James Morgan, University of Michigan

Martin David, U. S. Department of Treasury (On leave from University of Wisconsin) Harvey Brazer, U. S. Department of Treasury (On leave from University of Michigan)

We present here an analytical model used to explain family income in the United States, a brief description of the statistical methods, and some selected results. The analysis of family income is itself part of a larger study of income distribution and redistribution, and of patterns of intergenerational change, to be published by McGraw-Hill Book Company. The study was supported by a grant from the Ford Foundation and smaller supplementary grants from the Office of Education and the Federal Office of Vocational Rehabilitation. Wilbur Cohen was a fourth principal researcher on the total project.

In the United States, some of the forces affecting family income are decisions and acts of the individuals themselves. Income is no longer a predetermined, exogenous factor to the individual family. Hence it is important to know how past history, outside forces, and individual decisions jointly determine family incomes, and their distribution.

Family income is made up of components, each affected by its own set of forces. We think of this process as a series of steps. At each step something is being determined as the result of outside forces and the acts and decisions of the individuals. An individual's earned income is the product of three things, each of which can be analyzed separately: whether he works, how long he works, and at what hourly rate. We thus have a whole set of things to explain, namely the items in the boxes of Chart 1.

As you can see, we ran ten multivariate analyses, relying on two-way tables to examine the other less important components such as transfer income from outside the dwelling, income taxes, and the income added by secondary units living with relatives.

You can think of the results of the analysis as allowing the derivation of an expected value of the income of a given family as follows: using various factors, estimate the probability that the head worked. Then use a table of random numbers and decide whether that particular head worked. If he worked, find expected values for hourly wage rates and hours and multiply one by the other to estimate the head's earnings. A similar estimate of expected value can be made for capital income, and income (money saved) from growing food or fixing up the house. The sum of these plus some small earnings of minors provide an estimate of the factor income of the unit excluding the wife's earnings. That income and other explanatory factors are then used to predict whether the wife works, her wage rate and hours. The resulting total "gross factor income" then helps explain transfer income, income tax, and whether the unit lives with relatives, either becoming dependent upon them, or providing housing for them. The estimated family income then includes income added by secondary units.

The decisionsinvolved at the various stages are clearly not independent of one another. Many of them are alternatives, and hence jointly determined, such as whether the head of a unit should work longer hours or let his wife go to work. We cannot be so circular as to use the head's income to explain why his wife works, and use his wife's income to explain how long the head works. We treat these joint decisions by making one conditional on the other. Since one can always convert a joint probability into a set of conditional probability statements, the overall results will be the same whichever we treat as being determined first, and thence influencing the next step.¹

The particular sequence we have chosen makes some empirical sense, too, we feel. It assumes that transfer incomes are largely determined by other income of the unit, not the reverse, and that the decision to live with relatives is usually made in the light of what the unit's income is, though a few people might decide to live with relatives in order to loaf.

At each stage, then, the thing to be explained is seen as the result of various exogenous forces and of previous things already determined. In some cases the order is inevitable. We can analyze the decision of how long to work only for those who had decided to work (though the possibility of part-time work may have influenced the decision to work).

Space does not permit us to present more than a small sample of the results, a full set of which will appear in a book to be published by McGraw-Hill in 1962. They come from a national probability sample interviewed in the spring of 1960.

The analysis uses smaller units than the family: spending units, or where possible adult units -- adults or adult couples with their children, sometimes called nuclear families because the analysis of the decision to combine to form larger units with higher incomes was part of the scheme. A glossary of terms is provided.

The method of analysis at each stage was essentially multiple correlation of regression with dummy (one or zero) variables, though the IBM program is actually an iterative one, not a matrix inversion? It is an extension of multiple

¹That is, the overall predictions will be the same. The estimated effects of different factors will depend on the order, to the extent that the factors in the second analysis are correlated with the outcome at the previous stage which is used in the second analysis as one of the predictors.

²See Daniel Suits, 'Use of Dummy Variables in Regression Equations," <u>Journal of the American</u> <u>Statistical Association</u>, 52 (December, 1957), 548-551. correlation to the situation where the explanatory factors are membership in subclasses like age groups, rather than numerical variables. Like any regression, it minimizes the summed squared errors of predictions made by the derived formula. We use dummy variables even where it might be possible to form a numerical scale of the explanatory factor. This has the advantage that we need make no restrictive assumptions about the linearity of the effect.

The most important restriction of this form of analysis is the assumption that each explanatory factor affects the dependent variable in an additive manner, regardless of the value of the other explanatory factors. This assumption of additivity is, of course, only an approximation to reality. Where interaction effects seemed likely to be important, we built them in from the beginning by using joint classification on two dimensions at once, or checked later by rerunning the whole analysis for part of the population, e.g., white, nonfarmer males not yet retired.

The analysis used various number of predictors, depending on which of the available measures seemed appropriate. Where the same factor, such as race or education, is used at each stage, its total effect in the determination of family income can be inferred from its effect at each stage.

We shall discuss here only one of these analyses, that explaining the hours worked during 1959 by the 86 per cent of the heads of spending units who worked at all. Clearly, purposeful decisions, like taking on a second job or working overtime, and involuntary results of outside forces, like illness or unemployment, affect the hours a man works during a year. Indeed, our theoretical model groups the various characteristics used to explain hours worked into constraining factors (like local unemployment and the worker's physical stamina and capacity for long hours of work) and motivational factors (like the pull of wages and the need for money and the desire to get ahead).

Each of these has, as proxy measures for it, a number of measured characteristics of the individual, his family, or the situation he faces. Sometimes the same measure is a proxy for more than one theoretical construct. When we realize also that some are measured more accurately than others, it becomes clear that the interpretations of the results requires some judgment beyond the usual statistics.

However, the results are presented in terms of the importance and the significance of sets of dummy variables, and for each set, both the unadjusted and adjusted (multivariate) coefficients are presented.³

³For those not familiar with dummy variables they are merely variables which take on only one of two values, one or zero, one if the individual belongs to a particular class on a particular characteristic. In the case of "race," there is a dummy variable "nonwhite" which is equal to one if the individual is nonwhite, otherwise it is zero. Table 1 shows the sets of subclasses in order of their importance. We are clearly interested in the importance of each set, such as "hourly wage rate." The measure we use is analogous to the beta coefficient of multiple correlation.⁴

The significance test is somewhat less legitimate, treating the adjusted coefficient as a set of means and comparing the variance estimate derived from them with that derived from the varaince unexplained by the whole regression. The measure is probably somewhat nonconservative. The different ranking of race as to importance and significance is, of course, meaningful. A significant difference applicable only to a small proportion of the population aids prediction less than an equally significant difference affecting large groups.

We have dealt with the problems of an additive analysis where interactions exist, by building them into the characteristics from the beginning; for instance, education and age form a joint classification, as do sex, marital status, and children in the "adult unit composition," and a personality measure and an attitude in "needachievement and attitude toward hard work."

You will notice that we use both "factual" and attitudinal variables, as well as local conditions and family background measures.

We have selected only a few of the specific classifications in Table 1 to duscuss here because of the shortness of space.

Age and education interact in their effect on earnings, and the joint classification, used elsewhere in explaining wage rates was also used in the analysis of hours worked. (See Table 2.) It turns out that not only do those with more education earn more per hour, but particularly when they are older, they manage to work more hours. Most of these differences are the result of differential unemployment, not different motivation. The differences between the third and fourth columns represent the multivariate adjustments for the effects of the other predictors used, and are relatively small and unsystematic here. For instance, the top row shows that high school dropouts under 25 worked 267 hours less than the average of 2092, but that after adjustment for other things, their low age and education account for only 164 hours of the difference.

In regression, one cannot have dummy variables for each class of a characteristic without overdetermining the system, but the coefficients can always be converted into a set, one for each subclass, with a weighted mean of zero for each set, and this makes the constant term in the predicting equation equal to the grand mean. Our iterative program produces its output in this form in the first place.

⁴The beta coefficient is generally considered not so good as rerunning the analysis without that factor to observe the loss in predictive power, but better than a coefficient of separate determination.

Table 3 deals with an interesting economic problem of the supply of labor in response to wage rewards, and illustrates the advantage of multivariate adjustments for other factors (like age and education), for only after adjustment does the negative relation between wage rates and hours worked become apparent. Why is the adjustment so dramatic? Because there are a number of factors causing spurious positive relations between wages and hours: professionals work longer hours at higher wages, because they are professionals, not because of the wages. The disabled work shorter hours at lower wages, vet few would attribute the shorter hours to the lower wages. Similarly, the young and uneducated have more unemployment and lower wages, but we should not attribute the unemployment to the low wages.

Why the negative relationship? Is leisure a superior good with a high income elasticity so that increases in real income from a higher wage rate are partly used to purchase more leisure? Perhaps so. Or perhaps there are minimum standards, that people who earn less per hour than others in their subclass, feel compelled to work longer to achieve an acceptable standard.

Table 4 shows a small but probably significant tendency for those with plans for the future that will require funds, to work longer hours now. The reduction in the effect is the result of the fact that the middle aged, more highly educated people, have more plans. The "no plans group" is composed of some persons who do not plan and others who have no living parents or children in school for whom to plan. Plans to send children to college seem to have more impact than plans to help parents, and having both types of plans seem to have an effect roughly equal to the sum of the two independent effects.

Finally, because of intrinsic interest, and the possibility that with better measures the factor might prove still more important, we look at Table 5. It is a joint classification according to a personality measure, and an attitude, both of which appear to have some effect on hours worked. The personality measure is an index of the achievement motive, behind which there is a great deal of theory and laboratory experimental work, and some analysis of cultures and subcultures.⁵ The possibility is suggested that achievement motivation may affect economic progress through the behavior of the masses as well as through the initiative of the small entrepreneurial class.

Theoretically the achievement motive, the propensity to derive satisfaction from overcoming obstacles by one's own effort in situations where

⁵See John W. Atkinson (ed.), <u>Motives in Fantasy, Action and Society</u> (Princeton: D. Van-Nostrand, 1958); and David McClelland, <u>The Achieving Society</u> (Princeton: D. VanNostrand, 1961). one's own performance can be compared with some standard of excellence, is developed by early childhood factors, and changes only slowly if at all under the impact of subsequent experience. It seems to affect how many hours people work. In other parts of our study it also affected wage rate, education completed by the head, and by his children, and education planned for the children who had not finished with school.

The subjective probability or belief that hard work leads to success in this world, rather than luck or help from friends, is a measure of a more volatile attitude which is assumed to be subject to change according to one's experience. Atkinson's theoretical model has this attitude interacting with the achievement motive, their <u>product</u> being the resultant motivation to act. In our case, to act means to work long hours. Perhaps because of other constraints on hours worked, our data would indicate something less than a multiplicative relationship.

Summary

We have attempted to provide something of the outline of our analysis, and the flavor of its results. The multivariate dummy variable procedure with selected interaction effects built in is seen to be useful in distilling the effects of different factors. (Professor Orcutt and his colleagues make extensive use of this method in preparing data for their simulation model of the economic system, reported earlier at these meetings.)

The analysis does not deal with the fact that explanatory factors may be at different levels in the causal chain, even when we break decisions or results down as we have. For instance, if achievement motivation is indeed fixed in early childhood it may help determine how much education the individual completes but education could not affect achievement motivation. Putting them both in a simultaneous multivariate analysis allows education, through which the achievement motive operates, to take credit for something which is ultimately the result of prior motivation.

Ultimately what is needed in the analysis of data is more flexible multivariate analysis procedures which take account both of interaction effects, and of the logical sequences which are possible when one explanatory factor can affect a second explanatory factor, but cannot be affected by it.

⁶John Atkinson, "Motivational Determinants of Risk-Taking Behavior," <u>Psychological Review</u>, 64 (November, 1957), 339-372.



Table l

•

Characteristics used to explain hours worked by spending unit heads (heads who worked during 1959)

	Indexes of:			
Characteristics of spending unit heads	Relative importance (Beta coefficient)	Significance (F-ratio)		
adult unit composition	.291	42.4**		
education and age	.258	11.7**		
occupation	.254	56.5**		
hourly earnings	.234	24.0**		
physical condition	.115	15.4**		
plans to help parents or children	.084	8.2**		
need-achievement score and attitude toward				
hard work	.066	2.5*		
religious preference and church attendance	.063	2.3*		
race	.044	6.8**		
extent of unemployment in states	.032	0.9		
difference in education of heads and wives	.025	0.4		
immigration of heads and fathers	.018	0.6		

** significant at probability level of .01
* significant at probability level of .05

Hours worked: deviations for education and age (for spending unit heads who worked during 1959)

		Per cent of				
	Number	spending	Unadjusted	Adjusted		
Education and age	of cases	unit heads	deviations*	deviations*		
<u>0 - 11 grades</u>						
under 25	76	2.5	-267	-164		
25 - 34	251	7.8	72	46		
35 - 44	297	9.5	69	90		
45 - 54	332	10.4	59	106		
55 - 64	269	8.5	-167	-53		
65 and older	104	3.8	-776	-699		
12 grades						
under 25	72	2.5	-101	-34		
25 - 34	97	3.4	215	133		
35 - 44	129	4.5	150	160		
45 - 54	78	2.6	178	130		
55 - 64	32	1.1	-220	-90		
65 and older	5	.2	83	-205		
<u>12 grades and some college</u> or nonacademic training						
under 25	84	3.0	-404	-376		
25 - 34	142	5.0	134	50		
35 - 44	139	5.0	323	224		
45 - 54	99	3.5	185	177		
55 and older	77	2.8	-234	-159		
college degree						
under 35	98	3.4	-126	-312		
35 - 44	82	2.9	280	106		
45 - 54	53	1.9	90	45		
55 and older	53	1.8	111	95		

* deviations from grand mean of 2092 hours

Table 3

Hours worked: deviations for hourly earnings (for spending unit heads who worked in 1959)

Hourly earnings	Number of cases	Per cent of spending unit heads	Unadjusted deviations*	Adjusted <u>deviations*</u>
none**	23	0.7	-426	-82
\$0.01 - 0.74; negative amount	347	8.3	195	369
\$0.75 - 0.99	175	5.0	-17	228
\$1.00 - 1.49	397	12.5	-120	84
\$1.50 - 1.99	337	12.0	74	114
\$2.00 - 2.99	714	26.6	5	-52
\$3.00 - 3.99	331	12.1	-34	-219
\$4.00 - 4.99	130	4.8	-70	-269
\$5.00 and over	115	4.1	2	-301

* deviations from grand mean of 2092 hours

****** self-employed who just covered expenses

Table 4

Hours worked: deviations for plans to help parents or children (for spending unit heads who worked during 1959)

Plan pare	s to help nts or children	Number <u>of cases</u>	Per cent of spending unit heads	Unadjusted deviations*	Adjusted deviations*
no p	lans to send children to college or help				
	parents	1255	41.2	-153	-60
plan	to help parents	466	15.6	-13	-7
plan	to send children to college	520	17.8	18 8	68
plan	to send children to college and help parents	s 328	11.5	277	117

* deviations from grand mean of 2092 hours

Table 5

Hours worked: deviations for attitude toward hard work and need-achievement score (for spending unit heads who worked during 1959)

Need-achievement score and attitude toward hard work	Number of cases	Per cent of spending <u>unit heads</u>	Unadjusted deviations*	Adjusted deviations*
hard work is more important than luck; need- achievement score is in	:			
high range	631	22.3	28	28
middle range	921	31.4	39	25
low range	440	İ4.2	10	-29
hard work is less important than luck; need- achievement score is in	:			
high range	120	3.9	0	18
middle range	228	7.3	-151	-72
low range	139	4.0	-231	-181
need-achievement score not ascertained	90	3.0	22	73

* deviations from grand mean of 2092 hours

Glossary of Terms

<u>Adult Unit</u> A person 18 or older, his spouse if he is married, and any children under 18 who live with him and for whom he is responsible.

<u>Adult Unit Composition</u> Marital status, sex, number of children under 18 for spending unit heads.

Dependent Adult Unit Any adult unit which does not contain the head of a spending unit. An adult unit which does not contain the major earner in the spending unit.

<u>Family</u> All occupants of a dwelling unit who are related to each other by blood, marriage, or adoption.

<u>Gross Disposable Income</u> Disposable money income, nonmoney transfers, money saved by home production, 6 per cent on investment in owner-occupied home. An alternative definition: gross factor income, total nonfamily transfers, net intrafamily transfers, less income tax. Total momey and nonmoney income available to the unit after taxes.

<u>Home Production</u> Labor which the unit expended in growing some of its own food, performing repairs and improvements in the home. The value of home production is the respondent's estimate of the money that the unit saved by doing this work itself.

Hourly Earnings For heads and wives who are not self-employed businessmen or farmers, the ratio of wage and salary income to hours worked. For self-employed businessmen and farmers, wages and salaries were imputed by deducting a reasonable return on the capital investment in the enterprise from the business or farm income (6 per cent of capital in business enterprises, 6 per cent of livestock and equipment, 5 per cent of property in farms). Earning rates were calculated on the basis of these imputed wages and salaries. If the business or farm did not earn enough to cover the imputed return on capital, earning rates may be negative.

<u>Need-Achievement Score</u> A measure of the extent to which spending unit heads differentiate in favor of high status, high reward occupations, thought to be a measure of need-achievement.

<u>Physical Condition</u> Spending unit head's report of the presence or absence of an illness physical condition, or nervous condition which limits his work.