Estimation of Relationships for Limited

Dependent Variables

James Tobin

September 15, 1955
(As revised July 25, 1956)

"What do you mean, less than nothing?" replied Wilbur.
"I don't think there is any such thing as less than nothing
Nothing is absolutely the limit of nothingness. It's the
lowest you can go. It's the end of the line. How can some-
thing be less than nothing? If there were something that
was less than nothing, then nothing would not be nothing, it
would be something--even though it's just a very little bit
of something. But if nothing is nothing, then nothing has
nothing that is less than it is."

> E. B. White, Charlotte's Web
> (New York: Harper, 1952) p. 28.

In economic surveys of households, many variables have the following
characteristics: The variable has a lower, or upper, limit and takes on
the limiting value for a substantial number of respondents. For the
remaining respondents, the variable takes on a wide range of values
above, or below, the limit.

The phenomenon is quite familiar to students of Engel curves, rela-
tionships showing how household expenditures on various categories of
goods vary with household income. For many categories--"luxuries"--
zero expenditures are the rule at low income levels. A single straight
line cannot, therefore, represent the Engel curve for both low and high
incomes. If individual households were identical, except for income
level, the Engel curve would be a broken line like OAB in Figure 1.
But if the critical income level OA were not the same for all house-
holds, the average Engel curve for groups of households would look
like the curve OB. A similar kind of effect occurs under rationing of
a consumers' good. The ration is an upper limit; many consumers choose
to take their full ration, but some prefer to buy less.*

---

* For theoretical exposition of the effects of aggregate demand
functions of lower or upper limits on individual expenditure in combina-
tion with differences in tastes among households, see [4] and [6] and
the literature there cited.

As a specific example, many-- indeed, most--households would report
zero expenditures on automobiles or major household durable goods during

Figure 1
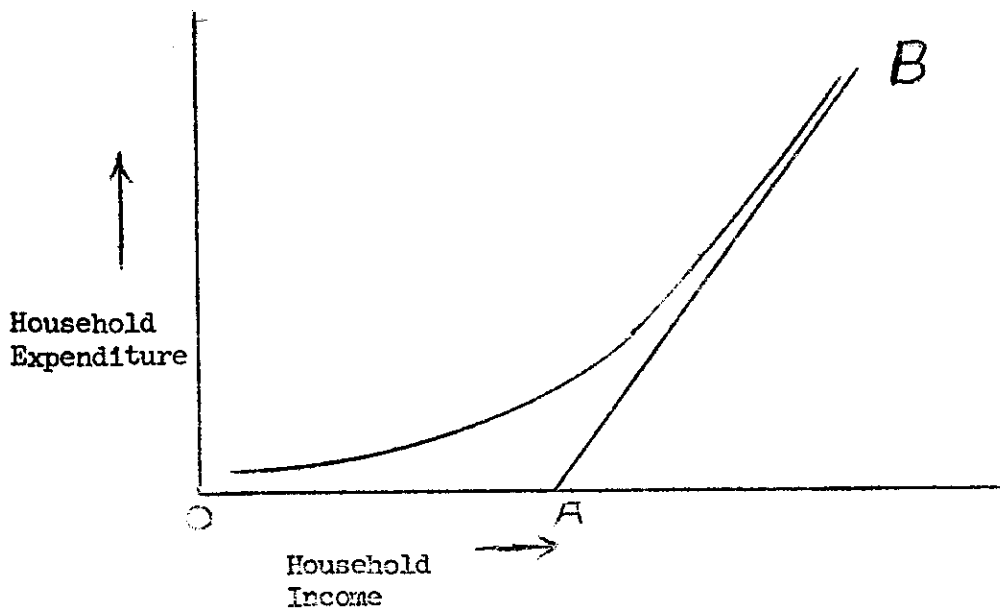
any given year. Among those households who made any such expenditure, there would be wide variability in amount.*

---

* For figures on frequency of purchases and on the distribution of amounts spent among purchasers, see [1], Part II; Supplementary Tables 1, 5, and 10.

In other cases, the lower limit is not necessarily zero, nor is it the same for all households. Consider the net change in a household's holding of liquid assets during a year. This variable can be either positive or negative. But it cannot be smaller than the negative of the household's holdings of liquid assets at the beginning of the year; one cannot liquidate more assets than he owns.

Account should be taken of the concentration of observations at the limiting value in statistical estimation of the relationship of a limited variable to other variables and in testing hypotheses about the relationship. An explanatory variable in such a relationship may be expected to influence both the probability of limit responses and the size of non-limit responses. If only the probability of limit and non-limit responses, without regard for the value of non-limit responses were to be explained, probit analysis provides a suitable statistical model. (See [5].) But it is inefficient to throw away information on

the value of the dependent variable when it is available. If only the value of the variable were to be explained, if there were no concentration of observations at a limit, multiple regression would be an appropriate statistical technique. But when there is such concentration, the assumptions of the multiple regression model are not realized. According to that model, it should be possible to have values of the explanatory variables for which the expected value of the dependent variable is its limiting value; and from this expected value, as from other expected values, it should be possible to have negative as well as positive deviations.

A hybrid of probit analysis and multiple regression seems to be called for, and it is the purpose of this paper to present such a model.

The Model

Let W be a limited dependent variable, with a lower limit of L. The limit may not be the same for all households in the population. Let Y be a linear combination of the independent variables $(X_1, X_2, \ldots X_m)$, to which W is by hypothesis related.

$$(1) \qquad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_m X_m$$

Households differ from each other in their behavior regarding W for reasons for which differences in the independent variables X and the lower limit L do not fully account. Those other differences are taken to be random and to be reflected in ε, a random variable with mean zero and standard deviation σ, distributed normally over the population of households. Household behavior is then assumed to be as follows:

$$(2) \qquad \begin{array}{ll} W = L & (Y - \varepsilon < L) \\ W = Y - \varepsilon & (Y - \varepsilon \geq L) \end{array}$$

Let P(x) represent the value of the cumulative unit-normal distribution function at x; let $Q(x) = 1 - P(x)$; let Z(x) be the value of the unit-normal probability density function at x. The distribution of W - L may be derived from the distribution of ε, as follows:

For given values of the linear combination Y and the limit L,

(3) $\Pr(W = L|Y, L) = \Pr(\varepsilon > Y - L) = Q(\frac{Y - L}{\sigma})$.

(4) $\Pr(W > x \geq L|Y) = \Pr(Y - \varepsilon > x)$

$= \Pr(\varepsilon < Y - x) = P(\frac{Y - x}{\sigma})$

Consequently, the cumulative distribution function for W, for given Y and L, is:

(5) $F(x; Y, L) = 0$ \hspace{2cm} $(x < L)$

$F(L; Y, L) = Q(\frac{Y - L}{\sigma})$

$F(x; Y, L) = Q(\frac{Y - x}{\sigma})$ \hspace{2cm} $(x > L)$

The corresponding probability density function is:

(6) $f(x; Y, L) = \frac{1}{\sigma} B(\frac{Y - x}{\sigma})$ \hspace{2cm} $(x > L)$

The expected value of W for given values of Y and L is:

$$E(W; Y, L) = L \, Q(\frac{Y - L}{\sigma}) + \int_{L}^{\infty} \frac{x}{\sigma} B(\frac{Y - x}{\sigma})dx$$

$$= L \, Q(\frac{Y - L}{\sigma}) + Y \int_{-\infty}^{\frac{Y - L}{\sigma}} B(x)dx + \sigma \int_{-\infty}^{\frac{Y - L}{\sigma}} -xB(x)dx$$

Since $- xB(x) = B'(x) = \frac{dB(x)}{dx}$, we have:

(7) $E(W; Y, L) = L \, Q(\frac{Y - L}{\sigma}) + YP(\frac{Y - L}{\sigma}) + \sigma B(\frac{Y - L}{\sigma})$

The Maximum Likelihood Solution

A sample includes q observations of households for whom W is at the limit L. Each observation consists of a limit L_i', to which the dependent variable W_i' is equal, and a set of values of the independent variables $(X_{1i}', X_{2i}', \ldots X_{mi}')$, where i is a subscript to denote the observation and runs from 1 to q. A sample also includes r observations for which W is above the limit L; each one may be described as $(W_j, L_j, X_{1j}, X_{2j}, \ldots X_{mj})$ where j runs from 1 to r

Let $(a_0, a_1, a_2, \ldots a_m, a)$ be estimates of $(\frac{\beta_0}{\sigma}, \frac{\beta_1}{\sigma}, \frac{\beta_2}{\sigma} \ldots \frac{\beta_m}{\sigma}, \frac{1}{\sigma})$.

Let $I_i' = Y_i' a = a_0 + a_1 X_{1i}' + a_2 X_{2i}' + \ldots a_m X_{mi}'$

and let $I_j = Y_j a = a_0 + a_1 X_{1j} + a_2 X_{2j} + \ldots a_m X_{mj}$.

The likelihood of a sample is:

$$(8) \qquad \phi(a_0, a_1, \ldots a_m, a) = \prod_{i=1}^{q} F(L_i'; Y_i', L_i') \cdot \prod_{j=1}^{r} f(W_j; Y_j, L_j)$$

$$= \prod_{i=1}^{q} Q\left(\frac{Y_i' - L_i'}{1/a}\right) \cdot \prod_{j=1}^{r} a\, Z\left(\frac{Y_j - W_j}{1/a}\right)$$

$$= \prod_{i=1}^{q} Q(I_i' - aW_i') \cdot \prod_{j=1}^{r} a\, Z(I_j - aW_j)$$

The natural logarithm of ϕ,

$$(9) \qquad \ln\phi = \phi^*(a_0, a_1, \ldots a_m, a) =$$

$$\sum_{i=1}^{q} \ln Q(I_i' - aW_i') + r\ln a - \frac{r}{2}\ln 2\pi - \frac{1}{2}\sum_{j=1}^{r}(I_j - aW_j)^2$$

Let X_0 and X_0' be identically 1 for all i and j. Then setting the derivatives of $\phi*$ equal to zero gives the following system of $m + 2$ equations.

$$\phi_k^* = \frac{\partial \phi^*}{\partial a_k} = \sum_{i=1}^{q} \frac{-S(I_i' - aW_i')\, X_{ki}'}{Q(I_i' - aW_i')} - \sum_{j=1}^{r} (I_j - aW_j)X_{kj} = 0$$

$$(k = 0,1,2,\ldots m)$$

(10)

$$\phi_{m+1}^* = \frac{\partial \phi^*}{\partial a} = \sum_{i=1}^{q} \frac{S(I_i' - aW_i')\, W_i'}{Q(I_i' - aW_i')} + \frac{r}{a}$$

$$+ \sum_{j=1}^{r} (I_j - aW_j)\, W_j = 0$$

These equations are non-linear. The quantity $\dfrac{-S(x)}{Q(x)}$ is tabulated as Δ_{min} in [2], pp. 185-88, where the argument for the table is $x + 5$.

The matrix of second derivatives, obtained by differentiating (10) is given by (11). Here $w_{min}'(x)$ is the derivative of $-\Delta_{min}(x)$, and may, like Δ_{min}, be found by entering the tables of [2], pp. 185-88 with the argument $x + 5$.

$$\phi^*_{kt} = \frac{\partial^2 \phi^*}{\partial a_k \partial a_t} = -\sum_{i=1}^{q} X'_{ki} X'_{ti} w'_{min} (I'_i - aW'_i)$$

$$- \sum_{j=1}^{r} X_{kj} X_{tj}$$

$$(k,t = 0,1,\ldots m)$$

(11)
$$\phi^*_{k,m+1} = \frac{\partial^2 \phi^*}{\partial a_k \partial a} = +\sum_{i=1}^{q} W'_i X'_{ki} w'_{min} (I'_i - aW'_i)$$

$$+ \sum_{j=1}^{r} X_{kj} W_j$$

$$(k = 0,1,\ldots m)$$

$$\phi^*_{m+1,m+1} = \frac{\partial^2 \phi^*}{\partial a^2} = -\sum_{i=1}^{q} W'^2_i w'_{min} (I'_i - aW'_i) - \frac{r}{a^2}$$

$$- \sum_{j=1}^{r} W^2_j$$

Newton's method (See [3]) for iterative solution of (10) may be applied as follows: Let $(a_0^{(0)}, a_1^{(0)}, \ldots a_m^{(0)}, a_{m+1}^{(0)})$ be a trial solution, where, for notational convenience, a_{m+1} represents what has previously been written as simply a. (The choice of an initial trial solution will be discussed below.) New estimates $(a_0^{(0)} + \Delta a_0, a_1^{(0)} + \Delta a_1, \ldots a_m^{(0)} + \Delta a_m, a_{m+1}^{(0)} + \Delta a_{m+1})$ can be found by solving the set of $m + 1$ linear equations (12) for the Δa, where all the ϕ^*_k are assumed to be linear between the trial solution and the real solution.

$$\phi_k^* \ (a_0^{(0)} + \Delta a_0, \dots, a_m^{(0)} + \Delta a_m, a_{m+1}^{(0)} + \Delta a_{m+1})$$

$$\phi_k^* \ (a_0^{(0)}, a_1^{(0)}, \dots a_m^{(0)}, a_{m+1}^{(0)})$$

$$+ \sum_{t=0}^{m+1} \Delta a_t \phi_{kt}^* \ (a_0^{(0)}, a_1^{(0)} \dots a_m^{(0)}, a_{m+1}^{(0)}) = 0 \ .$$

$$(k = 0,1,2,\dots m+1)$$

(12)
$$\sum_{t=0}^{m+1} \Delta a_t \ \phi_{kt}^* \ (a_0^{(0)}, a_1^{(0)}, \dots a_m^{(0)}, a_{m+1}^{(0)})$$

$$= - \ \phi_k^* \ (a_0^{(0)}, a_1^{(0)}, \dots a_m^{(0)}, a_{m+1}^{(0)})$$

The process may be repeated with the new estimates as provisional estimates until the Δa are negligible.

If the final estimates a_k are used to evaluate the matrix of second derivatives (11) at the point of maximum likelihood, the negative inverse of that matrix gives large-sample estimates of the variances and covariances of the estimates a_k around the corresponding population parameters.

Tests of Hypotheses

Hypotheses about the relationship of W to one or more of the independent variables X may be tested by the likelihood-ratio method. Consider for example, the hypothesis that $\beta_1 = \beta_2 = \dots = \beta_m = 0$. This is the hypothesis that neither the probability nor the size of non-zero responses depends on the X's. According to the hypothesis, there remain only two parameters, β_0 and σ, to be estimated so as to maximize (9), which now becomes:

(13)
$$\phi^*(a_0, 0, 0, \dots 0, a) = \sum_{i=1}^{q} \ln Q(a_0 - a W_i')$$

$$- \frac{r}{2} \ln 2\pi + r \ln a - \frac{1}{2} \sum_{j=1}^{r} (a_0 - a W_j)^2$$

The maximizing values of a_0 and a may be found by solving equations (10) similarly simplified by putting all other a_k equal to zero. If (13) is evaluated with these solutions, then the logarithm of the likelihood ratio λ is the difference between (13) and the value of (9) when it is maximized without the constraint of the hypothesis. The statistic $-2 \log \lambda$ is for large samples approximately distributed by chi-square with m degrees of freedom. In similar fashion other hypotheses about subsets of the β's may be tested.

Initial trial estimates

The speed of convergence of iteration by Newton's method depends of course, on the choice of the initial trial estimates. The following procedure for finding initial estimates relies on linear approximation of the function $\dfrac{-Z(x)}{Q(x)}$ or in other words on a quadratic approximation of $\ln Q(x)$. This approximation converts the first $m+1$ equation of (10) into linear equations in the a_k for given a. These equations may be solved to give the a_k as linear functions of a. When these solutions are substituted in the $\underline{m+2}^{nd}$ equation, it becomes a quadratic equation in a.

Let x_0 be the unit normal deviate such that $Q(x_0) = \dfrac{\hat{q}}{q+r}$, the proportion of cases in the sample for which the variable W takes on its limit-value. A linear approximation to $\Delta_{min}(x) = \dfrac{-Z(x)}{Q(x)}$ is $\Delta_{min}(x) = \Delta_{min}(x_0) + (x - x_0)\,\Delta'_{min}(x_0)$

Remembering that $\Delta'_{min}(x) = -w'_{min}(x)$, we have:

$$(14) \quad \Delta_{min}(x) = \Delta_{min}(x_o) + x_o \, w'_{min}(x_o) - x \, w'_{min}(x_o)$$

$$= A + Bx$$

Substituting (14) in the first $m + 1$ equations of (10) gives

$$\sum_{i=1}^{q} (AX'_{ki} + Ba_o X'_{oi}X'_{ki} + Ba_1 X'_{1i} X'_{ki} + \ldots Ba_m X'_{mi} X'_{ki} -$$

$$Ba\, W'_i X'_{ki})$$

$$-\sum_{j=1}^{r} (a_o X_{oj} X_{kj} + a_1 X_{1j} X_{kj} + a_m X_{mj} X_{kj} - a\, W_j X_{kj})$$

$$= 0$$

$$(15) \quad a_o \left[\sum_{j=1}^{r} X_{oj} X_{kj} - B \sum_{i=1}^{q} X'_{oi} X'_{ki} \right] + a_1 \left[\sum_{j=1}^{r} X_{1j} X_{kj} - B \sum_{i=1}^{q} X'_{1i} X'_{ki} \right] + \ldots + a_m \left[\sum_{j=1}^{r} X_{mj} X_{kj} - B \sum_{i=1}^{q} X'_{mi} X'_{ki} \right] =$$

$$a \left[\sum_{j=1}^{r} W_j X_{kj} - B \sum_{j=1}^{q} W'_i X'_{ki} \right] + A \sum_{i=1}^{q} X'_{ki}$$

$$(k=0,1,2,\ldots m)$$

Solving (15) gives numbers g_k and h_k such that:

$$(16) \quad a_k = g_k + h_k\, a \qquad (k = 0, 1, 2, \ldots m)$$

The final equation of (10) is, after using the approximation of (14):

$$(17) \quad \frac{r}{a} + a_o \left[\sum_{j=1}^{r} X_{oj}\, W - B \sum_{i=1}^{q} X'_{oi}\, W'_i \right] +$$

$$\ldots + a_m \left[\sum_{j=1}^{r} X_{mj}\, W_j - B \sum_{j=1}^{q} X'_{mi}\, W'_i \right]$$

$$- a \left[\sum_{j=1}^{r} W_j^2 - B \sum_{i=1}^{q} W_i'^2 \right] - A \sum_{i=1}^{q} W'_i = 0$$

When (16) is substituted in (17), it becomes a quadratic equation in a. The solution of (17) may then be used in (16) to obtain initial trial estimates of all the coefficients.

An Example

For purposes of illustration, an example has been worked out using data from the reinterview portion of the 1952 and 1953 Surveys of Consumer Finances conducted by the Survey Research Center of the University of Michigan for the Board of Governors of the Federal Reserve System.*

The data refer to 735 primary non-farm spending units* who were interviewed twice, once in early 1952 and once in early 1953. The

---

* Of the 1036 spending units in the reinterview sample, these 735 have been the subject for calculations for other purposes and are therefore a convenient group to use in this analysis. Excluded are all spending units who had one or more of the following characteristics: (a) farm; (b) secondary. i.e. not the owner or principal tenant of the dwelling; (c) total income for the two years 1951-52 zero or negative; (d) not ascertained as to age of head of spending unit, amount of expenditure on durable goods during 1951-52, or amount of liquid asset holdings in early 1951. In addition, one extreme observation was excluded, where the spending unit had such a low positive two year income that the ratio of durable goods expenditure and, especailly, liquid asset holdings to income were very high.

---

frequencies, averages and other statistics for the reinterview

sample should not be taken as representative of the population of

the Unites States. The Surveys of Consumer Finances do collect data

on distributions of income, liquid assets, and durable goods purchases

that are representative of that population; tables on these distri-

butions may be found in [1]. But the reinterview sample, on which the

calcuations of this paper are based, fails to be representative

insofar as it omits spedding units who moved between the two surveys.

Moreover, these calculations are based on simple counts of sampled spending units,

without allowance for the fact that the sampling design gave some

spending units greater probablilities of being included in the sample

than others. The purpose of this example is not to estimate popula-

tion frequency distributions, but only to examine the relationship of

durable goods expenditure to age and liquid asset holdings within

this sample. It is not necessary to consider here how the relation-

ship exhibited in this sample differs from the one that would be ex-

hibited in a complete enumeration But it may well be that the sample

gives unbiased estimates of the parameters of the relationship,

even though it gives biased estimates of the separate frequency

distributions of the variables

The variables are as follows:

W   Ratio of 1951-52 total durable goods expenditure to 1951-52 total

disposable income. Durable goods expenditure is the two-year

sum of outlays, net of trade-ins or sales, for cars and major house-

hold appliances and furniture. Two-year disposable income is the

sum of the two annual incomes reported by the spending unit less

estimated federal income tax liabilities. Both expenditure and

income were reported for 1951 in the interview in early 1952, and

for 1952 in the second interview, in early 1953. Since expenditure

is necessarily zero or positive, and since zero and negative incomes

have been excluded, the ratio is necessarily zero or positive.

X_1 Age of the head of the spending unit, as reported in

1953, on the following scale:

18-24 yrs.:	1
25-34 yrs.:	2
35-44 yrs.:	3
45-54 yrs.:	4
55-64 yrs.:	5
65 or more years:	6

X_2 Ratio of liquid asset holdings at beginning of 1951

to 1951-52 total disposable income. Liquid asset holdings include

bank deposits, savings and loan association shares, postal savings

and government saving bonds.

In this example, the lower limit L is zero for all cases.
Table 1 shows the basic data.

Table 1

Sums of Squares and Cross Products

183 limit observations

	$X_o' = 1$	X_1'	X_2'	W'
$X_o' = 1$	183			
X_1'	824	4056		
X_2'	102.15	552.03	402.3333	
W'	0	0	0	0

552 non-limit observations

	$X_o = 1$	X_1	X_2	W
$X_o = 1$	552			
X_1	1976	8060		
X_2	168.06	751.54	255.6740	
W	61.449	207.598	20.559	13.113087

Table 2 presents the estimates of the parameters obtained by the
initial approximation and reports the successive iterations leading
to the maximum likelihood estimates. Estimates are shown also in
Table 3, on the assumption that there is no relation between W and
liquid asset holdings X_2.

In the approximation used to obtain initial trial values, the function

$$- \frac{Z(x)}{Q(x)}$$

was approximated linearly about the point $x_o = .67$, so that $Q(x_o) = .25$, the proportion of non-zero cases in the sample. Thus the constants A and B in (15) and (17) were equal to $-.76003$ and $-.75771$ repectively.

Table 2

Iterative Estimation of Parameters

	a_o	a_1	a_2	a
Initial trial values	1.326	-.2200	.0330	7.984
First derivatives	-4.398	-21.759	-1.812	.898
Second derivatives a_o	-680.557			
a_1	-2542.416	-10,805.486		
a_2	-238.41	-1128.653	-535.223	
a	61.449	207.598	20.559	-21.772
Indicated changes	.0152	-.00507	.00199	.0376
Second trial values	1.3407	-.2251	.0350	8.022
First derivatives	-.047	.292	.064	.002
Second derivatives a_o	-680,260			
a_1	-2540.666	-10,795.522		
a_2	-238.26	-1,127.900	-535.893	
a	61.449	207.598	20.559	-21.688
Indicated changes	-.0015	.00037	.00001	-.00064
Final estimates	1.3392	-.2247	.0350	8.022
Standard errors	(.118)	(.0295)	(.0495)	(.252)

Table 3

Iterative Estimation of Parameters

Assuming that $\beta_2 = 0$

	a_0	a_1	a
Initial trial values	1.337	-.219	8.040
First derivatives	-2.841	-16.124	-.001
Second derivatives: a_0 -680.419			
a_1 2541.428		10,799.12	
a 61.449		207.598	-21.652
Indicated Changes	.010	-.004	-.010
Second trial values	1.347	-.223	8.030
First derivatives	-.456	-2.017	+.116
Second deriviatives: a_1 -680.179			
a_1 2539.988		-10,790.472	
a 61.449		207.598	-21.674
Indicated changes	.001	.0003	-.005
Final estimates	1.347	-.223	8.030
Standard errors	(.117)	(.028)	(.252)

Estimates of the variances and covariances of the parameter estimates can be obtained from the negative of the inverse of the final matrix of second derivatives. These are shown in Table 4. The corresponding standard errors of the coefficients are given in the final rows of Tables 2 and 3.

Table 4

Estimated Variances and Covariances of Parameter Estimates

	a_0	a_1	a_2	a
a_0	+.0139			
a_1	-.00318	+.000867	~~.0235~~	
a_2	+.000880	-.000454	+.00245	
a	+.00987	-.00115	+.000470	+.0635

On assumption that $\beta_2 = 0$:

	a_0	a_1	a
a_0	+.0136		
a_1	-.00302	+.000784	
a	+.00970	-.00106	+.0635

The size of the standard error of a_2 indicates that the hypothesis that $\beta_2 = 0$, that there is no net relationship between expenditure and liquid asset holding, cannot be rejected. This hypothesis can also be tested, with the same conclusion, by the likelihood-ratio method. At the point of maximum likelihood, unrestricted by this hypothesis, $\phi*$ in (9) has the value $722.5 - \frac{552}{2} \ln 2\pi$. The final estimates in Table 3 correspond to the point of maximum liklihood restricted by the hypothesis that $\beta_2 = 0$. At this point $\phi*$ has the value $721.8 - \frac{552}{2} \ln 2\pi$. The statistic $-2 \ln \lambda$ is thus equal to 1.4, which is not a significant value of chi-square with one degree of freedom.

A test of the hypothesis that neither age nor liquid asset holding ^was^any effect on expenditure on durable goods may also be made by the likelihood-ratio method. Assuming, in accordance with the hypothesis that $\beta_1 = \beta_2 = 0$, the values of a_0 and a that maximize (13) are found to be .4839 and 7.720. For these values, $\phi^* + \frac{552}{2} \ln 2\pi$ is equal to 692.7. Hence $-2 \ln \lambda$ is equal to 59.6, a significant chi-square for two degrees of freedom. The hypothesis must be rejected. Thus this test, as well as the size of the estimated standard error of a_1 indicates a significant relationship of durable goods expenditure to age.

The relationship of W to X_1 and X_2, as estimated in Table 2, is shown in Figure 2, as the broken line ABC. The expected value of W implied by this relationship may be computed from (7) in the manner illustrated in Table 5. These points are also shown in Figure 2. For comparison, the least squares multiple regression of W on X_1 and X_2 has also been plotted. The estimated effect of liquid asset holding X_2 has been illustrated by drawing two graphs relating W to X_1, the first, Figure 2-a, on the assumption that $X_2 = 0$ and the second, Figure 2-b, on the assumption that $X_2 = 2$.

The expected value locus, estimated by the method of this paper is nonlinear. It is always above the broken line ABC, asymptotic to AB at the left where the probability of not buying $(W = 0)$ approaches zero, and asymptotic to BC at the right where the probability of buying $(W > 0)$ approaches ^zero^. Multiple regression approximates this non-linear locus with a linear relationship. As Figure 2 shows, the approximation is fairly close for the central

Table 5

Calculation of Expected Values

| X_1	$X_2 = 0$				$X_2 = 2$			
	$I = 1.3392$ $-.2247X_1$	Calculated Probability of Buying $P(I)$	$Z(I)$	Calculated Expected Value $E(W) = \dfrac{IP + Z}{8.022}$	$I = 1.4092$ $-.2247X_1$	Calculated Probability of Buying $P(I)$	$Z(I)$	Calculated Expected Value $E(W) = \dfrac{IP + Z}{8.022}$
0	1.3392	.910	.163	.172	1.4092	.921	.148	.180
1	1.1145	.867	.214	.147	1.1845	.882	.197	.155
2	.8898	.813	.267	.123	.9598	.832	.252	.131
3	.6651	.747	.319	.102	.7351	.768	.304	.108
4	.4404	.670	.362	.082	.5104	.695	.350	.088
5	.2157	.585	.390	.064	.2857	.612	.383	.070
6	-.0090	.497	.399	.049	.0610	.524	.398	.054
7	-.2337	.408	.388	.037	-.1637	.435	.393	.040
8	-.4584	.323	.359	.026	-.3884	.350	.370	.029

Figure 2-a
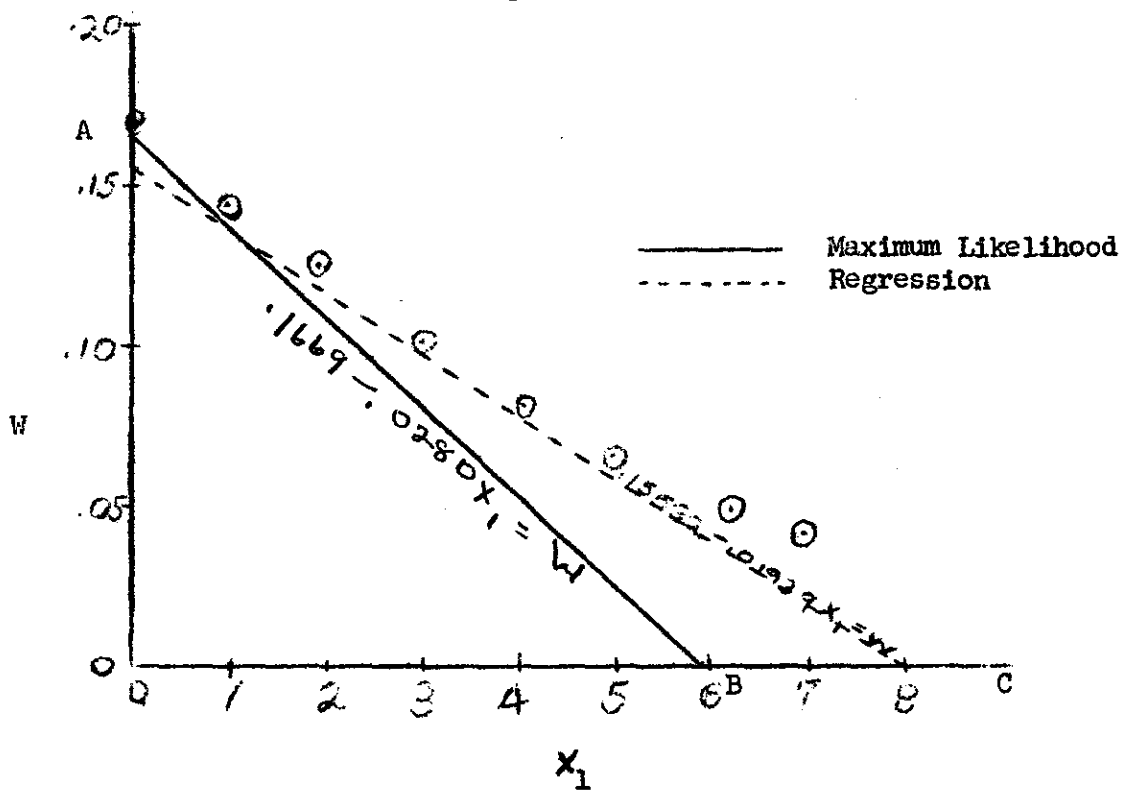
Figure 2-b
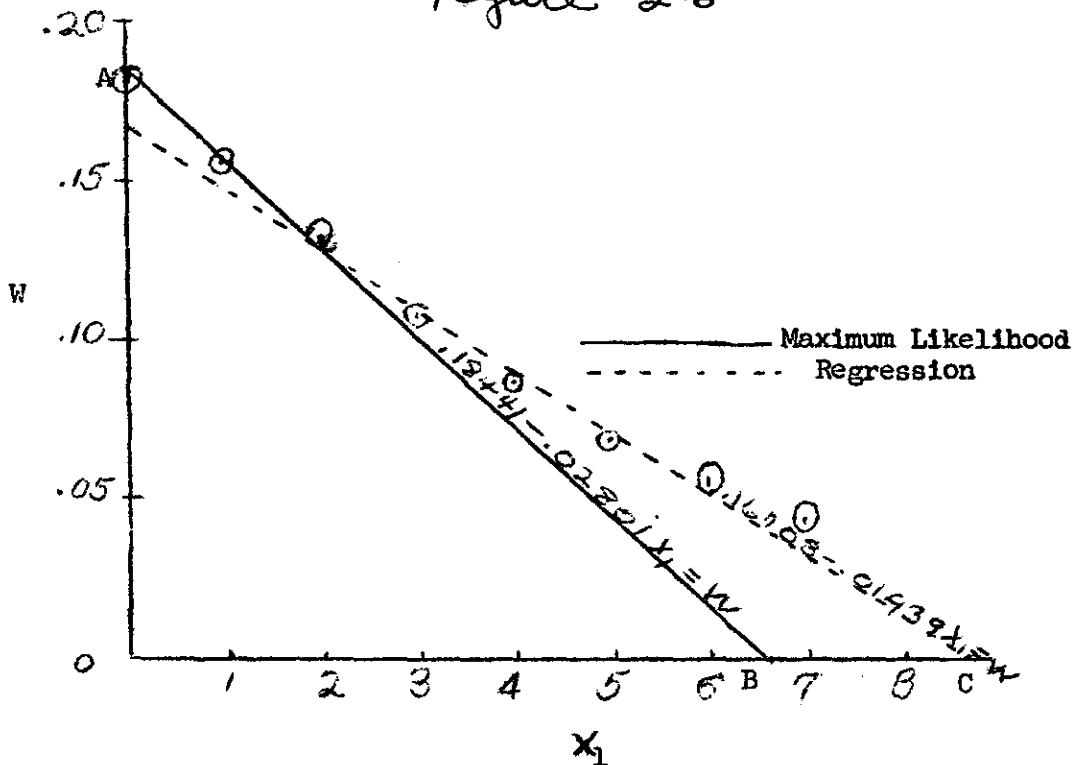
range of values of the sample. But outside the central range
there can be large discrepancies. There are indeed conceivable
values of the independent variables for which multiple regres-
sion would give negative estimates of the expected value of W.
It is true that the absence of negative observations in the
sample tends to keep the regression above the axis until extreme
values of the independent variables are reached. But this pro-
tection is purchased at the cost of making the regression line
so flat that expenditure is under-estimated at the opposite end.
These discrepancies could be important in predicting expenditure
for extreme cases or for aggregates which include extreme cases.

[1] Board of Governors of the Federal Reserve System, 1953 Survey of Consumer Finances, reprinted with supplementary tables from Federal Reserve Bulletin, March, June, July, August and September 1953.

[2] Cornfield, J. and Mantel, N., "Some New Aspects of the Application of Maximum Likelihood to the Calculation of the Dosage Response Curve," Journal of the American Statistical Association, 45 (1950), 181-210.

[3] Crockett, J. B. and Chernoff, H., "Gradient,Methods of Maximization," Pacific Journal of Mathematics, 5 (1955), 33-50. (Reprinted as Cowles Commission New Series Paper No. 92.)

[4] Farrell, M. J., "Some Aggregation Problems in Demand Analysis," Review of Economic Studies, 21 (1954) 193-203.

[5] Tobin, J., "The Application of Multivariate Profit Analysis to Economic Survey Data," CFDP No. 1.

[6] Tobin, J., "A Survey of the Theory of Rationing," Econometrica 20 (1952), 521-553.