Bulletin Number 87-9

July, 1987

ECONOMIC DEVELOPMENT CENTER

ESTIMATING THE INTRAFAMILY INCIDENCE OF HEALTH: CHILD ILLNESS AND GENDER INEQUALITY IN INDONESIAN HOUSEHOLDS

Mark M. Pitt

Mark R. Rosenzweig

ECONOMIC DEVELOPMENT CENTER

Department of Economics, Minneapolis

Department of Agricultural and Applied Economics, St. Paul

UNIVERSITY OF MINNESOTA

Estimating the Intrafamily Incidence of Health: Child Illness and Gender-Inequality in Indonesian Households

Mark M. Pitt

and

Mark R. Rosenzweig

University of Minnesota

July 1987

Abstract

In this paper, we demonstrate the difficulties of identifying both the own- and cross-effects of health on the allocation of time within a household, and develop and implement a method for estimating the effects of infant morbidity on the differential allocation of time by other family members based on discrete indicators of health and of activity participation commonly available in survey data. Estimates obtained from Indonesian household data indicate that inattention to problems of the measurement and endogeneity of health leads to a substantial underestimate of the effects of variations in child morbidity on the intrahousehold division of labor, and our estimates that take into account the "simultaneity" of health-activity associations indicate that increased levels of infant morbidity significantly exacerbate existing differentials in work-home time allocations across teenage boys and girls in Indonesia.

Mark M. Pitt
Mark R. Rosenzweig
Department of Economics
1035 Management & Economics
University of Minnesota
Minneapolis, MN 55455
USA

In recent years, a number of studies have provided quantitative evidence on the impact of a person's health on his or her time allocation, and on productivity, farm profits and wage rates in both developing and developed countries (Bartel and Taubman (1979), Lee (1982), Sickles and Taubman (1986), Strauss (1986), Pitt and Rosenzweig (1985)). Not all of the costs of ill health, however, are borne by the individual whose health is temporarily or permanently impaired. Within a household, the ill health of one person is likely to evoke resource adjustments by other persons in the household in which the illness occurs. For example, family members may spend more time with children when they are ill. To the extent to which time in family-provided health "care" is transferred from activities that contribute to the long-term health or development of those persons caring for the family member in bad health, estimates of the direct effects of health initiatives on health understate the beneficial role of such programs. Few estimates exist, however, of the consequences of illness for the composition of household activities.

In many developing countries, older children appear to contribute importantly to child care along with the mother.[1] In such settings, among older children, schooling, employment in productive activities, and household care are competing activities. Table 1 provides a description of the principal activities of Indonesian children in age groups 10-13 and 14-18, of mothers and of household heads in households in which there is an infant aged less than two years, obtained from a 1980 national probability survey (SUSENAS, 1980). There are three striking features of Table 1. First, almost all male heads in such households are in the labor force. Second, almost 25 percent of teenage girls aged 14-18 in households with

Table 1

Division of Labor by Household Membership, Sex and Age in Households with Children Less than Two Years Old[a]

Major Activity	Children Age 10-13 Boys	Children Age 10-13 Girls	Children Age 14-18 Boys	Children Age 14-18 Girls	Mothers	Male Heads
Labor force	3.8[b]	4.0	28.7	15.9	24.2	97.4
School	91.2	86.5	60.7	51.1	0.2	0.2
Household care	0.8	5.2	2.9	24.6	75.2	0.1
Other	4.2	4.3	7.7	8.3	0.3	2.3

a. Total number of households = 10,231.
b. Percent of group in activity.

infants have as their principal activity household care. Third, girls in the 14-18 year age group are 8 times more likely than similarly-aged boys to be engaged in household activities, and are 16 percent less likely than boys to be in school.[2]

A natural question is whether changes in levels of infant morbidity would significantly affect the disparities in the intrafamily allocation of activities by sex observed in a country like Indonesia, and thus whether observed gender inequality in human capital could be reduced through improvements in child health. Identification of the "third party" incidence of health--the effects of the health of person i on the behavior of person j--when the behavior of j may directly affect i's health (e.g., time in child care), however, is not straightforward. For example, it is difficult to find instruments that directly influence i's health but not that of j, net of the health of i. Perhaps for this reason, the few existing studies (e.g. Salkever (1982), Bartel and Taubman (1986)) of intrafamily health effects have not attempted to discriminate between the time-allocative consequences and causes of ill health.

In this paper, we demonstrate the difficulties of identifying both the own- and cross-effects of health within a household, and develop and implement a method for estimating the effects of infant health on the differential allocation of time by other family members using data from Indonesia. We are able to estimate how mothers and teen-age girls and boys reallocate their activities relative to each other in response to an infant's health and are able to test whether, in particular, the division of activities among teenagers by sex is influenced by the demand for child care responsibilities associated with infant morbidity. We also test whether the infant's gender and health programs, net of their direct effects on the

health of infants and children, alter the intrahousehold allocation of activities.[3] Section 1 contains a simple model used to illustrate the identification problem. Section 2 describes the data and the econometric methodology we use to identify the effects of child health on the family activity distribution in a context in which both the endogenous health and activity indicators are measured by discrete variables. Section 3 presents estimates, based on Indonesian household data, of the differential responses of family members to infant health. The results indicate that inattention to problems of the measurement and endogeneity of health leads to a substantial underestimate of the effects of variations in child morbidity on the intrahousehold division of labor, and our estimates that take into account the "simultaneity" of health-activity associations indicate that increased levels of infant morbidity significantly exacerbate existing differentials in work-home time allocations across teen-age boys and girls in Indonesia. Section 4 contains a summary and conclusion.

1. Household Activities and the Effects of Health

Consider a set of households each consisting of n member "types" (defined, say, by sex, age, etc.), in which the total time Ω of each household member i in household t is allocated between two activities, work (for pay) and home time ℓ_{it}. The home time of every member of household t influences jointly the healthiness of each member H_{it}, along with health goods X_{it} allocated to i. Each member's home time is thus a public good, while X_{it} is a "private" good. This private-public distinction is not essential to the model, as discussed below, but simplifies notation. The health technology is thus given by:

$$(1) \quad H_{it} = h(\ell_{1t},\ldots\ell_{nt},X_{it},H_{1t},\ldots H_{i-1t},H_{i+1t},\ldots,H_{nt})$$

where (1) also incorporates the possibility of intrafamily health externalities. Assume also that the household welfare function contains the home time and health of each individual member and a jointly consumed commodity Z_t. Households are heterogeneous in their preferences for these "goods," as expressed by the parameter ϵ_t, such that:

$$(2) \quad U_t = U(\ell_{it},\ldots,\ell_{nt},H_{1t},\ldots H_{nt},Z_t;\ \epsilon_t)$$

The household budget constraint is

$$(3) \quad F_t + \sum_i(\Omega-\ell_{it})\ w_{it} = p_x\sum_i X_{it} + p_z Z_t$$

where w_{it} = market wage rate of person-type i in household t, p_x = price of the health good, p_z = price of Z, and F_t = non-earnings income in household t.

Maximization of (1) subject to (2) and (3) yields n necessary first order conditions for the allocation of the home time of each household member:

$$(4) \quad \frac{\partial U}{\partial \ell_i} + \sum_k \frac{\partial U}{\partial H_k}\frac{\partial H_k}{\partial \ell_i} = \lambda w_i$$

where it can be seen that i's time in the home ℓ_i depends on the (marginal) effects of his (her) time on the healthiness of each family member. The n household demand equations for members' home time, solved in terms of all of the exogenous variables of the model, are given by

$$(5) \quad \ell_{it}\ \ell(p_x,p_z,w_i,w_1,\ldots w_{i-1},w_{i+1},\ldots w_n,\ F_t,\epsilon).$$

Note that if home time were a private good, allocatable to the production of health for each individual household member, there would be n^2 demand equations for the ℓ_{ikt}, i=1,...,n; k=1,...,n, with $\ell_{it} = \sum_k \ell_{ikt}$

Now consider the conditional demand equation for a particular ℓ_i in which the home time for i is solved, conditional on the health good $\bar{X}_j$ allocated to household member j, in terms of all of the remaining exogenous variables that can influence the allocation of resources other than $\bar{X}_j$ and other than via income effects:

$$(6) \quad \ell_{it} = \ell^c(p_x, p_z, w_i, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots w_n \; F_t^*, \; \bar{X}_{jt}, \; \epsilon),$$

where $F_t^* = F_t - p_x\bar{X}_j$. Conditional demand equation (6), in contrast to conditional demand equations from standard commodity demand systems in which there is a unique price for every commodity, contains all of the prices contained in the "reduced form" or unconditional demand equation (5). In particular, even when member j's health good is held constant, movements in the price of the health good X can still induce substitution effects with respect to member i's allocation of home time (or allocation of time devoted to each member's health production in the private good case), since a change in p_x may influence the allocation of the remaining X goods and thus the productivity of i's time in the production of other members (inclusive of i). There is no price that uniquely corresponds to the health good allocated to (any) family member j.

Because the number of commodities in the model (n person-specific time "goods," n health goods, X and Z) exceeds the number of prices (n wage rates, one p_z and one p_x), identification of the influence of j's healthiness on i's allocation of time cannot be achieved, when there is heterogeneity in household preferences, by the usual strategy of utilizing the price of the conditioned endogenous commodity (or its proxy) as an instrument. This same problem arises, of course, for estimating the own effect of health; there is no (exogenous) price that directly affects i's

health that does not also affect i's time allocation conditional on i's health, given that i allocates some of his/her time to the production of the health of other family members.[4] Only if there are no interdependencies among individuals (as in (1) and (2)), would p_x be a valid health instrument, i.e. if all individuals in a population were independent. But in that case, there are no cross-health effects other than via health externalities (contagion).

How has this identification problem been solved in prior work on the behavioral consequences of health? In Bartel and Taubman's (1986) study of mental health cross-effects between husbands and wives, healthiness is treated as an exogenous variable, uninfluenced by a spouse's work time or labor force participation; that is, the ℓ_{kt}, k=1,...,n, are absent in (1). Similarly, in Salkever (1982), the only study we know of to estimate the cross effects of child health, child health is assumed in the theoretical model to be influenced by mother's care but to be exogenous to mother's work time in the empirical analysis. In Pitt and Rosenzweig (1986), the health of the male head of the household is treated as (and is confirmed to be) an endogenous variable affected by and affecting his own supply of labor, and variables corresponding to health programs and sanitation conditions are employed as instruments. But what is interpreted to be the own effect of the head's health on his labor supply cannot be correct if the head's time in the market responds to say, his wife's health and wife's health is, as is likely, correlated with head's health.

To identify the consequences of changes in health in a household context in which health is influenced by behavior thus requires that additional model structure be imposed, in the absence of exogenous person-specific health prices. One strategy, for example, which readily permits

identification of the differential health effects on family members' time allocation, employed below, is to impose equality restrictions (rather than zero restrictions inconsistent with a general household model) on the effects of subsets of exogenous variables on subsets of person-specific behaviors. For example, consider linearized versions of the conditional demand equation (6) for two family members i and j:

$$(7) \quad \ell_i = \alpha_{0i} + \alpha_1 w_i + \alpha_2 \psi_i(w_j, W) + \alpha_{3i} p_z + \alpha_{4i} p_x + \alpha_{5i} H_k + \alpha_{6i} \epsilon$$

$$(8) \quad \ell_j = \alpha_{0j} + \alpha_1 w_j + \alpha_2 \psi_j(w_i, W) + \alpha_{3j} p_z + \alpha_{4j} p_x + \alpha_{5j} H_k + \alpha_{6j} \epsilon$$

where $\psi(\ )$ is some summary statistic for the wage rates of all household members and W = vector of members' wages other than i. If, net of the health of family member k, the effects (α) of a change in the price of, say, p_x on the time allocation of i and j are the same (e.g., if the health production function for i and j is the same), then, with $\alpha_{4i} = \alpha_{4j}$ the ij difference equation would be

$$(9) \quad \ell_i - \ell_j = \alpha_{0i} - \alpha_{0j} + \alpha_1(w_1 - w_j) + \alpha_2(\psi_i - \psi_j) + (\alpha_{3i} - \alpha_{3j}) p_z + (\alpha_{5i} - \alpha_{5j}) H_k + (\alpha_{6i} - \alpha_{6j}) \epsilon$$

and p_x, which affects the level of individual k's health, could serve as an identifying instrument. Restrictions based on the equality of paramters across household member types are thus one means of estimating how the healthiness of a family member influences the time allocation of the family members.

2. Data and Estimation Procedure

Identifying the effects of child health on the intrafamily allocation of time based on cross-member and activity-specific equality restrictions

imposes data requirements that are severe - information on child health, on the activities of all household members, on the prices of health goods, as well as a large sample size are required. Sample size is important since the existence of differences across teen-age siblings and the mother within families is used to identify the time-allocative consequences of child morbidity.

The 1980 National Socioeconomic Survey (SUSENAS) of Indonesia provides information on household-level food consumption and on the incidence of illness during the past year for all children less than 10 years of age in approximately 65,000 households distributed throughout the country. All members of households in the survey also provide information on their principal activity in the last week, divided among labor force, school, household care, and other, and on their age, sex, relationship to head, educational attainment, employment time (if any) and wage rates.[5]

The 1980 SUSENAS also identifies the county (kabupaten) in which each household resides. We constructed from the household data, kabupaten price medians for 14 individual foods and price indices for 7 food groups based on kabupaten average consumption shares. Prices vary substantially in the cross-section even for homogenous goods due to the island geography of Indonesia and its relatively underdeveloped infrastructure (Pitt and Rosenzweig (1986). We also obtained information on village-level health programs, drinking water sources, and sanitation facilities, which were aggregated to kabupaten village frequencies.[6] There are thus 525 areas defined in the data set, based on rural-urban distinctions within the 300 kabupatens. All of these kabupaten-level variables are potential instruments in an analysis in which differential activity effects are estimated, since all may potentially affect the levels of infant illness

(health), but may not <u>differentially</u> influence the time allocation of the family members net of their effects on infant health.

Since almost all children aged 6 to 10 are in school (or report that they are), to examine the impact of child health on the activities of older children requires that we look at the effects of the health of children less than 10 (since no health information is provided for children 10 years of age or older) on the time allocation of family members aged 10 and above. In particular, we will examine the effects of the health of infants aged less than 4 years on the activities of their siblings aged 10-18 and mother (since, as Table 1 indicates, virtually all heads are in the labor force).

Two important features of the data are that (i) all of the activity variables and the measure of health (sick or not) are discrete and (ii) health or illness refers to the past year while the activities correspond to the last week. Our choice of econometric technique is influenced both by the nature of the data and by the restrictions implied as sufficient by the model to identify the within-family activity effects of infant illness.

As noted above, our strategy, which permits identification of the consequences of health in a household context, is to impose equality restrictions on the effects of subsets of exogenous variables on subsets of person-specific behaviors. Consider linearized versions of the conditional activity equations for family members in a representative household containing a mother and her teenaged son and daughter:

$$(10)\quad I^*_{ij} = (\alpha_{iM} + \delta^A_{ij} D_j)' A_j + (\gamma_{iM} + \delta^h_{ij} D_j) h^* + (\beta_{iM} + \delta^X_{ij} D_j)' X + Z\lambda_{ij} + \epsilon_{ij}$$

where I^*_{ij} is the level at which household member j undertakes activity i; D_j has the value of one in the equation for j and zero otherwise in the son

and daughter equations; h^* is the endogenous health of the mother's infant child; A is a vector of member-specific exogenous variables; X and Z are vectors of household-specific exogenous variables, to be distinguished below; a, β, γ, δ and λ are vectors of coefficients, and ϵ_{ij} are error terms having a multivariate distribution with zero means and covariance matrix Σ. Furthermore, ϵ_{ij} is the sum of a household specific component υ_i and a person-specific component ψ_{ij}

$$(11) \quad \epsilon_{ij} = \upsilon_i + \psi_{ij}.$$

The household-specific term ν_i reflects unobserved variation in the health environment of households, the infant health technology and household preferences. The person-specific component ψ_{ij} represents unobservable individual attributes such as ability and health endowments.

The reduced form equation for the healthiness of the mother's infant is

$$(12) \quad h^* = X\Pi_x + Z\Pi_z + A_j\Pi_a + \eta$$

where $\Pi = (\Pi_x, \Pi_z, \Pi_a)$ is a vector of reduced-form parameters and η is an error term. The vector of exogenous variables Z are those for which equality restrictions are imposed:

$$(13) \quad \lambda_{ij} = \lambda_{ik} \quad i \neq k;\ i,k = \text{member type}$$

Identification requires that there are at least as many variables in Z as endogenous regressors.

Neither the health h^* nor the activity variable I^*_{ij} is observed in our data. What is observed is a set of three dichotomous indicators (I_{1j}, I_{2j}, I_{3j}) indicating which activity is the primary activity of the household member, and a dichotomous indicator for h^*. Estimation of the

parameters of the system (10) - (13) is thus complicated by the discrete nature of all of the endogenous variables, the endogeneity and discreteness of one of the regressors, and the selection of households - those having the mother of an infant of age 3 or below and two siblings 10 to 18 years of age and of opposite sex resident in the household. One approach to estimating the system is to extend Mallar's (1977) simultaneous equations probit model to the multivariate case with sample selection. Even for the simple case combining the binary choice of time activities among two individuals with sample selection, however, the likelihood is computationally intractable, involving the evaluation of the trivariate normal cumulative distribution. In addition, this likelihood would need to simultaneously estimate a large number of parameters: the parameters of the system (10) - (13) plus an additional set of parameters to account for selection.

Using a fixed effects approach such as in Chamberlain (1980) eases the computational problem greatly. First, note that the identifying restrictions (13), the equality restrictions imposed on the effects of subsets of exogenous regressors (Z) on <u>person-specific</u> behaviors, imply that the set of regressors Z have zero coefficients in a differenced conditional time allocation equation. By differencing the conditional time allocation equations for any pair of household member types (say daughter and sons), we thus reduce the number of parameters we need to estimate. That we cannot identify the parameters λ_{ij} is not of great concern since it is maintained that they do not <u>differentially</u> affect time allocation and it is differential effects in which we are primarily interested. Likewise, with this technique only the differential effects δ on time allocation are identified for the regressors X and h^* that do not vary across member types. For example, we cannot determine whether a change in an element of X, X_n,

increases or decreases the likelihood that any of the three person-types will be engaged in any of the alternative activities. We will be able to estimate how a change in X_n differentially affects the likelihood of engaging in any time-activity for any pair of person-types. Although identifying level effects on activities - that is the parameters β, γ, and λ - is more informative, the identification of the δ parameters is sufficient to test hypotheses about intra-household distributions.

A second advantage of a fixed effects procedure is that differencing eliminates the sample selection problem. Given a selection rule, the residuals in (10) can be rewritten as the sum of residuals having zero mean and another term which adjusts for the truncation of the distribution of behaviors. As is well known, in the case of normality, this term is proportional to the Mills ratio [Heckman (1974) and Lee (1976)]. An additive term analogous to the Mills ratio makes this adjustment for truncation if the error structure has a logistic rather than a normal distribution. The important point is that the term which adjusts the residual for its truncation is household-specific, that is, it is the same for all household-member types since it is households - rather than household members - that are selected into or out of the sample. These additive adjustment terms vanish when household member type equations are differenced.[7]

The fixed effects logit model, which conditions on sufficient statistics for the incidental parameters, can provide consistent estimates of this model while a model with normally distributed residuals cannot (Chamberlain 1980). Unlike in the conventional uses of fixed effects models, however, we do not assume that all of the parameters are identical across the household members. Our approach consists of estimating a set of

two-stage fixed effects binary logits - one such logit for each alternative time activity and each pair of household member types. Consider the case of two household member-types and one of the activity choices. Following Chamberlain, maximum likelihood estimation of a binary fixed effects logit model is accomplished as a standard logit problem with a rescaled dependent variable. This is equivalent to maximizing the conditional likelihood function, conditioning on a set of sufficient statistics for the incidental parameters υ_i. In the case of two persons (j=1,2) and an activty a, $\Sigma_j I_{aj}$ is a sufficient statistic for υ_i. If $\Sigma_j I_{aj} = 0$ or 2 then both I_{a1} and I_{2a} are determined given their sum. Thus the only case of interest is $\Sigma_j I_{aj} = 1$. The possibilities are $(I_{a1}=1, I_{a2}=0)$ or $(I_{a1}=0, I_{a2}=1)$, and thus the rescaled dependent variable y_{a12} has the value one if the first possibility is true and zero if the second is true. All observations for which $\Sigma I_{aj} = 0$ or 2 do not contribute to the likelihood and are dropped.

This fixed effects probability model can equivalently be written in the common latent variable formulation:

$$(14) \quad I^*_{a1} = (\alpha_{aM} + \delta^A_{a1})'A_1 + (\gamma_{aM} + \delta^h_{a1})h^* + (\beta_{aM} + \delta^x_{a1})'X + Z\lambda_{a1} + \epsilon_{a1}, \quad I_{a1} = 1 \text{ if } I^*_{a1} > 0,\ I_{a1} = 0 \text{ otherwise;}$$

and

$$(15) \quad I^*_{a2} = (\alpha_{aM} + \delta^A_{a2})'\ A_2 + (\gamma_{aM} + \delta^h_{a2})h^* + (\beta_{aM} + \delta^x_{a2})'X + Z\lambda_{a2} + \epsilon_{a2}, \quad I_{a2} = 1 \text{ if } I^*_{a2} > 0,\ I_{a2} = 0 \text{ otherwise.}$$

The fixed effects logit model uses two cases

Case I: $I^*_{a1} > 0$ and $I^*_{a2} \leq 0$,

so that

$$(16) \quad y^*_{a12} = I^*_{a1} - I^*_{a2} = \alpha'_{aM}(A_1 - A_2) + \delta^{A'}_{a2}A_1 - \delta^{A'}_{a2}A_2 + (\delta^h_{a1} - \delta^h_{a2})h^* + (\delta'_{a1} - \delta'_{a2})X + (\psi_{a1} - \psi_{a2}) > 0$$

and

Case II: $I^*_{a1} \leq 0$ and $I^*_{a2} > 0$,

so that

$$(17) \quad y^*_{a12} = I^*_{a1} - I^*_{a2} = \alpha'_{aM}(A_1 - A_2) + \delta^{A'}_{a1}A_1 - \delta^{A'}_{a2}A_2 + (\delta^h_{a1} - \delta^h_{a2})h^* + (\delta'_{a1} - \delta'_{a2})X + (\psi_{a1} - \psi_{a2}) \leq 0.$$

Thus, y^*_{a12} is a latent variable which underlies the rescaled dichotomous indicator y_{a12} resulting in a linear latent variable model statistically identical to that of Mallar's except for the (yet unspecified) error distribution. Extension of Mallar's results to the case of the logistic distribution is relatively straightforward, requiring only that the second-stage compound error $(\psi_{a1} - \psi_{a2}) + (\delta^h_{a1} - \delta^h_{a2})\eta$ have a logistic cumulative distribution.

Estimating the household choice problem as a set of binary fixed effects logits yields a multiplicity of parameter estimates.[8] For example, for any one choice, say, employed or not employed, coefficient vectors are estimated for all possible differenced pairs: daughter-mother, daughter-son and son-mother. However, it must be the case that knowledge of any two of the true coefficient vectors determines the third. This will not be <u>exactly</u> true for small sample estimates. Note that the fixed effects logit conditional likelihoods for each of the activities necessarily involves different samples since those household-member pairs for which $\Sigma_j I_{aj} = 1$ do not contribute to the likelihood for activity a. For example, if the choices are {a,b,c} for the ordered mother, daughter and son triad, this household will not contribute to the daughter-son likelihood for the choice

a versus not a. Likewise, the household will not contribute to the son-mother likelihood for the choice b versus not b, nor the daughter-mother likelihood for the choice c versus not c.

We can overcome the multiplicity of parameters problem and achieve more efficient estimates by carrying out minimum distance estimation as an additional stage in the estimation procedure. With the linear restrictions of the system, minimum distance estimation in this case is equivalent to generalized least squares (GLS) in which the vector of dependent variables is the complete (redundancy inclusive) set of two-stage binary fixed effects logit parameter estimates. Let $\Psi = (\hat{\alpha}', \hat{\delta}')$ where, for example, $\hat{\alpha}' = [(\hat{\alpha}_1 - \hat{\alpha}_2), (\hat{\alpha}_1 - \hat{\alpha}_3), (\hat{\alpha}_2 - \hat{\alpha}_3)]$, and the activity choice subscript is omitted for simplicity. The restrictions are thus of the form: $(\alpha_2 - \alpha_3) = (\alpha_1 - \alpha_3) - (\alpha_1 - \alpha_2)$, so the minimum distance "regressors" form a matrix containing only 0, 1, and -1 as elements.

Unique minimum distance estimates of the parameter differences $\hat{\Upsilon} = (\hat{\alpha}^+, \hat{\delta}^+)$ where, for example, $\hat{\alpha}^+ = [(\hat{\alpha}_1 - \hat{\alpha}_2)^+, (\hat{\alpha}_1 - \hat{\alpha}_3)^+]$, are given by

$$(18) \quad \hat{\Upsilon} = (M'\Omega^{-1}M)^{-1}M'\Omega^{-1}\Psi,$$

where M is a design matrix of restrictions, and Ω is the covariance matrix of the parameter vector Ψ.

Estimation of the covariance matrix Ω is not straightforward, since the endogeneity and latent nature of one of the regressors (h^*) must be accounted for. The formula that we use for computing the asymptotic covariance matrix $\hat{\Omega}$ is derived in the Appendix. The asymptotic covariance matrix of the minimum distance parameters $\hat{\Upsilon}$ is given by

$$(19) \quad \mathrm{Var}(\hat{\Upsilon}) = (M'\hat{\Omega}^{-1}M)^{-1}.$$

In the results presented below, t-ratios based on the usual information matrix formula typically output by a logit estimation program, Ω and Var($\hat{T}$) are presented in order to suggest both the magnitude of the biases resulting from not accounting for the stochastic regressor and the efficiency gained from minimum distance estimation.

In our data the vector of individual specific exogenous variables A consists only of age, in years. The vector X contains the education of the mother and her husband, in years, the predicted market wage of the mother and her husband, and the proportion of villages with public health clinics (PUSKESMAS and polyclinics) in the district (kabupaten) in which the household resides. While the possibility that education and market wages may differentially affect time allocation is obvious we also allow health clinics to differentially affect time allocation, net of the healthiness of the infant. This may be so if such clinics educate households about regimens of infant health care which may be more intensive in the time of one or the other household member-types. For example, if it is presumed that mothers rather than their teen-age siblings interact with health clinics with respect to the health of their infants, and are more adept at understanding and carrying out complex regimens of care, it would then seem likely that the prevalence of health clinics would increase relatively the time of the mother in homecare activities. On the other hand, the ready availability of health clinics may reduce the risk to an infant's health resulting from the substitution of sibling time for that of more "efficient" mother's time, thereby reducing the relative time of mothers in the home.

The vector Z, consisting of exogenous regressors which do not differentially affect the allocation of time among household member-types, net of infant morbidity, contains two subsets of regressors: the 14

commodity prices and 7 price indices and the community characteristics, including programs and health facilities (4 variables), public waste facilities (7 variables) and drinking water sources (5 variables).

3. Results

Table 2 reports, for comparison purposes, the estimates of the differential activity effects of child illness δ^h_{ij} obtained using (i) the standard, maximum-likelihood fixed effects model, which does not take into account the endogeneity of child health; (ii) the two-stage fixed effects model, and (iii) the minimum distance procedure, based on the consistent, two-stage logit parameter and covariance matrix estimates. There are seven estimates for each procedure, corresponding to the three activities and the three member-type pairs (schooling is an activity for only 12 of 5831 mothers). Table A in the appendix reports the first-stage, maximum-likelihood logit estimates of the reduced-form determinants of child illness used in the two-stage and minimum-distance models. In that equation, the sets of area-level food prices ($\chi^2(21) = 141$) and programs and health facilities ($\chi^2(11) = 76.1$) contribute significantly to the value of the likelihood--the environment in which each household resides thus significantly influences the probability that a young child will be ill.

The inconsistent single-stage logit estimates do not indicate any differences in family activity responses to child illness. However, the procedures that take into account the possibility that child illness is influenced by differential activity allocations indicate that variations in the levels of child morbidity significantly alter the division of time across mothers and their older children and across teenage sons and daughters. The test statistics obtained from the minimum-distance system

Table 2

Estimates of the Differential Effects of Child Morbidity on the Activities of Mothers and Teenage Children, by Estimation Procedure

Activity and Member Pair	ML Fixed Effects Logit	ML Two-Stage Fixed Effects Logit	Minimum-Distance	Number Contributing to Likelihood
Household Care				
Daughter vs. Mother	.121 (0.83)[a]	-.550 (3.00) [4.01][b]	-.593 (3.04)	3854
Son vs. Mother	-2.16 (0.60)	-1.24 (3.42) [4.60]	-1.46 (3.48)	4105
Daughter vs. Son	-.598 (1.46)	1.01 (1.25) [2.39]	.866 (2.51)	745
Employment in Labor Force				
Daughter vs. Mother	-.104 (0.66)	.599 (2.17) [3.52]	.549 (2.03)	1612
Son vs. Mother	.0962 (0.70)	.164 (0.75) [1.21]	.0422 (0.18)	1884
Daughter vs. Son	-.409 (1.56)	-.0645 (0.16) [0.25]	.506 (1.85)	979
Schooling				
Daughter vs. Son	.202 (1.03)	-.104 (0.36) [0.52]	-.198 (0.71)	1547

a. Asymptotic t-ratio in parentheses beneath coefficient.

b. t-ratio uncorrected for use of stochastic regressors (from information matrix), in brackets.

estimates indicate rejection of the hypothesis that child illness does not alter the division of time in household care among family members ($\chi^2(5) = 17.4$) or across teenage sons and daughters ($\chi^2(3) = 10.5$). In particular, mothers appear to shift into (out of) home time more (less) strongly than do their teenage daughters and sons when their young children are ill and teenage daughters are more (less) likely to increase (decrease) their time in household care than are their brothers in response to their younger sibling's illness. Moreover, the minimum-distance estimates suggest that child illness lowers the probability of daughters being in school compared to sons, although the differential is not statistically significant. The results thus suggest that reductions in child morbidity would diminish significantly existing sex-based inequalities across teenagers in household activities but only marginally the differentials in the accumulation of human capital acquired via schooling.

The two-stage and single-stage fixed effects logit estimates of differential illness effects are not directly comparable, since illness in the latter is measured by a dichotomous indicator of morbidity, while illness in the former is measured by the predicted latent measure of illness obtained from the first-stage logit illness equation. However, the logit model provides a mapping between a change in the latent index of morbidity and the probability of illness. For example, a one standard deviation decrease in the latent illness variable (.63) would be associated with a decline in the probability of illness, evaluated at its mean level (.52), of 0.15, or 29 percent. Roughly speaking, the estimated single-stage coefficient for illness incidence would have to be approximately 4 times the magnitude of the coefficient estimated using the predicted latent index (assuming both coefficients are of the same sign). However, of the set of

four δ^h coefficients which are estimated with reasonable precision using the predicted latent index, the signs of the single-stage and two-stage estimates differ in all but one case. Where the signs conform (the son-mother household care differential), moveover, the magnitude of the single-stage estimate is less than one-sixth that of the two-stage estimate. The unrestricted two-stage household care estimates also conform reasonably closely to the (unimposed) cross-equation restriction that the difference between δ^h estimated from the daughter-mother pairs and δ^h estimated from the son-mother pairs equals the independently estimated δ^h from the pairings of sons and daughters (1.01 versus .69). The single stage estimates do not conform closely (.337 versus -.598).

To assess the quantitative importance of the estimates of differential morbidity effects on cross-member activity distributions, it is necessary to impose additional restrictions. The effect of a change in child health h on the probability P^i_a of a teenage child i participating in activity a is:

$$(20) \quad \frac{dp^i_a}{dh} = p^i_a(1-p^i_a)\ [\gamma_{aM} + \delta^h_{ai}]$$

While δ^h_{aM}, the level effect of h on the latent activity index of the mother, is not identified, since for each activity a, P^i_a is known and the δ^h_{ai} are estimated, comparisons of probability effects across family members only require knowledge of any one person-specific level effect. Thus, for example, if it is known or thought to be highly likely that teenage sons do not take on child care responsibilities, then it is possible to identify, from our estimates, the level and differential effects of a change in child illness on the probabilities of mothers and daughters allocating their time to household care and to the labor force.

In Indonesia, where our data suggest that less than two-percent of male siblings aged 10-18 are involved in household care, the assumption that teenage boys do not alter their time devoted to household activities in response to the illness of a young sibling may be unheroic, certainly less so than one which imposes an arbitrary value for the response of the mothers' (daughter's) time to her child's (younger sibling's) illness. Our minimum-distance estimates indicate that with no response of teenage boy's time in the home to infant illness, the probability that the mother devotes her time to household care rises by .19 (26 percent) from its mean probability of .71 in response to a one standard deviation decrease in latent child health. The daughter's likelihood of participating principally in household care also increases, by .05 (38 percent) from its mean probability of .13. The corresponding single-stage estimates indicate that the mother's and daughter's rates of participation in home care also rise in response to the equivalent increase in child morbidity, but the change is only .007 and .006 respectively.

Since mother's allocate their time predominantly between only two activities (less than one percent (n=42) report other or school activities) the restriction on the response of the son's participation in household care to his sibling's illness also permits identification of all of the member-specific employment probability responses to child morbidity. The minimum distance estimates indicate that if it is true that teenage sons do not alter their (minimal) home care responsibilities, the one-standard deviation rise in latent morbidity, associated with an increase in the incidence of illness of 29 percent, would decrease the mother's labor force participation rate by .19 (65 percent), decrease the daughter's participation rate by .05, or by one-fourth the mother's decrease, and decrease the son's participation

rate by .11, slightly more than one-half the mother's decrease, at the sample means.

Table 3 reports the minimum-distance estimates of member age effects for the three principal activities and test statistics on differences in age effects across sons and daughters within activities. These estimates approximately replicate the patterns reported in Table 1--while there appears to be no age gradient for mothers, teenage daughters significantly increase their participation in home care as they age, while as sons age they significantly decrease their home care responsibilities. Both sons and daughters' labor force participation rates rise with age, but rise significantly faster for teenage boys. School attendance rates drop significantly with age, but the rates do not differ significantly between teenage sons and daughters. In households that have at least one teenaged child of each sex, the accumulation of skills via formal schooling thus does not appear to be differentiated by sex, but boys acquire significantly more labor market skills and girls significantly more skills associated with household responsiblities as they age, tendencies that are exacerbated evidently where child morbidity is high.

The point estimates in Table 3 indicate that, at the sample means, each year of age increases teenage girls' likelihood of participation in household care by 29 percent, while the probability of taking on household responsibilities <u>decreases</u> by 8 percent for each year their teenage male siblings age. Labor force participation rates rise by .035 percent (39 percent) and by .081 (57 percent), respectively, for girls and boys, while school rates diminish by .15 (21 percent) per year for girls and by .13 (16 percent) per year for teenage boys.

Table 3

Minimum Distance, Two-Stage Fixed Effect Logit Estimates: Mother and Teenage Child Age Effects on Activity Allocations

	Activity		
Age of	Household Care	Employment in Labor Force	Schooling
Mother	.00842 (0.67)[a]	.0143 (1.29)	-- --
Daughter	.380 (12.2)	.462 (14.8)	-.717 (16.0)
Son	-.0110 (0.19)	.708 (20.7)	-.674 (14.7)
Daughter-Son differential (t-statistic)	6.06	5.57	0.77

a. Asymptotic t-ratio in parentheses beneath coefficient.

The minimum-distance estimates of the differential effects of household characteristics on activity rates across mothers, sons and daughters are reported in Table 4. Table 5 contains the associated test statistics for the joint influence of each household characteristic on the distributions of activities across (i) all family member types and (ii) across only sons and daughters. Of the 9 characteristics, only the father's wage and the gender of the young child do not significantly affect the activity distribution across all member types. The latter finding thus suggests that the care received by young children, measured by who in the household predominantly takes on child care responsibilities, does not depend on the gender of the child, despite their being marked differentiation by activities among children by gender when children reach their teenage years.

Only three household characteristics significantly influence the distribution of activities across sons and daughters - child morbidity, the presence of a health program and the mother's wage, the latter at the 10 percent level. It is notable that the mother's wage matters while the wage of the husband has no significant effect on the household's activity distribution. The point estimates suggest that where adult female wages are high, teenage girls are more likely to assume child care responsibilities, are less likely to be in the labor force, and are less likely to be in school compared to teenage boys. These findings are consistent with findings from other developing countries (Rosenzweig (1986)) that girls substitute for the mother's time in the home when their mothers increase their labor force activity in response to increased to female wages, at the expense of girls' labor force experience and schooling.

The presence of health programs in the village appears also to have a direct influence on the division of household activities, net of the effects

Table 4

Minimum Distance, Two-Stage Fixed Effects Logit Estimates: Differential Effects of Household and Areal Characteristics on Activity Allocations

	Household Care			Employment			Schooling
Variable	Dtr.-Moth.	Son-Moth.	Dtr.-Son	Dtr.-Moth.	Son-Moth.	Dtr.-Son	Dtr.-Son
Child morbidity	-.593 (3.04)[a]	1.45 (3.48)	.866 (3.48)	.549 (2.03)	.0422 (0.18)	.506 (1.85)	-.198 (0.71)
Child age	.353 (3.59)	.573 (2.72)	-.221 (1.18)	-.303 (2.48)	-.131 (1.27)	-.172 (1.36)	.127 (0.96)
Child gender (female=1)	.135 (0.91)	-.488 (1.52)	.623 (2.02)	-.155 (0.96)	-.00314 (0.02)	-.158 (0.92)	.122 (0.66)
Health program	-.0855 (0.32)	1.58 (3.36)	-1.67 (3.78)	1.34 (4.37)	1.23 (4.22)	.219 (0.65)	.104 (0.30)
Rural	1.05 (4.91)	2.04 (4.29)	-.992 (2.21)	.961 (3.45)	.872 (8.89)	.887 (0.29)	.242 (0.94)
Mother's wage ($x10^{-3}$)	.736 (1.21)	.374 (0.27)	.362 (0.28)	-.703 (0.65)	.149 (0.20)	-.852 (0.68)	-1.14 (1.11)
Father's wage ($x10^{-3}$)	-.643 (0.57)	1.57 (0.71)	-2.22 (1.08)	-.442 (0.30)	-.629 (0.52)	.188 (0.12)	1.08 (0.66)
Mother's schooling	-.107 (1.08)	-.115 (0.72)	.0074 (0.05)	-.252 (3.04)	-.413 (5.44)	.161 (1.73)	-.157 (1.28)
Father's schooling	-.305 (4.29)	-.425 (2.23)	.120 (0.66)	-.236 (2.47)	-.216 (2.54)	-.0205 (0.20)	.202 (1.48)
Intercept	-9.26 (10.9)	-7.90 (5.04)	-1.36 (0.91)	-8.09 (8.72)	-11.0 (12.8)	2.29 (2.78)	-1.05 (0.94)

a. Asymptotic t-ratio in parentheses beneath coefficients.

Table 5

Chi Square Test Statistics: Joint Influence of Household and Areal Characteristics on Household Activity Allocations

Variable	Hypothesis	
	No Influence on Mother and Children Differentials	No Influence on Daughter-Son Differentials
Child morbidity	17.4	10.5
Child's age	18.0	3.85
Child's gender	7.60	3.10
Health program	45.1	14.5
Rural location	67.9	5.28
Mother's wage	8.56	6.59
Father's wage	4.67	3.91
Mother's schooling	36.5	3.87
Father's schooling	41.0	4.47

of such programs on young children's health. Our estimates suggest that health clinics diminish gender-based inequality in the household; first by (presumably) reducing the incidence of child morbidity, which tends to exacerbate inequalities, and also directly--where such programs are more prevalent, for given child morbidity levels, the likelihood that girls take on household responsibilities is reduced and the likelihood that girls are in the labor force and in school are increased relative to boys. The programs also appear to increase the probabilities that girls and boys are in the labor force relative to mothers.

4. Conclusion

In this paper we have estimated, from unique household and community-level data from Indonesia, the effects on the allocation of time among family members induced by changes in levels of child morbidity. The results, obtained using estimation procedures that take into account the possibility that infant health is both influenced by and influences how a household distributes its activities across its members, reject the hypothesis that the effects of infant health on the allocation of family members' time are the same for all family members. In particular, existing sex-based differences in the division of time between household and labor-force activities in Indonesia are worsened where child morbidity is at a higher level. We not only found that mother's labor force participation falls more than that of her husband and her teenaged daughters and sons when infant morbidity rises but teenaged daughters were significantly more likely to increase their participation in household care activities compared to teenaged sons in response to increases in infant morbidity. We also found that the distribution of activities across teenaged siblings, for given infant morbidity levels, varies with the wage of the mother such that where

mothers receive higher wage rates, daughters are even more likely than sons to be engaged in household care activities and are less likely to be in school.

Our results also suggest that estimates of the consequences of child health on the activities of family members need to consider both the endogeneity of health and the integrated nature of households, as estimates of the effects of variations in child morbidity obtained without accounting for the simultaneity between the household activity distribution and child health differed markedly from those estimates which did. Attention to the determinants of health in a household context thus not only should result in better estimates of the specific consequences of interventions designed to improve health but in a greater awareness of the differentiated effects of such interventions.

Footnotes

This research was funded in part by a grant from NIH, HD21096. We are grateful to Lung-Fei Lee for valuable suggestions and to the Biro Pusat Statistik for providing the data. Earlier versions of this paper were presented at the 1986 Econometric Society Winter Meetings, New Orleans, at Gadjah Mada University, Indonesia, and at the University of Illinois.

1. For example, in rural Philippines households (Evenson, et al., 1980), older siblings of children aged less than three contribute 15 percent of the total hours spent in child care by family members.

2. The division of labor by gender among teenagers is equally apparent in other Asian societies. In the Philippines (Evenson, et al., 1980) among children aged 12-14, girls' hours in the labor force were 52 percent those of boys, while boys' hours in household activities were 44 percent those of girls. In an intensive anthropological study of a village in Bangladesh (Cain, 1980), the number of hours per day in wage work spent by girls aged 13-15 was found to be 21 percent that of similarly aged boys; girls' hours in crop production was 41 percent that of boys, and the number of hours spent by girls in household work was almost 10 times that of boys in that age group.

3. In Evenson et al. (1980), the total number of hours spent in child care by family members was 23 percent less when the young child was a girl than when the young child was a boy. Mothers spent 23 percent more hours with boy than with girl infants, while older siblings of the child spent 23 percent less hours caring for their infant sisters compared to their infant brothers.

4. One possible set of person-specific instruments would include the characteristics of a person's non-resident relatives; i.e., relatives who do

not participate in household activities. For example, the healthiness or other traits of the mother's parents might influence the mother's propensity for illness but not, given her health, her own or other household member activity allocations. There is an important intergenerational asymmetry, however, as such variables would only permit identification of first generation (parents) health effects on their children, but not vice versa or among (second generation) children.

5. Six mutually exclusive principal activities are distinguished in the SUSENAS survey: work, temporarily not at work, school, home-care, retired or disabled, and other. Temporarily not at work is defined as having a job but not working during the last week for reasons such as strike, flooded fields or holiday. These first two activites are aggregated into one category. The home-care category excludes those who do household chores for pay. The "other" category is defined to include those activities not falling into the first five and includes a "life of leisure" ("hidupnya bermalas-malas"). Households in which the mother of an infant or her teenagers were disabled were dropped from the sample.

6. Information on village-level health, water and sanitation failities come from the data tapes of the Village Potential (Potensi Desa) census carried out as part of the 1980 Population Census (Sensus Penduduk). Health facilities are measured by the proportion of villages in each rural/urban kabupaten with a maternity hospital, a family planning clinic, a doctor, and a polyclinic or a public health clinic (PUSKESMAS). Water sources are measured by the proportion of villages whose principal water source is from a pipe, a pump, a well, a spring, a river or solely from rainfall. Waste disposal methods are the proportion of villages in which waste is buried or

burned. Sanitation facilities are village lavatories, distinguished by family/non-family and Presidential Decree (INPRES)/non-INPRES.

7. An additional advantage of the fixed effects procedure is that if the placement of health or other programs or facilities by government agencies is related to unobserved (by the researcher) time-persistent interregional differences in factors affecting health levels (preferences, endowments), program variables are still valid instruments for the second stage difference equations, as long as the unobservables influencing program placement only affect activity levels and not cross-member activity differentials. Moreover, consistent estimates of the effects of program variables appearing in the second stage can be obtained without recourse to instrumental variables techniques as would not be the case for time allocation levels equations. Evidence of endogenous health program placement, based on fixed-effects procedures, is presented in Rosenzweig and Wolpin (1986).

8. Alternatively, the household's choice problem could be considered a single fixed effects multinomial logit. Unfortunately, the econometric properties of two-stage multinomial fixed effects logit have not yet been derived.

References

Bartel, Ann and Paul Taubman (1979): "Health and Labor Market Success: The Role of Various Diseases," Review of Economics and Statistics 61, 1-8.

Bartel, Ann and Paul Taubman (1986): "Some Economic and Demographic Consequences of Mental Illness," Journal of Labor Economics 4, 243-256.

Cain, Mead T. (1980): "The Economic Activities of Children in a Village in Bangladesh," in Rural Household Studies in Asia, Eds., H. P. Binswanger; R. E. Evenson; C. A. Florencia; and B. N. F. White, Singapore: Singapore University Press, 218-248.

Chamberlain, Gary (1980): "Analysis of Covariance with Qualitative Data," Review of Economic Studies 47, 225-238.

Evenson, Robert E., Barry M. Popkin and Elizabeth K. Quizon, (1980): "Nutrition, Work and Demographic Behavior in Rural Philippines Households," in Rural Household Studies in Asia, Eds., H. P. Binswanger; R. E. Evenson; C. A. Florencia; and B. N. F. White, Singapore: Singapore University Press, 289-366.

Lee, Lung-Fei (1982): "Health and Wage: A Simultaneous Equations Model with Multiple Discrete Indicators," International Economic Review 23, 199-222.

Mallar, C. D. (1977): "The Estimation of Simultaneous Probability Models," Econometrica 45, 1717-1722.

Pitt, Mark M. and Mark R. Rosenzweig (1985): "Health and Nutrient Consumption Across and Within Farm Households," Review of Economics and Statistics 67, 212-223.

Pitt, Mark M. and Mark R. Rosenzweig (1986): "Agricultural Prices, Food Consumption and the Health and Productivity of Farmers," in Agricultural Household Models: Extensions and Aplications, Eds., Inderjit Singh, Lynn Squire and John Strauss, Baltimore: Johns Hopkins Press.

Rosenzweig, Mark R. (1986): "Program Interventions, Intrahousehold Distribution and the Welfare of Individuals: Modeling Household Behavior," World Development 14, 233-243.

Rosenzweig, Mark R. and Kenneth I. Wolpin (1986): "Evaluating the Effects of Optimally Distributed Programs: Child Health and Family Planning Interventions," American Economic Review 76, 470-482.

Salkever, David S. (1982): "Children's Health Problems: Implications for Parental Labor Supply and Earnings," in Economic Aspects of Health, Victor Fuchs, Ed., NBER, Chicago: University of Chicao Press.

Sickles, Robin C. and Paul Taubman (1986): "An Analysis of Health and Retirement Status of the Elderly," Econometrica 54, 1339-1356.

Strauss, John (1986): "Food Consumption and Agricultural Productivity," Journal of Political Economy 94,

Table A

Maximum Likelihood Logit Estimates: Determinants of the Probability of Illness Among Children Aged Less than Four

Variable	Coefficient	t-ratio
CONSTANT	1.29707	.955
MOMAGE	-.646546E-02	-.069
MAGESQ	.506342E-05	.004
MOTHER - DTR. AGE	-.506894E-01	-1.328
MOTHER - DTR. AGESQ	-.915668E-03	-1.103
MOTHER - SON AGE	.603272E-01	1.563
MOTHER - SON AGESQ	.110239E-02	1.301
MOTHER'S SCHOOLING	.107405E-01	.362
FATHER'S SCHOOLING	-.298856E-01	-1.009
MOTHER'S WAGE	.556766E-03	1.591
FATHER'S WAGE	-.753808E-04	-.149
KIDONE	.923330	9.740
KIDTWO	1.36489	14.169
KIDTHREE	1.31061	13.810
SEX(FEMALE=1)	.455054E-01	.821
URBAN	.357464E-01	.247
HEALTH PROGRAMS	.392496E-02	.023
Price of:		
RICE	.311470	1.818
SLTFISH	.544188E-01	3.800
STRBEANS	.670885E-01	.984
EGGPLNT	-.786679E-01	-.848
BANANA	-.257271	-3.096
SALT	1.76126	1.538
COCONUTS	-.351024	-1.902
COOKOIL	.287622E-02	.051
SUGAR	-8.18044	-5.568
TEA	.103315	1.085
COFFEE	.539092E-01	.793
KEROSENE	1.14508	3.432
EGGS	-.129248	-.391
SPINACH	-.137416	-1.462
Price index for:		
MEAT	.161709	1.057
MILK	.234919	.943
VEGETABLES	-.501199	-3.232
FRUIT	.152733	1.134
SPICES	-.184311	-1.367

(continued)

Table A (continued)

Maximum Likelihood Logit Estimates: Determinants of the Probability of Illness Among Children Aged Less than Four

Variable	Coefficient	t-ratio
Price index for:		
TOBACCO	-.117752	-.724
FISH	-.534664	-4.650
PEANUTS	-.243795E-01	-.996
NOODLES	.230735	1.862
GRAIN	-.430524	-2.199
LEGUMES	.514817	3.087
TUBERS	.107008	.771
Programs and toilets:		
MATERNITY HOSPITAL	-.192642	-.711
FAMILY PLANNING	.112515E-01	.079
TOILET - PRES.	-1.21457	-2.236
TOILET - NON PRES.	.241120	.385
FAMILY TOILET - PR	8.25312	2.338
DOCTOR	-.464891	-1.777
PUBLIC TOILET	.260026	.375
FAMILY TOILET	-7.64113	-2.153
Drinking water sources:		
PIPE	.226818	.755
PUMP	-.686404	-1.181
WELL	-.627654	-2.176
SPRING	-1.20324	-3.003
RIVER	-1.29208	-3.605
Waste disposal:		
BURY	-.154073	-.612
BURN	.155714	.838
Chi-Squared (57)	503.08	

Appendix

Derivation of the Two-Stage Logit Covariance Matrix and the Minimum-Distance Weighting Matrix Ω

The two-stage binary logit model can be expressed as

$$\text{(A.1)} \quad y_1^* = X_1\beta_1 - \epsilon_1$$

$$\text{(A.2)} \quad y_2^* = X_2\beta_2 + y_1^*\gamma + \epsilon_2 .$$

Substituting (A.1) into (A.2) yields

$$\text{(A.3)} \quad y_2^* = X_2\beta_2 + \gamma(X_1\beta_1) - u_2$$

where $u_2 = \gamma\epsilon_1 - \epsilon_2$ and both ϵ_1 and u_2 are logistic.

The MLE estimates of β_1 are obtained by maximizing the likelihood

$$\text{(A.4)} \quad L_1 = \Pi_{i=1}^{N} \frac{e^{y_1 X_1\beta_1}}{1+e^{X_1\beta_1}}$$

and the MLE estimates of $\theta = (\beta_2, \gamma)$ are obtained by maximizing the likelihood

$$\text{(A.5)} \quad L_2 = \Pi_{i=1}^{N} \frac{e^{y_2(X_2\beta_2+X_1\hat{\beta}_1)}}{1+e^{X_2\beta_2+\gamma X_1\hat{\beta}_1}}$$

The log derivatives of this likelihood are

$$\text{(A.6)} \quad \frac{\partial \ell n L_2(\theta, \hat{\beta}_1)}{\partial\theta} = \sum_{i=1}^{N} \begin{pmatrix} X_2' \\ X_1\hat{\beta}_1 \end{pmatrix} \left(y2 - \frac{e^{X_2\beta_2+\gamma X_1\hat{\beta}_1}}{1+e^{X_2\beta_2\gamma X_1\hat{\beta}_1}} \right)$$

and

$$\text{(A.7)} \quad \frac{\partial \ell n L_2(\hat{\theta},\hat{\beta}_1)}{\partial \theta} = \frac{\partial \ell n L_2(\theta,\hat{\beta}_1)}{\partial \theta} + \frac{\partial^2 \ell n L_2(\theta,\hat{\beta})}{\partial \theta \partial \theta'} (\hat{\theta}-\theta) = 0$$

so that

$$\text{(A.8)} \quad (\hat{\theta}-\theta) = - \left[\frac{\partial^2 \ell n L_2(\theta,\hat{\beta}_1)}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial \ell n L_2(\theta,\hat{\beta}_1)}{\partial \theta} .$$

Additionally,

$$\text{(A.9)} \quad \frac{\partial \ell n L_2(\theta,\hat{\beta}_1)}{\partial \theta} = \frac{\partial \ell n L_2(\theta,\beta_1)}{\partial \theta} + \frac{\partial^2 \ell n L_2(\theta,\beta_1)}{\partial \theta \partial \beta_1'} (\hat{\beta}_1 - \beta_1)$$

$$= \frac{\partial \ell n L_2(\theta,\beta_1)}{\partial \theta} - \frac{\partial^2 \ell n L_2(\theta,\beta_1)}{\partial \theta \partial \beta_1'} \left[\frac{\partial^2 \ell n L_1(\beta_1)}{\partial \beta_1 \partial \beta_1'} \right]^{-1} \frac{\partial \ell n L_1(\beta_1)}{\partial \beta_1}$$

so that

$$\text{(A.10)} \quad \sqrt{N}\,(\hat{\theta}-\theta) = V \sum_{i=1}^{N} W_i$$

where $V = -\left[\frac{\partial^2 \ell n L_2(\theta, \hat{\beta}_1)}{N \cdot \partial\theta\partial\theta'} \right]^{-1} \left\{ I \mid - \frac{\partial^2 \ell n L_2(\theta, \beta_1)}{\partial\theta\partial\beta_1'} \left[\frac{\partial^2 \ell n L_1(\beta_1)}{\partial\beta_1\partial\beta_1'} \right]^{-1} \right\}$

and

$$\sum_{i=1}^{N} W_i = \frac{1}{\sqrt{N}} \left[\begin{pmatrix} X_2' \\ X_1\hat{\beta}_1' \end{pmatrix} \left\{ y_2 - \frac{e^{X_2\beta_2+\gamma X_1\hat{\beta}_1}}{1+e^{X_2\beta_2+\gamma X_1\hat{\beta}}} \right\} \right.$$
$$\left. X_1 \left\{ y_1 - \frac{e^{X_1\hat{\beta}_1}}{1+e^{X_1\hat{\beta}_1}} \right\} \right]$$

The minimum distance weighting matrix between two activities Ω_{12} is

$$\text{(A.11)} \quad \Omega_{12} = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} \left[\sum_i \left[\begin{pmatrix} W_{1i} \\ W_{2i} \end{pmatrix}' (W_{1i} \; W_{2i}) \right] \right] \begin{pmatrix} V_1' & 0 \\ 0 & V_2' \end{pmatrix}$$