University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

Jere R. Behrman
Barbara L. Wolfe

HUMAN CAPITAL INVESTMENTS IN WOMEN AND FERTILITY IN A DEVELOPING COUNTRY: EXTENSIONS TO INCLUDE HEALTH AND NUTRITION AND TO DEAL WITH INCOMPLETE DATA PROBLEMS

DP #610-80

Human Capital Investments in Women and Fertility in a Developing Country: Extensions to Include Health and Nutrition and to Deal with Incomplete Data Problems

Jere R. Behrman
Department of Economics and
Population Studies Center,
University of Pennsylvania

Barbara L. Wolfe
Department of Economics and
Preventive Medicine, and
Institute for Research on Poverty,
University of Wisconsin-Madison

May 1980

This paper is one of a series resulting from a survey and research project to investigate the social, economic, and demographic roles of women in the developing country of Nicaragua. Funding for 1977-1978 was provided by the Population and Development Policy Research Program sponsored jointly by the Ford and Rockefeller Foundations. Further funding has been provided by the Agency for International Development Negotiated Contract AID/otr-C-1571 for 1977-1980. Behrman is also a Guggenheim Fellow for the 1979-1980 academic year.

The project is being jointly conducted by the Universities of Pennsylvania and Wisconsin, the Centro de Investigaciones Sociales Nicaraguenese (CISNIC), and the Banco Central de Nicaragua. Humberto Belli, Director of CISNIC, and Antonio Ybarra, head of the Division of Social Studies and Infrastructure, Banco Central de Nicaragua, are coprincipal investigators, with Behrman and Wolfe, for the whole project. Belli supervised the collection of all survey data.

The authors would like to thank, but not implicate, the funding agencies, their coprincipal investigators, Bryan Boulier of Princeton University and Mark Rosensweig of the University of Minnesota, and associates in the project, especially David Blau, Kathleen Gustafson, Michael Watts, and Nancy Williamson. Behrman and Wolfe equally share major responsibility for this paper.

ABSTRACT

The economic model of fertility emphasizes two important roles of increased human capital investments in women on fertility: (1) an income effect which tends to increase the demand for children and (2) a price effect which tends to decrease the demand for children because of the higher opportunity cost of childcare in terms of foregone labor force opportunities. Previous empirical estimates of these effects have focused on human capital investment in the form of formal schooling, and have found that the price effect dominates.

However, in developing countries a broader definition of human capital may be appropriate because of widespread nutrition and health deficiencies. We estimate the Easterlin, Pollak and Wachter (EPW) extended model of fertility, which includes endogenous tastes and biological factors, with a broader representation of women's human capital that encompasses health, nutrition and work experience in addition to schooling.

We also explore the implication of data incompleteness in two senses: (1) selectivity in providing the necessary data and (2) incomplete fertility in that many women in our sample may have further children. Our estimates suggest that selectivity, which is ignored in previous studies, is a significant factor. They also suggest that the form of control for incomplete fertility is important, but that a multiplicative control for duration of exposure has the same implications as the more sophisticated Boulier and Rosensweig procedure.

In more substantive respects, we find that income has a positive effect and that increased predicted earnings for the women has a negative

effect with our broader definition of human capital. We also find that human capital investments in nutrition and health have additional effects, beyond those incorporated in the opportunity costs of labor market earnings. For nutrition the additional impact is to reinforce the opportunity cost dimension of more human capital by lowering fertility. But improved health increases family size, apparently because of increased fecundity, lengthened productive spans and increased energy, which provides support for the EPW emphasis in biological factors. Moreover our estimate for the health and nutrition effects of fertility are more robust than are those for the traditional schooling variable. Therefore we conclude that fertility analysis and family planning, health and nutrition policy analysis in similar developing countries should be expanded to incorporate a broader definition of human capital to include health and nutrition status.

1. INTRODUCTION

Many economists emphasize the important role of human capital investments in women in the determination of fertility. Such investments have an income effect by increasing women's earnings potential, which in itself leads to an increase in the demand for children and other normal goods. But they also may have a price effect due to the increased opportunity cost of the woman staying at home to care for the children and thereby foregoing higher market earnings that would be commanded by her greater capital stock. Empirical explorations of these possibilities focus on investment in education through formal schooling. Available estimates suggest that the opportunity cost element often dominates in developing countries, so that increased education for women lowers fertility.[1]

In this paper we extend the empirical application of this model to incorporate a broader applied definition of human capital which includes the woman's health and nutrition status in addition to her formal schooling. We do so because of the prevalence of health and nutrition problems in most developing countries. Moreover, elsewhere we present evidence that health and nutrition status significantly affect other important socio-economic outcomes for women in developing countries, such as the probability of their labor force participation and their level of earnings.[2]

We estimate fertility determination relations for a random sample of 1063 women in 1977 in Managua, the central metropolitan area of the Central American developing economy of Nicaragua. Although recently population growth rates in many developing countries have declined significantly, those in prerevolutionary Nicaragua remained very high at 3.3% per year, a rate exceeded in only 4 of the 92 developing countries included in World Bank (1978).

In order to estimate the model, we must face up to two senses in which the data are incomplete. First, for 83 of the 1063 women, values of critical variables are missing. To simply drop these women from the sample might result in selectivity bias of the type that Heckman (1974, 1976) and others have examined extensively recently in other contexts. Second, the women in our sample range in age from 15 to 45 years. Except for the oldest group, the data are incomplete in the sense that fertility may be incomplete. If we do not control for this sense of incompleteness, our estimates also may be biased. For example, if more educated women have children at older ages than do less educated women because the former are in school longer, the use of data from women with incomplete fertility may bias the estimated effect of education on fertility.

We organize our presentation as follows. In Section 2 we discuss our model of complete fertility. In Section 3 we discuss the missing values of some variables problem and the possibility of selectivity bias. In Section 4 we present our estimates of fertility determinants with control for data selectivity and with the <u>a priori</u> preferred corrections for incomplete fertility. In Section 5 we consider some alternative approaches for dealing with incomplete fertility. In Section 6 we summarize our conclusions about the two types of data incompleteness and about the extension of human capital concerns to include health and nutrition.

2. MODEL OF COMPLETE FERTILITY

We adopt the basic economic model of fertility that has evolved from the work of Becker (1960), Easterlin (1968, 1973), Leibenstein (1957), Willis (1973) and that has been extended recently by Easterlin, Pollak, and Wachter (EPW, 1980) to incorporate endogenous tastes and natural fertility.[3]

We take as given a number of characteristics of the adults in the household: age, age of first cohabitation (C), schooling (S), family background (F), migratory status (MIG), marital status (M), health and nutrition status (H), ability and motivation (A), and income other than earnings of the woman (Y). In our empirical work we concentrate on the characteristics of the woman because of our interest in the role of women and their critical role in household production, which ties in integrally with fertility determination. More extended models could incorporate some of these characteristics as endogenous variables, but we do not adopt such models here for three reasons.[4]

First, we think that for the most part the human capital variables of primary interest--schooling and health and nutrition status--reflect recursive decisions in which the parents of the adults in the current household had a major role. Elsewhere we present estimates of the determinants of these variables which are largely consistent with this hypothesis (Behrman and Wolfe, 1980 a,b,c,d,g, and Wolfe and Behrman, 1980). Even current household nutritional inputs for which this assumption probably is strongest, for example, reflect women's training and capabilities more than current economic conditions. To the extent that our assumptions are too strong in this respect, of course, our estimates may be subject to simultaneity bias.

Second, we wish to keep our model manageable so that we can deal with the two dimensions in which the data are incomplete. The extension to a larger model would make this difficult, if not impossible.

Third, we wish to maintain comparability with other estimates of fertility determinants. For the most part these estimates have been made under the assumption that such characteristics are determined recursively (see Williams, 1974).

We posit for each household a one-period utility (U) function which is defined over commodities (Z), completed family size (N), frequency of intercourse (q), infant mortality (d), and the intensity and use of contraceptives (u), all conditional on norms regarding the consumption of goods (Z*) and completed family size (N*):

$$U = U(Z, N, q, d, u; Z^*, N^*).$$

The first three arguments (Z, N, q) all have positive derivatives. If frequency of intercourse (q) did not have a positive derivative, abstinence would be costless and a perfectly effective form of contraception, so the actual number of children would be the number desired in a perfect contraceptive society. Infant mortality (d) has a negative derivative--otherwise infanticide would be a much more attractive contraceptive practice than it appears to be in reality. Likewise the use of less extreme contraceptive practices (u) apparently involves some disutility and therefore is included among the arguments in the utility function.

The norms regarding consumption of goods (Z*) and completed family size (N*) may be dependent on the exogenous variables which are mentioned above and on characteristics of peers (say, coworkers or neighbors).

We use both representations in our empirical work, but note that both have problems. For the personal and family background variables, it is very hard to identify whether the channel of any impact is through changing such norms or through affecting abilities and schooling and other such intervening variables (or through some mixture). For the peer variables a simultaneity problem is inherent since the behavior of the peers also depends on their norms, which in turn depend upon the behavior of the original household.

Household utility is maximized subject to six constraints:

(1) The commodity collection vector (Z) is producible within the production set defined by the identical household technology T, given the market goods collection vector (X), the time allocation vector (t), ability and motivation (A), and health and nutrition status (H):

$$(Z, X, t, A, H) \in T.$$

(2) The time (t_{js}) that each of J individuals devotes to each of S activities equals the total time that each individual has ($\bar{t}_j$):

$$\sum_{s=1}^{S} t_{js} = \bar{t}_j, \quad h = 1, \ldots, J.$$

However we posit that the time in the paid work force of all individuals in the household other than the woman is fixed exogenously by social customs (e.g., prime age males work, small children do not) and by work options (e.g., 40-hour weeks).

(3) The total expenditure on market goods (the n elements of X) that are used to produce commodities Z and on contraceptives (r (u)) is less than or equal to the sum of nonwage income (v) and market wage earnings (for which s = m) of all household members:

$$\sum_{k=1}^{n} p_k x_k + r(u) \leq v + \sum_{j=1}^{J} w_j t_{jm} = w_w t_{wm} + Y_{other}.$$

Given the exogenously determined time in market activities for all but the woman and the exogenously given market wages for everyone, the last sum is equal to the woman's earnings plus exogenous other income (Y_{other}).

(4) The household biological births (b) production function depends upon the frequency of intercourse (q), health and nutrition (H), the household's consumption of commodities (Z) and purchases of goods (X) through their impact on fecundity and the reproductive span, a vector of practices such as lactation (L) which affect the probability of conception given exposure through intercourse, age of first cohabitation (C), and the length and intensity of use of contraception (u):

$$b = B\ (q, H, Z, X, L, C, u).$$

(5) The household biological infant mortality (d) function depends on the population at risk (b), household consumption of commodities (Z) and purchases of goods (X), health and nutrition status (H), ability (A), and practices such as lactation (L) which affect infant mortality:

$$d = D\ (b, Z, X, H, A, L).$$

(6) Completed family size (N) is given by the difference between births (b) and deaths (d):

$$N = b - d.$$

Each household maximizes utility with respect to these six constraints. The solution is an optimal set of endogenous decision variables for each household (Z^o, X^o, t^o, b^o, d^o, N^o, q^o, L^o, u^o) as functions of the variables and functions that the household considers to be given: goods prices (p); wages rates (w); nonwage income (v); household technology, T; the birth function, B; the death function, D; the market cost function for fertility regulation, r; ability (A); health and nutrition status (H); schooling (S); age of first cohabitation (C); and the norms for the consumption of goods (Z*) and for completed family size (N*).[5]

We are interested in the determinants of fertility or completed family size (N^o). We assume that the underlying functions are sufficiently well-behaved so that we can solve for fertility as a function of the exogenous variables:

$$N^o = N^o \; (p, w, v, A, H, S, C; Z^*, N^*).$$

We do not have observations on market prices (p), but expect that excluding this vector does not cause serious omitted variable problems. We use Y_{other} to represent the impact of nonlabor income and earnings other than from the woman. *A priori* we expect that the derivative with respect to other income is positive.

We do not have direct observations on ability (A), but posit that it depends on human capital investments such as those in schooling (S) and in health and nutrition (H) and on family background (F). We represent an important dimension of the woman's ability by her predicted earnings (Y_{pred}) as estimated from schooling, health and nutrition status, and work experience.[6] In addition we include as separate factors her schooling (S), health and nutrition status (H), and family background (F). We do not have strong priors on the signs of the derivatives with respect to her predicted earnings (Y_{pred}), schooling (S), health and nutrition status (H), and family background (F) because of the possible opposing opportunity cost versus income effects of all four variables. All of these may be representing effects on birth production or on infant mortality through the fourth and fifth constraints, while at least schooling and family background may help to determine norms for the consumption of goods (Z*) and for completed family size (N*).

We have observations on the age of first cohabitation (C). We expect that the derivative with respect to this variable is negative since the younger the age, the longer the reproductive span and the larger the number of births, ceteris paribus.

We do not have direct observations on the norms for the consumption of goods (Z*) and for completed family size (N*). However we posit that these norms depend primarily on schooling (S), migratory status (MIG), marital status (M), and various family background (F) variables: presence of male and of female raisers during childhood, occupational status of male and of female raisers during childhood, and number of siblings. Because of the simultaneity problem, we limit our representation of the possible impact of peers on these norms to the inclusion of one variable which represents median neighborhood income. We generally do not have strong priors on the derivatives with respect to these variables, once again, because they may work indirectly through ability in ways that counter their direct impact through the norms. Nevertheless in some cases we expect that the direct effect dominates any indirect ones. For example we expect that the derivative with respect to number of siblings is positive both because of the dominance of the direct effect on norms for family size, and possible genetic effects on fecundity.

We should note that for completed family size (N*) we have observations on two variables that <u>prima facie</u> might seem to be directly related to the norms: the number of children that the respondent would have if she could begin again and the ideal Nicaraguan family size. We are cautious about using these variables as right-hand side regressors to represent such norms, however, because of frequent hypotheses about them being contaminated by the actual fertility experience of the respondent. But we do explore the nature of these variables in Section 5.

Our theoretical model and the availability of data, thus, lead us to the following relation for completed family size:

$$N^o = N^o\ (Y_{other},\ Y_{pred},\ H,\ S,\ C;\ Z*,\ N*),$$

where the norms are represented by the various background and neighborhood variables that we have discussed. We also include a stochastic term to represent random elements in individual behavior. The appropriate functional form depends upon all of the functions in the constraints and the utility function. For simplicity we assume that it can be approximated by a linear form.

3. MISSING VALUES OF SOME VARIABLES

We note in Section 1 that one sense in which our data are incomplete is that critical values of some fertility variables are missing for 83 of the 1063 women in our sample. We can write the relation that we wish to estimate for the i^{th} household as follows:

$$N_i^o = W_{1i}\ G_1 + V_{1i}, \qquad (1)$$

where N_i^o is completed fertility for the i^{th} household, W_{1i} is a vector of right-hand side variables which are discussed in Section 2, G_1 is a vector of coefficients to be estimated, and V_{1i} is the disturbance term for the i^{th} household.

However we do not observe the data that are necessary to estimate relation (1) for all 1063 households. There is a selection rule that gives the probability of providing data as a function of the respondents' schooling, other human capital variables, marital status, and background variables (W_{2i}):

$$R_i = W_{2i} G_2 + V_{2i}, \qquad (2)$$

where R_i is an indication of whether or not complete data are provided.

There are complete data for the i^{th} household if and only if R_i has a positive value. The selection rule for responding, therefore, is

$$R_i > 0 \quad \text{or} \quad V_{2i} > -W_{2i} G_2. \qquad (2a)$$

We are able to estimate regressions for fertility only for the households which satisfy this selection rule. In other words, for which

$$E \; (N_i^o \mid R_i > 0) = W_{1i} G_i + E \; (V_{1i} \mid R_i > 0). \qquad (1a)$$

Of course many studies of fertility determinants have data that are incomplete for some respondents. The general practice is to drop such observations from the sample and to estimate the fertility relations from the remaining subsample of complete observations. If the expectation of the disturbance term in the fertility relation conditional on the selection rule for responding [i.e., the last right-hand side term in relation (1a)] is zero, such a practice does not cause biases. If this conditional expectation is not zero, however, simply dropping incomplete observations may cause biases. Such a practice is equivalent to excluding the last term in relation (1a) and may cause a selectivity bias that is akin to omitted variable bias from excluding the conditional expectation.

To our knowledge, no one who has estimated fertility determinants for developing countries has controlled for such selectivity bias. Yet it seems to us that selectivity bias in regard to providing data is a real possibility. A priori we think that women who are formally married, who have more education and other forms of human capital, and who come from better economic backgrounds are more likely to provide data. Therefore simply dropping women from the sample for whom data are not complete is not likely to be random with respect to the fertility relation.

To deal with this aspect of incompleteness in our data, we apply the Heckman (1976) procedure for selectivity problems.[7] First we estimate a probit relation for the probability that we have complete fertility observations for the i^{th} household. From that probit estimate we calculate the inverse of the Mill's ratio (λ_i) which can be used to control for the selection decision about reporting fertility data since relation (1a) can be rewritten as

$$E\left(N_i^o \mid R_i > 0\right) = W_{1i}\, G_i + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}}\, \lambda i. \tag{1b}$$

We include the estimate of the inverse of the Mill's ratio as a regressor in an ordinary least squares estimate of relation (1b) under the assumption that there also is an additive conditional disturbance term with desirable properties.

Table 1 gives our maximum likelihood estimates for the probit relation pertaining to whether or not a woman responds completely to the fertility-related questions. The overall relation is significantly nonzero at standard levels, so the availability of data apparently is related to the included characteristics of the woman, her household, and her background. The significantly nonzero coefficient estimates suggest that women in higher income households who are older, who have not had parasitic diseases, and who have cohabited previously are more likely to respond fully. The age and previously cohabited variables may suggest that those who have completed their fertility are more likely to respond. The coefficient estimates also are almost at the margin of standard significance levels for negative effects of having had therapeutically treatable diseases, the socioeconomic status of the female raiser, having always lived in Managua, and being currently single or in a civil marriage (the left-out categories).

Perhaps somewhat surprising, schooling and predicted earnings, and almost all of the family background and neighborhood characteristics are clearly not significant. Therefore women with more education, greater ability (as represented by higher predicted earnings), or better family background in terms of stability or socioeconomic status are no more (or less) likely to respond than others, ceteris paribus. The only channel through which human capital variables for the women seem to work is associated with avoiding diseases, particularly of a parasitic nature.

We also are somewhat surprised that those in common law arrangements or who have previously cohabited are no less (and quite possibly more) likely to respond than are those in religious or civil marriages or without previous cohabitation. Apparently the stigma, if any, associated with less formal living arrangements and with having had previous living companions does not lead to a reluctance to respond to questions about fertility.

We include the inverse of the Mill's ratio as estimated from this probit in our estimates below to control for selectivity. We obtain significant coefficient estimates for this variable in our preferred relations and in a number of alternatives. Therefore we conclude that selectivity bias from dropping incomplete observations may be a significant problem in many fertility studies.

4. OUR PREFERRED FERTILITY ESTIMATES

Our model in Section 2 is of the determinants of complete family size or fertility. However, as we note in Section 1, most of the women in our sample are young enough that they may have further children, so their fertility is not complete.

Of course this is a common problem in fertility studies. A frequent resolution is to subdivide samples by age cohorts, and sometimes to focus only on the age cohorts for whom fertility probably is basically complete. We do not find such a strategy very attractive for our sample of women in Managua because the resulting subsample would be quite small.

Instead we concentrate on alternative approaches for which we can use our entire sample. In our data set we have four fertility-related measures: current number of living children, expected number, number if began again, and ideal Nicaraguan family size. Ostensibly, the expected number best measures the desired construct of completed fertility. However it is widely believed that this measure is contaminated by actual fertility experience to date.

Therefore, we have more confidence in working with the actual current number of living children and correcting for exposure and relative position in the life cycle. Our preferred correction is the Boulier and

Rosensweig (1978) normalization of the current number of children by international standards for fertility patterns that are conditional on age of first cohabitation and on duration since first cohabitation. These standards reflect the average impact of biological and cultural-economic factors across a number of societies. We use this procedure for our preferred estimates, which we discuss in this section. In the next section we explore the sensitivity of our conclusions to the choice of this particular procedure for controlling for incomplete fertility by considering what estimates result if there is no control, if alternative controls are used, or if our other fertility-based measures are used.

We now turn to our preferred estimates. Column one in Table 2 gives coefficient estimates for our fertility model with this dependent variable.[8] We discuss these estimates with reference to the variable groups that are suggested by our model in Section 2.

Woman's Human Capital

We are primarily interested in the coefficient estimates of those variables and their implications regarding the opportunity cost versus the income effect of various human capital investments. Our estimates suggest that higher women's predicted earnings (higher opportunity costs) reduce fertility. Therefore those human capital factors that increase earnings indirectly reduce fertility. Thus, investments in women's health, nutrition, on-the-job training and schooling reduce fertility by increasing opportunity costs in terms of predicted earnings from paid labor force participation.[9] However there are additional significant effects for nutrition and health (but not for schooling) beyond those captured by predicted earnings. For nutrition these reinforce the dominance of opportunity costs and the reduction of fertility with an improved nutrition state. For the health variables, in contrast, greater prevention of therapeutically treatable and parasitic diseases directly increases fertility through reducing subfecundity and raising overall energy levels. The direct effects of those health variables, moreover, outweigh the opposing indirect effects which are captured by the predicted earnings variables.

Therefore, our estimates for coefficients of human capital variables suggest five important conclusions. First, a broader definition of human capital than just schooling is useful since nutrition and health status and on-the-job training all also have significant direct or indirect effects on fertility. Second, schooling works significantly only through the woman's predicted earnings, but the nutrition and health variables have additional direct significant impact. Thus, we do not have evidence that schooling significantly changes costs towards children once we control for its impact on opportunity costs. Third, for schooling, on-the-job training, and nutrition, the opportunity cost dimension dominates, so more investment reduces fertility. Fourth, for health, the income dimension dominates, so better health increases fertility. Fifth, increases in women's paid labor force opportunity costs through means other than human capital investments, such as by lessening sexual discrimination (see Behrman, Wolfe, and Tunali, 1979), are likely to lessen fertility.

Other Income

Our estimates suggest that other income significantly increases fertility. Thus, children are normal goods, with more being demanded as development occurs, ceteris paribus.

Woman's Background

We posit that the background variables may affect norms for family size. However we find evidence of significant effects for only two of our background variables.

Women who migrated from other areas to Managua have an average of 0.1 more children per year of exposure than do those who always have resided in Managua. One explanation for this result is that higher norms for family size prevail in the more traditional small cities and towns and rural areas than in Managua, and these immigrants into Managua tend to have higher norms than women who always have resided in Managua. A second explanation is that opportunity costs for having children are lower outside of Managua (see Behrman, Wolfe, and Blau, 1980), and immigrants have not adjusted completely to the higher opportunity costs.

The other significantly nonzero coefficient estimate is the negative one for the occupational status of the mother (or other female raiser) of the woman. We interpret this estimate to mean that family size norms of a woman have been affected inversely by the impact of the opportunity cost of her mother. This is an interesting intergenerational woman's opportunity cost impact on fertility. But we do not find evidence of an impact of the number of children that a woman's mother actually had (i.e., number of the woman's siblings), once we control for her mother's occupational status.

Marital Status

The coefficient estimates for the marital status variables basically suggest that women with greater exposure to conception tend to have more children. Such a result is consistent with the EPW emphasis on natural fertility, but also with a model in which women who have greater tastes for children are more likely to have been accompanied at the time of our survey. On an a priori basis we thought that the type of accompaniment might make a difference since women in a religious marriage might have more security and might be more children oriented than women in common law arrangements. However our estimates imply no significant effect of this difference in the institutional arrangements of the accompaniment on fertility.

Age of First Cohabitation

In our model of complete fertility in Section 2 we include the age of first cohabitation to reflect the extent of exposure to intercourse. Of course the Boulier and Rosensweig (1978) procedure purports to control for age of first cohabitation, so it might be reasonable to exclude this variable for cases in which we use their procedure. In fact we have estimated our model both with and without this variable. We report the estimates with it because the coefficient estimates of the other variables hardly change when it is introduced, and its coefficient estimate is of some interest.

If the Boulier and Rosensweig procedure corrected exactly for the age of first cohabitation, we would expect to find an insignificant coefficient estimate for this variable. That we obtain a significantly positive estimate of this coefficient suggests that the Boulier and Rosensweig normalization by international standards overcorrects for the effects of the age of first cohabitation. However, as we note above, the impact on other coefficient estimates is not substantial.

Response Selectivity

The coefficient estimate of the inverse of the Mill's ratio is highly significant. This means that selectivity into the sample is not random, but is determined significantly by the decision rule that we estimate in the previous section. The failure to incorporate the decision rule into the estimator, therefore, might cause selectivity bias in the estimated coefficients of our fertility model.

5. ALTERNATIVE TREATMENTS OF THE INCOMPLETE FERTILITY PROBLEMS

In this section we consider some alternatives to the Boulier and Rosensweig (1978) normalization by international standards for the control of incomplete fertility. Table 3 gives the means, standard deviations, and correlations among our preferred Boulier and Rosensweig variable and eight others. These eight include each of the four fertility-related measures which we mention at the start of the previous section and each of the same four variables normalized by duration of exposure (i.e., years since first cohabitation). In the regression context the latter normalization is equivalent to a multiplicative control since every term on the right-hand side effectively is multiplied by duration. A common alternative is an additive control for duration, which we also consider in our regression estimates.

Before we turn to some regression estimates, we consider the correlations in Table 3. The most striking aspect of these statistics is the low correlation between the Boulier and Rosensweig alternative and seven of the other eight measures. Only the correlation with live children/duration at 0.47 is higher than 0.20. This suggests that if the Boulier-Rosensweig alternative is a good control for incomplete fertility, most of the other measures (with the possible exception noted above) probably are not. Therefore estimates based on them may be quite misleading regarding the impact on fertility of human capital investments in women.

A related point is that the expected number of children, which supposedly might represent best completed fertility, is very highly correlated with the current number of live children (0.90) and hardly correlated at all with the Boulier and Rosensweig measure (0.09). Since many of the women in our sample are early in their childbearing years while others are near the end, we think that this pattern reflects the great inadequacy of the expected number as a proxy for complete fertility. Because the expected number is so contaminated by experience to date, it becomes a good proxy for complete fertility only near the end of the childbearing years--at which time, of course, it is not very necessary since actual number of children also almost approximates completed fertility.

We also note that most of the other relatively higher correlation coefficients are between one of the original four fertility-related measures and the same measure normalized by duration. Such a pattern is hardly surprising. What is of interest, however, is that this pattern does not hold for current live children. The correlation between current live children and live children/duration is only .13. In this case, apparently controlling for duration leads to a measure with much different variation than the actual number of children, as we would expect if such a control leads to a good proxy for completed fertility in a sample with a wide range of ages.

We now discuss regression results with various dependent variables from the eight alternative measures and with and without additive controls for duration. To present and to discuss in detail all of the permutations of estimates which we have obtained would be very long and tedious, so we only focus on the following important summary findings.

No control for incomplete fertility may contribute to quite misleading estimates. In column two of Table 2 we give the estimates with current live children as the dependent variable and with no control for duration. The contrast with our preferred estimates in column one is sharp. The signs of the estimated coefficients of all of the woman's human capital variables are reversed (except for nutrition), as is the sign for other income. The implications that fertility increases as women's opportunity costs rise and that children are an inferior good with respect to other income are not credible on *a priori* grounds nor consistent with most available estimates. They illustrate well the pitfalls of not controlling for incomplete fertility.

Control for incomplete fertility by normalization by duration works well, but additive control for duration does not. In columns three and four of Table 2 we present estimates with current live children as the dependent variables and with control by adding duration and by normalizing by duration, respectively.

The estimates with control by normalization by duration are quite similar to our preferred estimates in column one. The signs of the significant coefficients are the same and the implied elasticities are almost identical.[10] Therefore this alternative leads to the same implications as does the Boulier and Rosensweig procedure. Although it is a priori a somewhat less elegant way of controlling for incomplete fertility, if our finding that it leads to the same results is robust, it may be preferable since it is somewhat easier to implement.

The estimates with control by adding duration as a right-hand side variable, however, differ in some important respects from our preferred estimates. Some of the signs of important coefficients are different and/or insignificant (e.g., other income, woman's predicted income).

For our other measures, the patterns are similar to those that we obtain with dependent variables based on current live number of children. The estimates for dependent variables expected number, if could begin again, and ideal Nicaraguan family size differ sharply from our preferred ones if there is no control for incomplete fertility. If there is additive control for duration, they still differ in some important respects. If there is normalization by duration, they are very similar in sign and in implication to the preferred estimates. The most noteworthy difference is that the coefficient estimate of schooling indicates a significantly positive impact on fertility (instead of the insignificant positive one in column one), once there is control for the negative effect of schooling (larger in absolute value) through the woman's opportunity costs in terms of her predicted earnings. But much more important than such differences are the quite striking similarities in the pattern of coefficient estimates, despite the generally low correlations among the dependent variables in Table 3. This robustness reinforces our confidence in our preferred results.

Response selectivity quite generally may be a problem. For most (although not all--see column three in Table 2 for an exception) of our regressions, the estimated coefficients of the response selectivity terms are definitely significantly nonzero. This pattern, once again, reinforces our conclusion that the common practice of simply dropping incomplete observations may lead to selectivity bias in the coefficient estimates.

6. CONCLUSIONS

We consider two senses in which our data are incomplete which may affect our estimated impact of human capital investments in women on their family size.

The first sense of data incompleteness relates to selectivity into the sample. We find evidence that older women from higher income households with better health (but who have changed male companions) are more likely to provide the necessary fertility-related data. The significance of such factors in determining whether or not we have fertility data reinforces the possibility of selectivity bias in the fertility relations because *a priori* we would expect to find some of these same variables to be important in fertility determination. Our estimation of our fertility model with the incorporation of such a selectivity possibility generally leads to significant coefficients for the selectivity terms. Therefore the common practice of simply dropping incomplete observations before estimating fertility relations may lead to selectivity biases in the coefficient estimates.

The second sense of incompleteness pertains to the use of data relating to incomplete fertility to estimate the determinants of completed fertility. On a priori grounds we prefer the Boulier and Rosensweig (1978) normalization of current number of children by international standards conditional on age and duration of exposure to control for incomplete fertility. But we also explore a number of alternatives. From this investigation we conclude that the failure to control adequately for incomplete fertility may contribute to very misleading results. We also conclude that for our sample, normalization by duration since first exposure gives basically the same estimates as does the Boulier and Rosensweig procedure. If this finding is robust across samples, control for incomplete fertility by normalization for duration may be preferrable to the Boulier and Rosensweig procedure since it is easier.

We now turn to substantive conclusions about the determinants for fertility for women in Managua. We do not find much statistical support for the EPW extension of the fertility model to include endogenous tastes. We do find support for important roles of variables that traditionally have been emphasized by economists. Other income has a significantly positive effect, which means that family sizes would increase with development if child costs were constant. However child costs are not likely to be constant because of changing market prices and opportunity costs. For example, our estimates imply that increased earnings possibilities for women raise the opportunity costs of having children and reduce significantly fertility.

Of course this is a standard result in regard to the impact of schooling for women on fertility. But we emphasize that it holds for a broader than usual definition of human capital which includes health and nutrition and on-the-job training in addition to schooling. Moreover it implies that reduced sexual discrimination in the labor market also would induce lower fertility.

For human capital investments in health and nutrition, furthermore, our estimates suggest the existence of additional effects beyond that represented by earnings from labor force participation. For nutrition the additional impact is to reinforce the opportunity cost dimension of more human capital by lowering fertility. On the other hand improved health on net increases completed family size, apparently because of increased fecundity, lengthened reproductive spans, and increased energy. For our sample at least, these human capital effects are more robust than are those for the investments in schooling that usually are emphasized. Biological human capital may be more important than mental human capital. These results support the EPW extension of the fertility model to include biological factors. Since many policies in developing countries are directed towards changing health and nutrition status without reference to possible effects on fertility, we think that further exploration of the link between such human capital investments in health and nutrition may have high payoffs in terms of better policy design.

Table 1 Probit Estimates for Complete Responses for Fertility Data from 1063 Women in Managua, 1977

Variables	Coefficient Estimate	Estimated Coefficient / Estimated Standard Deviation
1. Woman's Human Capital		
Woman's Predicted Earnings	-.68	-0.8
Schooling	.021	0.6
Nutrition (Protein)	-.15	-0.8
Had Therapeutically Treatable Diseases	-.18	-1.3
Had Parasitic Diseases	-.29	-2.4
2. Other Income	.27	2.3
3. Background		
Male Raiser Present	.026	0.1
Female Raiser Present	.265	0.7
Male Raiser Occupational Status	-.068	0.1
Female Raiser Occupational Status	-1.09	-1.3
Number of Siblings	-.007	-0.3
Always in Managua	-.186	-1.5
Monthly Church Attendance	-.017	0.7
Neighborhood Median Income	.000	0.1
Neighborhood Population Density	-.001	0.4
4. Marital Status		
Religious Marriage	.27	1.5
Common Law Marriage	.27	1.6
Previous Cohabitation	.38	1.9
5. Age	.032	3.4
6. Constant	.78	1.5

-2 Log Likelihood Ratio = 40.7

Table 2. Fertility Model Regressions[a]

Right-Hand Variables	Current Children International Standards (Boulier & Rosenzweig) (1)	Current Children (2)	Current Children (3)	Current Children Duration (4)
1. Woman's Human Capital				
Woman's Predicted Earnings	-.61 (2.3)	4.39 (5.7)	-.66 (0.8)	-.32 (2.7)
Schooling	.02 (1.6)	-.20 (6.8)	-.01 (0.3)	.01 (2.2)
Nutrition (Protein)	-.17 (3.3)	-.25 (1.7)	-.31 (2.1)	-.08 (3.3)
Had Therapeutically Treatable Diseases	-.11 (2.2)	.66 (4.9)	-.26 (1.8)	-.04 (2.1)
Had Parasitic Diseases	-.27 (5.9)	1.59 (11.9)	-.03 (0.2)	-.13 (6.2)
2. Other Income	.09 (2.6)	-.12 (11.8)	-.17 (1.4)	.05 (3.2)
3. Background				
Male Raiser Present	.14 (1.5)	-.23 (0.9)	-.26 (1.1)	.05 (1.2)
Female Raiser Present	.20 (1.5)	-2.05 (5.3)	-.29 (0.8)	.10 (1.7)
Male Raiser Socio-Economic Status	-.00 (1.0)	.01 (0.9)	.01 (1.2)	-.00 (0.7)
Female Raiser Socio-Economic Status	-.01 (2.3)	.07 (7.7)	.01 (0.8)	-.00 (2.4)
Number of Siblings	.00 (0.2)	.05 (2.6)	.03 (1.6)	.00 (0.5)
Always in Managua	-.10 (2.4)	.83 (6.7)	-.07 (0.5)	-.05 (2.6)
Neighborhood Median Income	-.00 (1.0)	-.01 (1.9)	-.00 (1.1)	-.00 (1.0)
4. Marital Status				
Religious Marriage	.19 (2.8)	-1.15 (6.0)	.14 (0.7)	.09 (2.9)
Common Law Marriage	.17 (2.8)	-1.63 (9.4)	-.08 (0.4)	.08 (2.9)
Previous Cohabitation	.13 (1.7)	-2.52 (11.7)	-.76 (3.1)	.06 (1.8)
5. Age of First Cohabitation and Duration				
Age of First Cohabitation	.04 (7.1)	-.19 (10.4)	.00 (0.1)	.02 (6.8)
Duration			.21 (12.9)	
6. Response Selectivity	2.72 (7.1)	-22.85 (20.8)	-.13 (0.1)	1.32 (7.7)
Constant	-.17 (0.9)	10.50 (20.3)	1.8 (2.1)	-.10 (1.2)
$\bar{R}^2$	.09	.40	.49	.10
Standard Error	.59	1.69	1.57	.26

[a]Absolute value of t statistics are beneath present estimates in parentheses. 980 observations are used in each regression. Column one gives the preferred estimates (see text).

Table 3 Means, Standard Deviations, and Correlations Among Alternative Fertility Measures for 980 Women in Managua, 1977

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Live Children/International Standard (Boulier and Rosensweig)	(1)	.47	.13	1.0						
Current Live Children	(2)	1.0								
Live Children/Duration	(3)	.12	1.0							
Expected Number	(4)	.09	.90	.13	1.0					
Expected Number/Duration	(5)	.16	-.27	.60	-.06	1.0				
If Could Begin Again	(6)	-.06	.13	-.04	.23	-.06	1.0			
If Could Begin Again/Duration	(7)	.03	-.43	.36	-.25	.86	.22	1.0		
Ideal Nicaraguan Family Size	(8)	-.02	.19	.02	.33	.01	.38	-.01	1.0	
Ideal Nicaraguan Family Size/ Duration	(9)	.06	-.45	.44	-.29	.87	-.00	.89	.09	1.0
Mean		.78	3.2	.35	3.8	.54	2.3	.40	3.0	.52
Standard Deviation		1.1	2.2	.28	2.0	.57	1.1	.52	1.0	.61

Notes

[1]Williams (1974) surveys many of these estimates.

[2]See Behrman, Wolfe, and Tunali (1979), Behrman and Wolfe (1979, 1980 a,c,d,e,f,h,i), and Behrman, Wolfe, and Blau (1980).

[3]Another approach that attempts to incorporate biological considerations into the analysis of fertility is the use of "renewal models" that seek to account for fertility through factors such as age at sexual union, frequency of intercourse, probability of conception, length of the nonsusceptible period, and duration of reproductive union. See Crafts and Ireland (1976), David and Sanderson (1976), Keyfitz (1971), Leriodon (1976) and Michael and Willis (1976). EPW claim to be concerned with the more basic variables that underlie the intermediate variables in these renewal models.

[4]For an exploration of the determinants of schooling, age of first cohabitation, household demand for nutrition inputs, migratory status, wages, and health utilization, see Behrman and Wolfe (1980 a,b,c,d,e,g,h,i), Behrman, Wolfe, and Tunali (1979), Blau (1980), and Wolfe and Behrman (1980).

[5]EPW argue that different subsets of these relations apply to different households, depending upon whether fertility is determined entirely by the biological birth production function at one extreme, entirely by preferences and related constraints in a "perfect contraceptive" manner at the other, or by some intermediate combination. In this paper we assume that all households are in the intermediate categories. In Behrman and Wolfe (1980f) we explore the relevance of the EPW categories for a broader nationwide sample that includes many women from rural and small-town areas for whom the biological birth

production function *a priori* might seem to be much more relevant than for our present sample of Managuan women.

[6]The underlying estimates control for labor force participation due to the presence of small children, so there is not a selection bias nor a simultaneity problem. See Behrman, Wolfe, and Tunali (1979).

[7]An alternative approach would be to use a nonlinear maximization procedure to estimate the parameters that maximize the likelihood of obtaining complete fertility observations for the subsample of 980 women and incomplete observations for the other 83. We choose the Heckman procedure, however, because it is computationally easier and still consistent, although less efficient than the maximum likelihood alternative.

[8]Although children come only in positive integer values, we use ordinary least squares procedures and do not worry about this aspect of the dependent variables in our regressions. We do so because the range is fairly large, because we do not want to further complicate our estimation technique, and because this gives us more comparability with other studies.

[9]Once again, the predicted earnings are based on the estimates in Behrman, Wolfe, and Tunali (1979).

[10]The absolute magnitudes of the estimates in column four are about half of those in column one, but the implied elasticities at the point of means are about the same since the mean for the dependent variable for column one is about twice that for the dependent variable in column four (see Table 3).

REFERENCES

Becker, Gary S. 1960. An economic analysis of fertility. Demographic and Economic Changes in Developed Countries. Princeton: National Bureau of Economic Research.

Behrman, Jere R. and Barbara L. Wolfe. 1979. The impact of health and nutrition on the number of surviving children in a developing metropolis. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980a. Child health and nutrition determinants in a developing country. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980b. Determinants of health utilization in a developing country. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980c. The determinants of schooling for children in developing countries: Family background, number of siblings, sex, birth-order, and residence. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980d. Important early life cycle socioeconomic decisions for women in a developing country: Years of schooling, age of first cohabitation, and early labor force participation. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980e. Internal migration of women in a developing country. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980f. A more general approach to fertility determination: Endogenous preferences and natural fertility in a developing country. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980g. Parental preferences and intrafamilial allocations of human capital investments in schooling and health in a developing country. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980h. The returns to schooling in terms of adult health, occupational status, and earnings in a developing country: Omitted variable bias and latent variable-variance components estimates. Philadelphia: University of Pennsylvania, mimeo.

__________. 1980i. Wage rates for adult family farm workers in a developing country and human capital investments in health and schooling. Philadelphia: University of Pennsylvania, mimeo.

Behrman, Jere R., Barbara L. Wolfe, and David Blau. 1980. The impact of changing population composition on the distribution of income in a developing country. Madison: University of Wisconsin, mimeo.

Behrman, Jere R., Barbara L. Wolfe, and Insan Tunali. 1979. Determinants of women's earnings in a developing country: A double selectivity, extended human capital approach. Institute for Research on Poverty Discussion Paper #596-80. Madison: University of Wisconsin.

Blau, David. 1977. On the relation between child malnutrition and economic growth in less developed countries. Madison: University of Wisconsin, mimeo.

__________. 1980. Nutrition, fertility and labor supply in developing countries: An economic analysis. Ph.D. dissertation. Madison: University of Wisconsin.

Boulier, Bryan and Mark R. Rosensweig. 1978. Age, biological factors and socioeconomic determinants of fertility; A new measure of cumulative fertility for use in the empirical analysis of family size. Demography (November), 487-498.

Crafts, N.F.R. and N.J. Ireland. 1976. Family limitation and the English demographic revolution; A simulation approach. Journal of Economic History (September), 598-623.

David, Paul A. and Warren C. Sanderson. 1976. Contraceptive technology and family limiting behavior: Toward a quantitative history of the diffusion of contraceptive practices in America, 1850-1920. Stanford: Stanford University, unpublished manuscript.

Easterlin, Richard A. 1968. Population, Labor Force and Long Swings in Economic Growth: The American Experience. New York: National Bureau of Economic Research.

__________. 1973. Relative economic status and the American fertility swing. In Eleanor B. Sheldon, ed., Family Economic Behavior: Problems and Prospects. Philadelphia: Lippincott.

Easterlin, Richard A., Robert A. Pollak, and Michael L. Wachter (EPW). 1980. Towards a more general economic model of fertility determination: Endogenous preferences and natural fertility. In Richard A. Easterlin, ed., Population and Economic Change in Less Developed Countries. Chicago: National Bureau of Economic Research and University of Chicago Press.

Heckman, James. 1974. Shadow prices, market wages, and labor supply. Econometrica 42:2 (July), 670-694.

__________. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5:4 (Fall), 475-492.

Keyfitz, Natha. 1971. How birth control affects births. Social Biology (June), 109-121.

Leibenstein, Harvey. 1957. Economic Backwardness and Economic Growth. New York: John Wiley and Sons.

Leriodon, Henri. 1976. Human fertility: The basic components (Judith F. Helzner, trans.). Unpublished manuscript.

Michael, Robert T. and Robert J. Willis. 1976. Contraception and fertility: Household production under uncertainty. In Conference on Research in Income and Wealth, Household Production and Consumption, Studies in Income and Wealth (Vol. 40). New York: National Bureau of Economic Research.

Williams, Anne D. 1974. Effects of Economic Development on Fertility: Review and Evaluation of the Literature. Washington: TEMPO.

Willis, Robert J. 1973. A new approach to the economic theory of fertility behavior. Journal of Political Economy 81:2 (Part 2, March/April), 14-64.

NOTES

[1]We will assume that all households face the same prices; quantity measures are transformed to set prices equal to 1.

[2]In the case H household characteristics are distinguished, one vector δ_i of length H should be prespecified.

[3]For $P_i = 1$, $i = 1,N$, the intercept for a couple equals $\alpha_o + \Sigma\frac{\alpha_i}{1 + J_i}$.

[4]The factors $2\alpha_o$, $2\alpha_1$, etc. in LM's equation (9) are probably a mistake. See LM equation (5').

[5]Stemming from an augmented Linear Expenditure System estimated by Abbott and Ashenfelder (1976). LM treat these elasticities as constants.

[6]These expenditures and income data are estimates based on "reduced-form" expenditure functions (income is not a variable in these functions) and income functions based on individual personal characteristics.

[7]See Lluch et al., 1977, for the additional assumptions needed to derive at this result.

[8]Since the scales turned out to be about the same for a large range of values of U, only one scale is given. More detailed information in a paper discussing the equivalence scale derived is available from the author.