Estimation of the Probability of Labor Force Participation of the AFDC Population-At-Risk

ERIC SOLBERG*, GERALD BROWN**, AND HERBERT RUTEMILLER*

1. Introduction

The primary purpose of this paper is to apply maximum likelihood estimation (MLE) of non-linear specifications of the probability of labor force participation (LFP) of female family heads with dependent children present, the Aid to Families with Dependent Children (AFDC) population-at-risk. Despite the fact that the method of MLE has been available for decades, MLE of highly non-linear specifications is not common. Although methods for the analysis of qualitative data have been discussed for many years, analysts continue to use inappropriate methods to estimate inappropriate functional specifications. One explanation is that computer software for MLE solution of highly non-linear functions is not trivial.

A secondary purpose of this paper is to present a theoretical justification for the use of a sigmoid shaped function when estimating a probability like the labor force participation rate whether or not confronted with a dichotomous dependent variable. The sigmoid specification closely agrees with the shape expected for a labor force participation function and is logically consistent with a probability interpretation while the linear probability function is not.

A method which has frequently been used to help circumvent the inherent inconsistency of predicting a non-linear phenomenon with a strict linear model is the use of categorical explanatory variables. As a practical matter, use of dummy variables in linear regression is often easier and less costly. Also, much of the available data in the past has been reported categorically. Economists were forced to use these categorical data rather than a better continuous measure. Habits are hard to break. At the same time, goodness-of-fit does not solve the problem that standard tests of significance are not appropriate within the context of a dichotomous dependent variable even given robust tests; MLE solves that problem. Furthermore, the practical consideration between MLE and linear regression will disappear as large scale MLE programs become more readily available. Direct MLE of a nonlinear specification avoids inherent inefficiencies of transformations designed to smooth an estimated linear probability specification into a nonlinear form.

Many past applications have relied on ordinary least squares (OLS) to obtain estimates of a linear probability model in both the case of a dichotomous variable [9,11] and when using labor force participation rates [6, 8]. Nerlove and Press [14] presented a cogent theoretical argument for MLE of the logistic specification. They also presented a program for MLE of the logistic as well as some empirical applications. Gunderson [10] using a dichotomous variable, has recently compared the estimated probability of trainee retention after training, comparing the OLS linear probability, MLE probit, Orcutt transformation of the linear probability, Theil transformation of the logistic, and Warner transformation of the linear probability. Gunderson applied MLE only in the case of the probit while noting that the transformations do not eliminate the inherent inefficiencies of OLS estimation of the linear probability function.

Aigner [1] has recently applied MLE as an alternative to OLS and the use of instrumental variables when estimating a labor supply function from data similar to the CPS. Amemiya and Boskin [3] have applied MLE in the case where the dependent variable is truncated lognormal. But there have been very few applications of MLE of the probability of an event when the de-

*California State University–Fullerton. **Naval Postgraduate School.

pendent variable is binary.

The three specifications used in this study are presented in Section 2: the linear probability model; the logistic; and that known as Urban's curve. Section 3 presents a rationale for the signoid specification by aggregating the individual's labor force participation decision. Section 4 describes the data and variables used in this study. Section 5 presents the empirical results and a comparison between the specifications. Section 6 offers a brief summary of results.

2. The Models

Labor force participation is a qualitative characteristic. An observation consists of noting whether the characteristic is present. Thus, the dependent variable, designated as Y, is dichotomous and takes a value of 1 if the family head had a job or was looking for work and a value of 0 if not in the labor force. A natural way to proceed is to estimate $Pr(Y = 1; X) = LFPR(X) = \theta$, where $LFPR$ denotes the labor force participation rate, X a set of stimuli, and θ a probability. The predicted value of the dependent variable can be alternatively interpreted as the probability of participation for an individual or as the labor force participation rate for individuals with like characteristics.

The probability of labor force participation can be considered as the parameter θ_i in a family of distributions

$$f(y_i, \theta_i) = \theta_i^{y_i}(1 - \theta_i)^{1-y_i};\ y_i = 0, 1 \qquad (1)$$

where θ_i are assumed to be a function of the variables $X_{i1}, \ldots, X_{ik}$. The estimation of θ_i can be obtained from a series of N observations $[y_i; X_{i1}, \ldots, X_{ik}]$, $i = 1, \ldots, N$. For example,

$$\theta_i = \theta_i(X_{i1}, \ldots, X_{i1}) \qquad (2)$$

is the probability that the *ith* individual will be participating in the labor force when characterized by the variables $X_{i1}, \ldots, X_{ik}$; that is, $\theta_i = Pr(Y_i = 1; X_i)$ or, alternatively, the expected proportion from a set of persons confronted with like stimuli that will be participating in the labor force.

The implication of this model is that repeated trials on individuals with the same characteristics will produce some successes and some failures in accordance with the Bernoulli parameter, θ_i. This may be contrasted with a discriminant model where two immutable populations, successes and failures, exist and the problem is to classify individuals into one or the other.

The empirical problem is to obtain estimates for the θ_i in (2). A *linear probability model* specifies that

$$\theta_i = E(Y_i; X_i) = Pr(Y_i = 1; \underline{X})$$
$$= \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij},\ i = 1, \ldots, N \qquad (3)$$

However, there is nothing inherent in unconstrained linear regression estimation of θ_i that guarantees that the predicted values will fall in unit interval. The predicted value, $\bar{\theta}_i$, can be reconciled with the probability interpretation by applying the following rule, where θ_i^* is the predicted probability:

$$\theta_i^* = 1 \text{ if } \bar{\theta}_i \geqslant 1\ ;$$
$$\theta_i^* = \bar{\theta}_i \text{ if } 0 < \bar{\theta}_i < 1\ ;$$
$$\theta_i^* = 0 \text{ if } \bar{\theta}_i \leqslant 0\ .$$

This artificial rule circumvents the fact that the least squares estimate extends outside the unit interval, but the estimates are no longer unbiased for θ_i.

A second weakness in applying linear regression to (3) is that the error has discrete distribution and had a diagonal covariance matrix with elements $[\theta_i(1 - \theta_i)]$ along the diagonal. Because of the changing variance of the error, the OLS coefficients estimator, although unbiased, is not efficient. Under heteroscedasticity the standard tests of significance do not apply. McGillivary [13] has shown that $\bar{\theta}_i(1 - \bar{\theta}_i)$ is a consistent estimator of the variance of the error, but $\bar{\theta}_i(1 - \bar{\theta}_i)$ may be negative. An application of weighted least squares (WLS) is limited to taking those predicted values from the OLS estimates that lie inside the unit interval causing a loss in the number of observations.

In summary, empirical difficulties arise in treating the quantal response model as a linear probability regression model. Since the disturbances are non-normal and heteroscedastic, even the asymptotic use of the standard estimators and test statistics is questionable. The single advantage to the linear regression model is that the computational procedure is relatively simple.

Figure I

OLS Fit of True Sigmoid Function

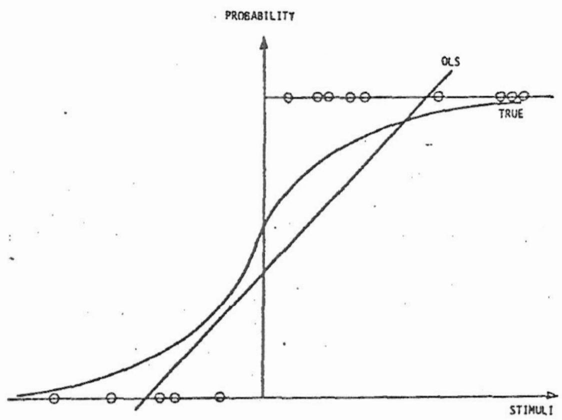

Figure I illustrates the seriousness of misspecification. If the true function is sigmoid, the dichotomous observations represented by the small circles will result in an OLS fit as shown. Not only will the estimated function extend outside the unit interval, but the estimated relationships between θ and the stimuli are likely to be seriously biased not only at the extremes.

The *logistic model* specifies that

$$\theta_i = 1/[1 + \exp[-(\beta_0 + \sum_{j=1} \beta_j X_{ij})]], \quad i = 1, \ldots, N \qquad (4)$$

which has been proposed by Berkson [5] and Theil [16] within the contexts of bio-assay and information theory respectively. The primary advantage of (4) is that θ_i is bounded by the values of zero and one.

To estimate β in (4) by linear regression, the transformation

$$\ln[\theta_i/(1-\theta_i)] = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}, \quad i = 1, \ldots, N \qquad (5)$$

requires sample observations of θ_i. When confronted with single observations of Y_i for each X_i, observations at different values of X_i must be combined into classes and the relative frequencies f_g for each class computed:

$$\ln[f_g/(1-f_g)] = \beta_0 + \sum_{j=1}^{k} \beta_j \bar{X}_{gj} + \nu_g; \quad g = 1, \ldots, G \qquad (6)$$

with G denoting the number of classes, and each $\bar{X}_{gj}$ as the mean of the observations X_{ij} in the gth class. This LOGIT specification exhibits heteroscedasticity due to unequal sized groups [12, 16]. This suggests that (6) be estimated by WLS. The grouping technique tends to drastically reduce the sample size, and the detail contained in micro-data will be reduced. Aggregation error may also become a problem. There is also a problem of appropriate grouping since f_g cannot be allowed to be zero or one, but large size groups reduce the effective sample size. Finally, this model does not provide a least squares solution for the θ_i, but rather for a quite arbitrary non-linear transformation, $\ln[\theta_i/(1-\theta_i)]$.

As an alternative to the logistic that is also mathematically constrained to the unit interval, the *Urban's curve* model specifies that

$$\theta_i = 0.5 + [\tan^{-1}(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})]/\pi, \qquad (7)$$

which leads to the transformation,

$$\tan[(2\theta_i - 1)\pi/2] = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}; \quad i = 1, \ldots, N \qquad (8)$$

Clearly, the same problems exist as with the logistic model, e.g., data must be combined into classes to use linear regression techniques. Ashton [4] compared the Urban's and logistic transformations, as well as the probit and sine transformations, and found that the Urban's curve approached the limits of the unit interval slowly compared to the other sigmoid transformations which were all similar over the whole range.

Maximum likelihood estimation seems to be an appropriate technique to estimate the parameters in (4) or (8) above. The MLE has many desirable large sample properties, including consistency, asymptotic unbiasedness and efficiency, and invariance [12]. Given a random sample, the likelihood function is given by

$$L(\beta_0, \beta_1, \ldots, \beta_k, y) = \prod_{i=1}^{N} f(y_i, \theta_i)\,, \tag{9}$$

or

$$\ln L = \sum_{i=1}^{N} \ln [f(y_i, \theta_i)]\,. \tag{10}$$

Given Bernoulli observations for the logistic (4) or the alternative Urban's curve (8) the log- likelihood is

$$\ln L(\theta) = \sum_{i=1}^{N} [y_i \ln \theta_i + (1 - y_i) \ln (1 - \theta_i)]\,. \tag{11}$$

The problem is to determine the values of $\hat{\beta}_0, \ldots, \hat{\beta}_k$ to maximize $\ln L$. The set of $(k = 1)$ equations $\partial \ln L / \partial \beta_j = 0, j = 0, \ldots, k$, are transcendental and there is no closed form solution for $\hat{\beta}$. This leads to the use of an iterative ascent method such as the first-order gradient method, Newton's second-order method using both the gradient and Hessian, or the method of scoring, where the Hessian is replaced by its expectation [12].

Experience with all three of these classes of ascent methods indicates that the method of scoring is computationally the most acceptable from the standpoint of successful convergence in reasonable time. The method of scoring provides successive estimates $\hat{\beta}_{(i)}, \hat{\beta}_{(i+1)}, \ldots$, according to the equations $\hat{\beta}_{(i)} = \hat{\beta}_{(i+1)} - \Delta' H$ where H is the $(k + 1)$ by $(k + 1)$ matrix whose elements are $E(\partial^2 \ln L / \partial \beta_j \partial \beta_m)$ and Δ is the gradient of $\ln L$ whose elements are $\partial \ln L / \partial \beta_j$.

Even with the best of methods, convergence to the MLE can be problematic as the size of β increases for a highly non-linear model such as (4) or (8). The topic of iterative MLE in highly non-linear models has been investigated by Brown [7]. He has developed computer methods leading to successful convergence for large dimensionality and highly non-linear models. These techniques were used for the present specifications.

3. The Participation Decision

Assume that a potential labor market entrant wishes to maximize his expected utility from the net present value between occupations. What is needed first is to rank order the present dollar value of any number of occupational alternatives to the alternative of not participating in the labor force. Consider a decision by the *ith* individual to participate in the labor force during the time interval (τ, T). Let Ω_j^i be the discounted net monetary gain from participation in the *jth* occupation for which that individual is qualified.

The net present value of the *jth* occupation can be expressed as:

$$\Omega_j^i = \int_\tau^T r(t) \exp(-\rho t) dt - \int_\tau^T w(t) \exp(-\rho t) dt - C \exp(-\rho t) \tag{12}$$

where

$r(t) = p(t) r^0 + [1 - p(t)] r^*$;
$p(t) =$ probability of employment in *jth* occupation;
$r^0 =$ market wage rate for *jth* occupation;
$r^* =$ unemployment compensation rate which may be zero;[1]
$\rho =$ subjective discount rate for *ith* individual;
$w(t) = p(t) w^*(t)$, the welfare payment reduction associated with the *jth* job if the *ith* person is employed; and
$C =$ fixed cost of entry into the *jth* occupation.

The value of welfare over the period for the *ith* person is

$$\Omega_{m+1}^i = \int_\tau^T W(t) \exp(-\rho t) dt \tag{13}$$

where $W(t)$ is the welfare payment.

While all of the above equations are expressed

[1] The unemployment benefit is discounted over the whole period since it represents a potential wage substitute even if never received. Moreover, it is paid by the employer and would most likely be passed on in the form of a higher money wage if it were not required.

in terms of dollars, the *ith* individual's utility may be substituted for dollars, assuming a monotonic utility function, without loss of generality.

If Ω_j^* is the largest value from the set Ω_1^i, ..., Ω_M^i, then the participation decision is

$$Y_i = 1 \text{ if } \Omega_j^* > \Omega_{M+1}^i ,$$
$$Y_i = 0 \text{ if } \Omega_j^* \leqslant \Omega_{M+1}^i ,$$

where Y_i stands for the dichotomous decision by the *ith* person to participate in the labor force. A decision maker compares the largest net present value among attainable occupations to his welfare alternative. If the largest return exceeds the welfare alternative, the agent chooses to participate in the labor force. That is, an individual will seek employment in the *jth* occupation if he expects a net monetary gain that is larger than any alternative.

The value of the welfare alternative and the welfare loss due to employment are such that they may be unique for each individual. The value depends on the particular welfare program requirements. For example, $W(t)$ may correspond to the AFDC full state standard and $w(t)$ may then correspond to the rate of reduction in the AFDC full state standard. For a description of the AFDC program see Solberg and Langille [15]. The decision to participate is dependent also on $r(t)$ which is determined by the level of the wage rate r^0, the level of r^* and the probability of employment in the *jth* occupation.

Associated with the LFP decision is a critical value of the wage rate, the reservation wage, above which $Y = 1$. In general there will exist a minimum level of the wage rate, r_{min}, below which no individual will decide to participate. As the wage rate rises above r_{min} a greater number of individuals will decide to participate where some differences in the critical wage exists since tastes vary as well as circumstance. Let N_j denote the total number of individuals participating in the *jth* occupation group, then

$$N_j = \sum_{i=1}^{N} Y_i(r^0); N_j \leqslant N \qquad (14)$$

Since Y_i is dependent on the wage rate, so is N_j. The value N represents the population of potential entrants. The N function is a step-function since Y is dichotomous; however, with large numbers of participants, this step-function can be approximated by a smooth curve like that in

Figure II

The Aggregate Participation Function

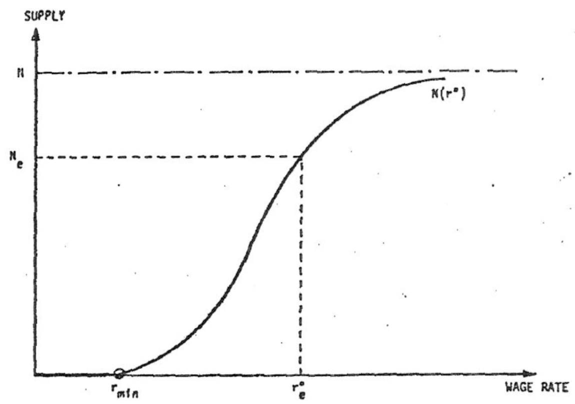

Figure II. The $N(r^0)$ curve in Figure II is that of a sigmoid curve and is consistent with a unimodal distribution, most people's tastes are more alike than different.

The labor force participation rate (LFPR) traditionally used to study LFP behavior can be computed directly from the aggregate participation relation. To find the LFPR, simply divide the equilibrium N_e, which is determined by the prevailing wage rate (r_e^0), into the total available population; thus, $LFPR = N_e/N$. The LFPR has often been used by researchers in their study of LFP, since LFP cannot be observed unless micro data is available. Note that since the aggregate participation function is sigmoid shaped, so the LFPR function must be also.

The labor force participation rate has the natural interpretation as a point estimate of the probability of labor force participation. If $\theta = Pr(Y = 1)$ and $1 - \theta = Pr(Y = 0)$, then the random variable

$$\sum_{i=1}^{n} Y_i$$

is binomial over n independent trials and $LFPR = \Sigma Y_i/n$ is an unbiased estimator of θ. Moreover, $\Sigma Y_i/n$ is asymptotically normally distributed which is unimodal and has a sigmoid shaped distribution function.

There are some important empirical implications which are also immediately obvious. First, the LFPR and LFP should always be non-negatively related to the wage rate.[2] Second, estimation of the LFPR calls for a sigmoid shaped functional form. Finally, the liberalization of the disregard criteria in welfare programs like the AFDC Program will cause an increase in the net present value of any occupation for those agents categorically and financially eligible and will therefore increase the LFPR. If the skill constraint restricts occupational choice to the secondary labor market, there will be little loss in empirical relevance if occupation is defined loosely allowing the comparison of the effects due to earned income, welfare income, unemployment benefits income, other independent income, and the probability of employment without regard to occupation.

4. The Data

A subfile was created from the March 1970 Person Family file of the Current Population Survey which selected observations if the family head was female and if dependent children were present, which may be viewed as the AFDC population-at-risk.[3] Only those family heads who were in the civilian non-institutionalized population whose primary source of income was not gained from self-employment in agriculture were included in the universe. Any observation which corresponded to family heads over the age of 70 years was omitted in order to limit the universe to those who could reasonably be expected to participate in the labor force. This restriction also tended to delete cases earning retirement or old age assistance. These limitations resulted in a sample of 2,222 observations, 1,284 with $Y=1$ and 938 with $Y=0$.

The dichotomous dependent variable was assigned a value of unity if the head was working, with a job but not working, or looking for employment. LFP was assigned a value of zero for those heads who were at home, in school, unable to work, or had other reasons for not participating.

The independent variables include: expected earnings (EARNINGS), total actual earned income of the head in hundreds of dollars multiplied by one-minus the unemployment rate, a proxy to measure the influence of the probability of employment;[4,5] welfare (WELFARE), the combined total of income in hundreds of dollars received from all public assistance programs, AFDC, Old Aid Assistance, or Aid to the Blind and Totally Disabled; expected unemployment benefits (UCB), the combined total of income in hundreds received from unemployment compensation, workman's compensation, government employee pensions, and veteran's payments; other income (OTHER INCOME), residual family income in hundreds derived by subtracting the prior income categories from total family income; a dummy variable (SMSA) to identify whether the family resided in a central city SMSA; a dummy variable (KIDS) which indicates the presence of children five years old or less; a dummy variable (RACE) indicating the head's race was Black; the actual age of the family head (AGE); the highest grade of school attended by the head (EDUCATION). In addition to the income variables categorized as finely as the CPS would allow, the other explanatory variables were included in order to control for differences in tastes between individuals or environ-

[2] If pensions are included and the pensions depend on wage rates, then higher wage rates may cause earlier retirement and reduce labor force participation at a later date. Current LFP is not affected.

[3] The AFDC unemployed parent category was not included in this study since this constitutes a special and minor fraction of the total AFDC population.

[4] While the unemployment rate does not in general measure the probability of not finding a job for an individual, the inverse variation between the probability of employment and the decision to participate in the labor force is important to capture. Wickens [17] has shown that "... it is better to use even a poor proxy than to use none at all and omit the unobservable variable."

[5] An attempt to create an instrument for earnings by regression using the characteristics of the family head as explanatory variables was abandoned because of the extremely low predictive ability of the estimated relations; therefore, it is true that the earnings variable used and labor force participation are subject to tautological relationship. This does not detract from the main point of the paper.

mental influences, that is, to destratify the sample.

5. The Results

The three specifications, linear, logistic, and urban, were used to obtain forecasts of θ_i, based upon all 2,222 sample observations of Y_i and the nine independent variables defined in Section 4. Results are shown in Table 1.[6] For all three specifications, the probability θ_i is an increasing function of the argument βX. It is interesting to note that for all three, the corresponding regression coefficients of each independent variable in Table 1 have the same sign. This is reassuring since at least the direction of influence implied by previous research is likely to be correct.

TABLE 1

REGRESSION COEFFICIENTS

	Linear Model	Logistic Model	Urban Model
Intercept	0.519	-0.131	-2.218
Earnings	0.010	0.125	0.319
Welfare	-0.010	-0.042	-0.043
UCB	-0.087	-0.264	-0.078
Other Income	-0.001	-0.009	-0.012
SMSA	-0.026	-0.416	-0.636
Kids	-0.118	-0.733	-0.804
Race	0.047	0.315	0.317
Age	-0.004	-0.013	-0.010
Education	0.009	0.117	0.153
Model, $\theta =$	βX	$\frac{exp[\beta X]}{1 + exp[\beta X]}$	$\frac{1}{2} + \frac{1}{\Pi} tan^{-1} [\beta X]$

To test the forecasting ability of the specifications, the same regressions were rerun with a randomly selected subset of 1,111 out of the 2,222 observations. Then the resulting equations were used to forecast θ_i for the remaining 1,111 observations. The 1,111 forecast $\hat{\theta}_i$'s from each specification were classified into 20 cells, and the actual frequency count of Y_i's in these cells was obtained. The linear specification resulted in 141 infeasible forecasts of θ_i. These were replaced by 0.0 for $\hat{\theta}_i < 0$ and 1.0 for $\hat{\theta}_i > 1$. Results are shown in Table 2. To compare the fit of the three specifications, a chi-squared statistic was calculated for each: for the linear, $\chi^2 =$ 103.66; for the logistic, $\chi^2 = 29.78$; and for the urban, $\chi^2 = 27.49$. Clearly, the latter two are superior. In fact, the linear specification is rejected at the 5 percent level ($\chi^2_{20} = 31.41$) by a goodness-of-fit test.

To emphasize the danger in making inferences from the linear probability model, the standard errors of the estimated coefficients and the corresponding T-ratios are reported in Table 3. It should be emphasized that the T-test is not valid for the linear model, unless the empirical distribution was shown to be mound shaped and an appeal were made to the robustness of the statistic. The test would indicate UCB to be sig-

[6] The logistic was also examined by a stepwise likelihood estimation technique akin to stepwise regression. The order of entry of the independent variables was: WELFARE, EARNINGS, EDUCATION, OTHER INCOME, and KIDS. The remaining independent variables were not significant by a likelihood ratio test. The results are available on request.

TABLE 2

FORECASTS OF θ_i FOR LAST 1,111 OBSERVATIONS BASED ON FIRST 1,111 OBSERVATIONS

	Linear Model		Logistic Model		Urban Model	
Forecast θ_i	$Y=0$	$Y=1$	$Y=0$	$Y=1$	$Y=0$	$Y=1$
0.0 – 0.05	29*	0	21	1	1	0
0.05 – 0.10	18	4	60	7	83	8
0.10 – 0.15	22	3	91	6	155	14
0.15 – 0.20	30	4	59	11	87	28
0.20 – 0.25	42	7	71	19	50	14
0.25 – 0.30	56	13	45	21	25	7
0.30 – 0.35	68	12	39	16	15	9
0.35 – 0.40	77	28	26	17	8	6
0.40 – 0.45	48	27	15	12	4	7
0.45 – 0.50	42	39	8	18	3	7
0.50 – 0.55	17	38	6	14	1	2
0.55 – 0.60	9	36	4	16	0	6
0.60 – 0.65	2	45	8	14	2	3
0.65 – 0.70	6	38	2	13	1	7
0.70 – 0.75	4	37	1	16	3	7
0.75 – 0.80	0	39	4	17	3	15
0.80 – 0.85	0	43	1	32	6	22
0.85 – 0.90	2	39	4	38	8	18
0.90 – 0.95	3	27	3	54	9	96
0.95 – 1.00	1	156**	8	293	12	359
Totals	**476**	**635**	**476**	**635**	**476**	**635**

*16 of these forecasts were less than 0.0
**125 of these forecasts were greater than 1.0

nificantly different from zero at 10 percent level of significance in the linear model. But UCB fails in the logistic or urban model. Further, SMSA is not significant in the linear model, but it is significant in both the logistic and urban model, but it is not significant at 5 percent in the logistic or urban model. The RACE variable was statistically insignificant only in the urban model. Except for the RACE variable, the logistic and urban model are in close agreement, but they are contradictory to the linear model in several important variables.

To facilitate comparison between the linear model and the highly non-linear logistic and urban models, the derivative of each function with respect to any explanatory variable was computed and evaluated at the means of the explanatory variables. These results are reported in Table 4. Except for the rates of change of the dummy variables SMSA, KIDS, and RACE, the rates of change are remarkably similar for the models with one important exception, the earnings and welfare variables. While the coefficients of the EARNINGS and WELFARE variables are of the same magnitude in the linear model, indicating equal subjective valuation of earnings and welfare income, the coefficients of EARNINGS are much greater in magnitude relative to the coefficients WELFARE in both the sigmoid specifications. Policy implications from the sigmoid curves would be quite different from those implied by the linear model.

TABLE 3

STANDARD ERRORS
(T-RATIO IN PARENTHESIS)

Variable	Linear	Logistic	Urban
EARNINGS*	3.90 E-04 (25.64)	6.75 E-03 (18.52)	3.10 E-02 (10.27)
WELFARE*	8.90 E-04 (11.23)	7.36 E-03 (5.70)	1.38 E-02 (3.12)
UCB	5.03 E-02 (1.73)	3.45 E-01 (0.77)	4.73 E-01 (0.16)
OTHER INCOME*	3.60 E-04 (2.78)	2.74 E-03 (3.28)	4.84 E-03 (2.48)
SMSA	1.70 E-02 (1.53)	1.37 E-01 (3.03)	2.31 E-01 (2.75)
KIDS*	2.17 E-02 (5.44)	1.71 E-01 (4.27)	2.76 E-01 (2.91)
RACE	1.81 E-02 (2.60)	1.43 E-01 (2.20)	2.26 E-01 (1.40)
AGE	9.10 E-04 (4.40)	6.93 E-03 (1.88)	1.00 E-02 (1.00)
EDUCATION*	3.05 E-03 (2.95)	2.43 E-02 (4.81)	3.95 E-02 (3.87)

*Significant by likelihood ratio test in the logistic.

6. Summary

In summary, the functional form makes quite a difference. An investigator should be quite wary of making generalizations based on any single specification or estimation technique. However, the above results have shown in striking fashion the superiority of MLE of the sigmoid specifications over the OLS estimation of the linear probability specification. Although the logistic or urban specification require iterative solution, this is no barrier on a modern digital computer, with appropriate special algorithms. A further advantage of the MLE is the asymptotic normality of the estimates of θ_i which permits large sample interval estimation, and the iteration method of scoring employed yields directly an estimate of the standard deviation of each normally distributed $\hat{\theta}_i$. Also standard tests of significance are now applicable.

Perhaps most importantly, the sigmoid specifications are consistent with a probability interpretation since the estimates lie inside the unit interval, and the sigmoid shape is consistent with the assumed unimodal distribution of the participation decision.

In conclusion, results reported in previous investigations of the probability of labor force participation or labor force participation rate which have relied on the least squares estimation of a linear probability specification are likely to be unreliable as to the magnitude of the response attributed to changes in explanatory variables.

REFERENCES

[1] D. J. Aigner, "An Appropriate Econometric Framework for Estimating a Labor Supply Function from the SEO File," *International Economic Review,* 15, No. 1, 1974, pp. 59-68.

[2] ——————, "MSE Dominance of Least Squares with Errors-of-Observation," *Journal of Econometrics,* 2, 1974, pp. 365-72.

[3] Takeshi Amemiya and Michael Boskin, "Regres-

TABLE 4

RATES OF CHANGE AT MEANS

Variable	Linear	Logistic	Urban
EARNINGS	0.010	0.027	0.015
WELFARE	-0.010	-0.009	-0.002
UCB	-0.087	-0.058	-0.004
OTHER INCOME	-0.001	-0.002	-0.001
SMSA*	-0.026	-0.091	-0.031
KIDS*	-0.118	-0.160	-0.039
RACE*	0.047	0.069	0.015
AGE	-0.004	-0.003	-0.001
EDUCATION	0.009	0.026	0.007
$\frac{\partial\theta}{\partial X_p}$	β_p	$\frac{\beta_p exp[\beta X]}{[1 + exp[\beta X]]^2}$	$\frac{\beta_p}{\Pi[1 + [\beta X]^2]}$

*These rates of change should be interpreted with caution since these are dummy variables which might best be thought of as location parameters. Here SMSA = 0.840, KIDS = 0.279, and RACE = 0.726.

sion Analysis when the Dependent Variable is Truncated Lognormal, with an Application to the Determinants of the Duration of Welfare Dependency," *International Economic Review*, 15, No. 2, 1974, pp. 485-96.

[4] W. D. Ashton, "The Logit Transformation," *Griffin's Statistical Monographs and Courses, No. 32*, Alan Stuart, Ed., Griffin, London, 1972.

[5] J. Berkson, "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, 39, 1944, pp. 357-65.

[6] W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation*, Princeton University Press, 1969.

[7] Gerald Brown, "Nonlinear Statistical Estimation with Numerical Maximum Likelihood," *Western Management Science Institute* (UCLA) Technical Report 222, 1974.

[8] Glen G. Cain, *Married Women in the Labor Force*, University of Chicago Press, 1966.

[9] M. S. Cohen, S. A. Rea, Jr., and R. I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, U.S. Department of Labor, 1970.

[10] Morley Gunderson, "Retention of Trainees, A Study with Dichotomous Dependent Variables," *Journal of Econometrics*, 2, 1974, pp. 79-93.

[11] L. J. Hausman, "The Impact of Welfare on the Work Effort of AFDC Mothers," *The President's Commission on Income Maintenance Programs-Technical Studies*, 1970, pp. 83-100.

[12] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 2, 3rd ed., Halner Publishing Company, 1973.

[13] R. G. McGillivary, "Estimating the Linear Probability Function," *Econometrica*, September 1970, pp. 775-76.

[14] Marc Nerlove and S. James Press, *Univariate and Multivariate Log-linear and Logistic Models*, Rand-1306EDA/NIH, December 1973.

[15] E. J. Solberg and F. Langille, "The Wage Rate, Potential Work Incentives, and Benefit Payment Reduction in the AFDC Program," *The Quarterly Review of Economics and Business*, Summer 1974, pp. 85-100.

[16] Henri Theil, *Economics and Information Theory*, Rand McNally and Co., 1967.

[17] M. R. Wickens, "A Note on the Use of Proxy Variables," *Econometrica*, July 1972, pp. 759-61.