Volume Title: Demographic and Economic Change in Developed Countries

Volume Author/Editor: Universities-National Bureau Committee for

Economic Research

Volume Publisher: Columbia University Press

Volume ISBN: 0-87014-302-6

Volume URL: http://www.nber.org/books/univ60-2

Publication Date: 1960

Chapter Title: An. Economic and Demographic Model of the Household Sector: A Progress Report

Chapter Author: Guy H. Orcutt, Alice M. Rivlin

Chapter URL: http://www.nber.org/chapters/c2391

Chapter pages in book: (p. 287 - 323)

An Economic and Demographic Model of the Household Sector: A Progress Report

GUY H. ORCUTT
HARVARD UNIVERSITY

ALICE M. RIVLIN
THE BROOKINGS INSTITUTION

In general, previous quantitative studies of demographic-economic interrelationships have concentrated on unidirectional relationships. Demographers have sometimes attempted to measure the influence of economic change on the vital rates, and economists have sometimes studied the impact of population growth or family size on economic variables, but the mutual interaction of economic and demographic factors has received more verbal than statistical acknowledgment.

Although the relevant data are still highly inadequate, we believe it is not too soon to try to construct a model of the household sector of the United States economy in which the interaction of economic and demographic variables is represented. The most promising approach to the problem lies in estimating the interrelations of economic and demographic variables at the level at which decisions to spend, save, work, and have children are actually made—that is, at the individual or household level. Certainly, we are also interested in the relationships between aggregates —total population, personal income, consumption, and so on—but it seems more feasible to obtain estimates of these relationships by simulating the behavior of a large number of individuals or households and aggregating the results of such behavior than it does to estimate relationships between aggregate time series directly.

Hence we have attempted to construct a model of the household sector formulated in terms of the demographic (birth, death, marriage, divorce) and economic (spending and saving) behavior of individual "decision units" and designed to be solved by computer simulation. The first

Note: The authors are heavily indebted to John Korbel, who programmed the calculations reported in Part III of this paper for the IBM 704, and to Fred Raines for his assistance in getting the basic survey data on magnetic tape. We are also deeply indebted to the Survey Research Center of the University of Michigan and the Board of Governors of the Federal Reserve Board for use of the 1956 Survey of Consumer Finances, and to the Ford Foundation, the Computation Center at M.I.T., the Brookings Institution, the Littauer Statistical Laboratory at Harvard, and the Wisconsin Alumni Research Foundation, for financial support, computational assistance and fellowships making possible the study of which this paper is one result. However, the conclusions and opinions in this article are those of the authors and not necessarily those of these institutions or individuals who have been helpful.

section of this paper will discuss the principal features of this type of model and its solution; the second will contain a brief description of the way in which we have estimated the demographic "outputs" of the household, and the last major section will give a more detailed discussion of the determination of mortgage debt, personal debt, liquid assets, and expenditures on selected durables.

The model presented is a far from finished product, and this paper should be regarded as a status report on research in progress, designed to explain the direction in which we are working, to illustrate what has been accomplished so far and the problems encountered, and to stimulate discussion and constructive criticism of the project. We believe, however, that our results thus far amply support the position that demographic and economic phenomena should be jointly considered, and we are prepared to offer the hypothesis that the interaction of economic and demographic variables at the household level provides a partial explanation of the business cycle.

Decision Unit Models and their Solution[1]

The most important feature of this type of model is that its units are decision-making units of the real economy—individuals, households, firms, financial institutions, and so forth. In the household sector, the units are individuals and combinations of individuals such as married couples, families, and spending units. These units have several kinds of possible behavior or "outputs." Individuals die, marry, set up new households; married couples have children and get divorced; spending units make purchases and acquire debts and assets.

Each unit has certain characteristics or "inputs." Inputs to individuals include sex, age, race, marital status; inputs to married couples include duration of marriage and the number of children born to the wife; inputs to spending units include their composition, asset, and debt position.

In addition to specific unit inputs, there are certain inputs to the system which are common to all units. These include the season and the calendar year and may include such aggregates as national income or employment. When a complete model of the economic system is constructed, these aggregates will be generated by the model, but we are obliged to "plug in" to our model of the consumer sector assumed values of some of the economic variables generated by other sectors.

[1] The general features of this type of model were more fully described in an earlier paper. See: Guy H. Orcutt, Martin Greenberger, and Alice M. Rivlin, "Decision-Unit Models and Simulation of the United States Economy," paper presented to the meetings of the Econometric Society in Philadelphia, December, 1957, mimeographed.

The outputs of a unit in a given time period depend on the inputs at the beginning of the period. This relationship between inputs and outputs is stochastic in nature; that is, it is the *probabilities* of occurrence of certain outputs, rather than the outputs themselves, which are regarded as functions of the inputs. The probability distributions which related inputs to outputs are called the "operating characteristics" of the units. For example, if the probability that a man will marry in a given period is taken to depend on his age and marital status, then the table (or function) which specifies the probabilities of marriage for males of various ages and marital conditions would be one of the operating characteristics of male units.

This stochastic feature of the model will be no surprise to demographers, who are quite used to estimating the probabilities of vital events happening to individuals of given characteristics and to using these probabilities together with the characteristics of the current population to predict the total number of such events one or more periods into the future. Birth, death, marriage, and divorce of individuals depend on such a myriad of factors that it is impossible to predict what will happen in individual cases. No one can say whether or not a particular individual will be alive a year from now, but, when large populations are "at risk," the vital events exhibit considerable regularity. An experienced life insurance company can determine within narrow limits and with a high degree of confidence the number that will survive for the next twelve months out of ten thousand insured persons of a given age.

The probabilistic approach seems similarly appropriate to analyzing the economic behavior of households. Here again, the actions of an individual household are influenced by hundreds of factors which we can neither ascertain nor measure. It is impossible to predict whether a particular household will or will not buy a house in the next month, but the probabilities of house purchase for various types of households can be estimated and these estimates used to predict aggregate house-buying in a large group of households with given characteristics.

Quite obviously, our model will be useful only if operating characteristics can be estimated which will prove to be stable or to change in a predictable way. The main research job of the project, in fact, is that of estimating these characteristics. It is a vast and challenging job, on which we have just got a good start.

The model is recursive, that is, the outputs of a unit in any period depend on its prior inputs, so that there is no simultaneous interaction between units and, hence, there are no simultaneous equations to be solved. This does not mean that units are conceived of as acting independently of each other, since the prior outputs or other units may be inputs to the unit in question; but it does mean that all interaction of units in the model is sequential rather than simultaneous. This recursive feature of the model necessitates the use of a rather short period of analysis, since many reactions, such as the response of marriage probabilities to changes in income, which may be sequential in fact, will appear simultaneous if a long period such as a year is considered. The period used in this model is the month. In most cases when monthly data were not available, annual data were converted to a monthly basis by applying a seasonal index.

Solution of the model is to be obtained by simulation on a large electronic computing machine. A population of several thousand units will be specified, the units being assigned ages, marriage durations, liquid assets, and other inputs in the proportions in which these characteristics appear in the base population. The simulation proceeds in one-month steps. In each month, each individual unit is considered in turn. For each possible output of the unit a probability of occurrence is specified by the relevant operating characteristics and the inputs to the unit at the beginning of the month. Whether the output occurs or not is determined by a random drawing from this probability distribution.

For example, suppose the simulation is started with the month of January, 1958, and the first individual considered is a single white male, age 34. We have already estimated the probability that a male with these characteristics will die in this month at, say, 0.0002; that is, this is the probability of death specified by the relevant operating characteristic. Then, in essence, we make a random drawing from a bag containing 10,000 balls, two of which are marked "die" and 9,998 of which are marked "not die." The man either dies and is eliminated from the population (and from his spending unit) or he lives through his month —depending on the outcome of the draw.

Random numbers generated by the computing machine provide us with a less cumbersome method of making the random drawings than balls in a bag. There may, of course, be more than two possible outcomes of the draw. The output "amount spent on durables," for example, might have four possible values—for example, $0, $0–100, $100–500, and over $500—or it might have many more.

When each possible output for each unit has been considered in this way, the first "pass" or month is complete. We enter the second month with a population of units which is slightly different both in size and composition from the initial one, since some individuals have died, or married, some couples have divorced, some babies have been born, some spending units have been created or destroyed, many have altered their asset and debt positions, and all surviving individuals are one month older. The whole procedure is then repeated for the second month and for as many more as desired.

In sum, the distinctive features of this model are its formulation in terms of decision units, the stochastic relation between inputs and outputs of units, and the solution by computer simulation. Bearing in mind the general outlines of the model, we turn now to a consideration of specific outputs and to the estimation of the operating characteristics associated with these outputs.

The Household Sector: Demographic Behavior

Estimating all the operating characteristics necessary for a decision unit model of the United States economy, or even of the consumer sector, is a rather formidable task. Fortunately, it can be broken down into manageable pieces. A logical place to start the work seemed to be to formulate a process by which the units in the consumer sector would be created and destroyed in the model in the proportions in which this creation and destruction takes place in the real world. Hence, our first efforts involved estimating the operating characteristics associated with birth, death, marriage, and divorce.

A model containing only these four outputs is of considerable interest in itself, since it is, in effect, a population projection model. As such, it differs from the population models in current use (such as the "cohort survival" method used by the Bureau of the Census, and others) in that: (1) it includes marriage and divorce (as well as birth and death) among the events to be projected, thus enabling us to project the number of married couples as well as the total population, and to make birth and marriage projections which are consistent with each other; (2) several new variables are included in predicting birth, marriage, and divorce probabilities. In particular, the probabilities of these outputs are made to depend in part on an income variable, which will ultimately be generated by the larger model when it is completed. Even in working with the four-output model, however, it is interesting to "plug in" various alternative values of this economic variable in order to generate alternative population projections.

A model limited to the four outputs of birth, death, marriage, and divorce is in the last stages of programming, and the results of actual runs should be available in the near future. The way in which the operating characteristics for this model were estimated has been described elsewhere and will be outlined only briefly here.[2]

DEATH

In the model, death is a possible output for all individuals, and when it occurs it simply eliminates the individual from the population. The operating characteristic associated with death gives the probability that an individual will die in a particular month—expressed as a function of his characteristics at the beginning of the month and any other relevant inputs such as the season of the year.

We projected probabilities of death by the simple expedient of fitting a trend to age-sex-race-specific mortality rates and extrapolating it into the future. A straight line was fitted to the logarithms of mortality rates in each age-sex-race group for the years 1933–1954, and the fit seemed quite good. Thus, we effectively assumed that the average annual rates of decrease in mortality rates in all groups observed in the past twenty years would continue for the next decade or so. Predicted annual mortality rates were converted to monthly rates by applying a seasonal multiplier computed from monthly mortality rates for 1946–1954 and assumed to remain constant.

MARRIAGE

Marriage is much more difficult to handle in a simulation model than is death, because marriage involves two individuals who will form a single new unit, a married couple. In order to accomplish this matching of individuals, the marriage process has been broken into two stages. In the first stage, it is decided whether or not a given individual will marry in the month. If marriage *is* predicted, then the characteristics of his mate are selected in the second stage. A mate with the requisite characteristics is then picked out of the model population and the two are united to form a married couple.

For the first stage, we needed an operating characteristic giving the probability that an individual with given characteristics will marry in the month. Marriage probabilities by sex, age, and marital status (that is, single and previously married) were estimated from census and vital statistics data for 1950. The relationship between marriage probabilities and economic conditions in the last three decades was then investigated. An index of the level of marriage probabilities for the years 1920 to 1955

[2] *ibid.*

was estimated and plotted against an index of economic conditions (personal disposable income per capita in constant prices). If the war years are omitted, the two series show remarkably similar movements over time.

In estimating the operating characteristic associated with the first stage of the marriage simulation, the historical relationship between the level of marriage rates and per capita income was taken into account. The operating characteristic estimated was of this form:

$$P \text{ (individual } i \text{ will marry in month)} = F_a \text{ (sex, age, marital status of } i) \cdot R \cdot F_b \text{ (month)}.$$

Here F_a is simply a table giving the marriage probabilities computed for 1950 by age, sex, and marital status, and F_b is a seasonal index computed for recent years and assumed to remain constant. R is a multiplicative factor which serves to raise and lower marriage probabilities in accordance with changes in the income variable. Actually R was of this form:

$$R = a + b \text{ (change in income)} - c \text{ (moving average of } R \text{ minus ``normal'' } R).$$

Here the final term serves to dampen the income effect by tending to return marriage rates to a level at which about 5 per cent of all women never marry, although a different "normal" could be substituted.

Incidentally, our model would have predicted the marriage down-turn associated with the 1957–1958 recession if we had got it running in time. Recent experience suggests, however, that marriage rates may be more sensitive to changes in employment than to changes in personal income.

The matching of marriage partners (stage two of the marriage process) was carried out on the basis of age and previous marital status only. The probability that a person of given age and marital status would pick a partner of a given age and marital status (and opposite sex, of course), were computed from data on marriages in 1955 and these probabilities were assumed to remain constant.

DIVORCE

Divorce is an output which dissolves the unit created by marriage into two separate units again. In our model, it is a possible output for any married couple in any month. At the same time, it is assumed that all married couples live together.

In estimating the operating characteristic associated with divorce we were severely hampered by lack of data, since, of all the vital events, divorce is the one about which we have least information. To get our divorce probabilities we took an approach similar to that used in estimating probabilities of marriage; that is, we first computed estimates of the probability of divorce (by duration of marriage) for a single year, but instead of holding these probabilities constant we introduced a multiplicative factor designed to raise and lower our divorce probabilities proportionately in response to changes in our income variable. The relationship between income and divorce rates was estimated from data for the period 1921–1955. However, although there appears to have been a positive correlation between divorce rates and economic conditions in the interwar period, this relationship is not very strong, nor has it persisted in the postwar period. Hence, while we have considerable confidence that the introduction of economic variables will improve marriage projections, we do not have the same confidence in this approach to divorce.

BIRTH

In our model, birth is an output of married couples only. Someone particularly interested in illegitimate birth rates might wish to introduce them into the model, but we did not feel that the present state of the data justified this refinement.

For our first model, we started by estimating age-parity specific birth rates for married women for a single year (1950). The data here are still inadequate, but the work of P. K. Whelpton and others is steadily filling the remaining gaps. We then investigated the relationship between income and age-parity-specific birth rates over the last three decades (to the extent that the data permitted) and made several alternative assumptions about the future course of this relationship. One such assumption is the following: (*a*) that first birth rates for zero-parity married women will remain constant, (*b*) that changes in birth rates for first through fourth parity women will exhibit the same relationships to changes of income that they did in the period 1920–1955 and (*c*) that rates for fifth and higher parity women will show a slow but steady decline. Our plan is to experiment with several such sets of assumptions in order to see what they imply with respect to completed family sizes with various patterns of income change. It should be noted that even if we hold age-parity specific birth rates for married women constant, our model will generate birth series which fluctuate in response to the business cycle especially in the lower birth orders, because our marriage rates depend on an income variable.

Future Work with the Four-Output Model

One obvious gap in our population model is that we have not yet introduced immigration and emigration, although this would not be difficult.

The possibilities of introducing more refinements into the estimation of birth, death, marriage, and divorce probabilities are almost limitless. In particular, we have not yet made any use of the growing body of information on differential vital rates, for example, data on the birth or mortality rates of different socio-economic groups. We plan to introduce these differential rates into our model as soon as possible, so as to avoid running into serious distortions in predicting economic outputs.

FAMILY AND SPENDING UNIT GENERATION

The model containing only birth, death, marriage, and divorce requires modification before it can be extended to include economic outputs such as spending and saving, because it generates only two kinds of units; namely, individuals and married couples (with their associated unmarried children). It is not clear that these are the units which make economic decisions.

Which units are relevant to the analysis of spending and saving decisions? This question seems to have no single answer. The logical unit of analysis for housing expenditures seems to be the household (that is, the person or persons occupying a dwelling). But for minor personal expenditures, it is probably the individual.

The most extensive and useful source of data on spending and saving behavior currently available to us is the Survey of Consumer Finances. These data relate to "spending units," which are defined as groups of one or more persons living together, related by blood, marriage, or adoption, and pooling their incomes for major expenditures. A grown child or other relative living in the family is counted as a separate ("secondary") spending unit if he earns as much as fifteen dollars a week and keeps more than half of it for his own purposes. The SCF data can also be tabulated by "family units" (that is, related persons living together, irrespective of financial arrangements). For our purposes, the family unit of the SCF is a particularly useful one, since it corresponds to the "family" as defined by the Census Bureau—a fact which gives us an additional source of information about the composition and demographic characteristics of such units.

We will eventually want to use households, families, spending units, and perhaps other units in our model of the consumer sector. For the present, however, we will concentrate first on family units and then on spending units in order to use the SCF data to the maximum extent possible.

There are several ways in which family formation might be simulated in the model. Perhaps the most obvious approach would be the introduction of a new matching process analogous to marriage. The marriage process itself would first have to be altered so that a new unit was *not* automatically created by every marriage—the newly married couple would remain associated with one set of parents. Then two new outputs would be introduced, "leaving a family" (analogous to divorce) and "joining a family" (analogous to marriage). Like marriage, the latter would involve two stages; that is, deciding whether or not a family would unite with another family in the period, and picking the other family which the first one would join.

However, it will be difficult to estimate the operating characteristics of families without collecting new information or at least making new tabulations. The basic problem is that statistics simply are not collected on movements of persons in and out of families. Birth, death, marriage, and divorce are well-defined events which are reported to "the authorities." By contrast, changes in family status (for example, a grown son moving out of the parental family) are not reported to anyone, and the probabilities of such events can only be inferred from cross-sectional data.

Although simulation of family formation by means of a matching and unmatching process seems feasible, an even simpler approach appears preferable as a first approximation. Since most children live with their parents and most "other adults" in families are either grown sons and daughters of the head or parents of the head or his wife,[3] it seems possible to generate family units of appropriate size by controlling the rate at which children leave the parental family without introducing any process other than marriage by which units are amalgamated.

The proportions of married and unmarried persons by age who were *not* living in families headed by a parent or other relative in 1950 can be estimated from the census (see Table 1). Can we use the changes in these proportions to measure the rate at which sons and daughters move out of the parental family as they marry and/or grow older? There are two difficulties with this interpretation. One is that children may become heads of families because the parents die, rather than because the children have left the family. This effect may be small at the younger ages (say, under 35) but it is certainly important as the "children" reach middle

[3] Paul C. Glick, *American Families*, Wiley, 1957, pp. 8, 36.

TABLE I

Proportions Not Living in Families Headed by Parent or Other Relative, by Age, Sex, Marital Status, 1950

Age	Single Male	Single Female	Married Male	Married Female
14–17	3.6	2.6	33.0	60.2
18–19	21.0	19.1	59.4	73.2
20–24	27.8	24.0	80.1	84.3
25–29	29.7	26.3	88.9	90.7
30–34	32.2	29.7	93.0	94.4
35–44	38.8	33.8	95.7	96.4
45–54	52.1	42.2	97.7	98.1

Source: Estimated from U.S. Bureau of the Census, *U.S. Census of Population: 1950*, Vol. IV, *Special Reports*, Part 2, ch. D, Table 1.

age. As a first approximation we will simply assume that no "children" leave the parental family after age 35. This arbitrary cut-off age can be raised or lowered after the results of the first few runs are examined. A second difficulty is that a substantial number of family heads have persons living with them who are reported as "parents of head or wife." These families cannot be ignored without distorting the size distribution of families in the model. Hence, we must introduce a new output which may be called "losing status as head of family." This is an output which as far as the model is concerned is possible only for parents whose children have not yet formed separate families. When it occurs, it involves no change in family composition, but merely a shift in the designation "head of household" from parent to child. The operating characteristic associated with this output can be only roughly estimated from census data and likewise may have to be adjusted in the light of preliminary results.

Once we have families in the model, the creation of spending units is not conceptually difficult. A new output is introduced which may be called "becoming a separate spending unit." This output is possible for any married couple or person over 18 except heads of families, and the operating characteristic may be estimated from the Survey of Consumer Finances.

Debt, Liquid Asset, and Spending Behavior of Spending Units

Up to this point we have been concerned with the demographic outputs of units—birth, death, marriage, divorce, family and spending unit formation—that is, those aspects of individual and family behavior which serve to determine and continually modify the size, composition, and groupings of the population of individuals. It seems clear that it will be useful to regard these probabilities as determined in part by prior demographic inputs to the units and in part by economic inputs. The specific inclusion of economic inputs shows promise of improving our ability to predict demographic events.

Now we turn the telescope around and focus on economic variables regarded as outputs of units. We will be concerned here with the determination of the probabilities associated with certain types of economic behavior of spending units. Of special interest is the extent to which the demographic variables generated by the model will prove to be useful in predicting these economic outputs.

OUTPUTS AND INPUTS

A list of all the spending unit outputs whose determination we would like to include in the model would be quite lengthy. It would certainly include the following fifteen variables and might well include some others: (1) mortgage debt incurred during month, (2) mortgage debt repaid during month, (3) personal debt incurred during month, (4) personal debt repaid during month, (5) change in liquid assets during month, (6) change in other financial assets during month, (7) expenditure on purchase of house during month, (8) expenditure on purchase of automobile during month, (9) expenditure on other durables during month, (10) expenditure on clothing during month, (11) expenditure on food during month, (12) expenditure on services during month, (13) receipts from sale of automobile during month, (14) receipts from sale of house during month, and (15) receipts from sale of other non-financial assets during month.

So far we have been able to make only very limited inroads on this list. In fact, the operating characteristics estimated in this paper refer only to the following four outputs, three of which are not even on our "desirable" list: (1) mortgage debt held, (2) personal debt held, (3) liquid assets held, and (4) expenditure on selected durables. The first three are stock variables, that is, the amounts of debt or liquid assets held by the spending unit on the survey date, *not* the changes in these amounts in the preceding period. We worked with them only because data on the corresponding flow variables were not gathered in the survey we were using. The last variable, however, is a flow variable covering expenditures by the spending unit on a selected list of durables during 1955.

Our work on these four output variables should be regarded mainly as illustrative of the estimation techniques with which we have been experimenting and which we would like to apply to other outputs as time, money, and data become available. The probabilities associated with many of the variables on our "desirable" list (such as automobile purchases) can be estimated from existing bodies of data, but some of the others may have to wait for bigger and better surveys.

A complete list of input variables which we eventually want to include would also be quite long. The following sets of demographic and economic inputs seem to merit testing and one could think of others.

Demographic input variables: marital status of head of spending unit*, interval since marriage if married couple*, age of head*, race of head*, sex of head*, number of adults*, number of children*, age of youngest*, age of oldest under 18*, education of head, veteran status of head, region, city type, household type, and number of dependents outside of spending unit.

Economic input variables: income of spending unit in current and previous years, employment status, occupation of head, mortgage debt at start of period, personal debt at start of period, liquid assets at start of period, other financial assets at start of period, stock of housing at start of period, stock of automobiles at beginning of period, and stock of other durables at beginning of period.

The starred demographic variables are of particular interest to us at the moment, because they can be generated by the demographic model described in the previous section. Partly for this reason and partly because of the limitations of the survey data, we concentrated on estimating the effect of these demographic inputs on the four economic outputs available. The survey data on the economic inputs were quite inadequate for our purposes. The spending unit's stock of a particular durable at the beginning of the period is undoubtedly an important determinant of probability that it will purchase that durable in the period. Unfortunately, the data necessary to estimate this were not available nor were adequate income histories of the spending units. Hence, our strategy is to explain as much as possible of the variation in outputs on the basis of the available demographic variables and then to rely on the use of aggregative time series data about incomes, prices, and credit terms to explain part of the residual variation.

AN APPROACH TO THE PROBLEM OF JOINT OUTPUTS

Use of a short time period does facilitate handling interactions between units, but it does not remove the need for regarding several outputs of each individual unit as simultaneously and jointly determined. One could, for example, estimate the parameters of a single equation giving the amount a unit will spend on an automobile in a month as a function of income, family size, duration of marriage, and a random variable, and also the parameters of a separate equation giving the amount of personal debt the unit will incur in the month as a function of the same three variables and another random variable. However, it would not be valid to consider that the two random variables are distributed independently.

In general terms our approach to this problem may be described as follows. Let the output variables which are to be simultaneously determined at the household level be $X_1, \ldots X_K$. The specification of values of these K variables may then be regarded as specifying a point in a K-dimensional space in which there are K orthogonal axes, with X_1 measured along the first, X_2 along the second, and so on. It is assumed that each combination of the relevant input variables determines a probability distribution defined over this output space for the point or vector of output variables. In simulation studies the output of a particular unit in a particular time period is to be obtained by a random drawing of a point from the multivariable probability distribution determined for that unit at that particular time.

The practical problems involved are how the joint probability distribution shall be specified and estimated and how a point shall be randomly drawn from the probability distribution.

Let the probability of any point in the output space for a given unit at a given point in time be represented by $P(X_1, \ldots X_K)$. We are assuming that this probability is a function of R input variables denoted by $Z_1, \ldots, Z_R$ and the coordinates of the point. That is, we are assuming that

$$(1) \qquad P(X_1, \ldots X_K) = F(Z_1, \ldots Z_R, X_1, \ldots X_K)$$

Now by the usual laws of probability we know that $P(X_1, \ldots X_K)$ can be expressed as the product of a marginal probability and a number of conditional probabilities as follows:

$$(2) \qquad P(X_1, \ldots X_K) = P(X_1) \cdot P(X_2 | X_1) \cdot P(X_3 | X_1, X_2) \ldots$$

$P(X_K | X_1, \ldots, X_{K-1})$. Which of the K output variables happens to be labeled X_1, which X_2, and so on, is immaterial except for small sample estimation problems.

The probabilities into which we have factored $P(X_1, \ldots X_K)$ will then be related to input variables $Z_1, \ldots, Z_R$ as follows:

$$
\begin{aligned}
(3)\qquad & P(X_1) = F_1^*(Z_1, \ldots Z_R, X_1) \\
& P(X_2|X_1) = F_2^*(Z_1, \ldots, Z_R, X_1, X_2) \\
& \quad " \quad " \quad " \\
& \quad " \quad " \quad " \\
& \quad " \quad " \quad " \\
& P(X_K|X_1, \ldots X_{K-1}) = F_K^*(Z_1, \ldots Z_R, X_1, \ldots X_K)
\end{aligned}
$$

The corresponding mean or expected values of $X_1, \ldots, X_K$, denoted by $E(X_1), \ldots, E(X_K|X_1, \ldots X_{K-1})$ will then be given as follows.

$$
\begin{aligned}
(4)\qquad & E(X_1) = F_1(Z_1, \ldots, Z_R) \\
& E(X_2|X_1) = F_2(Z_1, \ldots, Z_R, X_1) \\
& \quad " \quad " \quad " \\
& \quad " \quad " \quad " \\
& \quad " \quad " \quad " \\
& E(X_K|X_1, \ldots X_{K-1}) = F_K(Z_1, \ldots, Z_R, X_1, \ldots X_{K-1})
\end{aligned}
$$

In this study we approximated functions F_1 through F_K by either the sum or product of essentially free functions of the variables involved. At a later date, we hope to use the residual differences between expected values and observed values to explore additional features of the marginal and conditional probability distributions expressed in equation set 3.

ESTIMATION PROCEDURE

The basic estimation procedures used in this study may be characterized as a sequential application of single equation, least square techniques. The choice of a least square rather than a maximum likelihood criterion for the selection of appropriate parameter values has computational advantages. Moreover, it is our belief that it would be a preferable criterion even without these computational attractions. The primary justification for use of a least squares criterion is that its use does not require assumptions about the shape of the error term distribution function, whereas use of a maximum likelihood criterion does. Since we know very little about the shapes of error term distributions, if indeed stable and well defined error distributions are actually present, it seems advantageous to use estimation techniques which do not require assumptions about their shapes.

To estimate our parameters, we used a computer program designed to handle a wide variety of standard regression computations.[4] But since

[4] A preliminary write-up of this program is contained in "Description of General Correlation Package," by M. E. Callaghan, K. E. Kavanagh, and J. R. Steinberg, mimeographed distribution by M.I.T. Computation Center, CC-107, August 1, 1958.

we adapted this program to our particular needs by a device which is still somewhat unfamiliar, we will give an illustrative example.

Let the output or dependent variable be denoted by Y, its predicted value by Y^*, and the input or independent variables by X, Z, and W. Only three are used in this example, but the procedure is the same for more. The input variables are either classificatory by nature or are converted into such variables by scaling. For concreteness let us suppose that X may fall into any one of five categories numbered 0 through 4, and Z may fall into any one of eight categories numbered 0 through 7, and W may fall into any one of three categories numbered 0 through 2.

We selected the parameters in the following relations so as to minimize the sum of squares of the differences between Y and Y^* for the sample of spending units being used.

$$Y^* = a + F_1(X) + F_2(Z) + F_3(W) \tag{5}$$

where for each spending unit $F_1(X)$, $F_2(Z)$, and $F_3(W)$ are the appropriate values of these functions obtained by entering the following tables with the value of X, Z, and W that applies for the particular spending unit.

TABLE FOR $F_1(X)$

X	$F_1(X)$
0	0
1	b_1
2	b_2
3	b_3
4	b_4

TABLE FOR $F_2(Z)$

Z	F_2
0	0
1	c_1
2	c_2
3	c_3
4	c_4
5	c_5
6	c_6
7	c_7

TABLE FOR $F_3(W)$

W	F_3
0	0
1	d_1
2	d_2

The parameters to be estimated are thus seen to be a, b_1 through b_4, c_1 through c_7, and d_1 and d_2. If a particular spending unit has an X of 2, a Z of 6, and a W of 0, then the predicted value of Y, Y^*, will be $a + b_2 + c_6 + 0$. If X, Z, and W are all zero then the predicted value will be the intercept term, that is, a.

The above estimation problem could have been treated as an analysis of variance problem with three classificatory variables, unequal number of observations per cell, and with additivity being assumed. However, since an effective computer program for such a problem did not seem to be available, we sought to use the well developed programs which are available for linear regression problems. It should perhaps be added at this point that the only issue at stake here is computational ease, since the procedure we used does in fact yield mathematically identical estimates of the parameters involved.

To convert the above problem into a linear regression problem let us create a number of new variables $X_1, \ldots, X_4, Z_1, \ldots Z_7, W_1$, and W_2. These twelve new variables are defined as follows.

$$\begin{aligned} X_1 &= 1 \quad \text{if } X = 1 \\ &= 0 \quad \text{if } X \neq 1 \\ X_2 &= 1 \quad \text{if } X = 2 \\ &= 0 \quad \text{if } X \neq 2 \end{aligned}$$

and so on until we have

$$\begin{aligned} W_2 &= 1 \quad \text{if } W = 2 \\ &= 0 \quad \text{if } W \neq 2 \end{aligned}$$

Then it will be seen that the following linear regression equation will always yield exactly the same value of Y^* as did equation 5.

$$(6) \quad Y^* = a + b_1X_1 + \ldots + b_4X_4 + c_1Z_1 + \ldots c_7Z_7 + d_1W_1 + d_2W_2$$

Since the way in which the *a*, *b*, *c*, and *d* parameters enter into any computed value of Y^* is unchanged, and since it is these parameters in both cases which are being selected to minimize $(Y - Y^*)^2$, it is, of course, evident why the same results are achieved by estimating the parameters in equation 6 as by estimating them in equation 5. All that has been done is to translate the estimation problem into a form for which a good computer program exists.[5]

Summary of Results

The four basic variables used as dependent or output variables were mortgage debt held, personal debt held, liquid assets held, and annual expenditure on selected durables.

The study of each basic variable was broken down into two parts. In the first of these, a prediction equation for the probability that the dependent variable would be greater than zero was derived. In the second of these, attention was limited to spending units having a greater than zero value of the dependent variable, and a prediction equation, for the value of the dependent variable, was derived. Our results thus fall naturally into eight sections.

In each of these eight parts of our study, the independent or input variables included marital status and duration of marriage if married,

[5] The essential idea of the "dummy variable" approach may be found in Daniel B. Suits, "Use of Dummy Variables in Regression Equations," *Journal of the American Statistical Association*, Vol. 52, December, 1957, pp. 548–551.

age of head, education of head, and race of head. In the study of personal debt held, mortgage debt held also was used as an input variable. In the study of liquid assets held, both mortgage debt held and personal debt held were used as input variables in addition to the four input variables common to all eight parts. In the study of expenditure on durables, mortgage debt held, personal debt held, and liquid assets held were all used as input variables in addition to the basic four input variables.

In each of these eight parts, we computed successive regressions, adding in the input variables one at a time until all the input variables used in that part were included. Only these eight final multiple regressions are reported in this paper. A summary of these results is presented in the following eight sections. In each case a brief verbal statement is followed by results of the final multiple regression.

PROBABILITY SPENDING UNIT HOLDS MORTGAGE DEBT

Marital status is highly significant and exerts a substantial effect. Marriage increases the probability. The probability increases as duration of marriage increases up to between five and nine years and then decreases for longer durations.

Age of head is highly significant and exerts a substantial effect. The probability increases with age up to between 45 and 49 years of age and then decreases.

Education of head may be significant and appears to exert a modest effect. The probability increases with educational level attained by head.

Race is highly significant and exerts a modest effect. The probability is lower for Negroes.

The residuals were found to be significantly related to income, occupation, role of the spending unit in the dwelling unit, and to the predicted value of the dependent variable. This last result indicates that in this case a correction introduced to weaken the additivity assumption would be an improvement.

The residuals also were found to exhibit what may be a significant relation to number of adults, number of children, and veteran status of the spending unit.

If Y_1^* is used to designate the predicted probability a spending unit will hold mortgage debt, our equation for Y_1^* is

$$Y_1^* = -0.08 + F_{1.1} + F_{1.2} + F_{1.3} + F_{1.4} \tag{7}$$

Correlation coefficient = 0.39; Standard error of estimate = 0.41; Standard error of constant = 0.09

The values of functions $F_{1.1}$, $F_{1.2}$, $F_{1.3}$, and $F_{1.4}$ are to be obtained by using the following function tables. The bracketed figure to the right of each estimated value is the standard error of the estimated parameter.

$F_{1.1}$ (marital status)

Marital Status	$F_{1.1}$
Unmarried	0
1 year	0.07 (0.06)
2 years	.15 (.05)
3 years	.19 (.06)
4 years	.27 (.06)
5–9 years	.31 (.03)
10–20 years	.25 (.03)
Over 20 years	.14 (.02)

$F_{1.2}$ (age of head)

Age of Head	$F_{1.2}$
18–20	0
21–24	−0.02 (0.06)
25–29	.09 (.06)
30–34	.16 (.06)
35–39	.25 (.06)
40–44	.22 (.06)
45–49	.24 (.06)
50–54	.16 (.06)
55–59	.15 (.06)
60–64	.12 (.06)
65 and over	.07 (.06)

$F_{1.3}$ (education)

Education	$F_{1.3}$
None	0
Grammar	0.02 (0.07)
Some H.S.	.05 (.07)
H.S. Degree	.08 (.07)
Some College	.12 (.08)
College Degree	.13 (.08)

$F_{1.4}$ (race)

Race	$F_{1.4}$
White	0
Negro	−0.10 (0.03)
Other	−.14 (.09)

AMOUNT OF MORTGAGE DEBT HELD BY SPENDING UNIT

Marital status may be significant and seems to exert a substantial effect. Marriage increases the expected amount of mortgage debt, but the effect dwindles very rapidly as duration of marriage increases.

Age of head may be significant and may have a modest effect. The expected amount of mortgage debt seems to increase with increasing age until about thirty-four years of age and then it decreases.

Education of head is significant and exerts a large effect. The expected amount increases with the educational level achieved by the head.

Race does not seem very significant, although the expected amount is less for Negroes.

The residuals were found to be significantly related to income, occupation, and city type.

If Y_2^* is used to designate the predicted amount in dollars of mortgage debt for spending units having mortgage debt, our equation for Y_2^* is

$$Y_2^* = 3040 + F_{2.1} + F_{2.2} + F_{2.3} + F_{2.4} \quad (8)$$

Correlation coefficient = 0.432; Standard error of estimate = 4311; Standard error of constant = 5160.

The values of the functions $F_{2.1}$, $F_{2.2}$, $F_{2.3}$, and $F_{2.4}$ are to be obtained by use of the following function tables. The figure in brackets to the right of each estimated value of a function is the standard error of the estimated parameter.

$F_{2.1}$ (marital status)

Marital Status	$F_{2.1}$
Unmarried	0
1 year	4,550 (1,740)
2 years	1,380 (1,300)
3 years	1,530 (1,300)
4 years	167 (1,130)
5–9 years	49 (700)
10–20 years	744 (628)
Over 20 years	215 (625)

$F_{2.2}$ (age of head)

Age of Head	$F_{2.2}$
18–20	0
21–24	526 (4,620)
25–29	1,290 (4,490)
30–34	1,560 (4,480)
35–39	1,200 (4,480)
40–44	964 (4,480)
45–49	134 (4,490)
50–54	166 (4,510)
55–59	−1,270 (4,520)
60–64	−1,140 (4,540)
65 and over	−1,000 (4,530)

$F_{2.3}$ (education)

Education	$F_{2.3}$
None	0
Gr. School	835 (2,570)
Some H.S.	1,190 (2,590)
H.S. Degree	1,980 (2,580)
Some College	3,610 (2,600)
College Degree	4,930 (2,600)

$F_{2.4}$ (race)

Race	$F_{2.4}$
White	0
Negro	−1,040 (891)
Other	−1,010 (2,510)

PROBABILITY SPENDING UNIT HOLDS PERSONAL DEBT

Marital status is significant and exerts a large effect. Marriage increases the probability. The probability increases with duration of marriage up to about three years and then drops off.

Age of head of spending unit is significant and plays a large role. The probability decreases with age after about twenty-four and is very low if the head is over sixty-five years of age.

Educational status of head is probably significant and exerts a modest effect. The probability decreases with increasing education of head.

Race is significant and exerts a substantial effect. Negroes are more likely to have personal debt than whites.

Mortgage debt held by spending unit is significant and plays a modest role. Mortgage debt holders are more likely to have personal debt except for very large mortgage holders.

The residuals were found to be significantly related to income, occupation, number of children, age of eldest child, and region. The residuals exhibited what may be a significant relation to city type and to the predicted value of the dependent variable.

If Y_3^* is used to designate the predicted probability that the spending unit will have personal debt, our equation for Y_3^* is

$$Y_3^* = 0.63 + F_{3.1} + F_{3.2} + F_{3.3} + F_{3.4} + F_{3.5} \tag{9}$$

Correlation coefficient = 0.426; Standard error of estimate = 0.454; Standard error of constant = 0.098.

The values of the functions $F_{3.1}$, $F_{3.2}$, $F_{3.3}$, $F_{3.4}$ are to be obtained by use of the following function tables. Standard errors of the estimated parameters are given in brackets.

$F_{3.1}$ (marital status)

Marital Status	$F_{3.1}$
Unmarried	0
1 year	0.11 (0.06)
2 years	.22 (.06)
3 years	.29 (.07)
4 years	.16 (.06)
5–9 years	.17 (.03)
10–20 years	.18 (.03)
Over 20 years	.12 (.03)

$F_{3.2}$ (age of head)

Age of Head	$F_{3.2}$
18–20	0
21–24	−0.03 (0.07)
25–29	.00 (.06)
30–34	− .05 (.06)
35–39	− .14 (.06)
40–44	− .24 (.06)
45–49	− .18 (.06)
50–54	− .27 (.07)
55–59	− .30 (.07)
60–64	− .38 (.07)
65 and over	− .51 (.06)

$F_{3.3}$ (education)

Education	$F_{3.3}$
None	0
Gr. School	0.00 (0.08)
Some H.S.	.00 (.08)
H.S. Degree	− .10 (.08)
Some College	− .10 (.08)
College Degree	− .14 (.08)

$F_{3.4}$ (race)

Race	$F_{3.4}$
White	0
Negro	0.18 (0.03)
Other	.06 (.10)

The value of $F_{3.5}$ (mortgage debt) is zero if the unit does not have mortgage debt. If the unit has mortgage debt,

$$F_{3.5} = \underset{(0.03)}{0.14} - \underset{(0.000004)}{0.000007} \text{ (mortgage debt of unit in dollars).}$$

AMOUNT OF PERSONAL DEBT HELD BY SPENDING UNIT

Marital status does not appear to be very significant. The expected value of personal debt may be larger for units married over twenty years.

Age of head does not seem very significant. The expected value may reach a peak at about forty or forty-five years of age.

Education may be significant and seems to exert a large effect. The expected value increases with the educational level of the head.

Race is probably significant and the expected value is less for Negroes than for whites.

The level of mortgage debt is significant and the expected amount of personal debt increases with the amount of mortgage debt held. The mere holding of mortgage debt is not significant.

The residuals were found to be significantly related to income and occupation. They also appear to be related to number of adults and region.

If Y_4^* designates the predicted amount in dollars of personal debt for spending units having personal debt, our equation for Y_4^* is

$$(10) \qquad Y_4^* = 361 + F_{4.1} + F_{4.2} + F_{4.3} + F_{4.4} + F_{4.5}$$

Correlation coefficient = 0.280; Standard error of estimate = 1133; Standard error of constant = 382.

The values of $F_{4.1}$, $F_{4.2}$, $F_{4.3}$, and $F_{4.4}$ are to be obtained by use of the following function tables. Standard errors of the estimated parameters are given in brackets.

$F_{4.1}$ (marital status)

Marital Status	$F_{4.1}$
Unmarried	0
1 year	310 (192)
2 years	59 (173)
3 years	−110 (194)
4 years	87 (188)
5–9 years	81 (110)
10–20 years	97 (101)
Over 20 years	266 (109)

$F_{4.2}$ (age of head)

Age of Head	$F_{4.2}$
18–20	0
21–24	67 (211)
25–29	127 (204)
30–34	99 (205)
35–39	119 (207)
40–44	420 (210)
45–49	140 (210)
50–54	−37 (222)
55–59	2 (230)
60–64	−51 (240)
65 and over	55 (233)

$F_{4.3}$ (education)

Education	$F_{4.3}$
None	0
Gr. School	−57 (336)
Some H.S.	93 (340)
H.S. Degree	210 (340)
Some College	329 (347)
College Degree	543 (347)

$F_{4.4}$ (race)

Race	$F_{4.4}$
White	0
Negro	−223 (106)
Other	−178 (297)

The value of $F_{4.5}$ (mortgage debt) is zero if the unit does not have mortgage debt. If the unit has mortgage debt,

$$F_{4.5} = \underset{(105)}{-41} + \underset{(0.0134)}{0.0421} \text{ (the amount of mortgage debt of unit in dollars).}$$

PROBABILITY SPENDING UNIT HOLDS LIQUID ASSETS

Marital status is significant and exerts a modest effect. The probability increases with marriage and with marriage durations of over twenty years.

Age of head is significant and exerts a substantial effect. The probability increases with age of head up to about fifty-five to sixty years of age and then decreases.

Education of head is significant and exerts a very great effect. The probability increases with educational level achieved by head.

Race of head is significant and exerts a large effect. Negroes are less likely to have liquid asset holdings.

Mortgage debt of spending unit may be significant and appears to exert a modest effect. The amount of mortgage debt does not appear to be very significant, but units with mortgage debt are more likely to have liquid assets.

Personal debt is significant and exerts a modest effect. Units with personal debt are less likely to have liquid assets unless personal debt is very large.

The residuals were found to be significantly related to income, occupation, number of children, city type, region, and the predicted value of the dependent variable. They also may be related to age of eldest child and role of spending unit in its dwelling unit.

If Y_5^* designates the predicted probability that the spending unit holds liquid assets, our equation for Y_5^* is

$$Y_5^* = 0.20 + F_{5.1} + F_{5.2} + F_{5.3} + F_{5.4} + F_{5.5} + F_{5.6} \quad (11)$$

Correlation coefficient = 0.508; Standard error of estimate = 0.364; Standard error of constant = 0.08

The values of the functions $F_{5.1}$, $F_{5.2}$, $F_{5.3}$, and $F_{5.4}$ are to be obtained by use of the following function tables. Standard errors of the estimated parameters are given in brackets.

$F_{5.1}$ (marital status)

Marital Status	$F_{5.1}$
Unmarried	0
1 year	0.03 (0.05)
2 years	.06 (.05)
3 years	.04 (.05)
4 years	.07 (.05)
5–9 years	.05 (.03)
10–20 years	.01 (.02)
Over 20 years	.09 (.02)

$F_{5.2}$ (age of head)

Age of Head	$F_{5.2}$
18–20	0
21–24	0.01 (0.05)
25–29	.10 (.05)
30–34	.14 (.05)
35–39	.17 (.05)
40–44	.18 (.05)
45–49	.15 (.05)
50–54	.16 (.05)
55–59	.23 (.05)
60–64	.20 (.05)
65 and over	.16 (.05)

$F_{5.3}$ (education)

Education	$F_{5.3}$
None	0
Gr. School	0.28 (0.06)
Some H.S.	.44 (.07)
H.S. Degree	.54 (.07)
Some College	.61 (.07)
College Degree	.61 (.07)

$F_{5.4}$ (race)

Race	$F_{5.4}$
White	0
Negro	−0.40 (0.03)
Other	.04 (.08)

The value of $F_{5.5}$ (mortgage debt) is equal to zero if the unit does not have mortgage debt. If the unit has mortgage debt,

$$F_{5.5} = \underset{(0.02)}{0.04} + \underset{(0.000003)}{0.000002} \text{ (mortgage debt of unit in dollars).}$$

The value of $F_{6.6}$ (personal debt) is equal to zero if the unit does not have personal debt. If the unit has personal debt,

$$F_{5.6} = \underset{(0.02)}{-0.13} + \underset{(0.000008)}{0.000019} \text{ (personal debt of unit in dollars).}$$

AMOUNT OF LIQUID ASSETS HELD BY SPENDING UNIT

Marital status was found to be fairly significant and to exert a substantial effect. The expected value increases with marriage and is particularly high for marriages of ten- to twenty-year duration.

Age of head was significant and exerts a large effect. The expected amount of liquid assets increases with age up to about sixty-five years of age and then drops off.

Education level of head is significant and exerts a large effect. The expected amount increases with the educational level achieved by the head.

Race is not significant.

Mortgage debt presence or absence is significant, but its level is not. The expected amount of liquid assets is much less for mortgage debt holders.

The amount of personal debt is not significant, but its presence or absence is. Spending units holding personal debt have substantially smaller liquid asset holdings.

The residuals were found to be significantly related to income, occupation, and to the predicted value of the dependent variable. They also may be related to number of children, age of eldest child, city type, and role of spending unit in its dwelling unit.

If Y_6^* designates the predicted amount in dollars of liquid assets for units holding liquid assets, our equation for Y_6^* is

$$(12) \qquad Y_6^* = -4870 + F_{6.1} + F_{6.2} + F_{6.3} + F_{6.4} + F_{6.5} + F_{6.6}$$

Correlation coefficient = 0.251; Standard error of estimate = 12566; Standard error of coefficient = 4660.

The values of the function $F_{6.1}$, $F_{6.2}$, $F_{6.3}$, and $F_{6.4}$ are to be obtained by use of the following function tables. Standard errors of the estimated parameters are given in brackets.

$F_{6.1}$ (marital status)

Marital Status	$F_{6.1}$
Unmarried	0
1 year	463 (2,070)
2 years	1,380 (1,890)
3 years	492 (2,180)
4 years	928 (1,960)
5–9 years	1,540 (1,040)
10–20 years	2,970 (930)
Over 20 years	1,730 (833)

$F_{6.2}$ (age of head)

Age of Head	$F_{6.2}$
18–20	0
21–24	249 (2,280)
25–29	735 (2,180)
30–34	534 (2,150)
35–39	1,320 (2,160)
40–44	1,440 (2,160)
45–49	5,530 (2,170)
50–54	4,410 (2,210)
55–59	4,100 (2,220)
60–64	7,440 (2,240)
65 and over	5,220 (2,140)

$F_{6.3}$ (education)

Education	$F_{6.3}$
None	0
Gr. School	3,100 (4,250)
Some H.S.	5,260 (4,270)
H.S. Degree	6,270 (4,260)
Some College	7,280 (4,290)
College Degree	8,240 (4,290)

$F_{6.4}$ (race)

Race	$F_{6.4}$
White	0
Negro	−1,330 (1,800)
Other	12 (3,090)

The value of $F_{6.5}$ (mortgage debt) is equal to zero if the unit does not have mortgage debt. If the unit has mortgage debt,

$$F_{6.5} = \underset{(930)}{-2590} + \underset{(0.104)}{0.107} \text{ (mortgage debt of unit in dollars).}$$

The value of $F_{6.6}$ (personal debt) is equal to zero if the unit does not have personal debt. If the unit has personal debt,

$$F_{6.6} = \underset{(658)}{-2340} - \underset{(0.307)}{0.119} \text{ (personal debt of unit in dollars).}$$

PROBABILITY SPENDING UNIT MAKES EXPENDITURES ON CONSUMER DURABLES

Marital status is significant and exerts a large effect. Marriage increases the probability substantially. The probability increases until the second or third year of marriage and then decreases rapidly.

Age of head is significant and exerts a moderate effect. The probability increases with age until about fifty or fifty-five and then declines.

Educational status and race of head are not significant.

Presence or absence of mortgage debt is not in itself significant, but the amount of mortgage debt is significant. The probability increases with the amount of mortgage debt held.

The amount of personal debt does not appear to be significant; however, the presence or absence of personal debt is significant, and the probability of consumer durable expenditure is higher for spending units having personal debt.

The amount of liquid assets does not appear to be significant. However, the presence of liquid assets does not appear to be significant and increases the probability of expenditure on consumer durables.

The residuals were found to be significantly related to income, number of children, age of eldest child, rôle of spending unit in its dwelling unit, and to the predicted value of the dependent variable. They also may be related to number of adults, city type, and region.

If Y_7^* designates the predicted probability that the spending unit purchases consumer durables during year, our equation for Y_7^* is

$$(13) \quad Y_7^* = -0.048 + F_{7.1} + F_{7.2} + F_{7.3} + F_{7.4} + F_{7.5} + F_{7.6} + F_{7.7}$$

Correlation coefficient = 0.353; Standard error of estimate = 0.47; Standard error of constant = 0.10.

The values of the functions $F_{7.1}$, $F_{7.2}$, $F_{7.3}$, and $F_{7.4}$ are given in the following function tables. Standard error of the estimated parameters are given in brackets.

$F_{7.1}$ (marital status)

Marital Status	$F_{7.1}$
Unmarried	0
1 year	0.34 (0.06)
2 years	.46 (.06)
3 years	.45 (.07)
4 years	.24 (.07)
5–9 years	.18 (.03)
10–20 years	.13 (.03)
Over 20 years	.15 (.03)

$F_{7.2}$ (age of head)

Age of Head	$F_{7.2}$
18–20	0
21–24	0.07 (0.07)
25–29	.14 (.07)
30–34	.17 (.07)
35–39	.17 (.07)
40–44	.18 (.07)
45–49	.16 (.07)
50–54	.21 (.07)
55–59	.12 (.07)
60–64	.05 (.07)
65 and over	.10 (.07)

$F_{7.3}$ (education)

Education	$F_{7.3}$
None	0
Gr. School	0.07 (0.08)
Some H.S.	.14 (.09)
H.S. Degree	.05 (.09)
Some College	.06 (.09)
College Degree	.14 (.09)

$F_{7.4}$ (race)

Race	$F_{7.4}$
White	0
Negro	−0.02 (0.04)
Other	.05 (.10)

The value of $F_{7.5}$ (mortgage debt) is zero if the unit holds no mortgage debt. If the unit has mortgage debt,

$$F_{7.5} = \underset{(0.032)}{-0.021} + \underset{(0.000004)}{0.00011} \text{ (mortgage debt of unit in dollars).}$$

The value of $F_{7.6}$ (personal debt) is zero if the unit has no personal debt. If the unit has personal debt,

$$F_{7.6} = \underset{(0.022)}{0.149} - \underset{(0.000011)}{0.000016} \text{ (personal debt of unit in dollars).}$$

The value of $F_{7.7}$ (liquid assets) is zero if the unit has no liquid assets. If the unit has liquid assets,

$$F_{7.7} = \underset{(0.025)}{0.087} + \underset{(0.0000008)}{0.0000001} \text{ (liquid assets of unit in dollars).}$$

AMOUNT OF CONSUMER DURABLES EXPENDITURE IN A YEAR

Marital status, age of head, education of head, and race all failed to contribute significantly to explaining the amount of consumer durables expenditures.

The presence or absence of mortgage debt was not significant but the amount may be and appears to be positively related to the amount of expenditure on consumer durables.

Personal debt does not appear to be significant.

The presence or absence of liquid assets does not appear to be significant, but the amount of liquid assets does appear to be significant. The effect is modest, but the amount of liquid assets is positively related to the amount of expenditures on consumer durables.

The residuals were found to be significantly related to income. They also may be significantly related to occupation and role of the spending unit in its dwelling unit.

If Y_8^* designates the predicted amount in dollars of consumer durables purchases for spending units having made any amount of such purchases, our equation for Y_8^* is

$$(14) \quad Y_8^* = 145 + F_{8.1} + F_{8.2} + F_{8.3} + F_{8.4} + F_{8.5} + F_{8.6} + F_{8.7} + F_{8.8}$$

Correlation coefficient = 0.219; Standard error of estimate = 468; Standard error of constant = 216.

The values of $F_{8.1}$, $F_{8.2}$, $F_{8.3}$, and $F_{8.4}$ are to be obtained by use of the following function tables. Standard errors of the estimated parameters are given in brackets.

$F_{8.1}$ (marital status)

Marital Status	$F_{8.1}$
Unmarried	0
1 year	128 (89)
2 years	41 (77)
3 years	57 (87)
4 years	−23 (90)
5–9 years	31 (54)
10–20 years	7 (48)
Over 20 years	−3 (47)

$F_{8.2}$ (age of head)

Age of Head	$F_{8.2}$
18–20	0
21–24	118 (127)
25–29	167 (122)
30–34	110 (122)
35–39	120 (123)
40–44	145 (124)
45–49	93 (124)
50–54	101 (127)
55–59	70 (132)
60–64	0 (134)
65 and over	14 (127)

$F_{8.3}$ (education)

Education	$F_{8.3}$
None	0
Gr. School	61 (180)
Some H.S.	69 (182)
H.S. Degree	107 (182)
Some College	83 (185)
College Degree	108 (184)

$F_{8.4}$ (race)

Race	$F_{8.4}$
White	0
Negro	2 (61)
Other	112 (137)

The value of $F_{8.5}$ (mortgage debt) is zero if the unit has no mortgage debt. If the unit has mortgage debt,

$$F_{8.5} = \underset{(45)}{-18} + \underset{(0.005)}{0.01} \text{ (mortgage debt of unit in dollars).}$$

The value of $F_{8.6}$ (personal debt) is zero if the unit has no personal debt. If the unit has personal debt,

$$F_{8.6} = \underset{(35)}{-50} + \underset{(0.02)}{0.01} \text{ (personal debt of unit in dollars).}$$

The value of $F_{8.7}$ (liquid assets) is zero if the unit has no liquid assets. If the unit has liquid assets,

$$F_{8.7} = \underset{(40)}{48} + \underset{(0.002)}{0.005} \text{ (liquid assets of unit in dollars).}$$

Prospective Use of Aggregative Time Series

Decision-unit models are ideally set up to use data and relationships that apply at a micro-level. This is one of the major advantages of this type of model and should be fully exploited. Nevertheless, certain things should be kept in mind. In some cases data needed to determine appropriate micro-operating characteristics may not be available. In other cases the data may be available, but the particular need may not justify the added complexity or computing effort required fully to utilize the micro-information. However, there is nothing about decision-unit models or their simulation that restricts the model builder to the use of micro-data or micro-relations. He has the added opportunity to build at the micro level, but he retains an equal facility to incorporate the use of aggregative data and relations.

Our strategy is to build and test as completely as possible at the micro-level, but then to use aggregative data and relationships to complete our models and to bring them into alignment with historical aggregative data. This puts as much of the burden of testing and formulation as is feasible at the micro-level where it belongs. It retains the use of aggregative data as fully as possible for final testing and alignment of the over-all model. As data availability improves, and as our knowledge grows, decision-unit type models of the economic system may, and in fact should, place less and less reliance on aggregative data and relationships.

However, this should be a gradual evolution, and, since the final "payoff" of these models will usually be in predicting things about aggregates, it follows that final testing against aggregative data will always be necessary. Hopefully such testing will not require extensive gross adjustments aimed at bringing the model into line with the aggregative data used in testing.

We expect to conduct further studies at the micro-level aimed at predicting household behavior with respect to a more adequate set of output variables. We also hope to use, as inputs at the micro-level, some additional variables such as income and employment. We realize that this present study leaves an unduly heavy burden for aggregative data, particularly income and employment series. Nevertheless, a point will inevitably come at which it seems necessary to accept and use whatever results are available at the micro-level. How then should we proceed?

In the absence of further studies at the micro-level, our procedure will be to use the obtained household operating characteristics, such as they are. An initial population appropriate to a convenient historical date will be used, and the behavior of components of the over-all model, such as the household sector, will be simulated. The discrepancies between the aggregative time series generated and the actual historical pattern would be ascertained and related to other appropriate aggregative monthly time series. Results obtained would then be incorporated into the model in preparation for additional studies and eventual prediction.

Conclusion

While the major substantive findings of this study have already been presented, a few of their implications need to be stressed.

1. Life cycle changes are of great importance in predicting the economic and demographic behavior of households. However, it is not possible to describe the life cycle position of a spending unit by the value of a single variable. Such variables as marital status of heads of spending units, duration of marriage, age of heads, and number of children all exert an independent and significant effect on economic behavior.

2. Since duration of marriage plays an important role with respect to several economic outputs of spending units, household surveys ought to include a question on this point.

3. The effects of demographic variables such as duration of marriage and age of head are not linear in either the original or logarithmic forms. In many cases, they are not even monotonically increasing or decreasing functions of these variables. This fact points up the need to estimate and test relationships which are flexible enough to approximate the role played by such variables. Completely flexible relationships cannot be determined from finite amounts of data. However, in specifying functional forms we have assumed only a weakened sort of additivity and a variety of local continuity. While we would like to move in the direction of doing without an additivity assumption, we feel that the approach used

by us is far more satisfactory than approaches based on assumptions of linearity, cumulative normal, or other rigid and pre-specified functional forms. It should go without saying, by now, that single variable approaches to multivariate problems are completely unsatisfactory.

4. The interdependence problem does arise with respect to some outputs even at the spending unit level and even if very short time periods are used. Treatment of such outputs as change in personal debt, change in liquid assets, and expenditures on durables as independently determined is clearly inadmissible. We have presented one approach to this problem which is both acceptable and relatively simple to use.

COMMENT

ROBERT SOLOW, Massachusetts Institute of Technology

I have been trying for weeks without success to remember the source, but somewhere Maxim Gorki tells the story of a man who is walking through a Russian village showing off some remarkable machine. It does all sorts of things: peels potatoes, fixes shoes, fits multiple regressions for all I know. Finally one old peasant silently watches the gadget perform and asks: "Yes, but will it whistle?" The owner scornfully replies that it won't, and starts to recite the list of the machine's accomplishments, but all the peasant says is: "What good is it if it can't whistle." Of course this is the big question about Guy Orcutt's grandiose "decision-unit" model of the household sector: will it whistle? Will it tell us what we really want to know? It is still too early to say. Even the bulky document presented here is only a beginning, and it will be a few years before we can even hope for a genuine, loud, clear, piercing whistle.

In one way, at least, the Orcutt-Rivlin model appeals to one of the deep-seated prejudices of the economics profession—the belief in disaggregation. J. K. Galbraith has claimed that the most common cliché in economics is the one about the baby and the bathwater. I would also settle for a nickel every time someone tells someone else that he should disaggregate. Now I do think the micro-economic instinct, in part, is sound. We probably ought to theorize in terms of the actual decision-making units of the economic system. Even purely macro-economic theories probably ought to be capable of being rationalized in terms of the behavior of households and firms. But it is a *non sequitur* to draw from this the implication that empirical relationships based on or derived from micro-economic data will be more reliable than those based squarely on aggregates. That depends on many subtle statistical properties of the system being observed and it is not hard to construct examples, or find them in reality, in which the balance goes either way.[1] So one cannot be sure without evidence that decision-unit models will predict better or produce more insight than more aggregative approaches to the data. The only way to get some evidence is to try, and so far nobody but Orcutt has had the tenacity to try.

Since I have no special competence or experience in demography, I will confine my comments to the last two-thirds of the paper. There some experiments are described on the interaction of demographic and economic factors in determining the behavior of households with respect to mortgage and personal debt, liquid asset ownership, and expenditures on durable consumer goods. The estimation procedure was carried out in a rather peculiar way, but one which does shed some oblique light on the significance of the demographic inputs. Each of the four dependent variables was studied in two stages: first a regression analysis with the probability that a spending unit will have mortgage debt, or personal debt or liquid assets, or that it will spend on durables as dependent variable; followed by a conditional regression analysis to predict the amount of debt, assets, or expenditure for those units which had any. Initially the regressions were carried out with the four major demographic variables as the only independent variables: marital status, education, age of head, and race. Then one by one a whole list of additional demographic and economic inputs were tested as possible independent variables. It is this last step that I find peculiar, not so much because of the information it gives as because of what it conceals. There is no way of telling how much weight ought to be attributed to each of the possible independent variables since the outcome of this asymmetric procedure will depend on the order in which additional variables are tried. Nor is there any way at the end to assess how successful the whole estimation procedure has been.

One conclusion does stand out. The amount of predictive power contributed by the four original demographic inputs is pitifully small. The multiple correlations run around 0.4 with only one of the eight as high as 0.5 and several others as low as 0.2. It is true that one rarely gets good fits from cross-sectional survey data but this fact in no way increases the percentage of variance explained. A more dramatic indication of how little help one gets from the first four variables comes from the standard errors of estimate. In the four regressions which are supposed to predict probabilities, the standard errors of estimate range from 0.36

[1] This point is well made in an unpublished paper, "Is Aggregation Necessarily Bad?", by Zvi Griliches and Yehuda Grunfeld of the Department of Economics, University of Chicago, September, 1958.

to 0.47. Thus a 90 per cent two-sided prediction interval would in all cases have a length greater than unity. Of course in many cases a shorter interval would be given by the necessary bounds zero and one to the size of a probability. But even then in nearly all cases, two standard errors on the free side of the best estimate will carry almost to zero or to one.

Among the additional variables tested for influence on the predictands are several other demographic inputs mainly having to do with location and family structure, and some of them appear to add something to the explanation. But the inputs which seem consistently to have the most work to do are income and occupation, the only truly socio-economic variables which make an appearance in the model. This leads me to believe that it would be much more enlightening to do the estimation just the reverse of the way it was actually performed. Not only ought the socio-economic inputs to be introduced at the very start, but the additivity assumption ought to be abandoned from the beginning. My complaint is not that the *ad hoc* method actually used to deal with non-additivity won't work, but rather that whether it works or not it effectively masks the most interesting aspect of the final results: the way in which demographic variables *interact* (in the analysis of variance sense) with economic variables in the determination of household behavior. I would be happy to trade a little predictive accuracy for this bit of insight if it were necessary. But I don't think it is.

This brings me to a related point. I notice that the list of desirable output variables includes changes in various categories of debt, expenditures on various categories of commodities, and receipts from sales. But the list of economic input variables is limited to income, employment and occupational status, and initial conditions. Not a single price or interest rate or price expectation appears on the list. I would be happier if the system being simulated looked a little more like a collection of markets and a little less like a collection of mice. It is true that survey data taken at one point of time rarely gives information on supply and demand responses to price changes. But although the data may be collected statistically the application is to a process continuing over time, and over time prices change. And even if prices do not change much, I would guess that there are significant interactions among price variables and the demographic and socio-economic determinants of household spending behavior.

Let me conclude with a few comments on a more technical level. I have already mentioned what seems to me to be a real difficulty with the device used to cope with non-additivity. The same remark goes for the quite similar way of dealing with additional independent variables. Whether or not the two procedures lead to improved fits or good fits, they seem to me to be makeshift substitutes for structural estimation.

Finally there is the matter of interdependence and simultaneous equations in the estimation procedure. I do not pretend to know how important this complication is likely to be in models of this kind. I do not even know how important it is in econometric models in general. But if the difficulty bulks large in the present context, I am not certain that the authors have avoided it by the device of taking a short time period. They say: "If we take income as one of the variables determining the probability that a unit will purchase an automobile in the month, we feel we can assume that the purchase itself does not affect the income of that unit in the same month. This is all that is necessary to avoid simultaneous equation difficulties at the system level." I'm not sure that it is. What is required is that the *residuals* from the "true" structural automobile demand function be uncorrelated with the family's income. And it does not seem at all unlikely to me that economic events should create differential effects within a month which might tend, say, to push high-income families below the normal relation and low-income families above. Moreover, the choice of a short time period could also have the side effect of straining the Markovian character of the process, so that even given prices, incomes, and the like, the probabilities for any given month might depend on the remoter past, and not only on the conditions at the beginning of that month.

I have used my time for carping, because I presume that is what I am expected to do. But the truth of the matter is that decades of conventional econometric effort have not added much to our stock of reliable empirical knowledge. Maybe the newer methods of digital and analog simulation are on the right track. At least we can guarantee ourselves reliable "empirical" knowledge about artificial economic systems and maybe that is half the battle. Maybe they will even whistle.

Reply

That micro-economic data contains more information than the same data aggregated is obvious. In general, one might expect some of this extra information to be useful for purposes of testing and prediction. In any case, one always has the option of throwing away the extra information. Given the availability of the appropriate micro-data and the correct model underlying the behavior and interaction of micro-units, then the predictive power of this model will necessarily exceed, or at the very worst, match that of any model based on aggregates of the micro-data. If some data is available at an aggregative level but not at a micro-level or if account is taken of the fact that all models about the real world undoubtedly involve misspecifications, then it seems obvious that some models based in part or entirely on highly aggregative data will work better than some models based solely on micro-economic data. Our position is that some types of economic and demographic phenomena require the use of micro-data and relationships, and that it is important to develop models of socio-economic systems and methods of analysis that are appropriate for micro-data and relationships. Such models will be able to use aggregative data and aggregative relationships to the extent desired.

Solow errs in judging predictive power on the basis of correlation coefficients and standard errors of estimate derived from relationships pertaining to the behavior of micro-units. Correlation coefficients are seldom appropriate for such purposes for a variety of well-known reasons. The use of standard errors of estimate is frequently appropriate but only if the dependent variables in the estimated relationships are what one is actually trying to predict. The dependent variables in our relationships relate to the behavior of micro-units and in many cases can only assume the values zero or one. What we were estimating in these cases were *probabilities* and these probabilities are the *expected* or *average* values of the dependent variables. The standard errors of estimate refer to the standard deviation of the zeros and ones of the micro-units from the estimated probabilities of getting a one. To judge predictive ability with regard to the aggregates to be predicted, one would have to pay attention to the standard errors of the estimated *probabilities*. Judging from our study of residuals classified according to predicted probabilities, the standard errors of the estimated probabilities seldom exceed 0.03 in any range of probabilities that occurs frequently enough to be of much significance in terms of aggregative behavior. It might be noted also that many of our parameter estimates deviate from zero by between five and ten times their standard errors.

We agree with Solow that it would be better to introduce income at the start and thus at a micro-level. We hope to do so, but have not done so yet for two reasons. (1) Current income does not seem to be a reasonable variable for a causal explanation of something like present mortgage debt position. Income expectations at the time the debt was incurred would be relevant but were not available. Part of the idea in using such variables as education, race, and age was that they would be better proxies for the appropriate long-run income expectations than current income would be. (2) Before one can effectively utilize relationships making use of incomes of micro-units it is necessary to generate these micro-incomes. We hope to do this in a reasonable way, but we have not yet reached that stage. Therefore, we chose to use aggregate income in the way projected in our paper. Current micro-incomes as well as other variables were related to residuals out of curiosity, since it seemed that something might be learned and the added cost of doing so was minor.

Much the same comments could be made about our failure to include prices, interest rates, and expectations of these and other variables. We are as anxious as Solow that this should be done. The general framework of decision-unit or micro-analytic models and the methods available for studying such models are adequate. What is lacking is a proper knowledge of the role of such variables. Since existing bodies of cross-sectional data do not seem a promising base for estimating price and interest rate effects, our present strategy is to do what can be done with existing bodies of data and then use more or less aggregative time series to fill in some of the gaps.

We would be curious to see what kind of an operational definition Solow would give to the notion of, "the, 'true' structural automobile demand function." We see no objection to micro-analytic models in which the random components appearing in different relationships are not treated as independent. Nor do we see any reason why probabilities for any given month should not be made to depend on conditions or events prior to the beginning of that month. We do this in making use of date of birth and marriage and might equally well do it in other respects. Why this should strain the Markovian character of the process is not exactly clear, since prior events can be interpreted as part of the conditions at the beginning of the month. In any case, we would not regard straining the Markovian character of the process as identical with original sin.