University of the Philippines
SCHOOL OF ECONOMICS

Discussion Paper No. 8003 May 1980

An Auxiliary Model for Quantifying the Socioeconomic Impact of a Development Project

by

José Encarnación, Jr.

Abstract

One can in principle calculate the impact of a development project on various areas of concern to the policymaker. While some relationships among variables are specific to a particular project, there are also relationships that are common to all projects. A model of the latter would therefore be a common auxiliary to the specific models assessing the effects of different projects.

This paper presents a partial specification of such an auxiliary model and gives estimates (using the 1973 National Demographic Survey data set, which does not suffice for estimating more than a partial specification) of multiple regression equations with marital fertility, child mortality and wife's employment status as dependent variables.

An Auxiliary Model for Quantifying the Socioeconomic Impact of a Development Project[1/]

by José Encarnación, Jr.

1. Introduction

A development project is typically designed to advance only one or a few of the objectives that concern development planners. Different projects specialize, so to speak, in the pursuit of different objectives, and a project may even have a negative impact on another area of concern. (A rural road might, for example, increase productivity in the area but also increase urban unemployment by facilitating rural-urban migration.) With sufficient data and correct model formulation, one could calculate the impact of each project on all the areas of concern. For this purpose it would be useful to distinguish between: (a) relationships among variables that are specific to projects, and (b) relationships that are common to all projects. With a model of (b) in hand, impact analysis of a project could focus on (a) and then make use of the results already available from (b). A model of (b) is then auxiliary to (a).

This paper gives a partial specification of (b) using the 1973 National Demographic Survey (NDS), which data is incomplete for the purposes of comprehensive project impact estimates.

2. Data and Notation

The data is from the 1973 NDS of over 8,000 households. Our sample

---

[1/] Ms. Elizabeth Jacinto did the computations for this paper using the UPSE Computer.

size of 3,196 was obtained by selecting households satisfying the following criteria: the family is nuclear or extended vertically to the younger generation only; household head is male, working, and his noncash income (if any) is less than ₱1,000 annually; the wife married only once and age is 15-44 years; and information is provided on all the variables listed below.

AGn = 1 if wife is in age-group n, 0 otherwise,

where n = 4 if age is 15-19 years

5 if age is 20-24 years

6 if age is 25-29 years

7 if age is 30-34 years

8 if age is 35-39 years

9 if age is 40-44 years

AM = age of marriage of wife, in years

CEB = number of children born live

CMR = CND ÷ CEB

CND = number of children born live and now dead

DCM = 1 if CND > 0, 0 otherwise

DLR = 1 if rural residence, 0 otherwise

DMW = 1 if wife is a migrant whose place of residence is the same as in 1965 and different from place of birth, 0 otherwise

DRC = 1 if wife is Roman Catholic, 0 otherwise,

DWP = 1 if wife is working, 0 otherwise

EWm = 1 if wife has educational level m, 0 otherwise,

where m = 0 for no schooling

1 for one to four years of school

2 for five to seven years of school

3 for one to three years of high school

4 for high school graduate

5 for one to three years of college

6 for college graduate

MWk = 1 if wife is in category k, 0 otherwise,

where k = 0 for DMW = 0

1 for DMW = 1 and agricultural residence

2 for DMW = 1 and nonagricultural residence

PWRj = 1 if wife is in category j, 0 otherwise,

where j = 0 for DWP = 0

1 for DWP = 1 and place of work is at home

2 for DWP = 1 and place of work is away from home

YHi = 1 if husband's annual income is in category i, 0 otherwise,

where i = 1C for cash income less than ₱1000 and noncash income (if any) less than ₱1000

2C for ₱1000-2999 cash income and noncash income (if any) less than ₱1000

3 for ₱3000-4999 cash income

4 for ₱5000-6999

5 for ₱7000-9999

6 for ₱10000 and above

YWi defined the same way as YHi but with respect to wife's income

YW = 1 if YW1C = 1

2 if YW2C = 1

4 if YW3 = 1

6 if YW4 = 1

8 if YW5 = 1

11 if YW6 = 1, in thousand pesos.

Individual income data in the 1973 NDS is reported only in brackets and family income as such is not given. We therefore do not use a family income variable as this would have involved too many categories or else a summing of individual incomes by taking the midpoints of categories as estimates of individual incomes. While the latter procedure is of course possible (cf. Canlas and Encarnación, 1977), it makes income data appear more precise than may be warranted.

The means of the variables in the sample are given in Table 1.

Table 1. Means of the Variables

AG4: 0.0185	CND: 0.4143	EW3: 0.1126	PWR2: 0.1805
AG5: 0.1302	DCM: 0.2663	EW4: 0.0726	YH1C: 0.3457
AG6: 0.1990	DLR: 0.6815	EW5: 0.0379	YH2C: 0.4562
AG7: 0.2362	DMW: 0.2700	EW6: 0.0660	YH3: 0.1270
AG8: 0.2280	DRC: 0.8483	MEO: 0.7300	YH4: 0.0291
AG9: 0.1881	DWP: 0.2447	MW1: 0.1549	YH5: 0.0217
AM: 19.666	EW0: 0.0685	MW2: 0.1151	YH6: 0.0203
CEB: 4.8276	EW1: 0.2735	PWR0: 0.7553	
CMR: 0.0571	EW2: 0.3689	PWR1: 0.0641	

3. The Model

This is based on an earlier paper (Encarnación, 1979) which presented a model of choice where wife's fertility and her labor force participation are determined simultaneously by her educational level, husband's income, and other variables. Briefly, the model implies the existence of "threshold values" for wife's educational level, husband's income and family income, such that the qualitative effect of a variable changes when it passes the thresholds. Figure 1 illustrates.

In the upper panel of Figure 1, number of children C is measured on the vertical axis while family income Y and wife's educational level E--Y and E are assumed to be perfectly correlated for purposes of a simple diagram--are measured on the horizontal axis. Natural fertility or capacity to bear children C_k increases with Y and E for reasons of better health and nutrition; number of child deaths C_m decreases for opposite reasons. The number of children desired C^o falls with E and Y for a variety of reasons. What would then be observed for the number of children born would be the curve abC_b, and the number of surviving children the curve cdC^o. These two variables are thus nonmonotonic functions of E and Y, whose qualitative effects change at the threshold value E_c*.

In the lower panel, the proportion of wife's time spent at market work t is measured on the vertical axis while E and husband's income Y_h are measured on the horizontal axis. The wife's wage rate depends on E and we assume that the curve $c'd'e't'$ tells what is required of t if minimum consumption standards for the family are to be met. On the other hand, the curve $ee't^o$ indicates what t would be if the wife's choice

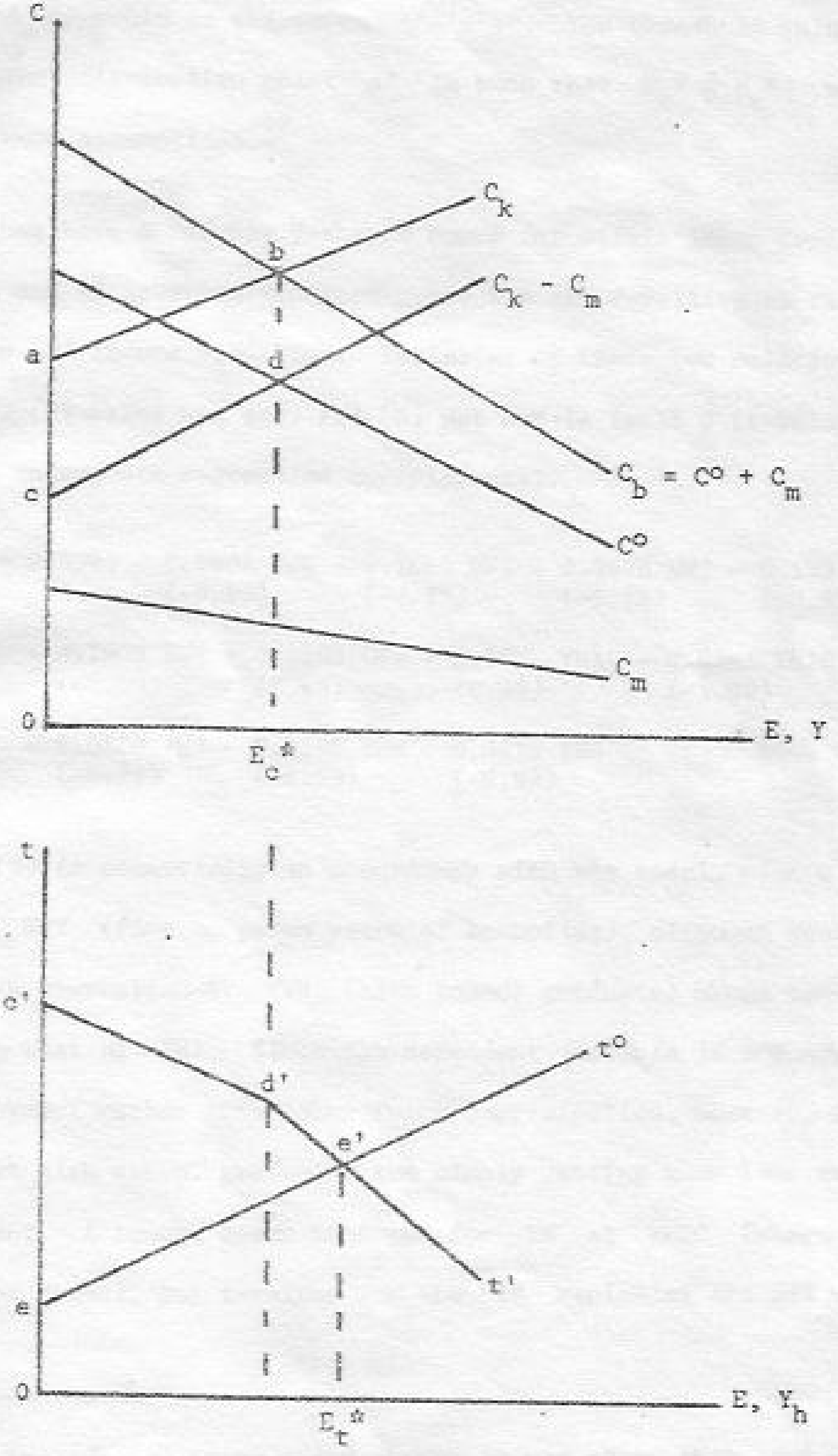

Fig. 1.

were not required to satisfy consumption standards. With this requirement, the observed t would be the curve c'd'e't°. The threshold value E_t* defined by the intersection point e' is such that $E_c^* \leq E_t^*$ under relatively weak assumptions.[2/]

We thus have a roughly V-shaped curve for wife's labor force participation rate and an inverted V-shaped curve for her fertility as functions of education and income variables. Estimates of these two relationships are given in eq. (1) below and eqs. (2)-(5) set out in Table 2 (t-values in parentheses underneath regression coefficients).

$$
\begin{aligned}
(1)\quad DWP = 0.3949 &- \underset{(-0.83)}{0.0400}\,EW0 - \underset{(-3.97)}{0.1642}\,EW1 - \underset{(-5.15)}{0.2078}\,EW2 - \underset{(-3.92)}{0.1722}\,EW3 \\
&- \underset{(-4.09)}{0.1900}\,EW4 + \underset{(5.63)}{0.2685}\,EW6 + \underset{(0.96)}{0.0521}\,YH1C - \underset{(-1.02)}{0.0541}\,YH2C \\
&- \underset{(-0.22)}{0.0120}\,YH3 - \underset{(-0.29)}{0.0189}\,YH4 - \underset{(-0.95)}{0.0675}\,YH6 \qquad \bar{R}^2 = 0.084
\end{aligned}
$$

Eq. (1) is essentially in accordance with the model, with a trough for EW at EW2 (five to seven years of schooling), although there is apparently an aberration at EW4 (high school graduate) whose coefficient is less than that of EW3. Since the dependent variable is a dummy for actual employment rather than labor force participation, however, it is possible that high school graduates are simply getting much less employment than they want. A trough seems to occur for YH at YH2C (where a poverty line could be drawn), but t-values for the YH variables are all weak.

[2/] In the earlier paper cited above, it was shown that $E_c^* = E_t^*$ under somewhat stronger assumptions.

Table 2. Regression Equations for CEB

	(2)	(3)	(4)	(5)
const.	11.543	11.457	11.539	11.454
AM	-0.2962 (-34.01)	-0.2960 (-33.94)	-0.2956 (-33.88)	-0.2954 (-33.81)
AG4	-6.1980 (-25.14)	-6.1899 (-25.04)	-6.1966 (-25.13)	-6.1854 (-25.02)
AG5	-4.7382 (-41.33)	-4.7555 (-41.36)	-4.7369 (-41.32)	-4.7528 (-41.34)
AG6	-3.0293 (-30.63)	-3.0392 (-30.66)	-3.0276 (-30.61)	-3.0362 (-30.63)
AG7	-1.3509 (-14.47)	-1.3539 (-14.50)	-1.3517 (-14.47)	-1.3544 (-14.49)
AG9	1.0110 (10.25)	1.0189 (10.32)	1.0105 (10.23)	1.0181 (10.31)
EW0	0.3835 (1.84)	0.3291 (1.54)	0.3688 (1.76)	0.3238 (1.51)
EW1	0.6199 (3.52)	0.5600 (3.07)	0.6055 (3.42)	0.5547 (3.04)
EW2	0.4615 (2.67)	0.4012 (2.26)	0.4505 (2.60)	0.3976 (2.34)
EW3	0.4606 (2.43)	0.3981 (2.07)	0.4591 (2.42)	0.4023 (2.09)
EW4	0.0913 (0.45)	0.0533 (0.26)	0.0929 (0.46)	0.0587 (0.29)
EW6	-0.0844 (-0.41)	-0.0160 (-0.08)	-0.0941 (-0.45)	-0.0288 (-0.14)
YH1C		0.1035 (0.44)		0.0901 (0.38)
YH2C		0.2065 (0.89)		0.2016 (0.89)
YH3		0.1095 (0.46)		0.1134 (0.48)

Table 2 (continued)

	(2)	(3)	(4)	(5)
YH4		-0.1025 (-0.36)		-0.1025 (-0.36)
YH6		-0.3176 (-1.03)		-0.3046 (-0.98)
DRC	0.1939 (2.13)	0.1894 (2.08)	0.1936 (2.13)	0.1888 (2.07)
DMW	0.1956 (2.71	0.1898 (2.62)		
DWP	-0.2548 (-3.31)	-0.2455 (-3.19)		
MW1			0.2561 (2.87)	0.2461 (2.76)
MW2			0.1094 (1.07)	0.1070 (1.03)
PWR1			-0.3104 (-2.38)	-0.3061 (-2.34)
PWR2			-0.2308 (-2.62)	-0.2193 (-2.47)
$\bar{R}^2$	0.563	0.563	0.562	0.563

In Table 2, eqs. (2)-(5) for number of children born live are set out as columns; (2) omits the YH, MW and PWR variables while the other three equations similarly omit some of the variables listed in the first column. In all four equations, a peak for EW is seen at EW1 (one to four years of schooling). A peak for YH is apparent at YH2C in (3) and (5) even though t-values are weak. (The model calls for family income here and not husband's income, but we use the latter as a rough proxy.)

Of interest are the dummies for religion (DRC), migrant status (DMW) and current employment (DWP), which are all significant. Apparently, looking at (2) and (3), being a Catholic adds about 0.2 children to a couple, as also being a migrant, while being employed reduces family size by 0.25. A closer look at the migrant and employment variables shows finer detail.

Eqs. (4) and (5) use MW1, MW2, PWR1 and PWR2 in place of the more crude DMW and DWP. Here it is migrants to agricultural areas (who are likely to have come from other agricultural areas) who have higher fertility, while other migrants exhibit a smaller increase not significantly different from zero. This would be consistent with the model if one considers that agricultural migrants probably improve their livelihood relatively more than do other migrants. (Cf. Encarnación, 1977, for similar suggestive results in Southeast Asia.) As for the employment dummy variable, a breakdown of this to PWR1 and PWR2 gives results that appear to go against usual expectations. Here we find that ceteris paribus, wives who work at home have apparently less children than those whose place of work is away from home. A possible explanation may be that wives working at home find it difficult to hold down a job away from home because of poorer health; their working at home and having less children would then be due to the same set of circumstances.[3/]

As the CEB equations involve age-at-marriage AM, we have the following equation:

$$\text{(6)}\quad AM = 21.575 - \underset{(-4.45)}{1.9811\ EW0} - \underset{(-5.15)}{1.9765\ EW1} - \underset{(-5.13)}{1.9149\ EW2} - \underset{(-3.70)}{1.4952\ EW3}$$

$$+ \underset{(0.19)}{0.0831\ EW4} + \underset{(4.91)}{2.1397\ EW6} - \underset{(-4.63)}{0.7416\ DLR} \qquad \bar{R}^2 = 0.101$$

This is quite in conformity with usual expectations, with AM advancing with EW and lower for rural women.

Finally, the NDS data permit the estimation of several equations concerning child mortality. Eqs. (7) and (8) below have CMR, the ratio of child deaths to children ever born, as a function of educational level and, in the case of the latter equation, of husband's income (as proxy for family income) as well. Its relationship to these two variables is generally

$$\text{(7)}\quad CMR = 0.0395 + \underset{(5.04)}{0.0775\ EW0} + \underset{(3.21)}{0.0423\ EW1} + \underset{(2.17)}{0.0281\ EW2} + \underset{(1.10)}{0.0158\ EW3}$$

$$- \underset{(-1.18)}{0.0028\ EW4} - \underset{(-1.13)}{0.0175\ EW6} \qquad \bar{R}^2 = 0.023 \qquad F = 13.76$$

$$\text{(8)}\quad CMR = 0.0291 + \underset{(4.38)}{0.0692\ EW0} + \underset{(2.54)}{0.0347\ EW1} + \underset{(1.66)}{0.0221\ EW2} + \underset{(0.77)}{0.0111\ EW3}$$

$$- \underset{(-0.41)}{0.0062\ EW4} - \underset{(-0.99)}{0.0155\ EW6} + \underset{(1.41)}{0.0251\ YH1C} + \underset{(0.65)}{0.0115\ YH2C} + \underset{(0.93)}{0.0167\ YH3}$$

$$- \underset{(-0.11)}{0.0023\ YH4} - \underset{(-0.04)}{0.0009\ YH6} \qquad \bar{R}^2 = 0.025 \qquad F = 8.34$$

[3/] However, it should be noted that if the regression coefficients of PWR1 and PWR2 are treated as means in a standard test of the difference between two means, we find that the difference between them is not large enough to reject the null hypothesis.

monotonic as one might expect. Perhaps more useful, however, are eqs. (9) and (10), where the dependent variable DCM is a dummy equal to one if a child has died. DCM could be interpreted as the probability (approximately)

$$\text{(9)} \quad \begin{aligned} DCM = {} & 0.7637 - \underset{(-11.54)}{0.0233\,AM} - \underset{(-6.75)}{0.3851\,AG4} - \underset{(-11.66)}{0.3069\,AG5} - \underset{(-9.19)}{0.2099\,AG6} \\ & - \underset{(-4.90)}{0.1060\,AG7} + \underset{(3.21)}{0.0734\,AG9} + \underset{(3.38)}{0.1608\,EW0} + \underset{(2.42)}{0.0985\,EW1} + \underset{(1.58)}{0.0629\,EW2} \\ & + \underset{(0.73)}{0.0322\,EW3} - \underset{(-0.35)}{0.0162\,EW4} - \underset{(-1.02)}{0.0485\,EW6} \qquad \bar{R}^2 = 0.122 \end{aligned}$$

$$\text{(10)} \quad \begin{aligned} DCM = {} & 0.7530 - \underset{(-11.44)}{0.0231\,AM} - \underset{(-9.96)}{0.3976\,AG4} - \underset{(-11.81)}{0.3116\,AG5} - \underset{(-9.38)}{0.2145\,AG6} \\ & - \underset{(-4.98)}{0.1076\,AG7} + \underset{(3.27)}{0.0747\,AG9} + \underset{(2.64)}{0.1288\,EW0} + \underset{(1.63)}{0.0684\,EW1} + \underset{(0.94)}{0.0383\,EW2} \\ & + \underset{(0.26)}{0.0117\,EW3} - \underset{(-0.67)}{0.0313\,EW4} - \underset{(-0.77)}{0.0371\,EW6} + \underset{(1.08)}{0.0587\,YH1C} + \underset{(0.43)}{0.0229\,YH2C} \\ & + \underset{(0.49)}{0.0272\,YH3} - \underset{(-0.60)}{0.0398\,YH4} - \underset{(-1.28)}{0.0915\,YH6} \qquad \bar{R}^2 = 0.124 \end{aligned}$$

of a child death as a function of the variables on the right-hand side. It could then serve as a crude proxy for health.

4. Using the Model

With due caution, one can use the regression equations reported above[4/] for purposes of estimating the impact of a development project on some variables of concern: fertility, labor force participation, and health. Accepting the usual interpretation of cross-section regression results as long-term relationships among the variables, the procedure would be simply

---

[4/] These are ordinary least-squares estimates since the model can be taken as recursive.

the following: Calculate the changes in the "independent" variables resulting from the project; then use the regression equations to estimate the changes in the dependent variables. The latter changes are then imputable to the project as its impact.

For example, suppose that one long-term effect of a project is to raise male family heads' incomes in the region from YH1C to YH2C. From eqs. (1), (5) and (10), the coefficients of these two dummies and the differences between them are given in the following table:

Table 3. Coefficients of YH1C and YH2C from Eqs. (1), (5) and (10)

	YH1C	YH2C	Difference
(1) DWP	.0521	-.0541	-.1062
(5) CEB	.0901	.2016	.1115
(10) DCM	.0587	.0229	-.0358

Accordingly, we obtain estimates of a reduction in wives' labor force participation, an increase in births and a decrease in child deaths by multiplying the last column in Table 3 by the number of families involved. Comparability among projects can then be had by expressing the estimates per peso of project costs.

If a project affects other "independent" variables, similar computations can be made and then added to get the total impact of a project, since the effects of the independent variables are additive in the regression equations.

Several observations might be made regarding the estimates so obtained. First, they are not predictions of changes between the present and the future (after the installation of a project), since some changes will occur with or without the project. One is here simply estimating the difference that a project makes, ceteris paribus. Second, the estimates have to do with long-term, not short-term, results. Finally, it is on the basis of some model which one considers correct that one justifies any particular interpretation of statistical observations--one cannot discuss the latter in a theoretical vacuum. This last would not be worth remarking were it not that statistical data are sometimes erroneously thought to be capable of "establishing" a causal relationship or empirical generalization.

References

Canlas, D. B. and J. Encarnación, 1977. "Income, Education, Fertility and Employment: Philippines 1973," Philippine Review of Business and Economics 14, no. 2, pp. 1-27.

Encarnación, J., 1977. "Population and Development in Southeast Asia: A Fertility Model," Philippine Economic Journal 16, no. 4, pp. 319-40.

______________, 1979. "Fertility Behavior and Labor Force Participation: A Model of Lexicographic Choice," U.P. School of Economics Discussion Paper 7908 (rev. July 1979).