Approaches that address fertility decisions based on microeconomic models were developed by Becker (1965), Becker and Lewis (1973), and Lancaster (1966). The central idea in these models have been, that households allocate their resources to maximize their utility function. Becker provided the microeconomic framework of analyzing household fertility decisions in his pioneering work *An Economic Analysis of Fertility.* Becker (1960) and Becker and Lewis (1973) argued that fertility is determined by the interaction between *quantity* and *quality* of children. In line with the neoclassical theory, he suggested that children can be viewed as durable goods and parents derive utility out of them as children are seen as a source of future income. The key feature of Becker's model was that the shadow price of children with respect to their number is greater, higher the quality of them. Similarly, the shadow price of children with respect to their quality is higher, greater the number of them. He assumed that parents maximize their lifetime utility function, which depends on the number of children, their quality and a composite household commodity. They also face a nonlinear budget constraint. The demand for children is determined by the interplay of price and income effects. Jacob Mincer (1963) introduced the concept of opportunity cost of time that is incurred to bring up a child in the context of demand for children. Since the price of time is generally linked with wage rate, Mincer illustrated his point with reference to fertility by regressing number of children on proxies for husband's and wife's earnings. The regression coefficient on the husband's earnings was interpreted as the income effect and the difference between the coefficients of wife's and husband's earnings were interpreted as compensated price effect. Using four sets of U.S cross-section data he showed that the absolute value of the negative price elasticity generally exceeded the positive income elasticity.

Socioeconomic studies on the determinants of fertility can be grouped according to the level of study. For example it could be a macro level study with countries as the unit of analysis or it could be household level study using sample surveys. In the household level studies fertility is arrayed against explanatory variables such as education and occupation of the parents, household income etc. The empirical results of these studies are often predictable. Davis and Blake (1956) and Bongaarts (1983), in this context, gave a simple but powerful framework for analyzing the proximate determinants of fertility. Davis and Blake pointed out that socioeconomic variables must operate through behavioral and biological factors to affect the number of children ever born to a woman. They called the biological factors *intermediate fertility variable* and identified a set of 11 such variables: age of entry into sexual relationship, fecundity, use of contraception, etc. John Bongaarts (1982) listed seven biological factors or *proximate determinants* of fertility through which the environmental and socioeconomic factors influence fertility. The factors identified by him were marriage, natural fecundability or frequency of intercourse, spontaneous intrauterine mortality, post partum infecundability, effectiveness and use of contraceptives, induced abortion and permanent sterility. As the available evidence suggested that there was little variation among populations in the proximate determinants of natural fertility, he formulated a model expressing TFR (Total Fertility Rate) as the product of four indices. They were index of proportion of woman married, index of contraception, index of induced abortion and index of lactational infecundability. His model could explain 96% of the variance in fertility level among societies. However, many empirical studies on determinants of fertility, as well as the theoretical models on fertility ignored the constraints put on fertility by the biological variables pointed out by Davis and Blake and Bongaarts. A multivariate analysis of the impact of socioeconomic factors on marital fertility in developing countries by German Rodriguez and John Cleland showed that woman's participation in the labour force is the most important among six other factors considered by them: urban or rural residence, husband's education, occupation and work status, wife's education and work status. The analysis was based on 20 developing countries. In most populations they found rural fertility to be substantially higher than urban population. But only half of this fertility gap was attributable to differences in education and employment opportunities. They concluded that presumably the other half was due to weakening of traditional high fertility values and greater access to family planning techniques. They observed wife's educational attainment to have a greater influence on fertility than that of husband. Their findings provide strong evidence that increased employment opportunities for women can have a major contribution in bringing down the fertility rate.

Finally, a large number of studies on fertility have considered the effect of psychology on fertility and fertility decision. This approach laid considerable importance on *latent demand*, which measures the desired family size. Later there were studies trying to explore the relationship between *individual modernization* and fertility and between *intraspousal communication* and adoption of contraceptives.

This study analyzes fertility using sociological (religion, caste), economical (education and occupation of husband and wife) and biological (age) factors. To avoid endogeneity problem I have included an economic status index in the analysis as a proxy for the income of the family. The multivariate linear regression results show that all the factors are statistically significant in explaining variation in number of children ever born to a woman, age of the wife being the most important one. Occupation and education of wife have greater impact on family size than that of husband. Families with the head of household belonging to backward castes have a greater family size. Similarly, on the religion front, controlling for other factors, Muslims have greater size of family compared to Hindus and Others. Interestingly the size of the family decreases as the family income increases or as the economic status index assumes higher values. Since family size is not continuous the results have also been cross-checked using an ordered logit model. The results of both the models corroborate with each other.

The paper proceeds as follows: Section 2 describes the dataset (Second round of National Family Health Survey, 1998-99); Section 3 gives the empirical evidence including the construction of the asset index using principal component analysis, and finally Section 4 concludes.

2. The dataset:

The National Family Health Survey (1998-99) was carried out by the Indian Institute of Population Sciences, Mumbai. The principal objective of the survey was to provide state and national-level estimates of fertility and infant mortality, and to shed light on the practice of family planning, the state of maternal and child health, the utilization of health services provided to mother and children, and so on.

The survey covered a representative sample of more than 90000 eligible women of age 15-49 years from 26 states of India. The data collection took place in two phases: the first stage started in November 1998, and the second stage started on March 1999. A uniform questionnaire, sample design, and field procedure was followed so that data across states are comparable. As indicated already, this study deals with four states of India: Bihar, Madhya Pradesh (M.P.), Uttar Pradesh (U.P.) and Rajasthan.

2.1 Some facts:

On the basis of the NFHS dataset, Table 1 contrasts certain socioeconomic indicators from the four chosen states with the all-India average. Notice that there is a dearth of awareness and literacy among women in these four states compared to what prevails in India on average.

At an all-India level, the percentage of women in the age group 15-49 years not exposed to mass media is 40.3; the corresponding figure for Bihar is 72.7. Likewise, at an all-India level, 52.8 percent of women in the age group 15-49 years are illiterate; the corresponding figure for Bihar is 76.6. Table 1 makes another interesting point. The fertility of women in the four chosen states exceeds the national average (observe the first row of Table 1, which gives the Total Fertility Rate figures). Therefore, an enquiry into the determinants of fertility in these four states is a worthwhile exercise.

For each woman in the 15-49 years age group, I first determine the number of children ever born to her. This done, Table 2 provides the average of the variable for each state, broken down by caste, religion, and education levels of husband and wife. Several interesting patterns emerge. Notice that if the head of the household belongs one of the backward caste categories (Scheduled Caste, Scheduled Tribe, or Other Backward Castes), the average number of children per women is slightly higher than if the household head belongs to non-backward castes. A religion-wise classification of the observations also reveals that Muslim women on an average have more children than any other religion. Finally, for each state, I have divided the observations into three groups according to the education of the wife and the husband. The first group consists of women with 0 to 5 years of education; the second group comprises women with 6 to 10 years of education, while the third group has women with education weakly exceeding 11 years. The average number of children ever born falls dramatically as the education of the wife increases. Repeating the procedure of dividing the observations according to the years of education of the husband I get the result that the average number of children ever born drops off as well when the husband's education rises.

2.2 Variable description:

In this study I explain the fertility decisions of women (that is, number of children ever born) using three sets of variables: sociological, economic and biological.

The caste and religion of the head of the household to which a woman belongs are sociological variables. The survey provides caste data broken down by four broad categories: Scheduled Caste, Scheduled Tribe, Other Backward Castes and Others. The government of India recognizes the first three caste categories as underprivileged. It is interesting to see whether woman belonging to these castes have more children ever born than women belonging to privileged classes. Accordingly, I have created a dummy variable, *bwcast,* which takes the value 1 if the respondent belongs to any one of the three backward castes, and is 0 otherwise. To study the effect of religion on fertility decisions, I have created three dummy variables: *relgh, relgm,* and *relgo.* Here, *relgh* takes the value 1 if the respondent is Hindu, and is 0 otherwise; *relgm* takes the value 1 if the respondent is Muslim, and is 0 otherwise; while *relgo* takes the value 1 if the respondent is neither Hindu nor Muslim, and is 0 otherwise.

To investigate the effects of the education and employment profiles of husband and wife on the number of children ever born, I have constructed four variables: *edum, eduf, occm* and *occf.* The variables *edum* and *eduf* are, respectively, the years of education of the wife and the husband. The other two variables, *occm* and *occf,* are dummy variables for the occupation of the wife and the husband respectively. If the respondent (wife) is engaged in a high-paid job (see Table 3 for details), *occm* takes the value 1; otherwise, *occm* assumes the value 0. Likewise, *occf* takes the value 1 if the husband has a high-paid job (see Table 3 for details), and is 0 otherwise.

The only biological factor that I consider is age, the respondent's age (in years). When *age* increases, a woman has had more years of fertility. Ceteris paribus, this means that the number of children ever born and the respondent's age should be positively related.

3. Empirical results:

The empirical results section comprises three subsections. In the first subsection, the construction of an economic status index is outlined. This index is used as a proxy for family wealth. The second subsection discusses the econometric model used to study the determinants of fertility and presents the empirical results obtained. The final subsection deals with the robustness of the empirical findings.

3.1 Economic status index:

The long-run economic status of a family (measured, for example, by family wealth) is one of the potential determinants of its size. One of the problems with the NFHS dataset is that it does not provide any data on the income or wealth of a family. But, the NFHS dataset provides detailed data on household asset ownership and household characteristics. I aggregate the household asset ownership data and the household characteristics data to construct an economic status index which will be used as a proxy for the economic status of a household. The economic status of a family will thereafter be approximated by the economic status of the household to which it belongs.

An index reflecting the economic status of a household can be constructed in several ways. A simple linear index constructed by giving equal weight to every variable which has information on the economic status of a household might seem interesting due to its simplicity. But, this method has serious drawbacks. Equal weightage given to every asset ownership variable implies that every asset is seen at par, as almost all of these ownership variables are binary. In other words, an equal-weightage index ensures, incorrectly of course, that a household with a car and without a bicycle is as well off as a household without a car and with a bicycle.

Another approach to index construction could be to estimate the current value of the assets possessed by a household. The index in this case would be a linear one with weights being the prices of the assets. Constructing such an index would require knowledge of purchasing prices, the dates of purchase, and depreciation rates for each of the assets. The NFHS data set provides only binary asset ownership variables and household characteristics, but not the details of purchase. Therefore it is not possible to calculate the current value of household assets.

The index that I have constructed is along the lines suggested by Pritchett (2001). I first consider a large set of asset ownership variables and household characteristics variables. Now, principal component analysis linearly transforms this original set of variables into a substantially smaller set of uncorrelated variables that represents most of the variation present in the original dataset. These derived variables are called principal components. The first principal component gives the linear combination of the original variables which has the largest sum of squared correlation with the original set of variables. The second principal component gives another linear combination of the original variables which is independent of the first principal component and has the second largest squared correlation with the original variables; and so on (see Appendix 1 for details).

Here, I will use only the first principal component to construct the economic status index. Suppose there are p asset ownership variables. Since the principal components are calculated using the standardized values of the original p variables, the index for the n'th family is expressed as follows:

$$y_{1n} = a_{11}\frac{(x_{1n}-\bar{x}_1)}{\sigma_1} + a_{12}\frac{(x_{2n}-\bar{x}_2)}{\sigma_2} + ... + a_{1p}\frac{(x_{pn}-\bar{x}_p)}{\sigma_p}$$

$$= \left(a_{11}\frac{x_{1n}}{\sigma_1} + a_{12}\frac{x_{2n}}{\sigma_2} + ... + a_{1p}\frac{x_{pn}}{\sigma_p}\right) - \left(a_{11}\frac{\bar{x}_1}{\sigma_1} + a_{12}\frac{\bar{x}_2}{\sigma_2} + ... + a_{1p}\frac{\bar{x}_p}{\sigma_p}\right)$$

$$= \left(a_{11}\frac{x_{1n}}{\sigma_1} + a_{12}\frac{x_{2n}}{\sigma_2} + ... + a_{1p}\frac{x_{pn}}{\sigma_p}\right) - Const., \quad (1)$$

where y_{1n} is the economic status index of the n'th family, a_{1i} is the coefficient of the i'th asset ownership variable derived from the first principal component, x_{in} is the value of the i'th asset ownership variable for the n'th family, $\bar{x}_i$ is the sample average of the i'th asset ownership variable, and σ_i is the sample standard deviation of the i'th asset ownership variable.

Since the constant in equation 1 is the same for every family, following simplified index leaves the ranking of the families unaltered:

$$y_{1n} = \left(a_{11}\frac{x_{1n}}{\sigma_1} + a_{12}\frac{x_{2n}}{\sigma_2} + ... + a_{1p}\frac{x_{pn}}{\sigma_p}\right). \quad (2)$$

The most important assumption required for the construction of the economic status index is that long run wealth of a household explains the variations present in the asset ownership variables. There is however no direct way to test this assumption.

3.1.1 Construction of the economic status index:

There were a total of 30069 respondents in the survey from the four states. The maximum number of respondents were from the state of Uttar Pradesh (9291) and the least number of observations were from the state of Rajasthan (6813).

I have considered two sets of asset ownership variables while constructing the economic status index: variables indicating the possession of consumer durables, and variables on household dwelling characteristics. There are eight dummy variables on consumer durable ownership (clock or watch, bicycle, radio, television, sewing machine, refrigerator, car, motorcycle) and ten variables on characteristic of household dwelling (three variables on drinking water, three variables on toilet facilities, and a variable each on source of lighting, number of rooms, kitchen characteristics, and main cooking fuel).

The first column of Table 4 gives the coefficients of the asset ownership variables for the first principal component. Thus, for example, the variable indicating whether the family owns a clock/watch is assigned a weight of 0.25 in the first principal component. The second column of Table 4 gives the means of the 18 variables (for example, the mean of the variable indicating ownership of clock/watch is 0.634) while column 3 records the standard deviation of the same. Since all the variables are binary (except Number of Rooms), it is easy to calculate the contribution of each variable to the economic status index. Column 4 which computes Coeff./S.D. (that is, column 1/column 3), gives the contribution of each variable to the economic status index. I refer to these values as scoring factors. Notice that owning a car increases the economic status index by the highest units (1.088), followed by owning a refrigerator (1.078).

3.1.2 Checking the reliability of the index:

I check the reliability of the economic status on two fronts: internal coherence and robustness.

3.1.2.1 Internal coherence: To check the internal coherence of the index, I have sorted the families according to the index in descending order. I refer to the lowest 40 percent of families as the *poor*, upper 20 percent of the families as the *rich*, and 40 percent of families in the middle as the *middle class*. Columns 5-7 of Table 4 show the average of the asset variables for these three groups of families. Notice that the ownership of assets vary considerably among the groups. For example, 23 percent of the poor own a clock or watch; the corresponding figure for the middle class and the rich are 86 percent and 99 percent respectively. Similarly, among the poor families 15 percent have houses with a separate kitchen whereas 75 percent of the rich families have this facility. A clear pattern emerges across the various asset categories: the probability of owning an asset increases with the status of the family as measured by the constructed index.

3.1.2.2 Robustness: To check the robustness of the economic status index, I will examine how the ranking of the families changes when a subset of the asset ownership variables are used to construct the index. Table 5 shows that the pattern of assignment of families into the three categories (poor, middle class and rich) does not change drastically even if I use a subset of the asset ownership variables to construct the economic status index. Consider, for example, how the poorest 40 percent (determined by the economic status index when all the 18 asset ownership variables are used to construct it) are reclassified if I use a subset of the 18 asset ownership variables to construct the economic status index. Column 1 is the base case where the poorest 40 percent based on the status index constructed using 18 asset variables, are considered to be 100 percent. Column 2 shows that if variables relating to drinking water sources and toilet facilities are dropped while making the index, 3.8 percent of the poor are promoted to the middle class group. Finally if only the consumer durable asset ownership variables are used to construct the index, 15.9 percent of the poor move up the ladder to the middle class group. But, in both the cases, none of the poor group would be reclassified as a member of the rich group. Therefore, it can be concluded that the economic status index survives the robustness test. Similar robustness results are obtained for the remaining two groups as well.

To get a clearer picture of the rankings produced by the three different indices, I calculated the Spearman rank correlation coefficient between the base case and each of the other two cases. The rank correlation between the ranks when all the variables are used and when I drop the drinking water and toilet facility variable to construct the index is 0.928. Even when the index is constructed using the consumer durable assets alone, the rank correlation with the base case is 0.797.

For an additional check of robustness, I have used the method of factor analysis to determine the weights to be assigned to the asset variables while constructing the economic status index. After ranking the families using the first factor, the Spearman rank correlation of the rankings with those derived earlier using the first principal component is 0.997. Clearly, the results are robust.

3.2 Results:

To analyze the determinants of the number of children ever born to a woman, I have used three sets of variables as regressors. As described in the data section these are sociological, economic and biological variables. To investigate the effect of these variables I have used a simple linear regression model.

The linear regression equation is as follows:

$$\begin{aligned} ebornch = {} & \beta_1 + \beta_2 \times relgm + \beta_3 \times relgo + \beta_4 \times bwcast + \beta_5 \times edum + \beta_6 \times eduf + \\ & \beta_7 \times occm + \beta_8 \times occf + \beta_9 \times ain + \beta_{10} \times age + \varepsilon, \end{aligned} \tag{3}$$

where ε is the error term accounting for the unobserved variables. ε is assumed to be orthogonal to each of the control variables. Not ruling out the fact that there might be endogeneity problem associated with such a model, it is still interesting to observe the results.

The results of the linear regression for the states of Bihar, Rajasthan, Madhya Pradesh and Uttar Pradesh are summarized in Table 6. Column 1 in Table 6 shows the coefficients of the regressors used in equation 3. Since I have dropped the variable *relgh* (dummy variable for religion Hindu) in equation 3, the positive coefficient (0.63) of *relgm* (dummy variable for religion Muslim) and the negative coefficient (-0.38) of *relgo* (dummy variable for other religions) implies that, in the four states, a Muslim woman have more children compared to a Hindu woman and a women belonging to other religions have less children compared to a Hindu women. In fact the absolute value of the coefficient of *relgm* is the highest in magnitude among all the control variables (except the constant). The coefficient of the dummy for backward castes, *bwcast*, despite being significant at 99 percent has a very small positive value. Therefore, it can be concluded that among the sociological determinants of fertility, religion is the most important determinant in the four concerned states.

Table 6 shows that all the economic determinants have negative coefficients, among them occupation of mother, *occm*, have the strongest impact on the number of children ever born. Both education of wife and husband have significant negative coefficients. The coefficient of *edum* and *eduf* indicate that years of education of a mother is six times more effective than that of a father in reducing fertility. The coefficient of occupation of the wife and the husband are -0.45 and -0.09 respectively. It implies that compared to a woman engaged in a low pay job, a woman having a high paid job has less number of children. This is also true for the husband. As the coefficient of economic status index, *ain,* is negative the economic status of a family also has a negative impact on number of children ever born.

Finally, the impact of the single biological factor, the age of wife in years, has a positive coefficient with the highest level of significance.

Column 3 in Table 6 shows the correlation coefficient of the explanatory variables with the number of children ever born. The biological factor, age of wife, has highest correlation with the number of children. Dummy variables for the religion Muslim and backward castes have positive correlation while all the other variables have a negative correlation with number of children ever born to a woman.

3.3 Robustness:

I have used two methods to check the robustness of the results reported in the previous subsection. Since the number of children ever born takes only non-negative integral values, I have estimated an ordered logit model using the same dataset. Secondly, in order to show that the results are not driven by any subset of the dataset, I have estimated both the linear regression model and the ordered logit model for each of the state and compared the results.

To estimate an ordered logit model (see Appendix 2 for details), I have created a new variable, *noc,* having three ordered categories, based on number of children ever born to a woman; *noc* takes the value 0 if the number of children ever born (*eborneh*) is 0, it takes the value 1 if *eborneh* takes the values 1, 2 or 3 and finally it takes the value 2 if *eborneh* is greater than 3. Thus I have divided the respondents into three ordered categories: not a single child ever born, one to three children ever born, and more than three children ever born. The same regressors as in equation 3 (except the constant) have been used to estimate the logit model.

The results of the ordered logit model are summarized in Table 7. The first column of Table 7 shows the coefficients. The sign of every coefficient is same as the coefficients obtained in the linear regression model. The coefficients of *relgm, bwcast* and *age* bear positive signs whereas those of the other control variables have negative sign. Therefore, the directions in which the regressors influence number of children are same in both the models. To check the robustness of the ordered logit model I have used the same model for each of the four states separately. These four ordered logit regressions yield four coefficients for each regressor. Column 2 and column 3 shows the maximum and the minimum among the four coefficients for each variable. Except for the dummy variable on occupation of the father, *occf,* the maximum and the minimum of the four coefficients bear the same sign as that of the coefficients recorded in column 1. Clearly, the results of the ordered logit model are robust. Further, column 4 shows in how many of the four logit regressions a regressor was found significant.

Since the coefficients of ordered logit models are hard to interpret I have calculated the probabilities of some hypothetical respondents to have no children (*noc* = 0), one to three children (*noc* = 1), and more than three children (*noc* = 2) ever born. This will enable us to understand the effect of the control variables on fertility predicted by the ordered logit model. In Table 8, column 1 shows the case numbers, column 2 – 11 are the values of the control variables taken by some hypothetical respondents. Column 12 – 14, give the probabilities that a respondent with certain characteristics, which is described in column 2 – 11, will have no children (*p0*), one to three children (*p1*), and more than three children (*p2*). First six regressors (column 2 to 7) are binary; they assume the value 1 if true and 0 otherwise. For example, if *relgh* takes the value 1 then the respondent is Hindu; otherwise 0 (see Table 3 for a detailed description). In the first four cases the variables *edum, eduf, ain, age* have been assigned their average values. In the first three cases I predict the probabilities *p0, p1 and p2*, ceteris paribus, for three women belonging to different religions. The distributions of predicted probabilities are identical if their religion is Hindu or other, whereas for the Muslim woman the distribution is skewed to the right. The probability to have one to three children is higher by 0.1 for a Hindu woman compared to a Muslim. Whereas, the probability to have more than three children is higher by 0.12 for a Muslim women compared to a Hindu woman. Therefore, Muslim women are more probable to have more than three children ever born. A comparison of case 1 and case 4 illustrates the effect of caste on number of children, which is negligible, as the probability distribution remains almost the same in both the cases. Cases 4 to 7 can be used to see how the increase in years of education of a woman can affect the number of children ever born to her. Similarly, cases 4, 8 and 9 can be used to evaluate the effect of educational years of a husband on the probability distribution of *noc*. It is evident from the comparisons that both *edum* and *eduf* have negative impact on *noc,* as the predicted probabilities of having more than 3 children reduces as the years of education of the wife and the husband increase. Cases 10 and 11 compare the effect of nature of occupation of the wife and the husband on *noc.* Clearly, the high paid occupation of the wife is more effective in bringing down the fertility than that of the husband. Finally, the cases 11 to 13 show the impact of increase in economic status index on the probability distribution of *noc*. Increase in economic status index decreases the probability for a woman to have more children.

To test the robustness of the results obtained from the linear regression model specified in equation 3, I estimated the same model for each of the four states. The sign and the magnitude of the coefficients obtained from the four regressions on each state are almost similar to that of the coefficients obtained earlier. The results are summarized in Table 9. Column 1 in Table 9 gives the coefficient of the regressors from Table 6. Column 2 and 3 in Table 9 gives the maximum and the minimum values among the four coefficients for each regressor. Except for the variable occupation of husband, *occf*, signs of all other control variables remain unchanged if the linear regression is run separately for each state. Column 4 in table 9 shows the number of times a regressor was found to be statistically significant, among the four regressions.

4. Conclusion.

Using multivariate analysis and NFHS data, this study explains fertility in four backward states of India. The extant literature explains fertility on the basis of economic and biological variables. In this study, I have incorporated certain sociological variables as well. In accord with conventional wisdom from developed country studies, I find that a woman's educational attainments and job status affect her fertility. Somewhat surprisingly sociological variables (especially religion) also influence fertility decisions.

References

Becker, Gary S., (1960), "An Economic Analysis of fertility", in Becker, ed., *Demographic and Economic Change in Developed Countries*, Princeton, NJ: Princeton University Press.

Becker, Gary S., and H. Lewis, (1974), "Interaction between Quantity and Quality of Children", in T.W.Schultz, ed., *Economics of Family,* Chicago: University of Chicago Press.

Becker, Gary S., and R. J. Barro, (1988), "A Reformulation of the Economic Theory of Fertility", *The Quarterly Journal of Economics,* Vol. 103, No. 1, 1-25.

Bongaarts, John, and Robert G. Potter, (1983), *Fertility, Biology, and Behaviour: An Analysis of the Proximate Determinants,* New York: Academic Press.

Bongaarts, John, (1978), "A Framework for Analyzing the Proximate Determinants of Fertility", *Population and Development Review,* Vol. 4, No. 1, 105-132.

Dunteman, G. H., (1989), *Principal Component Analysis,* Sage Publications: New Delhi.

Davis, Kingsley, and Judith Blake, (1956), "Social Structure and fertility: An Analytical Framework", *Economic Development and Cultural Change,* Vol. 4, 211-235.

Mincer, Jacob, (1963), "Market Prices, Opportunity Costs and Income Effects", in C. Christ, ed., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in memory of Yehuda Grunfeld,* Palo Alto: Stanford , 67-82.

Maddala. G. S, (1986), *Limited Dependent and Quantitative Variables in Econometrics,* Cambridge University Press.

Pollak, Robert A., and Wachter, Michael L., (1975), "The Relevance of Household Production Function and Its Implication for the Allocation of Time", *The Journal of Political Economy,* Vol. 83, No. 2, 255-278.

Pritchett, Lant, and Deon Filmer, (2001), "Estimating Wealth Effects without Expenditure of Data-or Tears: An Application to Enrollments in States of India." *Demography,* Vol. 38, No. 1 (February 2001), 115-132

Rodreguez, G., and John Cleland, (1981), "The Effects of Socioeconomic Characteristics on Fertility in 20 Countries" in *International Family Planning Perspective,* Vol. 7, No. 3, 93-101.

Roy, T., Jayachandran, V., and Banerjee, Sushanta K., (1999), "Economic Condition and Fertility Is there A Relationship?", *Economic and Political Weekly,* October 16-23.

Schultz, T. P., (1969), "An Economic Model of Family Planning and Fertility", *The Journal of Political Economy,* Vol. 77, No. 2, 153-180.

Schultz, T. P., (1973), "A Preliminary Survey of Economic Analyses of Fertility", *The American Economic review,* Vol. 63, No. 2, Papers and Proceedings of the Eighty-fifth Annual meeting of the American Economic Association, 71-78.

Visaria, P., and Rajan S. Irudaya, (1999), "National Family Health Survey: A Landmark in Indian Surveys", *Economic and Political Weekly,* October 16-23.

Table 1

Socioeconomic indicators for Rajasthan, Madhya Pradesh (M.P.), Uttar Pradesh (U.P.), and Bihar along with the All-India level

	Rajasthan	M.P.	U.P.	Bihar	**All-India**
Total Fertility Rate (TFR)	3.78	3.31	3.99	3.49	**2.85**
Avg. no. of children ever born to a woman	3.34	3.30	3.58	3.28	**2.99**
Avg. No. of surviving children to a woman	2.84	2.71	2.97	2.86	**2.59**
Mean age at marriage (female)	22.3	23.5	23.3	23.8	**24.9**
Percentage illiterate (female)	75.5	68.5	70.2	76.6	**58.2**
Percentage of females not regularly exposed to mass media	63.1	45.2	54.7	72.7	**40.3**

Note: Only females in the age group 15-49 years are considered in the calculations.

Source: calculations based on NFHS 1998-99.

Table 2

Average number of children ever born to a woman classified by caste, religion and education levels.

	Bihar	M.P.	U.P.	Rajasthan
Backward caste	3.28	3.37	3.67	3.46
Non backward caste	3.25	2.99	3.45	3.21
Hindu	3.19	3.28	3.48	3.29
Muslim	3.81	3.66	4.12	3.95
Others	3.02	3.06	2.78	2.63
Years of education of wife (0-5 yrs)	3.46	3.56	3.89	3.55
Years of education of wife (6-10 yrs)	2.52	2.22	2.46	2.35
Years of education of wife (11 yrs and above)	1.94	2.02	1.95	1.85
Years of education of husband (0-5 yrs)	3.53	3.76	4.26	3.89
Years of education of husband (6-10 yrs)	3.15	2.89	3.19	2.93
Years of education of husband (11 yrs and above)	2.64	2.53	2.84	2.58

Note: All calculations use proper weights.

Source: Calculations based on NFHS 1998-99.

Table 3

Description of the variables

Variable Description:	
ebornch	Number of children ever born to a woman
relgnh	Dummy variable: takes value 1 if the head of the household is Hindu; otherwise 0.
relgnm	Dummy variable: takes value 1 if the head of the household is Muslim; otherwise 0.
relgno	Dummy variable: takes value 1 if the head of the household is neither Hindu nor Muslim; otherwise 0
bwcast	Dummy variable: takes value 1 if the head of the household is either Scheduled Caste, Scheduled Tribe, or belongs to the Other Backward Castes; otherwise 0.
bwcast1	Dummy variable: takes value 0 if the head of the household is either Scheduled Caste, Scheduled Tribe, or belongs to the Other Backward Castes; otherwise 1.
occm	Dummy variable: takes the value 1 if the respondent is employed in a professional, technical ,managerial, or clerical job or works in the service sector; otherwise 0.
occm1	Dummy variable: takes the value 1 if the respondent is not working or belongs to the following three categories of workers: (1) agricultural or manual worker, (2) household or domestic worker, or (3) sales worker; otherwise 0.
occf	Dummy variable: takes the value 1 if the husband of the respondent is employed in a professional., technical ,managerial, or clerical job or works in the service sector; otherwise 0.
occfl	Dummy variable: takes the value 1 if the husband of the respondent is not working or belongs to the following three categories of workers: (1) agricultural or manual worker, (2) household or domestic worker, or (3) sales worker; otherwise 0.
eduf	Husband's education in years.
edum	Respondent's education in years.
age	Respondent's age in years.

Note: All calculations use proper weights.

Source: Calculations based on NFHS 1998-99.

Table 4

Scoring factors and descriptive statistics of the variables used in the construction of the economic status index

	(1) Coeff.	(2) Mean	(3) S.D.	(4) Coeff./ S.D.	Means (5) Poor 40%	(6) Middle 40%	(7) Rich 20%
Own Clock/Watch	0.25	0.634	0.482	0.52	0.231	0.863	0.987
Own Bicycle	0.154	0.547	0.498	0.309	0.303	0.69	0.751
Own Radio	0.23	0.333	0.471	0.489	0.052	0.434	0.699
Own Television	0.336	0.278	0.448	0.751	0.002	0.25	0.889
Own Sewing Machine	0.298	0.234	0.424	0.704	0.005	0.218	0.726
Own Refrigerator	0.27	0.067	0.251	1.078	0	0.009	0.319
Own Car	0.113	0.011	0.104	1.088	0	0.002	0.05
Own Motorcycle	0.279	0.105	0.307	0.909	0	0.041	0.447
Drinking Water from Pump/Well	0.04	0.967	0.178	0.24	0.947	0.978	0.987
Drinking Water from Open Source	-0.041	0.027	0.163	-0.25	0.044	0.018	0.01
Drinking water from Other Source	-0.012	0.005	0.073	-0.17	0.008	0.004	0.003
Flush Toilet	0.327	0.164	0.37	0.883	0	0.059	0.702
Pit Toilet/ Latrine	0.1	0.072	0.259	0.387	0	0.092	0.177
None/ Other Toilet	-0.346	0.764	0.424	-0.82	1	0.85	0.121
Source of Lighting	0.26	0.476	0.499	0.52	0.154	0.571	0.933
Number of Rooms	0.197	3.016	2.266	0.088	2.018	3.283	4.493
Separate Kitchen	0.211	0.394	0.489	0.433	0.152	0.459	0.755
Main Cooking Fuel (LPG, Electricity, Bio gas)	0.322	0.141	0.348	0.925	0	0.036	0.636
Economic Status Index		1.522	2.288		-0.033	1.707	5.764

Note: Except for number of rooms, all variables are binary (value equals 1 if true, 0 otherwise). Coeff. is the weight assigned to a variable in the first principal component.
Source: Calculations based on NFHS 1998-99.

Table 5

Classification difference of the poorest 40 percent

	(1) Base Case (18 Variables)	(2) All variables except drinking water sources and toilet facilities	(3) Only asset ownership variables
Poorest 40%	100.0	96.2	84.0
Middle 40%	0.0	3.8	15.9
Richest 20%	0.0	0.0	0.0
Total	100.0	100.0	100.0
Spearman rank correlation coefficient	1.0	0.928	0.797

Note: All correlation coefficients are significant.

Source: Calculation based on NFHS 1998-99.

Table 6

Results of the linear regression and correlation of each determinant with number of children ever born

Explanatory variables	(1) Coefficient	(2) p values	(3) Correlation with no. of children ever born
constant	-1.71*** (-37.00)	0.000	--
relgm	0.63*** (19.11)	0.000	0.09
relgo	-0.38*** (-4.46)	0.000	-0.02
bwcast	0.05** (2.26)	0.024	0.02
edum	-0.06*** (-17.13)	0.000	-0.24
eduf	-0.01*** (-5.74)	0.000	-0.20
occm	-0.45*** (-5.40)	0.000	-0.03
occf	-0.09*** (-2.88)	0.004	-0.04
ain	-0.09*** (-13.43)	0.000	-0.11
age	0.18*** (150.38)	0.000	0.65
N	29354	--	29354
R^2	0.4781		

Note: Values in parenthesis in column 1 are the t values. All the calculations are made using proper weights. * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$.
Source: Calculations based on NFHS 1998-99.

Table 7

Results of state-wise ordered logit regressions

Explanatory variables	(1) Ordered logit regression on the four states Coefficients	(2) Among the four regressions run separately for each state Max Coeff.	(3) Min Coeff.	(4) No. of states with significant coefficient
relgm	0.45*** (10.80)	0.63 ^{R}*** (6.35)	0.38 ^{M}*** (3.20)	4
relgo	-0.41*** (-4.04)	-0.03 ^{M} (-0.20)	-0.71 ^{B}*** (-3.58)	2
bwcast	0.07** (2.53)	0.18 ^{M}** (2.50)	0.12 ^{U}** (2.36)	4
edum	-0.07*** (-16.35)	-0.04 ^{B}*** (-4.58)	-0.08 ^{M}*** (-10.15)	4
eduf	-0.01** (-2.41)	-0.00 ^{B} (-0.34)	-0.02 ^{R}*** (-3.58)	2
occm	-0.36*** (-3.60)	-0.17 ^{R} (-0.86)	-0.49 ^{U}*** (-2.65)	1
occf	-0.11*** (-2.75)	0.12 ^{R} (1.45)	-0.27 ^{U}*** (-3.76)	1
ain	-0.08*** (-9.51)	-0.06 ^{R}*** (-3.47)	-0.09 ^{U}*** (-6.43)	4
age	0.19*** (93.96)	0.21 ^{U} (52.0)	0.17 ^{M}*** (42.98)	4
N	29354			
Pseudo R^2	0.2613	0.2900 ^{U}	0.2284 ^{M}	

Note: All the calculations use proper weights. Values in parenthesis in are the z values. When the dataset for the four states together is used cut points of the model are 2.595 and 5.863. R, M, U, B in col. 2 and col. 3 stand for Rajasthan, M.P., U.P., and Bihar, respectively. Column 4 is constructed considering 95 percent level of significance. * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$.
Source: Calculations based on NFHS 1998-99.

Table 8

Probability distribution of *noc* for some hypothetical cases

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Determinants										Probabilities		
Cases	*relgh*	*relgm*	*relgo*	*bwcast*	*occm*	*occf*	*edum*	*eduf*	*ain*	*age*	*p0*	*p1*	*p2*
1.	1	0	0	1	0	0	2.32	6	1.762	30.19	0.05	0.53	0.42
2.	0	1	0	1	0	0	2.32	6	1.762	30.19	0.03	0.43	0.54
3.	0	0	1	1	0	0	2.32	6	1.762	30.19	0.05	0.53	0.42
4.	1	0	0	0	0	0	2.32	6	1.762	30.19	0.05	0.54	0.41
5.	1	0	0	0	0	0	5	6	1.762	30.19	0.09	0.64	0.27
6.	1	0	0	0	0	0	10	6	1.762	30.19	0.09	0.63	0.28
7.	1	0	0	0	0	0	15	6	1.762	30.19	0.12	0.66	0.22
8.	1	0	0	0	0	0	2.32	10	1.762	30.19	0.05	0.55	0.40
9.	1	0	0	0	0	0	2.32	15	1.762	30.19	0.06	0.56	0.39
10.	1	0	0	0	1	0	15	15	1.762	30.19	0.17	0.67	0.15
11.	1	0	0	0	0	1	15	15	1.762	30.19	0.14	0.67	0.19
12.	1	0	0	0	0	1	15	15	5	30.19	0.17	0.67	0.15
13.	1	0	0	0	0	1	15	15	10	30.19	0.24	0.65	0.11

Note: Averages of *edum, eduf, ain* and *age* are 2.32, 6, 1.762 and 30.19 respectively.
p0 = Probability (*noc* = 0), *p1* = Probability (*noc* = 1), *p2* = Probability (*noc* = 2)
Source: Calculations based on NFHS 1998-99 using results of the ordered logit model (table 7, column 1).

Table 9

Results of state-wise linear regressions

Explanatory variables	Regression on the four states Coefficients (1)	Among the four regressions run separately for each state Max Coeff. (2)	 Min Coeff. (3)	No. of states with significant coefficient (4)
constant	-1.71*** (-37.0)	-1.24^{M}*** (-12.21)	-1.94^{B}*** (-19.22)	4
relgm	0.63*** (19.11)	0.71^{U}*** (12.36)	0.52^{M}*** (5.46)	4
relgo	-0.38** (-4.46)	-0.13^{M} (-0.92)	-0.5^{R}*** (-3.41)	2
bwcast	0.05*** (2.26)	0.17^{U}*** (4.05)	0.13^{M}** (2.17)	4
edum	-0.06*** (-17.13)	-0.04^{B}*** (-4.38)	-0.07^{M}*** (-9.0)	4
eduf	-0.01*** (-5.74)	-0.01^{B}* (-1.88)	-0.02^{M}*** (-4.55)	3
occm	-0.45*** (-5.40)	-0.25^{R} (-1.58)	-0.53^{U}*** (-3.35)	3
occf	-0.09*** (-2.88)	0.01^{B} (0.14)	-0.17^{U}*** (2.92)	1
ain	-0.09*** (-13.43)	-0.07^{M}*** (-5.71)	-0.11^{U}*** (-9.09)	4
age	0.18*** (150.38)	0.19^{U}*** (86.69)	0.16^{M}*** (66.77)	4
N	29354			
R^2	0.4781			

Note: Values in parenthesis are the t values. R, M, U, B in col. 2 and col. 3 stands for Rajasthan, M.P., U.P., and Bihar, respectively. Column 4 is constructed considering 95 percent level of significance. * p ≤ .1; ** p ≤ .05; *** p ≤ .01.
Source: Calculations based on NFHS 1998-99.

Appendix I

This appendix explains the method of principal component analysis used in the study to derive the economic status index. Suppose we have p variables ($x_1, x_2, x_3 ... x_p$) in a dataset, which are too many to be used independently in a statistical model. Principal component analysis reduces these p variables into a smaller number, k ($p \geq k$), of *uncorrelated* variables. These k derived variables are called principal components.

The *first principal component,* y_1, is a linear combination of standardized values of the original p variables ($x_1, x_2, x_3 ... x_p$). Specifically,

$$y_1 = a_{11} \times \hat{x}_1 + a_{12} \times \hat{x}_2 + \text{L} + a_{1p} \times \hat{x}_p = \sum_{i=1}^{p} a_{1i} \times \hat{x}_i, \quad \text{(A1.1)}$$

where $\hat{x}_i = \left(\frac{x_i - \overline{x}_i}{\sigma_i} \right)$, $\overline{x}_i$ is the mean of x_i, and σ_i is the standard deviation of x_i. The elements of the weight vector $(a_{11}, a_{12}, ..., a_{1p})$ are chosen to maximize the variance of y_1 subject to the restriction that the sum of the squared weights is equal to 1; that is, $\sum_{i=1}^{p} a^2_{1i} = 1$.

The *second principal component,* y_2, is also a linear combination of the p standardized variables ($\hat{x}_1, \hat{x}_2, ..., \hat{x}_p$). So,

$$y_2 = a_{21}.\hat{x}_1 + a_{22}.\hat{x}_2 + \text{L} + a_{2p}.\hat{x}_p = \sum_{i=1}^{p} a_{2i}.\hat{x}_i. \quad \text{(A1.2)}$$

Now, the elements of the weight vector $(a_{21}, a_{22}, ..., a_{2p})$ are chosen to maximize the variance of y_2 subject to two conditions: the sum of the squared weights is equal to 1 (that is, $\sum_{i=1}^{p} a^2_{2i} = 1$) and y_2 is uncorrelated with y_1 (this requires $\sum_{i=1}^{p} \sum_{i=1}^{p} a_{1i} a_{2i} = 0$)

The first principal component, y_1, has the largest sum of squared correlations with the original p variables; the second principal component, y_2, has the second largest sum of squared correlations with the original p variables, and so on. Sum of the squared correlations with the original p variables get smaller as the subsequent principal components are extracted. Generally, the first few principal components account for the most variation present in the original variables.

It also turns out that the sum of the variances of the p principal components is always equal to the sum of the variances of the original p variables. In other words,

$$\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \sigma^2_i , \tag{A1.3}$$

where λ_i is the variance of the i'th principal component and σ^2_i is the variance of x_i. It can be easily verified that the proportion of variation in the original p variables accounted for by the first k principal components is $\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \sigma^2_i}$.

Appendix II

This appendix provides details of the structure of an ordered logit model and its estimation using maximum likelihood methods.

There are circumstances when the dependent variable, y, assumes discrete values which are ordered. For instance, y might be a rating on a scale from zero to five. For such dependent variables, an ordered logit model (conditional on regressors, x) can be derived from a latent variable model.

While the discrete random variable, y, is observed, assume that there is an unobserved underlying continuous latent variable, y_i^*, defined by the following regression relationship:

$$y_i^* = x_i\beta + u_i \,, \tag{A2.1}$$

where β is a k×1 vector of parameters, x_i is a 1×k random vector (regressors), and u is the error term that is logistically distributed with unit variance. Let $\alpha_1 < \alpha_2 < \alpha_3 < ... < \alpha_J$ be unknown cut points. The observed discrete dependent variable, y, takes values based on the following definitions

$$\begin{array}{lll} y=0 & \text{if} & y^* \le \alpha_0 \\ y=1 & \text{if} & \alpha_0 < y^* \le \alpha_1 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y=J & \text{if} & y^* > \alpha_{J-1} \,. \end{array}$$

If one defines $\alpha_{-1} = -\infty$ and $\alpha_J = \infty$, the conditional distribution of y given x is as follows:

$$P(y=j/x) = P(\alpha_{j-1} < y* \le \alpha_j / x) = \Lambda\left(\alpha_j - x_i\beta\right) - \Lambda\left(\alpha_{j-1} - x_i\beta\right), \tag{A2.2}$$

where $\Lambda(x) = \dfrac{1}{1+e^{-x}}$.

The parameters α and β can be estimated by the maximum likelihood method. The likelihood function of the model assuming observations are independent of each other is:

$$L = \prod_{j=0}^{J}\prod_{i=1}^{n}\left[\left(\Lambda(\alpha_j - x_i\beta) - \Lambda(\alpha_{j-1} - x_i\beta)\right)\right]^{Z_{ij}} \tag{A2.3}$$

Where $Z_{ij} = 1$ if y_i falls in the j'th category, and $Z_{ij} = 0$ otherwise $(i = 1,2,3,\text{L} \,, n;\ \text{j}=1,2,3,\text{L} \ \ \text{J})$.