NBER Working Paper Series

A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women

James J. Heckman
University of Chicago and NBER

Robert J. Willis
NBER and Stanford University

Working Paper No. 112

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR
AND SOCIAL INSTITUTIONS

National Bureau of Economic Research, Inc.
204 Junipero Serra Boulevard, Stanford, CA 94305

August, 1975

(Revised November 1975)

Research for this paper was supported by a grant to the National Bureau of Economic Research from the Center for Population Research, National Institute of Child Health and Human Development, PHS, Department of Health, Education, and Welfare. We wish to thank Ralph Shnelvar and Louis Garrison for computational assistance. An earlier draft of this paper was presented at the Third World Congress of the Econometric Society, Toronto, Canada, in August, 1975. The present draft benefited from the comments of R. Gronau, S. Haberman, and L. Lillard.

Abstract: "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women"

by

James J. Heckman
University of Chicago and NBER

and

Robert J. Willis
NBER and Stanford University

In this paper, we discuss statistical problems that arise in studying sequences of quantal responses (e.g., labor force participation) in panel data on heterogeneous populations (i.e., populations in which there is unobserved variation in response probabilities). Assuming that response probabilities are governed by a beta distribution, we derive a generalization of the cross-section logit model to enable it to deal with sequences of discrete events in panel data. This model is applied to panel data on labor force participation of married women. One of our findings is that the distribution of participation probabilities is U-shaped, indicating that most women have participation probabilities near zero or one.

INTRODUCTION

Many important aspects of household behavior involve choices among discrete alternatives or decisions that lead to discrete outcomes. Recognition of this fact has recently led to a considerable development of statistical models appropriate to the analysis of such "quantal response" problems in cross section data.[1] Quantal response problems also arise in the study of life cycle behavior with panel data. The timing or time path of discrete events or decisions such as school leaving, labor force participation, migration, marriage, divorce, births and death provide examples of such problems. In this paper we argue that statistical models appropriate in the analysis of quantal response problems in cross section data are less useful in the analysis of panel data. We generalize the cross-section logit model to enable it to deal with sequences of discrete events in panel data. Our model is then applied to panel data on the labor force participation of married women.

The basic reason that the conventional logit model is misleading in the analysis of panel data stems from the so-called "mover-stayer" or "heterogeneity" problem. In a pioneering paper, Blumen, Kogan and McCarthy (1955) found that the conditional probability that a "representative" individual moves from a given occupation decreases with the length of time he has stayed in the occupation. They also found that individuals who changed occupations most frequently in the past were more likely to change in the future. One explanation for this phenomenon is that the population is heterogeneous in the sense that some individuals have persistently higher propensities to change occupation than do others. That is, some individuals are "movers" and others are "stayers". As time passes, stayers tend to become a larger fraction of the sample remaining in a given occupation. Hence, a group's conditional probability of changing occupations appears to decline with duration of stay even if each individual's transition probability is constant. Similar patterns are observed in data on the monthly probability of conception, marital instability, geographic mobility and other types of sequential quantal response.

Statistical models such as the logit model that may be appropriate for the analysis of quantal response problems on cross-section data are less appropriate in an analysis of sequential responses in panel data from heterogeneous populations. We demonstrate below that the reason for this inadequacy is that the conventional logit model estimates parameters that generate the mean response probability conditional on the values of exogenous variables, but gives no information about the higher moments of the distribution of probabilities among individuals in the sample. Under the heterogeneity hypothesis, however, it is the higher moments of the distribution of response probabilities that provide the observed patterns of sequential response.

In this paper, we extend the conventional logit model to deal with quantal response problems in panel data from heterogeneous populations and apply our model to data on sequential labor force participation of married women. Our interest in applying the model to labor force participation was stimulated by Ben-Porath's (1973) observation that cross-section estimates of labor force participation functions are inherently ambiguous with respect to their implication for lifetime attachment to the labor force because a sample mean conveys no information on higher moments of distribution.

Consider the following example, due to Ben-Porath. Suppose that a group of women are found to have an average yearly participation rate of 50 percent. At one extreme, this might be interpreted to imply that each woman in a homogeneous population has a 50 percent chance of being in the labor force in any given year, while at the other extreme, it might imply that 50 percent of the women in a heterogeneous population always work and 50 percent never work. In the first case, each woman would be expected to spend half of her married life in the labor force and half out of the labor force and job turnover would be expected to be frequent, with an average job duration of two years. In the second case, there is no turnover and current information about work status is a perfect predictor of future work status. There are, of course, an infinite number of intermediate possible interpretations of the cross-section participation rate.

In addition to their different implications for life cycle labor supply, these two extreme interpretations have different implications for other aspects of life cycle behavior as well. For example, in the homogeneous case, married women and their employers have lower incentives to invest in human capital, both general and specific, than men because of high job turnover and a smaller amount of time in the labor force over the life cycle to capture returns on investments. As the average level of female participation rises, investment incentives increase. In the second case, however, the investment incentives of married women who work are identical to those of men. If such women can distinguish themselves to prospective employers, their wages should be identical to those of men. Growth in average participation rates simply increases the fraction of women who have these incentives. Put differently, relationships estimated on aggregated cross-section data for the Marshallian "representitive (or average) women" will have no relationship to the true relationship for every individual in the case of a heterogeneous population but may closely reflect the true relationship for all women in the case of a homogeneous population.

The plan of the paper is as follows. In Section I, a simple model of sequential labor force participation is presented. In Section II we show that heterogeneity among women arises if there are unobserved permanent components that affect the probabilities of participation and which persist through time. The implications of these sources of heterogeneity for the distribution of participation probabilities and for observed patterns of sequential participation are explored. In Section III, we assume that participation probabilities are governed by a beta distribution. Under a plausible parameterization of this distribution, we derive a likelihood function for sequential participation which reduces to the likelihood function of the conventional logit model in the case of cross-section data. For this reason, we call our model a "beta-logistic model". Empirical estimates of the labor force participation model are presented and analyzed in Section IV. Some remarks on the limitations of the beta-logistic model and a summary conclude the paper.

I. A Model of Sequential Labor Force Participation.

Labor force participation decisions are one aspect of the more general problem of household allocation of time among competing market and non-market uses (Becker, 1965). In each year of marriage, a woman must choose how much time to allocate to market work and how much to allocate to house-work, child care, leisure and other nonmarket uses. These time allocation decisions are the outcome of the household's attempt to maximize a well-behaved lifetime utility function defined for its time and goods consumption constraints (see, e.g., Heckman, 1974, and Ghez and Becker, 1975).

Given the assumption of maximizing behavior, the supply of labor for each household member may be derived as a function of the household's prices, wages, assets and other constraints. In particular, for horizon T an optimal time path for the wife's hours of labor, h_t^* , $t = 1,\ldots,T$ is associated with the constrained utility maximization. The wife's optimal sequence of labor force participation may be represented as the binary variable, y_t^*, where

$$y_t^* = \begin{cases} 1 & \text{if } h_t^* > 0 \\ 0 & \text{if } h_t^* = 0 \end{cases} \tag{1}$$

Following conventional labor supply theory, a woman's labor force participation decision in year t depends on a comparison of the marginal benefit and marginal cost of taking a job. Marginal benefit of an hour of work is measured by the woman's market wage, w_t. The marginal opportunity cost of the first hour of work is measured by the woman's shadow price of time,

w_t^*, evaluated at zero hours of work. Put differently, w_t^* is equal to the subjective marginal rate of substitution between the woman's nonmarket time and market goods when the woman spends all of her time in nonmarket activities. If w_t exceeds w_t^*, the woman will take a job and if w_t^* exceeds w_t, she will stay out of the labor force. Hence, we may express the optimal labor force participation sequence as

$$y_t^* = \delta(w_t - w_t^*) \qquad t = 1,\ldots,T$$

$$\text{where } \delta[\cdot] = \begin{cases} 1 \text{ if } w_t - w_t^* > 0 \\ 0 \text{ if } w_t - w_t^* \leq 0 \end{cases} \tag{2}$$

If we could observe both w_t and w_t^*, equation (2) could be tested directly. Unfortunately, this is not the case. The shadow price of time cannot be observed and the market wage cannot be observed for women who do not work. While it is possible to utilize sample information on working and non-working women to form consistent estimates of functions determing w_t and w_t^* (see Heckman, 1974), we shall pursue a simpler "reduced form" approach by deriving a labor force participation function in terms of observable variables and unobserved components that determine w_t and w_t^*.

The market wage a woman may command depends on many variables including education, training, intelligence, motivation, local labor market conditions and chance events. In a given body of data, let us assume that we observe for the ith woman in year t a subset of these variables designated by the vector z_{it} and that the remaining unobserved variables and chance events contribute a percentage amount η_{it} to the woman's wage. We may then write her wage function as

$$\ln w_{it} = f(z_{it}) + \eta_{it}. \tag{3}$$

A shadow price of time function is associated with the household's constrained utility maximum. Thus, w_t^* may be written as a function of household parameters such as assets, the husband's wage rate, market prices of goods, the rate of interest, and expected values of future wage and prices. It is also a function of other constraints reflecting past choices and chance events including the number and age of children, etc. Again, some of these variables can be observed, but others, especially those reflecting a given household's tastes, technology or expectations about the future are not observed. Thus, in analogy with the market wage function, we shall write the shadow price function of the ith woman in year t as

$$\ln w_{it}^* = g(z_{it}^*) + \eta_{it}^* \tag{4}$$

where z_{it}^* is a vector of observed variables and η_{it}^* reflects the percentage contribution to the shadow price of unobserved variables and chance events. (Note that some variables in z_{it}^* and η_{it}^* may also belong in z_{it} and η_{it}).

The "reduced form" labor force participation function of the ith woman in year t is obtained by substituting equations (3) and (4) into equation (2):

$$y_{it}^* = \delta[f(z_{it}) - g(z_{it}^*) + \eta_{it} - \eta_{it}^*] \tag{5}$$

where, $\delta[\cdot]$ equals one if the expression in brackets is positive and equals zero otherwise. Assuming that $f(z_{it}) - g(z_{it}^*)$ is linear and suppressing the subscript i, we may rewrite (5) in more compact notation as

$$y_t^* = \delta[x_t'\beta - S_t] \tag{6}$$

where β is a column vector of coefficients, x_t is the column vector of observed variables belonging to z_{it} or z_{it}^* or both and $S_t = \eta_{it}^* - \eta_{it}$.

The unobserved component S_t may be viewed as a random index function, so that the probability that a woman worker works in year t is

$$\Pr(y_t^* = 1) = \Pr(x_t'\beta > S_t). \tag{7}$$

If S_t is from the logistic distribution, the probability in equation (7) may be written as the logit function

$$\Pr(y_t^* = 1) = \frac{e^{x_t'\beta}}{1+e^{x_t'\beta}} = \frac{1}{1+e^{-x_t'\beta}} \tag{8}$$

Similarly, we may think of the sequence of unobserved components $S_1, S_2, \ldots, S_T$ as a sequence of random index functions. For example, the probability that a woman works for three years may be written as

$$\Pr(y_1^* = 1,\ y_2^* = 1,\ y_3^* = 1) = \Pr(x_1'\beta > S_1,\ x_2'\beta > S_2,\ x_3'\beta > S_3). \tag{9}$$

If we assume that the $S_t (t=1,2,3)$ are independently and identically distributed, then equation (9) may be written as

$$\Pr(y_1^* = 1,\ y_2^* = 1,\ y_3^* = 1) = \Pr(x_1'\beta > S_1)\Pr(x_2'\beta > S_2)\Pr(x_3'\beta > S_3). \quad (10)$$

Assuming that the S_t are independently, identically logistically distributed, (10) may be expressed simply as the product of three independent logit functions. The parameters of these functions, the vector β, could be estimated by maximum likelihood methods from a single year of labor force participation data. The predicted probability of working in each of the three years would then be estimated by using the values of the x_t in each year to evaluate $(1 + e^{x_t'\beta})^{-1}$. If these probabilities are π_1, π_2 and π_3, then the probability of working in all three years is $\pi_1\pi_2\pi_3$. The predicted probabilities of other possible participation paths are, of course, equally easy to compute.

II. Heterogeneity and Its Implications

The assumption that the random indices S_1, S_2,... are statistically independent makes little sense on theoretical grounds and implies patterns of labor force participation that are dramatically different from observed patterns. Turning first to empirical evidence, consider some data on labor force participation paths for two years, 1967 and 1968, for a group of 1583 married women from the University of Michigan Panel Study of Income Dynamics.[2] In 1967, 41 percent of the women participated in the market. Assuming independence, we would expect that $(.41)^2 = .168$ of the women would work in both 1967 and 1968 while, in the sample, 35 percent worked both years. Similarly, assuming independence, we would expect $(.59)^2 = 0.348$ of the women would not work in either year while, in the sample, 50 per cent did not work in either year. Looking at these results another way, the conditional probability of working in 1968 was 0.86 for those who worked in 1967 and 0.15 for those who did not work in 1967, in contrast to a probability of 0.41 that would be expected for both groups under the independence assumption.

One possible explanation for the discrepancy between predicted and observed behavior in this example is that our theoretical model of labor force participation fails to specify the appropriate stochastic process followed by each individual. For instance, the evidence that conditional probabilities of work in 1968 depend on work status in 1967 might suggest that each person in the sample follows a two-state Markov chain with transition probabilities given by the observed conditional frequencies. While we cannot rule out the possibility that individuals follow a stochastic process involving state dependendence, we pursue another approach in which individual transition probabilities are independent of initial state but aggregate transition probabilities are state dependent.[3]

It is not plausible to assume that the random index functions S_1, S_2,... are independent. Recall that the S_t measure the impact on a woman's market wage and shadow price of time of unobserved variables, as well as chance events. It is reasonable to suppose that many of these unobserved variables remain reasonably constant over time, but differ considerably among women. For instance, among the factors determining a woman's market wage, variables such as ability, motivation and labor market structure are likely to be unmeasured and to remain essentially unchanged over time. Similarly, unmeasured factors which affect the shadow price of time such as the household's wealth, its preference function and household technology tend to remain stable over time. This implies that random sequence S_1, S_2,... will tend to be serially correlated rather than independent.

Following a convention in the analysis of covariance, we decompose the S_t for the ith woman in year t into a "permanent component," ε_i, and a "transitory component," U_{ti}. Thus, let

$$S_{ti} = U_{ti} + \varepsilon_i \tag{11}$$

where U_{ti} is a random variable with mean zero and variance σ_u^2 and ε_i is a random variables with mean zero and variance σ_ε^2. Suppressing the index i for notational convenience, we assume that U_t are serially independent, and independent of the ε so that

$$E(U_t U_\tau) = 0 \qquad t \neq \tau \tag{12}$$

and

$$E(U_t \varepsilon) = 0 \qquad t = 1,2,\ldots,\infty \tag{13}$$

Thus S_t is a random variable with mean

$$E(S_t) = 0 \tag{14}$$

and

$$\begin{aligned} E(S_t S_\tau) &= \sigma_\varepsilon^2 & t \neq \tau \\ &= \sigma_\varepsilon^2 + \sigma_u^2 & t = \tau \end{aligned} \tag{15}$$

Thus, the correlation coefficient between the S_t in any two years, ρ, is then defined as

$$\rho = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_u^2} \ . \tag{16}$$

If intercorrelation caused by persistent unobserved variables is present, the probability that a woman works for three years cannot be written as the product of the probabilities that she works in each year, as it was in (9). However, conditional on a given value of the permanent component, $\tilde{\varepsilon}$, we may write the probability statement for three years of work as

$$\begin{aligned} &\Pr(x_1'\beta > S_1,\ x_2'\beta > S_2,\ x_3'\beta > S_3 | \tilde{\varepsilon}) \\ &\quad = \prod_{t=1}^{3} \Pr(x_t'\beta > S_t | \tilde{\varepsilon}) \end{aligned} \tag{17}$$

because, holding ε fixed at $\tilde{\varepsilon}$, the conditional values of the S_t are independent. Allowing ε to vary over the real line, we may write the unconditional (or expected) probability of working three years as

$$\int_{-\infty}^{\infty} \prod_{t=1}^{3} \Pr(x_t'\beta > S_t | \varepsilon) h(\varepsilon) d\varepsilon$$

where $h(\varepsilon)$ is the density function of ε.[4] Similar probability statements may be written for other sequences of labor force participation.

We are now in a position to investigate the implications of serial correlation caused by unobserved permanent components. For simplicity, assume that we observe a sample of women in a stationary environment in which the vector of exogenous variables, x_t , remains constant over time so that $x_1'\beta = x_2'\beta = \ldots = x_t'\beta = x'\beta$. This implies that the expected participation rate in the sample,

$$\pi_t = \Pr(y_t^* = 1) = \Pr(x'\beta > S_t) = \bar{\pi}$$

remains constant over time.

One implication of correlation in the random sequence S_1, S_2,... due to unobserved components is that participation probabilities differ among women who are homogeneous in terms of their <u>observed</u> characteristics. To see this, consider a sample of women who are observationally identical in the sense that they share a common value of the vector of observed exogeneous variables, say, $x = \bar{x}$. The probability density function of participation probabilities in the sample, $f(\pi)$, may be written:

$$f(\pi) = \Pr(\bar{x}'\beta - \varepsilon > U_t|\varepsilon)h(\varepsilon) = \int_{-\infty}^{x'\beta} \Phi(U_t + \varepsilon|\varepsilon)h(\varepsilon)d(U_t + \varepsilon) \quad (19)$$

where $h(\varepsilon)$ is the probability density function of ε and $\phi(U_t + \varepsilon|\varepsilon)$ is the probability density function of S_t conditional on ε.

First consider the implications for the distribution of participation probabilities in the two extreme cases of zero and perfect serial correlation. Recall that the serial correlation coefficient is $\rho = \sigma_\varepsilon^2/(\sigma_\varepsilon^2+\sigma_u^2)$. If there are no permanent differences among women (i.e., $\sigma_\varepsilon^2 = 0$) so that $\rho = 0$, equation (19) reduces to

$$f(\pi) = \int_{-\infty}^{\bar{x}'\beta} \phi(U_t)dU_t \quad . \quad (20)$$

Since U_1, U_2 ... are independent, each woman in the sample has an identical probability of participation in each year. At the other extreme, if there are only permanent differences among women (i.e. $\sigma_u^2 = 0$) so that $\rho = 1$, equation (19) reduces to

$$f(\pi) = \Pr(\bar{x}'\beta - \varepsilon > 0). \tag{21}$$

In this case, the fraction of women for whom $\bar{x}'\beta - \varepsilon > 0$ will always work, while those for whom $\bar{x}'\beta - \varepsilon \leq 0$ will never work.

We call populations characterized by zero serial correlation "homogeneous" because, conditional on the observed characteristics measured by x_t, all women in such populations have identical participation probabilities. In effect, the only source of variation in participation for women in a homogeneous population is caused by transitory shifts in their budget constraints or indifferences curves measured by U_t. Populations in which ρ is greater than zero will be called "heterogeneous" because participation probabilities differ among women who are observationally identical. As we have seen, perfect serial correlation implies an extreme form of heterogeneity in which individual women have either zero or unitary participation probabilities. In this case, although each household's indifference curves and budget contraint remain perfectly stable over time, there are unobserved differences in preferences or constraints among observationally identical households. The correlation coefficient is a measure of the <u>relative</u> importance of unmeasured permanent and transitory differences in preferences and constraints among households.

In the general case, the distribution of probabilities depends on the relative size of transitory variance (σ_u) to the permanent component (σ_ε), and on the value of $\bar{x}'\beta$. In the appendix, we demonstrate that if $\sigma_u > \sigma_\varepsilon$ $(\rho < 1/2)$, the density of the population probabilities is a unimodal distribution while it is "U shaped" if $\sigma_u < \sigma_\varepsilon$. If $\delta_u = \delta_\varepsilon$ $(\rho = 1/2)$ the density is either monotonically increasing or decreasing depending on whether or not

$\bar{x}\beta'$ is greater or less than zero. In the special case of $\bar{x}\beta' = 0$, the density is uniform over the unit interval. These cases are illustrated in Figure 1, which graphs the density of probabilities against the probability for the special case of $\bar{x}\beta' = 0$.

We have shown that an observationally homogeneous group of women will be heterogeneous in labor force participation probabilities if there are wage unmeasured permanent differences among these women in either market wage rates or shadow prices of time. The distribution of participation probabilities depends on the relative importance of variation in permanent and transitory components as measured by ρ, the relative importance of transitory and permanent factors as measured by the serial correlation coefficient of the random index functions S_1, $S_2, \ldots$, and by the mean participation probability, $\bar{\pi}$. We now investigate the implications of heterogeneity for the observed time path of labor force participation.

In a heterogeneous population, it is important to distinguish a model of the behavior of individuals from a model of the average behavior of a group of individuals. For instance, our model of labor force participation behavior implies that the probability that a woman works in a given year is independent of her prior work experience. However, in a heterogeneous (but observationally identical) group of women, average behavior appears to contradict this model. In a heterogeneous population the conditional probability of remaining in a given state tends to increase, the longer the group has been in that state.[5] Thus, the conditional probability that a "representative" woman works appears in increase the longer she continues working. Similarly, the longer a representative woman has been out of the labor force, the more likely she is to stay out in a given year. A corollary of rising conditional

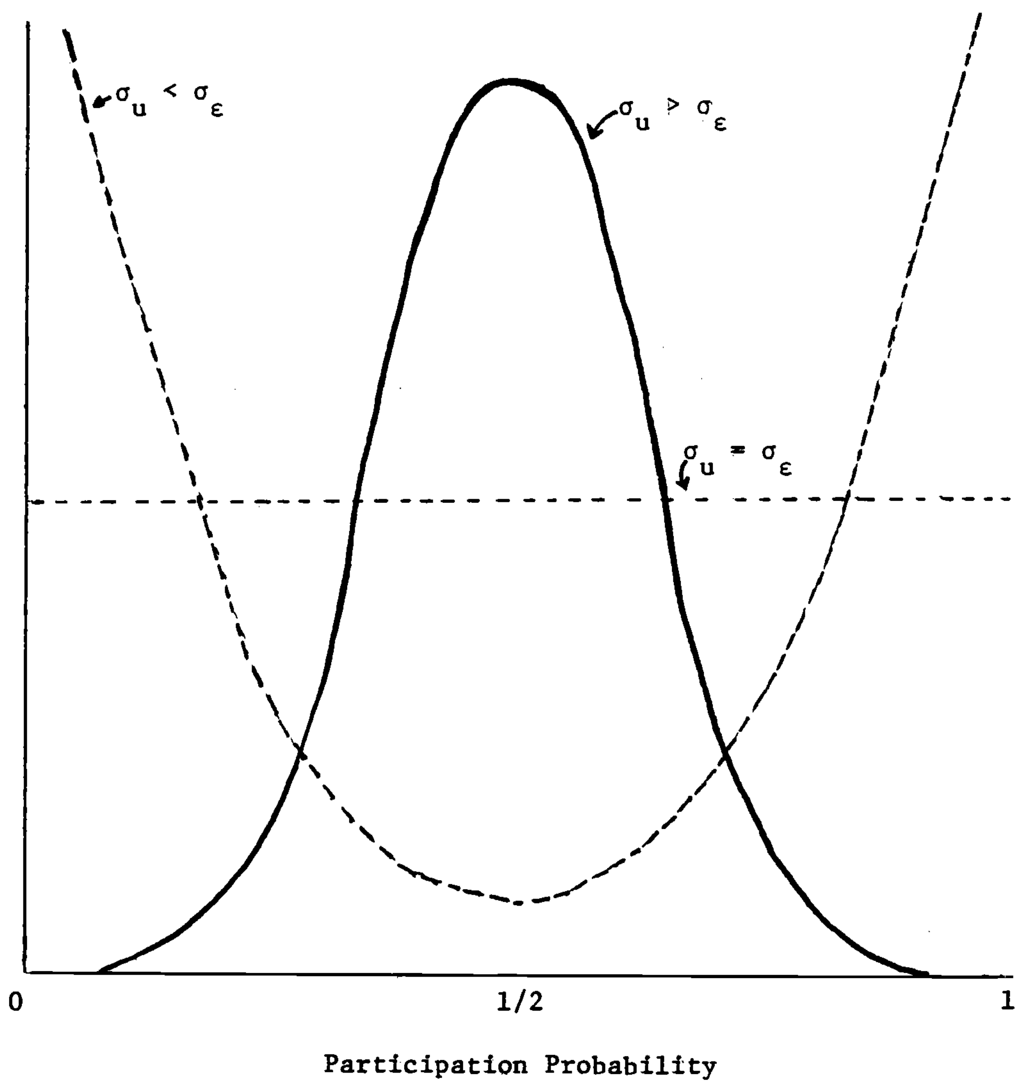


Figure 1: Distributions of Participation Probabilities with $X'\beta = 0$ for alternative values of σ_u and σ_ε .

probabilities of remaining in a given state is apparent state dependence of participation probabilities. That is, of otherwise identical women, those women who worked in year t-1 are more likely to work in year t than are those women who did not work in year t-1.[6]

The reason for the apparent contradiction between individual and average behavior in a heterogeneous population is due to a selection process. As time goes on, the women who have the highest (lowest) participation probabilities are most likely to be found in the subsample of continuous workers (nonworkers). Accordingly, as the sample composition of the group of working women changes, the conditional probability of participation (nonparticipation) appears to rise. Similarly, apparent state dependence in participation probabilities arises because the probability that an "average" individual will occupy a given state depends on his transition probability. Note that the line of causation here is just the reverse of that postulated in Markov chain models in which the transition probability is assumed to depend on the state an individual occupies.[7]

III. The Beta-Logistic Model

One way to deal with heterogeneity empirically is to assume a functional form for the distribution of participation probabilities and estimate the parameters of this distribution from panel data.[8] The beta distribution with the probability density function

$$f(\pi) = \frac{1}{B(a,b)} \pi^{a-1}(1-\pi)^{b-1} \qquad 0\leq\pi\leq 1,\ a,b>0 \tag{22}$$

where

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{23}$$

is an attractive choice of functional form for several reasons.[9] First, as is appropriate for a distribution of probabilities, the range of the beta distribution lies in the unit interval. Second, the distribution has only two parameters, a and b. Third, the shape of the distribution is flexible. It is unimodal if $a > 1$ and $b > 1$, U-shaped if $a < 1$ and $b < 1$, uniform is $a = b = 1$ and J-shaped if $a \geq 1$ and $b < 1$ or $a < 1$ and $b \geq 1$. As we demonstrated in the previous section, all of these are possible shapes of the distribution of participation probabilities.[10]

We now derive the expected probability of any participation path under the assumption that the yearly participation probability, π, has a beta distribution. If π were a constant, the probability that a woman works j years and does not work k years out of a total of $n = j + k$ years is

$$p(j,n) = \binom{n}{j} \pi^j (1-\pi)^k \qquad , \quad j = 0,1\ldots; \; k = n-j \; .$$

Now letting π be a random variable with the p.d.f. in (22), the expected probability of working j out of n years is

$$\begin{aligned} E(p(j,n)) &= \int_0^1 \binom{n}{j} \pi^j (1-\pi)^k f(\pi) d\pi \\ &= \binom{n}{j} \frac{1}{B(a,b)} \int_0^1 \pi^{a+j-1} (1-\pi)^{b+k-1} d\pi \qquad (24) \\ &= \binom{n}{j} \frac{B(a+j,b+k)}{B(a,b)} \\ &= \binom{n}{j} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+j)\Gamma(b+k)}{\Gamma(a+b+j+k)} \; . \end{aligned}$$

The properties of the model are easily derived from equation (24) using the recurrence relationship $\Gamma(x+1) = x\Gamma(x)$. The mean participation rate in any year is

$$E[p(1,1)] = \frac{a}{a+b} \quad , \qquad (25)$$

with variance,

$$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} \quad , \qquad (26)$$

which is a decreasing function of a and b . The apparent state dependence induced by heterogeneity may be seen by comparing the conditional probability of working in year t of women who worked in year $t-1$ with that of women who did not work in $t-1$:

$$\Pr(y_t^*=1|y_{t-1}^*=1)-\Pr(y_t^*=1|y_{t-1}^*=0) = \frac{a+1}{a+b+1} - \frac{a}{a+b+1} = \frac{1}{a+b+1} \quad (27)$$

This difference ranges from zero under homogeneity (i.e. $\sigma^2 \to 0$ as $a,b \to \infty$ holding $a/(a+b)$ constant) to unity under extreme heterogeneity (i.e. $\sigma^2 \to 0$ as $a,b \to 0$ holding $a/(a+b)$ constant). It may also be shown that the conditional probability of working rises as the duration of time in the labor force increases. Thus,

$$\Pr(y_t^*=1|y_{t-1}^*=1,\ldots.) = \frac{p(j=t,\ n=t)}{p(j=t-1,\ n=t-1)} = \frac{a+t-1}{a+b+t-1} \quad (28)$$

which is a positive monotonic function of t that approaches unity as t approaches infinity. A similar expression may be derived showing that the conditional probability of not working increases with the duration of time out of the labor force.

The model of equation (24) would be appropriate if the only factors causing differences in labor force participation among women were unobserved. However, we argue that the theory of labor supply suggests that a number of variables such as wife's education, the number and ages of children, husband's income influence participation and are typically observable. If there were a limited number of such variables, it would be possible to partition a sample to allow separate estimation of the parameters of the distribution of participation probabilities for groups of women with different values of the exogenous variables. However, if there are many exogenous variables or many values of a given set of variables, partitioning may be impractical.

An alternative approach that we pursue here is to postulate that the parameters a and b are functions of the vector of exogenous variable, x, as follows:

$$a = e^{x'\alpha}$$
$$b = e^{x'\beta} \qquad (29)$$

where α and β are column vectors of coefficients. This parameterization assures the non-negativity of a and b, as is required in (22).

The mean probability of participation, obtained by substituting (29) into (25), is

$$E[p(1,1)] = \frac{e^{x'\alpha}}{e^{x'\alpha} + e^{x'\beta}} = \frac{1}{1 + e^{-x'(\alpha-\beta)}} \quad . \qquad (30)$$

Note the (30) is nothing more than a logit function with coefficient vector $\gamma = \alpha - \beta$. With cross-section data (i.e. data on participation for only one year), β and α cannot be identified separately. Hence, the ordinary logit function can be used only to predict the mean participation rate in a population conditional on the x's, but cannot determine the higher moments of the distribution of participation probabilities. However, with participation data on the same individuals for two or more years both α and β can be identified.[11]

IV. Empirical Results

In this section, we present estimates of the model of sequential labor force participation of married women presented in section I based on data from the University of Michigan Panel Study of Income Dynamics, 1968-1972.[12] Our sample consists of 1583 white women who were continuously married to the same husband during the five-year period of observation, 1967-1971. In each year, we define a woman to have participated in the labor force if the respondent answered affirmatively the question: "Did your wife do any work for money last year?" This measure differs from the usual census in two respects: (1) we treat women who experience only unemployment as nonparticipants (2) the time frame is a year rather than the usual Census week.

Over the five-year period of observation, a woman may have followed any of 32 (= 2^5) possible participation paths. Assuming that participation probabilities among women follow a beta distribution, equation (24) in Section III implies that the contribution to sample likelihood of a woman who works j(=0, 1, ..., 5) years and does not work k = n-j years (n=1,...,5) is

$$\frac{B(a + j,\ b + k)}{B(a,b)} = \frac{\Gamma(a + b)}{\Gamma(a)\ \Gamma(b)} \ \frac{\Gamma(a + j) \quad \Gamma(b + k)}{\Gamma(a + b + j + k)} \qquad (31)$$

where $a = e^{x'\alpha}$ and $b = e^{x'\beta}$. Given a vector of independent variables, x', the coefficient vectors, α and β, are estimated by maximum likelihood.

The independent variables, their mean values and the estimated values of α and β are presented in Table 1. Since the likelihood function in (31) assumes that each individual has a constant participation probability over time, the values of the independent variables pertain to 1967, the initial year of observation. This is not a problem for variables such as education which remain constant over time, but is an undesirable restriction of the model for time-varying variables such as income and the number and age of children.[13]

As discussed in Section I, we take a "reduced form" approach to the estimation of the labor force participation function in terms of variables that determine the market wage (w) and/or the shadow price of time ($\hat{w}$). For instance, the wife's education is known to be positively related to her market wage, and therefore, is expected to have a positive effect on her probability of labor force participation.[14] The presence of children, especially young children, tends to increase the shadow price of time, thereby reducing the participation probability. Similarly, increases in family income (excluding wife's income) raises $\hat{w}$ and reduces partipation. Finally, the wage of unskilled labor measures the cost of substitutes for the wife's housework and is expected to have a negative effect on participation.

Maximum likelihood estimates of the coefficient vectors, α and β, of equation (31) together with associated asymptotic normal statistics are presented in columns (2) and (3) of Table 1. Of the sixteen parameter estimates, eleven are statistically significant at conventional levels (i.e., "t" > 1.9). The only variable that fails to have any appreciable effect on participation is husband's education. The other variables have effects on mean yearly participation in the hypothesized directions. To see this, recall that α and β are related to the parameters of the "cross-section" logistic function

$$\pi = E[p(1,1)] = \frac{1}{1+e^{-x'\gamma}}$$

Table 1: Maximum Likelihood Estimates of B(a,b)

Variable	(1) Mean	(2) α_i	(3) β_i	(4) $\frac{\partial \bar{p}}{\partial x_i} = (\alpha_i - \beta_i)\bar{\pi}(1-\bar{\pi})$
Intercept		-2.50 (6.2)*	-1.73 (4.2)	
Wife's education	11.32	.078 (2.4)	-.0810 (2.4)	.0385
# children ever born	2.75	.057 (1.5)	.0611 (1.6)	-.0010
# children not living at home	.85	-.200 (3.5)	-.117 (2.0)	-.0200
# children less than 6	.49	.177 (1.9)	.552 (5.9)	-.0907
Wage of unskilled labor in county	1.86	.054 (0.4)	.290 (2.1)	-.0571
Family income excluding wife ($10,000)	.8111	-.272 (2.1)	.252 (2.0)	-.0657
Husband's education	11.46	.0211 (0.8)	.029 (1.0)	-.0002

*Asymptotic "t" in parentheses

by the relationship $\gamma = \alpha - \beta$. Using this relationship we have evaluated the partial effect of each exogenous variable on mean yearly participation at the sample mean,

$$\frac{\partial\pi(1,0)}{\partial x_i} = (\alpha_i - \beta_i)\bar{\pi}(1 - \bar{\pi})$$

where $\bar{\pi}$ is the mean participation rate in the sample and α_i and β_i are the coefficients associated with the i^{th} independent variable, x_i. These effects are presented in column (4) of Table 1. As expected, wife's education has a strong positive effect on participation while children (especially young children), the county unskilled wage rate, and family income have negative effects on participation.

The empirical results indicate the presence of a considerable heterogeneity. To show this, we evaluate the parameters of the beta distribution using mean values of the exogeneous variables to conclude that $a = e^{\bar{x}'\alpha} = 0.232$ and $b = e^{\bar{x}'\beta} = 0.294$ which implies that the distribution of participation probabilities for women with "average" characteristics is U-shaped. That is, in a hypothetical population with a mean participation rate of .44, relatively few women have a probability of participation near the mean. This conclusion is essentially unaltered when we evaluate $a_i = e^{x_i'\alpha}$ and $b_i = e^{x_i'\beta}$ for each of the $i = 1,\ldots,N$ women in the sample. About 96 percent of the women have values of a and b less than unity. Thus, it appears that the distribution of participation probabilities is U-shaped for women of almost all socio-economic characteristics.

The total variance in participation probabilities across women in the sample is the sum of the variance in probabilities caused by differences in measured characteristics and the variance caused by unmeasured permanent components. To see this, let

$$\pi_i = E(\pi_i) + v_i$$

where π_i is the participation probability of the ith woman, $E(\pi_i) = e^{x_i'\alpha}/(e^{x_i'\alpha} + e^{x_i'\beta})$ is her expected probability based on the estimated coefficient vectors, α and β, and her measured characteristics, x_i, and v_i is a component due to permanent effects. The total variance in π_i may be defined as

$$\frac{\Sigma(\pi_i - \bar{\pi})^2}{N} = \frac{\Sigma[E(\pi_i) - \bar{\pi}]^2}{N} + \frac{\Sigma v_i^2}{N}$$

The variance in the predicted probabilities, $\mathrm{Var}[E(\pi_i)] = [E(\pi_i) - \bar{\pi}]^2/N$ is 0.016. The variance of the unobserved component, $\mathrm{Var}(V_i) = \Sigma v_i/N$, which is obtained from direct application of equation (26), is 0.1489. Hence, an estimated R^2 for our model is

$$\frac{\mathrm{Var}[E(\pi_i)]}{\mathrm{Var}[E(\pi_i)] + \mathrm{Var}(v_i)} = 0.097$$

This implies that over ninety percent of the total variation in participation probabilities in the sample is due to unmeasured permanent components. Put differently, the empirical distribution of expected probabilities based on measured characteristics is unimodal, with most women having expect probabilities near the mean for all women. In contrast, as we have seen, the distribution of actual participation probabilities, holding measured characteristics constant, is U-Shaped.

The high degree of heterogeneity in participation probabilities that is indicated by the small estimated values of the parameters a and b implies a correspondingly high degree of selection of women into "working" and "non-working" categories according to their participation probabilities. Thus, the predicted conditional participation probability in year t of a woman with average values of the exogenous variables who worked in year t-1 is $(a + 1)/(a + b + 1)=0.82$, while the conditional participation probability of a woman who did not work in year t-1 is $a/(a + b + 1) = 0.15$.

These estimates imply that knowledge of a woman's current work status is of considerable value in predicting the amount of time she is likely to spend in the labor force in the future. For instance, the estimates imply that a woman with average characteristics who is currently working will spend about eight out of every ten years in the labor force while a woman who is not currently working would be expected to spend only about one and one-half years out of ten at work. The accuracy of such predictions can be checked in the data for the four year period from 1968 to 1971 for women whose work status is known in 1967. The predicted mean years of work during this period for women who worked in 1967 is 4 x 0.81 = 3.24 years, and the actual average years worked is 3.19. Of the women who did not work in 1967, the predicted mean years worked is 4 x 0.15 = 0.60 years and the actual value is 0.89 years.

The fit of the estimated model may be examined by comparing the sample frequencies of possible labor force participation paths from 1967 to 1971 with the predicted probability of each path both of which are presented in Table 2. The actual frequencies are the numbers above each horizontal line in the

Table 2: Actual and Predicted Probabilities of Labor Force Participation Paths: 1967-71

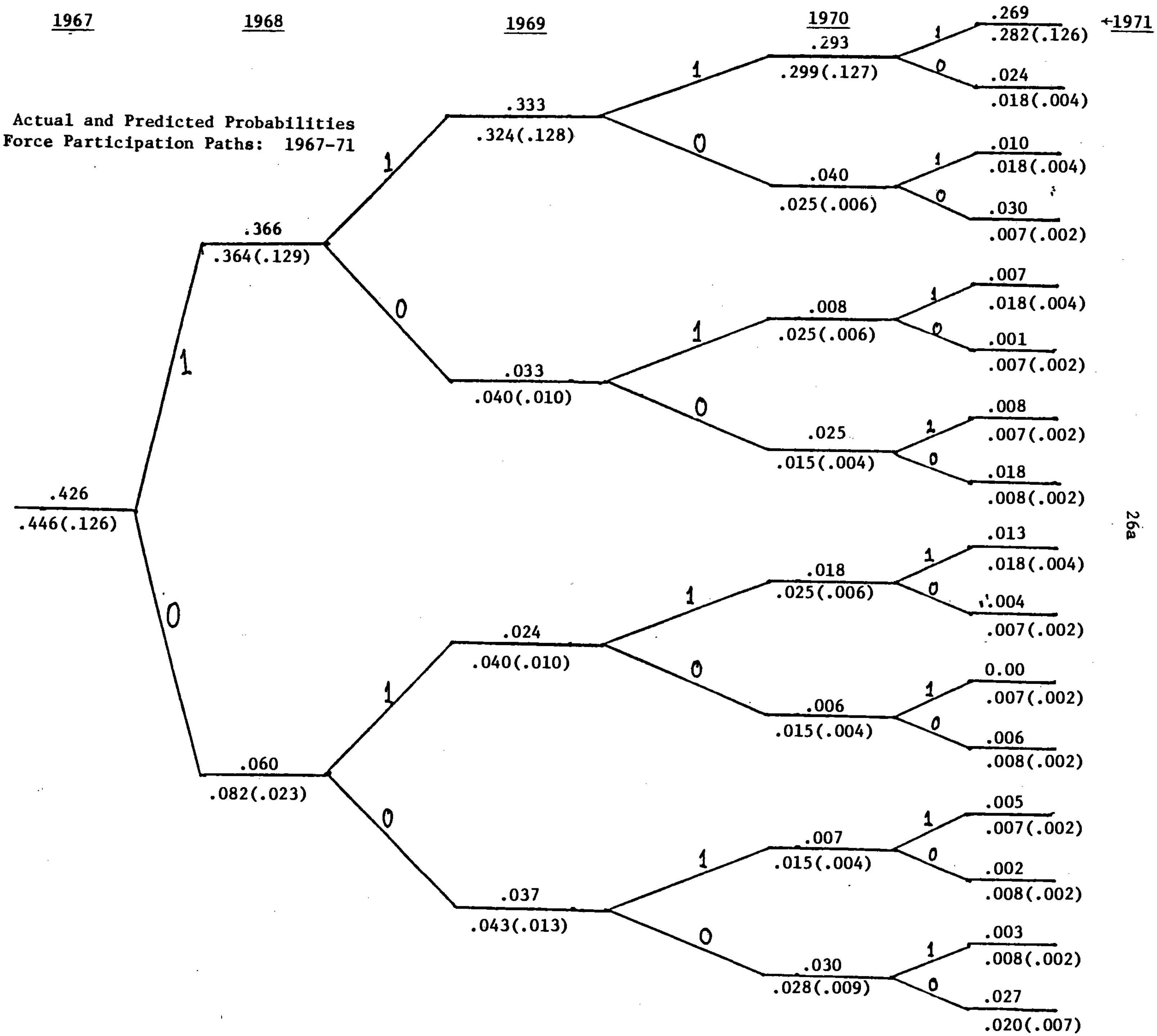
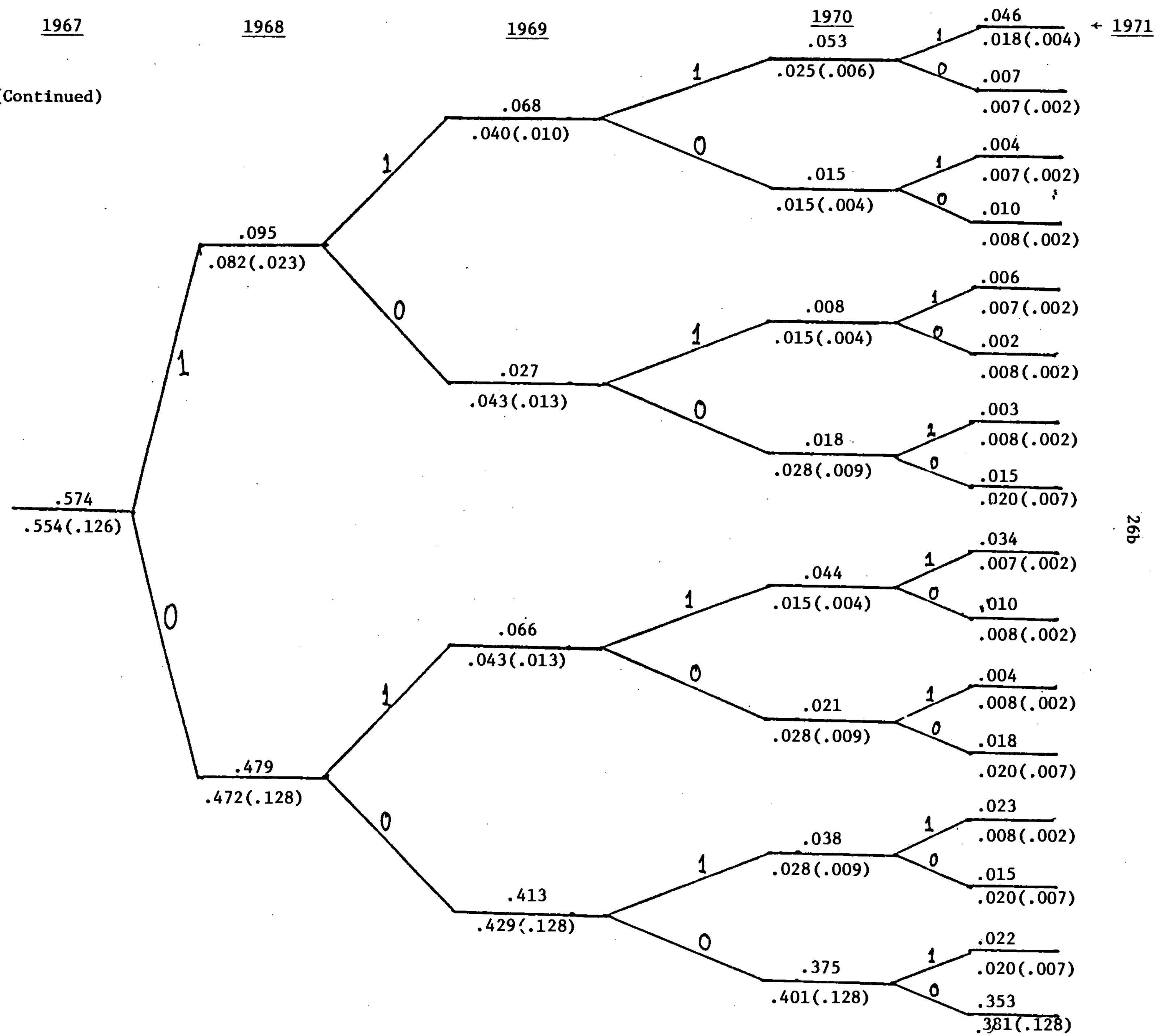


Table 2 (Continued)



tree diagram. The first page of the table displays the possible paths of the 42.6 percent of the women who were working in the initial year, 1967, and the second page gives the paths of the 57.4 percent of the women who did not work in 1967. The upper branch from each node (labeled "1") indicates "work" and the lower branch (labeled "0") indicates "not work". For example, 36.6 percent of the women worked in 1967 and 1968, 26.9 percent worked in all five years and 35.3 percent did not work in any of the five years.

The predicted probabilities of each participation path in Table 2 were computed for each woman in the sample by evaluating

$$p_i(j, n) = \frac{B(a_i + j, b_i + k)}{B(a_i, b_i)}$$

where $a_i = e^{x_i'\alpha}$, $b_i = e^{x_i'\beta}$, x_i' is the vector of exogenous variables for the i^{th} woman, and $n=j+k = 1,...,5$. The mean and standard deviation (in parentheses) of the predicted probabilities are presented below each horizontal line in the tree diagram. Visual inspection of Table 2 suggests fairly close agreement between actual and predicted participation probabilities, especially for continuous participation and continuous non-participation. The latter probabilities, extracted from Table 2, are tabulated for convenience in Table 3. However, on the basis of the chi-square statistics presented in Table 4, the hypothesis that the observed participation frequencies were generated by the beta-logistic model must be rejected for all but the first year.

Another way of examining the fit of the model is to compare the predicted and actual distributions of years worked by women in the sample. These distributions, based on information in Table 2, are presented in Table 5. Both the predicted and actual distributions are concentrated in the continuous

Table 3

		1967	1968	1969	1970	1971
Probability of working until year t	Actual	.426	.366	.333	.293	.267
	Predicted	.446	.364	.324	.299	.282
Probability of not working until year t	Actual	.574	.479	.413	.375	.353
	Predicted	.554	.472	.429	.401	.381

Table 4: Test of Goodness of Fit for Table 2

Number of years	1	2	3	4	5
X^2	2.56	12.79	74.66	222.23	506.38
$X^2_{.05}$	3.84	7.82	14.06	25.00	43.77
d.f.	1	3	7	15	31

Table 5. Actual and Predicted Distributions of years worked from 1967 to 1971

Years Worked	Percent of Sample	
	Actual	Predicted
5	.269	.282
4	.100	.090
3	.099	.070
2	.081	.080
1	.097	.100
0	.353	.381
Mean	2.300	2.240

$X^2 = 25.1$

$X^2_{.05} = 11.07$

work and continuous non-work categories with the remaining portion of both distributions nearly uniformly distributed across one to four years of work experience. Again, however, a chi-square test indicates that the discrepancy between predicted and actual years worked is statistically significant.

Both a priori considerations and evidence of systematic patterns in the deviations between actual and predicted participation probabilities indicate that the beta-logistic model fails to capture all the forces that generate the observed pattern of sequential labor force participation. In the economic theory of labor force participation presented earlier, we assumed that a woman would work if her market wage exceeded her shadow price of time, and, otherwise, that she would not work. This theory leads us to formulate the statistical model of participation at the individual level as a Bernoulli process in which the probability that a woman works in year t is independent of her labor force status in the previous year.

It is more realistic to assume that there are transactions costs in taking a job that result from the costs of search incurred by the job seeker and fixed hiring costs incurred by the employer. To the extent that such costs exist, the condition that the market wage exceed the shadow price of time becomes a necessary but not sufficient condition for a woman who is out of the labor force to take a job. In particular, the woman (and her prospective employers) must expect that the total benefits of taking the job will exceed these fixed costs. This total condition is more likely to be fulfilled, the greater is the excess of the market wage above the shadow price of time because, the greater the excess, the longer hours per year and the greater number of years the woman will expect to work given that she takes a job. Once a job has been taken, however, these fixed costs are sunk and the condition for remaining on the job is simply that the market wage exceeds the shadow price of time. Hence, the existence of transactions costs leads us to expect that a given woman is more likely to participate in year t if she worked in year t-1 than if she did not work in year t-1. This, in turn, implies that there may exist true state dependence in individual labor force behavior in addition to the apparent state dependence caused by selection in a heterogeneous population.

Although it is not easy to modify the beta-logistic model to allow direct estimation of the extent of true state dependence, an informal test for its presence is available. Let $\hat{p}^{(n)}$ be the predicted conditional probability of working in year t + n given that a woman worked in year t. Similarly, let $\hat{q}^{(n)}$ be the predicted conditional probability that a woman works in year t + n given that she did not work in year t. The predicted n-step transition matrix is then

$$\hat{T}^{(n)} = \begin{pmatrix} \hat{p}^{(n)} & 1 - \hat{p}^{(n)} \\ \hat{q}^{(n)} & 1 - \hat{q}^{(n)} \end{pmatrix}$$

Since the hypothesis underlying the beta-logistic model is that each woman in a heterogeneous population follows a Bernoulli process, it is obvious that $\hat{p}^{(1)} = \hat{p}^{(2)} = \ldots = \hat{p}^{(n)}$ and that $\hat{q}^{(1)} = \hat{q}^{(2)} = \ldots = \hat{q}^{(n)}$ so that $\hat{T}^{(n)}$ is constant with respect to n.[15] Using the predicted probabilities from Table 2,

$$\hat{T}^{(n)} = \begin{pmatrix} .816 & .184 \\ .148 & .852 \end{pmatrix}.$$

Empirical n-step transition matrices, $T^{(n)}$ (n = 1, ..., 4) computed from the observed probability in Table 2 are given in Table 6.

Table 6. Observed n-step Transition Matrices

1967-68	1967-69	1967-70	1967-71
$T^{(1)}$	$T^{(2)}$	$T^{(3)}$	$T^{(4)}$
$\begin{pmatrix} .859 & .141 \\ .166 & .834 \end{pmatrix}$	$\begin{pmatrix} .838 & .164 \\ .233 & .767 \end{pmatrix}$	$\begin{pmatrix} .765 & .237 \\ .249 & .747 \end{pmatrix}$	$\begin{pmatrix} .739 & .263 \\ .247 & .749 \end{pmatrix}$

It can be seen that, contrary to the constancy of $T^{(n)}$ implied by the beta-logistic model, the value of $p^{(n)}$ decreases and the value of $q^{(n)}$ increases systematically as n increases.

One explanation for such a pattern is that the true process at the individual level is Markovian rather than Bernoulli. If there were no heterogeneity, the observed probabilities $p^{(1)} = .859$ and $1-q^{(1)} = .834$ would be consistent estimates of the diagonal elements in the Markov transition matrix $T = T^{(1)}$. Given the assumption that the true process is a homogeneous Markov chain with transition matrix T, the n-step transition matrix is $T^{(n)} = T^n$.

Using $T = T^{(1)}$ from Table 6,

$$T^{(4)} = T^4 = \begin{pmatrix} .646 & .354 \\ .416 & .584 \end{pmatrix}.$$

We note that the predicted rate of decrease in $p^{(n)}$ and rate of increase in $q^{(n)}$ according to the homogeneous Markov chain hypothesis is much faster than the empirical rates of change in $p^{(n)}$ and $q^{(n)}$ in Table 6. As is well known, this is symptomatic of heterogeneity in Markov chains (see Blumen, Kogan and McCarthy, 1955 or Goodman, 1961) caused by the tendency of those with high probabilities of entering a given state (i.e. work or not work) being observed in that state in the initial period. The rates of change of $p^{(n)}$ and $q^{(n)}$ are not significantly reduced when variation in individual transition probabilities resulting from variations in observed independent variables (i.e. the variables listed in Table 1) are taken into account.[16] Hence, it appears that the data on sequential labor force participation reveal the coexistence of state dependence and heterogeneity caused by unobserved permanent differences among women.

Concluding Remarks

The main lesson of this paper is that a complete study of quantal response problems in panel data requires that explicit attention be given to the distribution of response probabilities in a sample, not just the mean response probability. Heterogeneity in responses will be present if all sources of persistent variation in the determinants of behavior cannot be held constant by the introduction of exogenous variables. Since exogenous variables in cross-section studies typically "explain" only a small fraction of the total variance in behavior, it is clear that there is plenty of scope for individuals to be persistently different from one another. However, it is obvious that "permanent" and "transitory" components of unexplained variance cannot be distinguished in a single cross-section. Hence, it is neither necessary nor possible to deal with the entire distribution of response probabilities in cross-section data; only the mean probability is of interest.

In panel data, however, the existence of unobserved permanent components of variance leads to serially correlated responses, a distribution of response probabilities with positive variance, and associated phenomena such as apparent state dependence and apparent time trends in response probabilities. In the paper, we assume that each individual has constant, state-independent response probabilities over time and that the distribution of response probabilities across individuals is a beta distribution. Given a particular parameterization of the exogenous variables, this model reduces to a conventional univariate or multivariate logistic model in the case of cross-section data. Accordingly, we have called this the beta-logistic model.

In our empirical application of the model to panel data on labor force participation by married women we found evidence of considerable heterogeneity. According to parameter estimates, the distribution of participation probabilities is U-shaped. Thus, loosely speaking, there tend to be two groups of women in the sample: "workers" whose participation probabilites are near unity and "non-workers" whose participation probabilities are near zero. Relatively few women in the sample have probabilities near the mean participation rate of 40 per cent.

In terms of the economic theory underlying the statistical model, this finding implies that unexplained variation in the budget constraints and indifference curves of individual women displays a high degree of stability over time. The empirical results also suggest that sample selection by work status is an important phenomenon. Those women who are working in a given year have a predicted participation probability of about .82 while those who are not working have a predicted probability of .15. In effect, the sample of currently working women is largely composed of "workers" and the sample of non-working women is largely composed of "non-workers." An implication of this is that knowledge of current work status is of considerable utility in predicting a woman's future labor force behavior.

Certain limitations of the beta-logistic model should be recognized explicitly in judging our empirical results, their implications, and the usefulness of the model in analyzing other types of quantal response problems. The two most important limitations are 1) the assumption that individual reponse probabilities are constant through time and 2) that response probabilities are independent of the current or past states occupied by an individual.

Clearly, in the case of labor force participation, some of the significant determinants of participation such as number and age of children, family income and so on do, in fact, vary over time. The beta-logistic model does not allow for the effects of such variation and is, therefore, misspecified. Analysis of the participation data also suggests the existence of true state dependence in addition to apparent state dependence caused by heterogeneity. Again, this indicates that the beta-logistic model is not a completely correct specification of the stochastic process generating the data. Despite these limitations, the empirical fit of the model appears to be quite good. Thus, we have some confidence that a major component of observed variation in sequential labor force participation rates is the result of heterogeneity.

BIBLIOGRAPHY

Becker, Gary S. "A Theory of the Allocation of Time." Economic Journal 75 (September 1965): 493-517.

Ben-Porath, Yoram. "Labor Force Participation Rates and the Supply of Labor." Journal of Political Economy 81 (May/June 1973): 697-704.

Blumen, Isadore; Kogan, Marvin; and McCarthy, Philip J. The Industrial Mobility of Labor as a Probability Process. Cornell Studies in Industrial and Labor Relations: Volume VI. Ithaca, N.Y.: Cornell University, 1955.

Boskin, M. and Nold, F.C. "A Markov Model of Turnover in Aid to Families with Dependent Children." (Mimeo, 1974) Journal of Human Resources (forthcoming).

Ghez, Gilbert R. and Becker, Gary S. The Allocation of Time and Goods Over the Life Cycle. N.Y.: Columbia University Press for NBER, 1975.

Goodman, Leo A. "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications." Journal of the American Statistical Association 65 (1970): 226-256.

__________. "Statistical Methods for the Mover-Stayer Model." Journal of the American Statistical Association 56 (1961): 841-868.

Kiefer, H. and Wolfowitz, J. "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters." Annals of Mathematical Statistics, Vol. 27, No. 4, December 1956.

Hall, Robert. "Turnover in the Labor Force." Brookings Papers on Economic Activity, No. 3 (1973): 709-764.

Heckman, James J. "Shadow Prices, Market Wages and Labor Supply," Econometrica 42 (July 1974): 679-694.

Heckman, James J. and Willis, Robert J. "Estimation of a Stochastic Model of of Reproduction: An Econometric Approach." In Household and Production Consumption, edited by Nester E. Terleckyj. N.Y.: Columbia University Press for NBER, 1975, (forthcoming).

________________________________. The Problem of Heterogeneous Populations in Quantal Response Models," Annals of Measurement, Spring, 1976.

Johnson, Norman L. and Kotz, Samuel. Discrete Distributions. Boston: Houghton Mifflin Co. 1969.

McFadden, Daniel. "Quantal Choice Analysis: A Survey." Department of Economics, University of California, Berkeley. March 1974.

Maritz, J. S. Empirical Bayes Methods. London: Metheun Press, 1970.

Morgan, James, et. al. Five Thousand American Families - Patterns of Economic Progress. Volumes I and II. Institute for Social Research, University Michigan, 1974.

Nerlove, Marc and Press, S. James. "Univariate and Multivariate Log-Linear and Logistic Models." R-1306-EDA/NIH. Santa Monica, California: RAND Corp., December 1973.

Sheps, M.C. and Menken, J. Mathematical Models of Conception and Birth. Chicago: University of Chicago Press, 1973.

Spillerman, Seymour. "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain." American Sociological Review 37 (June 1972): 277-294.

Wilks, S. S. Mathematical Statistics. N.Y.: Wiley and Sons, 1962.

FOOTNOTES

[1]See Daniel McFadden (1974) - for a survey of developments in this field and references to the literature.

[2]We shall describe this sample more fully later.

[3]*A priori* reasons for expecting state dependence and evidence for its existence are discussed below in Section IV.

[4]See Kiefer and Wolfowitz (1956) and Maritz (1970).

[5]See Heckman and Willis (1975) for a proof of this result. The result itself has a long history.

[6]Given the result that conditional probabilities of remaining in a given state increase, the proof of this proposition is quite simple. In an observationally homogeneous group of women with a constant x_t, the average probability of participation in year t-1 is $\bar{\pi}$ and the probability of nonparticipation is $1-\bar{\pi}$. Of those women who worked in the t-1, the conditional probability of working in year t is $\pi_t^A > \bar{\pi}$. Of those women who did not work in year t-1, the conditional probability of not working in year t is $1-\bar{\pi}_t^B > 1 - \bar{\pi}$. Hence, $\pi^B < \bar{\pi}$ and $\pi^A > \bar{\pi}$ so that $\pi_t^A > \pi_t^B$.

[7]Of course both lines of causation may be present in a given situation. If they are, an appropriate model might be, say, a Markov chain in which individuals have different transition probabilities. Stayer-mover models originated by Blumen, Kogan and McCarthy (1955) and elaborated by Goodman (1961) Spillerman (1972) and others represent attempts to deal with heterogeneity in Markov chains. Apparently, however, these writers have not noticed that the Markovian (i.e., state dependent) appearance of their data may itself arise from heterogeneity.

[8]An alternative method pursued in Heckman and Willis (1975) is to assume, as we did for illustrative purposes in the previous section of this paper, that U_t and ε are normally distributed. In this approach, the probability of a sequence of events is represented as a bivariate normal distribution with a constant correlation coefficient $\rho = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_u^2)$.

[9]The beta distribution is widely used as a "mixing distribution". (See, for example, Johnson and Kotz, 1969, pp. 78-79 and Sheps and Menken, 1973.)

[10]It is important to emphasize, however, that if U_t and ε are assumed to be normally distributed the resulting distribution of participation probabilities is not a beta distribution. (See the Appendix for a derivation of the distribution under these assumptions.) Thus, if the assumption that U_t and ε are normally distributed is maintained, the beta distribution may be regarded as an approximation to the true distribution of participation probabilities. Alternatively, it may be assumed that the distribution of U_t and ε are such as to lead to a beta distribution of participation probabilities. Unfortunately, we have been unable to find functional forms for the distributions of U_t and ε that lead to this result.

[11]This model may be readily generalized to extend the multinomial multivariate logistic model discussed by Goodman (1970) and Nerlove and Press (1973) to panel data on heterogeneous populations. This generalization and other methodological issues related to quantal response problems in panel data are discussed in a forthcoming paper by the authors. Essentially, the beta density is replaced by the Dirichelet density, with a parametization identical with that suggested in equation 29 in the text. Note further that it is straightforward to prove concavity of the likelihood function and to prove asymptotic normality of the estimates. See Heckman and Willis (1976).

[12]This five-year panel survey included about 2000 families from the 1967 Survey of Economic Opportunity and an additional 3000 families from a cross-

section of dwellings in the U.S. It contains information on a wide variety of demographic, social and economic variables on families and individuals over the period 1967-71. A detailed description of the data is contained in Morgan, et. al. (1974).

[13]We defer discussion of this and other limitations of the model to the concluding section of the paper.

[14]It may be the case that education also affects the wife's productivity in home production. However, even if it does, it is uncertain whether the shadow price of time is increased, decreased or left unchanged by an increase in education. For example, if education increases the productivity of time and goods in household production by the same proportion and the income elasticities of demand for goods and leisure are unitary, education has no effect on the shadow price of time.

[15]Using equation (47)

$$\hat{p}^{(n)} = \frac{\Pr(y^*_{t+n} = 1 \wedge y^*_t = 1)}{\Pr(y^*_t = 1)} = \frac{a+1}{a+b+1}$$

which is independent of n. Similarly, $\hat{q}^{(n)} = a/(a+b+1)$ is independent of n.

[16]The adjustment referred to follows the suggestion of Spillerman (1972), Hall (1973) and Boskin and Nold (1974) who use the independent variables to estimate regression or logit functions of the form $p_i^{(1)} = f(x_i) + u$ and $q_i^{(1)} = g(x_i) + v$ which are used to predict the transition probabilities of each individual in the sample. Then the predicted n-step transition matrix is simply $\bar{T}^n = \frac{1}{N}\sum_{i=1}^{N} T_i^n$, the mean of the individual n-step matrices. From our previous discussion, it is clear that this procedure will solve the heterogeneity problem only if u and v are not serially correlated. This will be the case only when the set of independent variables, x_i, include all persistent determinants of

labor force participation. The failure of the adjustment to substantially modify the predictions of the homogeneous Markov model indicates that most heterogeneity is the result of unobserved components. This is consistent with our finding reported earlier that the measured variables account for less than 10% of the total variance in participation probabilities.

APPENDIX

For convenience, define Z_t as random variable

$$Z_t = U_t - \varepsilon' - b$$

where $\varepsilon' = -\varepsilon$ in the text, and b is the value of the regression function $x'\beta$ at some artibrary point. U_t and ε' are each normally distributed with mean zero and variances σ_u^2 and $\sigma_{\varepsilon'}^2$, respectively. In this notation, a woman works if $Z_t < 0$.

For given ε' and b, the probability (with respect to U_t) that a woman works is

$$\Pr(Z_t < 0) = \Pr(\frac{U_t}{\sigma_u} < \frac{\varepsilon'+b}{\sigma_u}) = \pi(\varepsilon',b) = F(\frac{\varepsilon'+b}{\sigma_u})$$

Where F(x) is the cumulative univariate normal

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

We seek to derive and characterize the distribution of $\pi(\varepsilon,b)$. Note that $0 \leq \pi(\varepsilon,b) \leq 1$.

With respect to the distribution of ε', the cumulative distribution of $\pi(\varepsilon',b)$ is

$$G(j) = \text{Prob}_{\varepsilon'}(\pi(\varepsilon',b) < j) = \text{Prob}_{\varepsilon'}(F(\frac{\varepsilon'+b}{\sigma_u} < j) = \text{Prob}_{\varepsilon'}(\frac{\varepsilon'+b}{\sigma_u} < F^{-1}(j))$$

Since the distribution of $\frac{\varepsilon'}{\sigma_{\varepsilon'}}$ is univariate normal,

$$F(j) = F(\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}})$$

Thus if $\sigma_u = \sigma_{\varepsilon'}$ and $b=0$,

$$G(j) = j$$

so that the distribution of participation probabilities is uniform with median and mean identical at 1/2.

In the general case, the median is the value of j such that

$$1/2 = F(\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}})$$

$$F'(1/2) = 0 = \frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}}$$

so that the median $j_{m\varepsilon}$ is

$$j_{m\varepsilon} = F(\frac{b}{\sigma_u}) \quad .$$

The larger b, or the smaller the variance in the transitory component, the larger the median.

The density of j exists and is given by

$$G'(j) = \frac{\sigma_u}{\sigma_{\varepsilon'}} F'(\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}}) \; (F^{-1}(j))' = \frac{\sigma_u}{\sigma_{\varepsilon'}} \frac{F'(\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}})}{F'(F^{-1}(j))} \quad .$$

Since F(x) is normal,

$$G'(j) = g(j) = \frac{\sigma_u}{\sigma_{\varepsilon'}} \exp 1/2 \left[(F^{-1}(j))^2 - (\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}})^2\right]$$

and

$$\ln g(j) = \ln(\frac{\sigma_u}{\sigma_{\varepsilon'}}) + 1/2 \left[(F^{-1}(j))^2 - (\frac{\sigma_u}{\sigma_{\varepsilon'}} F^{-1}(j) - \frac{b}{\sigma_{\varepsilon'}})^2\right]$$

$$= \ln(\frac{\sigma_u}{\sigma_{\varepsilon'}}) + 1/2 \left[-(\frac{b}{\sigma_{\varepsilon'}})^2 + (F^{-1}(j))^2 (1 - (\frac{\sigma_u}{\sigma_{\varepsilon'}})^2) + \frac{2b\sigma_u}{(\sigma_{\varepsilon'})^2} F^{-1}(j)\right]$$

so that

$$\text{(A-1)} \qquad \frac{g'(j)}{g(j)} = \frac{2}{F'(F^{-1}(j))} \left[F^{-1}(j) (1 - (\frac{\sigma_u}{\sigma_{\varepsilon'}})^2) + \frac{b}{\sigma_u} (\frac{\sigma_u}{\sigma_{\varepsilon'}})^2 \right]$$

The term in the denominator is never negative.

Suppose that b=0 so that the median is $j_{m\varepsilon} = 1/2$ and the density posesses a unique critical point at the median if $\sigma_u \neq \sigma_{\varepsilon'}$. Since $F^{-1}(j) = -F^{-1}(1-j)$ the density g(j) is symmetric around j = 1/2, so that the median, mean and the critical point coincide. If transitory variance exceed permanent variance ($\sigma_u > \sigma_{\varepsilon'}$), the critical point is the mode (i.e., the distribution is "humped"), while if the opposite case is obtained the density if "U-shaped". These cases are illustrated in Figure 1 in the text.

In the general case, the critical point is given by

$$j_c = F\left(\frac{\frac{b}{\sigma_u} (\frac{\sigma_u}{\sigma_{\varepsilon'}})^2}{1 - (\frac{\sigma_u}{\sigma_{\varepsilon'}})^2}\right)$$

for $\sigma_u \neq \sigma_\varepsilon$. No critical point exists for $\sigma_u = \sigma_\varepsilon$.

In the special case $\sigma_u = \sigma_\varepsilon$, inspection of equation (A-1) reveals that if $b > 0$ (so that the median is greater than 1/2) the density g(j) begins at the origin and is monotonically increasing in j. While if $b < 0$ it is monotonically decreasing in j with the density becoming zero at $j=1$.

In the general case with $\sigma_u \neq \sigma_\varepsilon$, and $b \neq 0$, the distribution is "U-shaped" (or twisted "U-shape") as long as $\sigma_\varepsilon < \sigma_u$ and is "humped" as long as $\sigma_\varepsilon > \sigma_u$. If $b < 0$, and $\sigma_u > \sigma_\varepsilon$, the median is less than 1/2 while the mode (or critical point) exceeds 1/2. Similarly, if $b > 0$ and $\sigma_u > \sigma_\varepsilon$ the mode is less than 1/2 while the median exceeds 1/2. In the other cases, the critical value may lie to the left or the right of the sample median.