University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

Jere R. Behrman
Barbara L. Wolfe
F. Insan Tunali

DETERMINANTS OF WOMEN'S EARNINGS IN A DEVELOPING COUNTRY: A DOUBLE SELECTIVITY EXTENDED HUMAN CAPITAL APPROACH

DP #596-80

Determinants of Women's Earnings in a Developing Country:

A Double Selectivity Extended Human Capital Approach

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

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

F. Insan Tunali
Department of Applied Statistics
Middle East Technical University
Ankara, Turkey
Presently Ph.D. Candidate in the Department of Economics,
University of Wisconsin-Madison

Revised February, 1981

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, colleagues--in particular David Crawford at Pennsylvania and Arthur Goldberger at Wisconsin--and associates in the project, especially David Blau, John Flesher, Kathleen Gustafson, Michael Watts, and Nancy Williamson. Behrman and Wolfe equally share major responsibility for this paper. Tunali resolved the double selection problem and is responsible for the content of the appendix.

ABSTRACT

The determinants of labor force participation and earnings among women in a developing country are explored. A double selectivity procedure is developed and used to deal with the possible selectivity problems of (1) who selects into the labor force and (2) who reports earnings. A broad definition of human capital which includes health and nutrition in addition to education and experience is used. Sexual discrimination is investigated by comparing returns to men and women. Analysis is extended to the pluralistic nature of the labor market by dividing it into sectors--formal, informal, domestic--to analyze selection into sectors and compare returns to human capital factors across sectors.

Determinants of Women's Earnings in a Developing Country: A Double Selectivity Extended Human Capital Approach

In this paper we study the determinants of earnings for adult women in an urban area of a developing country, and make a number of contributions. We add to the very limited evidence about the impact of selectivity bias in estimates for developing countries,[1] by considering not only selective labor force participation, but also selective report of earnings. We use a broader definition of human capital than often is the case by including health and nutrition in addition to education and experience. We investigate the possibility of sexual discrimination. We consider labor market segmentation, which is widely hypothesized to be a critical feature of developing countries. Finally, we investigate the impact of varying family statuses and child care responsibilities on the shadow wage of women, given that these responsibilities are more frequently fulfilled by adults from extended families and by older children, and given the greater possibilities for on-the-job child care in the informal sector of the labor market than in more developed economies.

This study is part of a large, multiyear, international, interdisciplinary project, the purpose of which is to gain better understanding of the social, demographic, and economic role of women in developing countries. The primary data base is a random sample of 4104 women in the developing Central American country of Nicaragua; they were interviewed in 1977-1978. This sample is one of the few available for women in developing countries that includes current and retrospective integrated economic, demographic, and sociological information for all degrees of urbanization in a country.[2] It also includes data on a subsample of matched sisters, which enables us to control for childhood and adolescent background better than has been possible heretofore for large socioeconomic samples from developing countries.[3]

In the present study we focus on 1247 women who reside in Managua. Managua is the capital, and the most commercialized part of the country; its 500,000 inhabitants constitute almost a quarter of the country's total population. To explore possible sexual discrimination we also consider 643 men who live there.

We utilize a statistical model that extends Heckman's [24] treatment of selectivity, and relies on the formulation of the choice process as a trichotomy, with selection made sequentially. It enables us to resort to a computationally tractable, consistent estimation procedure which reduces the problem to the level of simplicity of the single selection treatment, involving univariate probit analysis and linear regression. A detailed discussion of the statistical model and of some alternative estimation procedures can be found in Tunali, Behrman and Wolfe [34].

We begin by laying out the model, paying particular attention to the treatment of selectivity. Next, we present empirical results of the selectivity equation. In Section 2 we discuss empirical results with additional human capital determinants in the earnings equation. Sexual discrimination is discussed in Section 3. Pluralism as the breakdown into segmented labor markets is the focus in Section 4. Family status and child care are discussed in Section 5. Conclusions follow. The essential elements of the statistical methodology are included in the appendix.

1. DOUBLE SELECTIVITY

We begin with a standard model in which ln earnings depend on formal education and linear and quadratic terms in experience. We note that the employment conditions in urban areas in Latin America such as the one from which our sample is drawn apparently satisfy at least one of the assumptions of most models of labor force supply better than do the conditions in labor markets in the United States: that hours worked can be adjusted to equate the market wage and the shadow wage (e.g., Heckman [26]). Casual empiricism suggests that there is much more flexibility in hours of employment in the labor markets that we study than is the case for most samples used from the United States and other developed economies.

Regression 1 in Table 1 is the OLS estimate of this basic ln earnings function for the 535 women in our sample who participated in the labor force and reported earnings. Under the necessary assumptions for such an interpretation,[4] the estimates imply a fairly high return to women's education--13%--and a significantly nonzero linear return to experience.

However, the estimates in this regression may suffer from two types of selectivity bias. First, there is the frequently analyzed question of labor force participation, or "work inclination." Of the 1247 women in our sample, only 579 participated in the labor force. A second possible selectivity problem that is generally ignored has to

Table 1

Various Earnings Functions for Men and Women for Managua, 1977

Sample	Education	Experience	Experience²	Protein	Days Ill	Always in Managua	Selection Variables: Labor Force Participation	Selection Variables: Reporting Earnings	Constant	R^2 / Sample Size
1. Women	.13 (12.8)	.04 (2.4)	-.000 (0.8)						4.58 (45.4)	.23 535
2. Women	.15 (13.1)	.09 (4.5)	-.002 (2.9)				.52 (3.9)	.45 (0.9)	3.77 (15.8)	.26 535
3. Women	.14 (12.9)	.09 (4.5)	-.002 (2.9)	.19 (2.1)	-.003 (1.9)	.26 (3.7)	.54 (3.8)	-.04 (.2)	3.51 (11.7)	.27 535
4. Men	.13 (8.9)	.09 (3.9)	-.002 (3.5)	.31 (3.9)	-.003 (1.4)		2.65 (2.6)	14.03 (1.9)	3.94 (10.0)	.38 600
5. Women and Men	.15 (18.3)	.10 (9.0)	-.001 (6.5)	.16 (2.6)	-.005 (3.1)		.50 (4.2)	-.66 (1.4)	3.89 (18.4)	.37 1135
6. Women, Formal Sector	.17 (4.8)	.06 (2.5)	-.001 (1.4)	.10 (0.8)	-.002 (0.6)	.21 (2.4)	.43 (2.2)	-.18 (0.5)	3.75 (7.3)	.25 193
7. Women, Informal Sector	.00 (0.1)	.15 (3.3)	-.003 (2.6)	.47 (3.2)	-.004 (1.6)	.22 (1.8)	1.12 (2.6)	-.03 (0.1)	2.20 (2.9)	.11 231
8. Women, Domestics	-.00 (0.2)	.02 (1.3)	-.000 (0.2)	-.13 (1.0)	.002 (0.7)	-.14 (1.1)	.10 (1.1)	-.15 (0.5)	4.87 (14.8)	.08 111

NOTE: For an extensive description of the data see Behrman, Belli, Gustafson and Wolfe (1979) and Behrman, Gustafson and Wolfe (1980). The dependent variable is the ln of earnings in the previous two weeks in terms of 1977 cordobas (7 cordobas equals 1 U.S. dollar). Education is measured by the highest grade of formal schooling completed. Experience is actual labor force experience in years (and not age minus years of schooling minus 6, nor related calculations). Protein is the percentage of protein requirements by international norms that is satisfied, on the average, per family member by the previous week's diet. Days ill is the number of days missed from work or from other similar activity since the previous Christmas. Always in Managua is a dummy variable, with a value of 1 for women who have lived all their lives in Managua, and 0 for all others. The selection variables are the inverses of the Mills ratios for labor force participation and for reporting earnings, respectively, that are discussed in sections 1 and 4 and in the Appendix. Men is a dummy variable with a value of 1 for men and a value of 0 for women. Absolute values of the t-statistics in parentheses.

do with "report inclination." Among the 579 women in our sample who participated in the labor force, only 535 reported earnings. Unless those reporting earnings constitute a random sample of the labor force participants, there will be a second source of bias.[5]

To substantiate this argument, consider the following system of equations for the i^{th} individual in our original sample (we have dropped the subscript i to avoid notational clutter):

$$(1) \quad Y_1^* = \underline{\beta}_1'\underline{X}_1 + U_1 \qquad \text{"work inclination"}$$

$$(2) \quad Y_2^* = \underline{\beta}_2'\underline{X}_2 + U_2 \qquad \text{"report inclination"}$$

$$(3) \quad Y_3 = \underline{\beta}_3'\underline{X}_3 + \sigma_3 U_3 \qquad \text{earnings}$$

Here $\underline{X}_j$ is a vector of regressors, $\underline{\beta}_j$ is a vector of unknown coefficients, j = 1,2,3 and σ_3 an unknown scale parameter. The residuals U_1, U_2 and U_3 are assumed to have zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho_{13} \\ \rho & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}.$$

Y_1^* and Y_2^* are unobservables determining the subsample for which observations on earnings are available. Providing the "work inclination" of the individual is sufficiently large, he or she will participate in the labor force. Given that he or she is in the labor force, earnings will be observed if the individual's "report inclination" is strong enough. Individuals in the labor force may not report earnings either because they are not employed or because they elect not to respond to inquiries about earnings in their interview. Introducing the dichotomous variables Y_1 and Y_2 to indicate the possible outcomes, this sequential selection process can be summarized as follows:

$$Y_1 = \begin{cases} 1 \text{ if } Y_1^* > 0 & \text{"work"} \\ 0 \text{ if } Y_1^* \leq 0 & \text{"not work"} \end{cases} \tag{4}$$

$$Y_2 \begin{cases} 1 \text{ if } Y_2^* > 0 \text{ and } Y_1 = 1 & \text{"report" and "work"} \\ 0 \text{ if } Y_2^* \leq 0 \text{ and } Y_1 = 1 & \text{"not report" and "work"} \\ \text{unobserved if } Y_1 = 0 & \end{cases} \tag{5}$$

We observe Y_3 if and only if $Y_2 = 1$, that is if and only if

$$Y_1^* > 0 \text{ and } Y_2^* > 0. \tag{6}$$

Using the above representation, we can write the regression equation of interest as

$$\begin{aligned} E(Y3 \mid Y_2 = 1) &= \underline{\beta_3'X_3} + \delta_3(U_3 \mid Y_2 = 1) \\ &= \underline{\beta_3'X_3} + \delta_3 E(U_3 \mid Y_1^* > 0, Y_2^* > 0). \end{aligned} \tag{7}$$

Providing $E(U_3 \mid Y_3^* > 0, Y_2^* > 0) \neq 0$, ordinary least squares will result in inconsistent parameter estimates, or "selectivity bias." Consistent estimation of the parameters of the earnings equation requires knowledge of the form of the conditional expectation

$E(U_3 \mid Y_1^* > 0, Y_2^* > 0)$, and hence the conditional distribution of the error term. This calls for imposing additional structure onto the model. One such structure is provided by the trivariate normal specification, as shown in the appendix. This enables us to estimate the unknown conditional expectation on the right-hand-side of equation (7) up to a constant of proportionality, using the sample separation information. The constant of proportionality and the parameters of the earnings equation can then be estimated using linear regression.

We assume that the two selection rules are independent in our analysis ($\rho = 0$).[6] That is, we assume that the unobserved variables in the selection rule for labor force participation, such as unobserved market-rewarded abilities, are not correlated with the desire for privacy and other unobserved variables in the selection rule for reporting earnings. In the appendix, we show that when the two selection rules are independent, the stochastic version of equation (7) has the form

$$Y_3 = \underline{\beta}_3'\underline{X}_3 + \sigma_3\rho_{13}\lambda_1 + \sigma_3\rho_{23}\lambda_2 + W_3 \tag{8}$$

with W_3 as the residual term and

$$\lambda_1 = \frac{f(\underline{\beta}_1'\underline{X}_1)}{1-F(\underline{\beta}_1'\underline{X}_1)} \tag{9a}$$

$$\lambda_2 = \frac{f(\underline{\beta}_2'\underline{X}_2)}{1-F(\underline{\beta}_2'\underline{X}_2)} \tag{9b}$$

where $f(\cdot)$ and $F(\cdot)$ denote the standardized univariate normal density and distribution functions respectively. Equations (9a) and (9b) are the familiar univariate expressions of the selection literature, which can be estimated using probit analysis.

In Table 2, probit 1 is a significantly nonzero relation for selection of women into the labor force. The significantly positive effect of more education and the significantly negative effect of other income both are standard results. The other estimates in the aggregate probit for women's labor force participation indicate correlation in labor force experience and effects of nutrition, child care, and marital status that also are a priori plausible and that we discuss below in Sections 2 and 5, respectively.

Probit 1 in Table 3 is a significantly nonzero relation for selection on reporting earnings. Determinants include education, the linear and quadratic experience terms[7] and nutritional status--positive effects of which probably reflect the fact that those who have more education, job experience and adequate diet are more likely to have jobs from which to report earnings (given labor force participation). However, none of these effects is significantly nonzero at standard levels.

Regression 2 in Table 1 gives the estimates that are obtained when the two selectivity terms are added to the core regression. The estimates indicate that selection on labor force participation is significant, but not that on reporting earnings. The latter result suggests that reporting is random, although it may only reflect the weakness of our probit for the report inclinations. Comparison between regressions 1 and 2 suggests that selectivity bias in regression 1 results in some underestimate of the positive impact of education

Table 2

Probits for Labor Force Participation or Work Inclination

Variables	(1) Women	(2) Men	(3) Women and Men	(4) Women in Formal Sector	(5) Women in Informal Sector	(6) Women in Domestic Sector
Education	.07	.07	.10	.26	-.07	-.12
	(5.4)	(2.5)	(8.8)	(13.2)	(4.6)	(5.1)
Experience	.20	.09	.17	.12	.13	.08
	(12.6)	(4.2)	(18.1)	(5.7)	(7.6)	(3.0)
Experience^2	-.005	-.002	-.003	-.004	-.003	-.003
	(7.8)	(3.7)	(11.7)	(3.8)	(4.2)	(2.6)
Protein	.46	-.14	.31	-.43	.10	1.22
	(4.3)	(0.6)	(3.4)	(3.0)	(0.9)	(7.5)
Medically preventable				-.24	-.07	.15
				(2.3)	(0.8)	(1.1)
Therapeutically treatable				.37	-.15	-.08
				(3.5)	(1.5)	(0.6)
Other income	-.43	-.15	-.51	-.19	-.12	-1.99
	(7.5)	(0.9)	(10.2)	(3.1)	(1.6)	(8.6)
Children under 5	-.60	.06	-.37	-.48	.10	-1.28
	(5.1)	(0.3)	(3.7)	(2.8)	(0.8)	(4.2)
Home child care	.34	-.13	.22	.29	.09	.81
	(2.9)	(0.6)	(2.3)	(1.7)	(0.7)	(2.6)
Single				-3.60	4.62	-.48
				(0.4)	(0.6)	(0.0)
Previously accompanied				2.30	-4.18	4.02
				(0.3)	(0.6)	(0.1)

Table 2--continued

Variable	(1) Women	(2) Men	(3) Women and Men	(4) Women in Formal Sector	(5) Women in Informal Sector	(6) Women in Domestic Sector
Constant	-1.49	.57	-1.32	-.84	-1.72	-5.58
	(7.9)	(1.3)	(8.1)	(1.3)	(2.5)	(0.4)
2*Ln Likelihood Ratio	418.3	22.4	814.1	307.8	190.7	286.9
Sample size	1247	643	1890	1247	1247	1247
No. participants	579	601	1180	203	257	119

NOTE: Medically preventable is a dummy variable, with a value of 1 if the individual ever has had such a disease and 0 otherwise (so is therapeutically treatable). Other income refers to earnings from other household members who are working in the labor force, plus all nonearning income (including transfers). Children under 5 is a dummy variable, with a value of 1 if there are children under five and 0 otherwise. Home child care is a dummy variable, with a value of 1 if other adults (e.g., extended family members) or children over 14 are available for home child care and 0 otherwise. Single is a dummy variable with a value of 1 if the individual never has been accompanied, and 0 otherwise. Previously accompanied is a dummy variable with a value of 1 if the individual is not currently accompanied but has been previously (and currently is separated, divorced, or widowed), and 0 otherwise. Other variables are defined in Table 1.

Table 3

Probits for Reporting Earnings for Men and Women in Managua, 1977[a]

Variables	(1) Women	(2) Men	(3) Women and Men	(4) Women in Formal Sector	(5) Women in Informal Sector	(6) Women in Domestic Sector
Education	.04 (1.5)	- .10 (0.1)	.07 (2.8)	.03 (0.4)	-.02 (0.4)	-.06 (0.5)
Experience	.04 (1.3)	-1.70 (0.6)	.03 (1.5)	-.13 (1.3)	.07 (1.6)	.15 (0.7)
Experience²	-.002 (1.5)	.029 (0.4)	-.000 (0.1)	.002 (0.7)	-.003 (1.8)	.000 (0.0)
Protein	.31 (1.5)	-4.44 (1.0)	.25 (1.4)	.34 (0.5)	.08 (0.2)	1.93 (2.3)
Days ill	.01 (1.3)	.07 (0.4)	.01 (1.2)			
Medically preventable				.91 (1.6)	-.12 (0.5)	.42 (0.8)
Therapeutically treatable				-.70 (1.6)	-.21 (0.8)	-1.23 (1.7)
Always in Managua				.18 (0.4)	-.24 (1.0)	6.02 (0.1)
Other income	-.03 (0.2)	-1.36 (0.1)	-.13 (1.3)	.13 (0.5)	-.12 (0.6)	-2.16 (1.5)
Children under 5	.42 (1.3)	5.03 (0.7)	.70 (2.6)	3.90 (0.0)	.50 (1.3)	2.12 (0.0)
Home child care	-.15 (0.5)	2.90 (0.3)	-.26 (0.9)	.57 (0.0)	-.22 (0.6)	-2.00 (0.0)

Table 3--continued

Variables	(1) Women	(2) Men	(3) Women and Men	(4) Women in Formal Sector	(5) Women in Informal Sector	(6) Women in Domestic Sector
Single				-1.05 (0.0)	.53 (0.0)	33.09 (0.1)
Previously accompanied				-3.05 (0.0)	-2.83 (0.1)	-4.55 (0.0)
Median neighborhood income	.03 (0.5)	1.39 (0.1)	.04 (0.5)	.00 (0.0)	.67 (1.0)	.09 (0.6)
Neighborhood population density	-.00 (1.2)	.22 (1.5)	-.00 (0.7)	-.01 (1.4)	.00 (0.1)	.00 (0.2)
Age	.00 (0.1)	.38 (0.7)	-.01 (0.7)	.04 (0.9)	.01 (0.4)	-.07 (1.4)
Number of siblings				-.11 (1.3)	-.11 (2.9)	-.03 (1.0)
Both raisers				.06 (0.1)	.26 (1.0)	1.23 (1.8)
Constant	.53 (1.1)	14.4 (0.4)	.73 (1.6)	5.39 (0.0)	3.07 (0.1)	-28.8 (0.1)
2*Ln Likelihood Ratio	17.2	14.0	37.4	28.2	20.1	24.4
Sample size	579	601	1180	203	257	119
Number reporting	535	600	1135	193	231	111

[a]The first 13 variables are defined in Tables 1 and 2 above. The additional neighborhood and family background variables generally are self-explanatory. Both raisers is a dummy variable with a value of 1 if the individual had two adult raisers (e.g., father and mother or some combination of parents, step-parents, other adults) during childhood and 0 otherwise.

and of experience--particularly the initial years of experience--on women's earnings.

2. ADDITIONAL HUMAN CAPITAL DETERMINANTS: HEALTH, NUTRITION, AND MIGRATORY STATUS

The literature for the developed countries heavily emphasizes human capital investments in education and experience in the determination of earnings. For the developing countries, however, emphasis has been equally great on other factors, particularly on health, nutrition and migratory status. Leibenstein [29] and many others have posited that poor health and nutrition status cause low productivity and low earnings for many in the developing countries.[8] Migration is often viewed as a form of investment in order to obtain higher wages (e.g., Harris and Todaro [23]).

To our basic double selection ln earnings model, we add variables that represent health status (days ill) and nutrition status (family protein intake per capita). We do not investigate in this paper the returns to migration (see Behrman and Wolfe [17]), but we do see if the earnings function shifts for women who have always been in Managua. A priori, we might expect such women to receive higher earnings, ceteris paribus, than immigrants because they have better connections with a labor market in which personal contacts are very important.

Regression 3 in Table 1 gives the resulting estimates. Protein input has a significantly positive impact on women's earnings, just as on their labor force participation and their reporting of earnings. Clearly, this dimension of nutrition seems to be important through several channels. The estimated coefficient for illness is negative, as expected, but not quite significantly nonzero at standard levels. Finally, those women who have always lived in Managua have a significant earnings increment, presumably either for the reasons that we discuss above or because this variable represents a background of higher socioeconomic status. Thus these estimates support the incorporation of a wider spectrum of human capital for the developing countries than is often done for the developed ones.

3. SEXUAL DISCRIMINATION

A number of observers have claimed that sexual discrimination is rampant in labor markets of developing countries (see Burvinić [21]). We explore this question by estimating our extended double selection ln earnings model for men (Table 1, regression 4) and for women and men combined (Table 1, regression 5; the variable for always living in Managua is not available for men and is excluded).

A variable-by-variable comparison across regressions 3, 4, and 5 suggests some interesting possibilities. The returns to education, if anything, appear to be higher for women than for men--perhaps because relatively few women have much education and labor markets are somewhat segmented by sex. For nutrition, the returns are higher for men. The pattern of the coefficient estimates for family protein intake per capita may reflect that the men tend to have jobs in which there is more pay-off to strength, or that they obtain a better than average share of the household food, a common pattern in traditional societies.[9] Also of some interest is the fact that selectivity in terms of labor force participation apparently is important for men as well as for women.[10]

To test for differences between earnings functions for men and women we conducted F tests for the set of variables in regression 4,[11] and found that there is indeed a highly significant difference. In another formulation, we also included a dummy variable for sex (male) on combined run: The coefficient was approximately 1 and significant at the 1% level. On average, thus holding these other factors constant, men earn more than women. We conclude, therefore, that there is significant evidence consistent with discrimination against women in the form of lower ln earnings.

4. PLURALISM

A long-acknowledged characteristic of many markets in developing countries is fragmentation, or pluralism. Systematic treatment of such pluralism dates back at least to Lewis's [30] seminal article on dualism.

For our study we divided the Managuan labor market into three sectors: (1) a formal sector, in which there are implicit or explicit ongoing wage contracts, usually defined working hours, often explicit formal fringe benefits such as social security, and often large-scale employers; (2) an informal sector, in which there are no contracts nor benefits like social security; here the production units are usually small, and often operate out of the home, on the streets, in open markets, or in other transitory quarters, frequently with many family workers; (3) a domestic sector, in which women work in households at domestic tasks, often receiving room and almost always board as part of their payment.

In our sample we have 203 women who are in the formal sector, 257 in the informal sector, and 119 in the domestic sector. We are interested in what determines selection into a particular sector and whether or not the returns to various human capital variables differ across sectors. We estimate our extended double selection model for each of these groups, redefining the first selection to refer to selection into a particular sector instead of into the undifferentiated labor force. Probits 4,5, and 6 in Table 2 refer to this selection. Probits 4,5 and 6 in Table 3 refer to the inclination to report earnings in the three sectors, respectively. Regressions 6,7, and 8 in Table 1 are the estimated double-selection ln earnings functions for the three sectors.

Examination of these relations leads to the conclusion that there are significant differences among the three sectors. In general, the double-selection ln earnings function is substantially more consistent (using the adjusted R^2) with variance in ln earnings in the relatively commercialized formal sector than in the other two, and somewhat more consistent with ln earnings variance in the informal sector than for domestics.

On a variable-by-variable basis there are some interesting patterns. Average education ranges from 7.5 years for women in the formal sector to about 3.6 years for women in the informal and domestic sectors. More education increases the probability that a woman is in the formal sector--as opposed to being out of the paid labor force;

informal sector and domestic employment are less likely than nonparticipation for the more educated. The returns to education in terms of earnings are also much higher in the formal sector than in the other two--and, in fact, are not significantly nonzero for the informal and domestic sectors.

N, B

Labor force experience varies from an average of 10 or 11 years for women in the domestic and formal sectors to over 17 for those in the informal sector. The combination of the linear and quadratic labor force experience terms increases the probability of labor force participation in one of these three sectors as opposed to being out of the labor force. Among the three sectors, greater experience points toward a lower probability of being a domestic. In terms of earnings, the highest returns to experience are in the informal sector with the formal sector next, but there is no significantly nonzero effect for domestics.

A better nutritional state, as represented by family protein intake per capita, appears to lead to higher probability of selection into the domestic sector and out of the formal sector, as opposed either to nonparticipation in the labor force or participation in the informal sector. This result is at first glance somewhat surprising; it is explained in part by simultaneity or reverse causality for domestics, who have relatively good diets because they receive board in the generally higher-income households in which they work. For them, for example, the average protein index is 18% above the average for the other two sectors. In terms of ln earnings the returns to nutrition are significantly positive only for the informal sector. This is probably because domestics tend to receive relatively good diets, as we note above, and formal sector workers tend to be well enough off to be above the threshold of gross malnourishment.

The health variables appear to have somewhat of a differential impact across sectors. Having had a disease that is preventable by medical measures (e.g., vaccination) reduces the probability of being in formal sector employment, as opposed to being out of the labor force or in the domestic sector. Those who report having had a disease that is therapeutically treatable (e.g., high blood pressure) are more likely to be in the formal sector, as opposed to being a domestic in the informal sector, or out of the labor force. We expect that this pattern does not directly reflect selectivity among the sectors so much as differential knowledge regarding the identification of therapeutically treatable diseases--knowledge in part acquired in paid or unpaid work activity. Coworkers and employers in the formal sector are likely to be better informed than are those in the other sectors. However, none of these disease categories nor a measure of days ill have significantly nonzero coefficient estimates in the sectoral ln earnings functions (the negative coefficient for days ill for the informal sector is closest).

Half of the women in the formal sector and 45% of those in the informal sector, but only 16% in the domestic sector, have always lived in Managua. The domestic sector, thus, is dominated by immigrants from smaller urban and rural areas. Always having been in Managua has a significantly positive coefficient estimate only in the ln earnings function for the formal sector (and one about as large, but not quite significant, for the informal sector). The returns for knowledge of the local labor market network (or having experience of "higher quality" in Managua), therefore, appear to be greatest for the most formal sector. The former may seem to be somewhat surprising, if one believes that recruitment on the basis of quality rather than on the basis of connections tends to become more important in more modern sectors.

The migration results may reflect higher quality experience of those in the formal sector who always worked in Managua. It may also reflect the greater availability of rents in the formal versus informal sector, which may be distributed in the form of earnings. Krueger [28] argues that such rents are quite important in developing countries. It seems plausible that they may be more concentrated in the formal sector in which education requirements, union membership, and other barriers to entry are much more likely to be effective.

In the double selection ln earnings functions for all three sectors, selection terms have significantly nonzero coefficient estimates only for labor force participation in the formal and informal sectors.

Thus we find some interesting patterns across the sectors. Returns to education and to always having been in Managua are significant only in the formal sector. Returns to experience are significant in both the formal and informal sectors. Improved nutrition increases productivity and earnings primarily in the informal sector. Domestics are primarily migrants from other parts of the country. Some of these factors also affect the selection into particular sectors, as do family and child care status. To these we now turn.

5. FAMILY STATUS AND CHILD CARE

The literature on women's labor supply for developed economies places great emphasis on the opportunity costs of married women in terms of household production, particularly where child care responsibilities are involved. A priori, such considerations would seem to need modification for developing countries, because conditions differ. The presence of other adults in extended families, of older children, and of domestic employees means that the opportunity costs well may be less. There would, however, seem to be significant differences among sectors, in that on-the-job child care is often a possibility in the informal sector, but not in the formal sector. Moreover many domestics sleep at their employer's, and are allowed to keep neither their companions nor any or more than one or two of their children with them. For these domestics, the opportunity costs of employment in terms of child care and family interaction may be quite high. Finally, many families are so poor that women may participate in the labor force no matter what the opportunity costs are in terms of child care, in hopes of keeping the family out of extreme poverty.

The probits on labor force participation in Table 2 shed some light on these issues. Probits 1 and 2 refer respectively to women and to men. The presence of children under 5 significantly reduces the probability of women participating in the labor force and the presence of home child care alternatives in the form of other adults or older children significantly increases this probability. Neither of these factors has a significantly nonzero impact on the labor force participation of men. Thus this aggregate pattern is similar to that found for women and men in more developed countries, although home child care is more widely available from other adults in extended families and from older children.

Probits 4, 5 and 6 refer to women's labor force participation in formal, informal, and domestic sectors, respectively. These disaggregate relations have several interesting features related to family status and child care. The presence of children under 5 particularly lowers the probability of participating in the domestic sector, as we expected; it also lowers the probability of participating in the formal sector because child care provisions are absent, and it is impossible to combine work and on-the-job child care in that sector. In contrast, the impact is not significantly nonzero for participation in the informal sector, also as we expected. For the same reasons, the impact of home child care is different across the three sectors; the largest impact is in the domestic sector and the smallest in the informal sector.

Similar considerations might seem to underlie the impact of marital status. But none of the marital status variables have significantly nonzero coefficient estimates. However, having higher income from a companion (or from other sources) greatly reduces the probability of participation in the domestic sector and somewhat reduces that of participation in the formal sector, since it lessens the need for additional income.

Thus the standard child care and marital status effects on the labor force participation of women are modified by considering the different options among the three sectors, and by the more common possibility of extreme poverty.

6. CONCLUSION

We have gained many insights into the factors determining labor force participation and ln earnings for women in the major metropolitan area of a developing country. We have considered the possibility of double selectivity for ln earnings estimates. Selectivity in regard to labor force participation may be important, but that in reporting earnings is generally not.

We have extended the standard human capital considerations to include factors beyond formal education and experience. Nutritional intake, especially of protein, has significantly positive effects on earnings of both men and women in the aggregate, although it is somewhat larger for the former. Health has a more marginal negative impact on women's and men's earnings. These results suggest that poor nutrition and health lower productivities and earnings for many adults in our sample. Programs that led to better diets for the poorer members of the society, therefore, would have some pay-off in terms of increased productivities and greater equalization in the distribution of earnings.

We have explored the possibility of sexual discrimination and have found that women receive significantly lower returns from the various human capital investments than do men.

We have found evidence that the presence of small children has a negative impact on the probability of labor participation of women, but not of men. However, this effect is offset in many more households than is the case in developing countries by the presence of other adults in extended families or of older children who fulfill home child care responsibilities.

Many of these effects are illuminated or modified, finally, by consideration of the pluralistic nature of the labor market. The returns to women's education are large for selection into and earnings in the formal sector, but not elsewhere. Knowledge of a local labor market network may also be rewarded more in that sector. The returns to experience also are significant in the formal sector, but are even larger in the informal sector. The returns to better nutritional (and perhaps health) status are greatest in the informal sector, since women in the formal sector tend to have above-minimal nutritional levels and health standards owing to higher family incomes, and those in the domestic sector tend to be above such standards owing to food and care provided by their employers. Extreme poverty, in the form of low income from other sources, tends to drive poorer women to participate in the domestic sector. Child care needs and the lack of home child care alternatives lead to selection out of the labor force, but particularly out of the formal and domestic sectors, since on-the-job child care is generally a possibility in the informal sector.

In developing countries, women play a large role in determining the current income distribution and in shaping the conditions under which the next generation is being raised. Even leaving aside questions of efficiency, productivity, and equality of opportunity for the present generation, therefore, in casting light on the factors that may affect their labor force participation and their earnings, we are providing information that is directly relevant to a critical policy area.

APPENDIX

In this appendix, the problem of estimation under two sample selection rules is tackled within a missing data framework, with the truncated normal distribution providing the distributional specification. Our approach is based on an extension of Heckman's [24] model and relies on the formulation of the choice process as a trichotomy, with selection made sequentially. The qualitative structure of our model generalizes Amemiya's [2] univariate sequential unordered normal model by accounting for possible dependence between the two selection rules. Catsiapis and Robinson [22] treat the problem in essentially the same manner, and arrive at our constrained model through a direct extension of Heckman's two-step procedure. Poirier [32] analyzes a slightly different two-selection problem, where two individuals facing the same choice set arrive at separate, possibly interrelated, but individually unobservable decisions, the joint outcome of which takes the form of a dichotomous observable variable.

We begin by reproducing the model in the text. For the i^{th} individual in our original sample we have:

$$Y_1^* = \underline{\beta}_1{}'\underline{X}_1 + U_1 \quad \text{"work inclination"} \tag{A1}$$

$$Y_2^* = \underline{\beta}_2{}'\underline{X}_2 + U_2 \quad \text{"report inclination"} \tag{A2}$$

$$Y_3 = \underline{\beta}_3{}'\underline{X}_3 + \sigma_3 U_3 \quad \text{earnings} \tag{A3}$$

where $\underline{X}_j$ is a vector of regressors, $\underline{\beta}_j$ is a vector of unknown coefficients and σ_3 denotes a scale parameter. The residuals are assumed to be normally distributed with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho_{13} \\ \rho & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

Our main objective is to estimate the parameters of equation (3), with the unobservable continuous random variables Y_1^* and Y_2^* determining the subsample (or selecting individuals) for which complete observations satisfying equation (A3) are available. Using the dichotomous variables Y_1 and Y_2 to indicate the outcome of the selection processes in equations (A1) and (A2), we can classify the individuals in the original sample as follows:

$$Y_1 = \begin{cases} 1 \text{ if } Y_1^* > 0 & \text{"work"} \\ 0 \text{ if } Y_1^* \leqslant 0 & \text{"not work"} \end{cases} \tag{A4}$$

$$Y_2 = \begin{cases} 1 \text{ if } Y_2^* > 0 \text{ and } Y_1 = 1 & \text{"report" and "work"} \\ 0 \text{ if } Y_2^* \leqslant 0 \text{ and } Y_1 = 1 & \text{"not report" and "work"} \\ \text{unobserved if } Y_1 = 0 & \end{cases} \tag{A5}$$

We observe earnings (Y_3) if and only if $Y_2 = 1$, that is if and only if:

$$Y_1^* > 0 \text{ and } Y_2^* > 0. \tag{A6}$$

Under selection rules (A4) and (A5), probability P_j that the i^{th} individual will fall into the j^{th} subsample is given by:

$$P_1 = \Pr(Y_1 = 0) = \Pr(Y_1^* < 0) = \Pr(U_1 < -\underline{\beta}_1'\underline{X}_1) = 1-F(\underline{\beta}_1'\underline{X}_1) \quad (A7)$$

$$P_2 = \Pr(Y_2 = 0) = \Pr(Y_1^* > 0,\ Y_2^* < 0) \quad (A8)$$

$$= \Pr(U_1 > -\underline{\beta}_1'\underline{X}_1,\ U_2 < -\underline{\beta}_2'\underline{X}_2) = G(\underline{\beta}_1'X_1,\ -\underline{\beta}_2'\underline{X}_2;\ \rho)$$

$$P_3 = \Pr(Y_2 = 1) = \Pr(Y_1^* > 0,\ Y_2^* > 0) \quad (A9)$$

$$= \Pr(U_1 > -\underline{\beta}_1'\underline{X}_1,\ U_2 > -\underline{\beta}_2'\underline{X}_2) = G(\underline{\beta}_1'X_1,\ \underline{\beta}_2'X_2;\ \rho)$$

where $F(\cdot)$ and $G(\cdot)$ denote the standardized univariate and bivariate normal distribution functions respectively. Note that the partitioning of the original sample is indeed complete:

$$\sum_{j=1}^{3} P_j = 1 - F(\underline{\beta}_1'\underline{X}_1) + G(\underline{\beta}_1'\underline{X}_1,\ -\underline{\beta}_2'\underline{X}_2;\ \rho) + G(\underline{\beta}_1'\underline{X}_1,\ \underline{\beta}_2'\underline{X}_2;\ \rho)$$

$$= 1 - F(\underline{\beta}_1'\underline{X}_1) + F(\underline{\beta}_1'\underline{X}_1) = 1$$

Equations (A7) and (A8) contain all available information for the individuals who do not work and for the individuals who work but do not report, while as for individuals who work and report, in addition to (A9), the dependent variable Y_3, earnings is observed. Under our trivariate normal specification, the probability density function for Y_3 is given by:

$$\phi(Y_3) = \frac{1}{P_3}\int_{-\infty}^{\underline{\beta}_2'\underline{X}_2}\int_{-\infty}^{\underline{\beta}_1'\underline{X}_1} \frac{1}{\sigma_3}\, h(U_1, U_2, \frac{Y_3 - \underline{\beta}_3'\underline{X}_3}{\sigma_3})\, dU_1\, dU_2$$

where $h(\cdot)$ denotes the trivariate density for $\underline{U}$, and P_3 is defined in (A9). Denoting the subsample of those who do not work by S_1, those who work but do not report by S_2 and those who work and report by S_3, the likelihood function for the entire sample has the form

$$L = \prod_{S_1} [1 - F(\underline{\beta}_1'\underline{X}_1)] \cdot \prod_{S_2} G(\underline{\beta}_1'\underline{X}_1, -\underline{\beta}_2\underline{X}_2; \rho). \tag{A10}$$

$$\prod_{S_3} \int_{-\infty}^{\underline{\beta}_2'\underline{X}_2} \int_{-\infty}^{\underline{\beta}_1'\underline{X}_1} \frac{1}{\sigma_3} h(U_1, U_2, Z_3) dU_1 dU_2$$

where

$$Z_3 = \frac{1}{\sigma_3} (Y_3 - \underline{\beta}_3'\underline{X}_3).$$

The complicated nature of the likelihood function and the large number of parameters to be estimated make the full information procedure extremely difficult. With this in mind, we now turn to a computationally simpler two-step procedure in the spirit of Heckman [24].

The sequential selection process partitions the original random sample into three mutually exclusive nonrandom subsamples, containing those with $Y_1 = 0$, those with $Y_2 = 0$, and those with $Y_2 = 1$. Since S_3 consists of individuals for whom Y_3 is observed, the regression equation of interest may be written as:

$$E(Y_3 \mid Y_2 = 1) = \underline{\beta}_3'\underline{X}_3 + \sigma_3(U_3 \mid Y_2 = 1) \tag{A11}$$

$$= \underline{\beta}_3'\underline{X}_3 + \sigma_3 E(U_3 \mid Y_1^* > 0,\ Y_2^* > 0).$$

Hence, providing $E(U_3 \mid Y_1^* > 0, Y_2^* > 0) \neq 0$, ordinary least squares will result in inconsistent parameter estimates, or "selectivity bias." We utilize the normality assumption to rewrite the conditional expectation on the right hand side (cf. [27], pp. 86-87) as:

$$E(U_3 \mid Y_1^* > 0, Y_2^* > 0) = E(U_3 \mid U_1 > -\underline{\beta}_1'\underline{X}_1, U_2 > -\underline{\beta}_2'\underline{X}_2) \qquad \text{(A12)}$$

$$= \rho_{13\cdot2}E(U_1 \mid U_1 > C_1, U_2 > C_2) +$$

$$\rho_{23\cdot1}E(U_2 \mid U_1 > C_1, U_2 > C_2)$$

where $C_j = -\underline{\beta}_j'\underline{X}_j$, $j = 1,2$ and

$$\rho_{ij}\cdot k = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{[(1-\rho_{ik}^2)(1-\rho_{jk}^2)]^{1/2}}.$$

The two expectations on the right hand side of equation (A12) may in turn be expressed as (c.f. [33], p. 406)

$$E_1 = E(U_1|U_1 > C_1, U_2 > C_2) = \frac{1}{P_3}\left\{f(C_2)[1-F(C_1^*)] + \rho f(C_1)[1-F(C_2^*)]\right\} \qquad \text{(A13)}$$

$$E_2 = E(U_2|U_1 > C_1, U_2 > C_2) = \frac{1}{P_3}\left\{f(C_1)[1-F(C_2^*)] + \rho f(C_2)[1-F(C_1^*)]\right\} \qquad \text{(A14)}$$

where

$$C_1^* = \frac{C_1 - \rho C_2}{[1 - \rho^2]^{1/2}},$$

$$C_2^* = \frac{C_2 - \rho C_1}{[1 - \rho^2]^{1/2}}$$

In view of (A13) and (A14), equation (A11) is highly nonlinear. As a convenient short cut, we first exploit the qualitative structure of the model to obtain estimates of E_1 and E_2, and then substitute the estimated values of the conditional expectations into equation (A12) to obtain estimates for the remaining parameters of (A11) using linear regression. We now describe the two steps explicitly, first for the original model, then for a constrained version with $\rho = 0$.

TWO-STEP ESTIMATION - THE UNCONSTRAINED MODEL

(1) Maximum Likelihood Estimation: We utilize the sample separation information (that is, data on Y_1 and Y_2) together with specification (A7) - (A9), and obtain the likelihood function

$$L_o = \prod_{S_1} [1 - F(\underline{\beta}_1'\underline{X}_1)] \cdot \prod_{S_2} G(\underline{\beta}_1'\underline{X}_1, -\underline{\beta}_2'\underline{X}_2; \rho) \cdot \prod_{S_3} G(\underline{\beta}_1'\underline{X}_1, \underline{\beta}_2'\underline{X}_2; \rho) \quad \text{(A15)}$$

which depicts the qualitative structure of our model.[13]

Subject to the identification condition that $\underline{X}_1$ includes one variable excluded from $\underline{X}_2$ (see Tunali, Behrman, Wolfe [34], pp. 9-10), maximization of (A15) will yield consistent estimates $\hat{\underline{\beta}}_1$, $\hat{\underline{\beta}}_2$, $\hat{\rho}$, hence $\hat{C}_1$, $\hat{C}_2$, $\hat{C}_1^*$, $\hat{C}_2^*$ and $\hat{P}_3$. Substituting these into (A13) and (A14) gives the estimated expectations $\hat{E}_1$ and $\hat{E}_2$.

(2) Linear Regression: Inserting the estimated expectations into the stochastic version of (A11) we get

$$Y_3 = \underline{\beta}_3'\underline{X}_3 + \sigma_3\rho_{13\cdot 2}\hat{E}_1 + \sigma_3\rho_{23\cdot 1}\hat{E}_2 + \sigma_3\tilde{V}_3 \quad \text{(A16)}$$

$$= \underline{\beta}_3'\underline{X} + \gamma_1\hat{E}_1 + \gamma_2\hat{E}_2 + \sigma_3\tilde{V}_3$$

where $\tilde{V}_3 = V_3 + \gamma_1(E_1 - \hat{E}_1) + \gamma_2(E_2 - \hat{E}_2)$ and $E(V_3 \mid Y_1^* > 0,\ Y_2^* > 0) = 0$. This is then fitted by linear regression of Y_3 on $\underline{X}_3$, $\hat{E}_1$ and $\hat{E}_2$ for the individuals in S_3. Consistency of the estimates follows from Slutsky's theorem.

Under selection, the standard least squares estimator of the population variance for Y_3 (see [34], p. 8 for the exact variance) will be inconsistent, in view of the fact that

$$V(Y_3 \mid Y_1^* > 0,\ Y_2^* > 0) = \sigma_3^2\ V(V_3 \mid Y_1^* > 0,\ Y_2^* > 0) \neq \sigma_3^2 = Var(Y_3).$$

This implies that σ_3, ρ_{13} and ρ_{23} cannot be identified using the estimated standard error of the regression together with $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\rho}$.

TWO-STEP ESTIMATION - THE CONSTRAINED MODEL ($\rho = 0$)

We now consider a constrained form of our structural model, where we assume $\rho = 0$. The assumed independence between the two decision rules reduces the trichotomy in (A7) - (A9) to the normal unordered response model discussed by Amemiya [2], p. 366:

$$P^* = 1 - F(\underline{\beta}_1'\underline{X}_1) \tag{A17}$$

$$P_2^* = [1 - F(\underline{\beta}_2'\underline{X}_2)] \cdot F(\underline{\beta}_1'\underline{X}_1) \tag{A18}$$

$$P_3^* = F(\underline{\beta}_2'\underline{X}_2) \cdot F(\underline{\beta}_1'\underline{X}_1) \tag{A19}$$

For this model, a two-step procedure for estimating the structure would run as follows:

(1) Maximum Likelihood Estimation (Probit Analysis): With (A17) - (A19), the likelihood function in (A15) factors, giving the special form:

$$L^* = \prod_{S_1} [1 - F(\underline{\beta}_1'\underline{X}_1)] \cdot \prod_{S_2+S_3} F(\underline{\beta}_1'\underline{X}_1) \cdot \prod_{S_2} [1 - F(\underline{\beta}_2'X_2) \cdot \prod_{S_3} F(\underline{\beta}_2'\underline{X}_2) \quad \text{(A20)}$$

maximization of which is equivalent to doing two independent probit analyses, without any further loss in efficiency. The first maximization involves the full sample, split on Y_1:

$$L_1^* = \prod_{S_1} [1 - F(\underline{\beta}_1'\underline{X}_1)] \cdot \prod_{S_2+S_3} F(\underline{\beta}_1'\underline{X}_1) \quad \text{(A21)}$$

and provides a consistent estimate of $\underline{\beta}_1$. The second involves the subsample with $Y_1 = 1$, split on Y_2:

$$L_2^* = \prod_{S_2} [1 - F(\underline{\beta}_2'\underline{X}_2)] \cdot \prod_{S_3} F(\underline{\beta}_2'\underline{X}_2) \quad \text{(A22)}$$

giving a consistent estimate for $\underline{\beta}_2$.

The expectations in (A13) and (A14) reduce to familiar univariate expressions

$$\lambda_1 = \frac{f(C_2)}{1 - F(C_1)} \quad \text{(A23)}$$

$$\lambda_2 = \frac{f(C_2)}{1 - F(C_2)} \quad \text{(A24)}$$

consistent estimates for which can be calculated using $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ to construct $\hat{C}_1$ and $\hat{C}_2$. Note that estimates of λ_1 are obtained from the entire sample, while individuals that fall into S_2 and S_3 alone are utilized to estimate λ_2.

(2) Linear Regression: Denoting those estimates of (A23) and (A24) by $\hat{\lambda}_1$ and $\hat{\lambda}_2$ respectively, we use linear regression on subsample S_3 to fit

$$Y_3 = \underline{\beta}_3'\underline{X}_3 + \sigma_3\rho_{13}\hat{\lambda}_1 + \sigma_3\rho_{23}\hat{\lambda}_1 + \sigma_3 W_3 \qquad \text{(A25)}$$

$$= \underline{\beta}_3'\underline{X} + \gamma_1^*\hat{\lambda}_1 + \gamma_2^*\lambda_2 + \sigma_3 W_3$$

where $W_3 = V_3 + \gamma_1^*(\lambda_1 - \hat{\lambda}_1) + \gamma_2^*(\lambda_2 - \hat{\lambda}_2)$ and $E(V_3 \mid Y_1^* > 0,\ Y_2^* > 0) = 0$ as before. Consistency follows from Slutsky's theorem.

It can be shown that for the constrained model with $\underline{\rho} = 0$,

$$\text{Var}\,(V_3 | Y_1^* > 0,\ Y_2^* > 0) = [(1 - \rho_{13}^2 - \rho_{23}^2) + \rho_{13}^2(1 + C_1\lambda_1 - \lambda_1^2) \qquad \text{(A26)}$$
$$+ \rho_{23}^2(1 + C_2\lambda_2 - \lambda_2^2)]$$

with

$$0 \leq 1 + C_j\lambda_j - \lambda_j^2 \leq 1 \qquad j = 1,2 \qquad \text{(A27)}$$

implying Var $(V_3 \mid Y_1^* > 0,\ Y_2^* > 0) < \sigma_3^2)$. Denoting the standard least squares estimate of the variance at the second step by $\hat{V}(W_3)$, a consistent estimate for σ_3 can be obtained by rearranging equation (A26) and recalling $\gamma_2^* = \sigma_3\rho_{23}$, $\gamma_1^* = \sigma_3\rho_{13}$:

$$\hat{\sigma}_3^2 = V(\hat{W}_3) - \hat{\gamma}_2^{*2} \cdot \frac{1}{T_3} \sum_{S_3} (\hat{C}_1\hat{\lambda}_1 - \hat{\lambda}_1^2) - \hat{\gamma}_1^{*2} \cdot \frac{1}{T_3} \sum_{S_3} (\hat{C}_2\hat{\lambda}_2 - \hat{\lambda}_2^2)$$

where T_3 denotes the number of observations in S_3. This estimate is guaranteed to be positive in view of (A27). Consistent estimates for ρ_{23} and ρ_{13} can now be obtained using $\hat{\gamma}_2^*$, $\hat{\gamma}_1^*$ and $\hat{\sigma}_3$.

Finally, we consider the relationship of the constrained model with Heckman's single-selection set-up. The assumption of independence enables us to treat the selection problem along the traditional line, with two separate first-step probits that lead into a second step where we include two constructed variables along Heckman's lines to correct for selectivity in the regression equation. Note, however, that the sequential aspect of the selection process is preserved despite independence. In fact, the constrained model can be extended to account for any number of selection rules, providing selection is sequential and sample separation information at each stage is available. Estimation under this specific kind of multiple selection will follow the lines discussed above. It should be kept in mind that collinearity in the regression equation is likely to be a problem unless nonoverlapping exclusions are available in the selection equations.

The limiting distribution of the two-step estimator for the constrained model can be developed along the line pursued in Heckman [25]. Given the expression for the exact variance, extension of Heckman's [24] approximate GLS procedure to the multiple selection problem is straightforward. Note that this procedure does not utilize the relationship between the coefficients of λ's in the regression equation and the residual variance. The resulting estimates are not asymptotically efficient, and the gain in efficiency over the OLS estimates is not likely to be worth the computational burden.

NOTES

[1]In recent years there has been a flood of estimates of such relations for women in the United States and other developed countries, with particular emphasis on the sample selectivity problem. Maddala [31] and Wales and Woodland [35] provide surveys of this literature.

[2]We describe the sample in some detail in Behrman, Belli, Gustafson, and Wolfe [3] and Behrman, Gustafson and Wolfe [4].

[3]We use this special "sisters" feature of the data in Behrman and Wolfe [14, 15, 16]. Other project studies that are completed or in process include Behrman and Wolfe [6, 7, 8, 9, 10], Blau [19, 20], Wolfe, Behrman and Flesher [36], and Ybarra [37].

[4]These assumptions include the existence of near-perfect capital markets and the absence of important omitted variables that are correlated with schooling. See Behrman, Hrubec, Taubman and Wales [5] for an extensive discussion and for evidence that the latter does not hold for white U.S. males.

[5]Often such missing data are assumed to be random. For exploration of alternative approaches to the randomly missing data problem, see Wolfe, Behrman and Flesher [36].

[6]The correlation between the predicted values of our selection variables in fact is quite low (0.14).

[7]The coefficient estimate of the quadratic experience term is negative, but the total impact of experience becomes negative only after 20 years of experience, which is more than most of the women in the sample have.

[8]In Behrman and Wolfe [11, 12], we consider the determinants of health and nutrition status.

[9]The second, intrafamilial distribution possibility is probably consistent with some other regressions (which we do not reproduce here), in which we find evidence that religiously married women receive higher returns to the family protein per capita variable than do other accompanied (but not religiously married) women. In unions that are formally sanctioned by religious marriages, the role of the woman may be stronger and she may receive a larger share of the better food than in other unions. Another interesting factor is that the data are collected by surveying women in this traditional society and women's responses may be affected by the stability of their relationship.

[10]However this coefficient estimate is less robust under specification change than are the others.

[11]The regression for women is identical to that in panel 3 except that the variable for always in Managua is dropped. The regression for the combined group is the same as in panel 5. If an additive dummy variable is included in addition, the F-test indicates no significant difference.

[12]While at first glance, it may appear that we should rank the sectors, and do a series of sequential choices, we do not see an unambiguous way of carrying this out. Instead, we treated the choice

problem involving selection into the three sectors separately, to be able to isolate the differential role of the selection variables in the three sectors.

[13]Note that this likelihood function is different from that of a bivariate probit. In Tunali, Behrman and Wolfe [34] we further simplify the first step and reduce it to two independent probits, without constraining the model to the case where $\rho = 0$. For a discussion of this modified approach, see pages 11, 12, and 23 of [34].

REFERENCES

[1] Amemiya, T.: "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," Econometrica (1974) 42:6, 999-1012.

[2] ______: "Qualitative Response Models." Annals of Economic and Social Measurement (1975) 363-372.

[3] Behrman, J.R., H. Belli, K. Gustafson and B.L. Wolfe: "How Many? How Much? The Determinants of Demoeconomic Roles of Women in a Developing Country Metropolis." Madison: University of Wisconsin, mimeo, 1979.

[4] Behrman, J.R., K. Gustafson and B.L. Wolfe: "Demoeconomic Characteristics of Women and Different Degrees of Urbanization in a Developing Country." Madison: University of Wisconsin, mimeo, 1980.

[5] Behrman, J.R., Z. Hrubec, P. Taubman and T.J. Wales: Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment and Schooling. Amsterdam: North-Holland Publishing Company, 1979.

[6] Behrman, J.R. and B.L. Wolfe: "Alternative Measures of Fertility and the Impact of Human Capital Investments in Schooling, Health and Nutrition in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1979.

[7] _____: "The Impact of Health and Nutrition on the Number of Surviving Children in a Developing Metropolis." Philadelphia: University of Pennsylvania, mimeo, 1979.

[8] _____: "A More General Approach to Fertility Determination: Endogenous Preferences and Natural Fertility in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1979.

[9] _____: "Child Health and Nutrition Determinants in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1980.

[10] _____: "The Demand for Nutrition in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1980.

[11] _____: "Determinants of Health Utilization in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1980.

[12] _____: "The Determinants of Schooling for Children in Developing Countries: Family Background, Number of Siblings, Sex, Birth-Order, and Residence." Philadelphia: University of Pennsylvania, mimeo, 1980.

[13] _____: "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, 1980.

[14] _____: "Parental Preferences and Intrafamilial Allocations of Human Capital Investments in Schooling and Health in a Developing Country." Philadelphia: University of Pennsylvania, mimeo, 1980.

[15] _____: "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, 1980.

[16] _____: "Wage Rates for Adult Family Farm Workers in a Developing Country and Human Capital Investments in Health and Schooling." Philadelphia: University of Pennsylvania, mimeo, 1980.

[17] _____: "Women's Migration in a Developing Country and Human Capital Investments in Schooling, Health, and Nutrition." Madison: University of Wisconsin, mimeo, 1980.

[18] Behrman, J.R., B.L. Wolfe and D. Blau: "The Impact of Changing Population Composition on the Distributions of Income and Socioeconomic Status in a Developing Country." Madison: University of Wisconsin, mimeo, 1980.

[19] Blau, D.: "On the Relation Between Child Malnutrition and Growth in Less Developed Countries." Madison: University of Wisconsin, mimeo, 1977.

[20] _____: "Nutrition, Fertility and Labor Supply in Developing Countries: An Economic Analysis." Madison: University of Wisconsin, Ph.D. dissertation, 1980.

[21] Burvinić, M.: Women and World Development: An Annotated Bibliography. Washington: Overseas Development Council, 1976.

[22] Catsiapis, G. and C. Robinson: "Sample Selection Bias with Two Selection Rules: An Application to Student Aid Grants." London, Ontario: University of Western Ontario, mimeo, 1978.

[23] Harris, J.R. and M.P. Todaro: "Migration, Unemployment and Development: A Two-Sector Analysis," American Economic Review, 60 (1970), 126-142.

[24] Heckman, J.J.: "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simpler Estimator for Such Models," Annals of Economic and Social Measurement (1976), 475-492.

[25] _____: "Sample Selection Bias as a Specification Error," Econometrica, 47 (1979), 153-161.

[26] _____: "Shadow Prices, Market Wages, and Labor Supply," Econometrica, 42 (1974), 679-694.

[27] Johnson, N.L. and S. Kotz: Distribution in Statistics: Continuous Multivariate Distributions. New York: John Wiley, 1972.

[28] Krueger, A. O.: "The Political Economy of the Rent-Seeking Society," American Economic Review (1974), 291-303.

[29] Leibenstein, H.: Economic Backwardness and Economic Growth, New York, 1957.

[30] Lewis, W.A.: "Economic Development with Unlimited Supplies of Labor," Manchester School (1954), 139-191.

[31] Maddala, G.S.: "Selectivity Problems in Longitudinal Data," Annales de l'INSEE 30-31, (1978), 423-450.

[32] Poirier, D.J.: "Partial Observability in Bivariate Probit Models," Journal of Econometrics (1980), 12:2, 209-219.

[33] Rosenbaum, S.: "Moments of Truncated Bivariate Normal Distribution," Journal of Royal Statistical Society, Series B, 23 (1961), 405-408.

[34] Tunali, F.I., J.R. Behrman and B.L. Wolfe: "Identification, Estimation and Prediction Under Double Selection." Paper presented to the Joint Statistical Meetings, Houston, Texas, 1980.

[35] Wales, T.J. and A. D. Woodland: "Sample Selectivity and the Estimation of Labour Supply Functions," International Economics Review, forthcoming.

[36] Wolfe, B.L., J.R. Behrman and J. Flesher: "A Monte Carlo Study of Alternative Approaches for Dealing with Randomly Missing Data." Institute for Research on Poverty Discussion Paper #587-79, 1980.

[37] Ybarra, R.A.: La Estructura Occupational de la Fuerza de Trabajo Feminina en Nicaragua, 1950-1977. Managua: Banco Central, 1978.