ROBERT E. HALL

Effects of the Experimental Negative Income Tax on Labor Supply

The negative income tax experiment in New Jersey and Pennsylvania has provided much new evidence on the decisions of poor families about their supplies of labor. Even though the income and substitution effects of the negative income tax act in the same direction—making work both less remunerative and less necessary—the average family in the experiment reduced its hours of work only slightly when subject to the tax and transfer provisions of the negative income tax.

I shall discuss three lessons that economic students of labor supply have learned.

First, the finding of small effects has an impact on the way theoretical models of labor supply are specified. I argue that a good theory of labor supply must take account of the diversity of behavior of apparently identical workers, as well as of their average behavior. A theory of the representative individual is inadequate to predict the effect of a negative income tax because of its differential appeal to those whose ability or propensity to work is low.

Second, I examine the relation between the effect of the negative income tax in the experiment with predictions from econometric studies of labor supply in cross sections. The theory of the first part of the paper suggests that a naive application of the cross-sectional studies will overstate the effect of the negative income tax. I argue that adjustment for these biases makes the cross-sectional results consistent with the experiment, at least within sampling error.

Third, I examine the data from the experiment from a point of view somewhat different from that of the extensive empirical analysis carried out by the economists associated with it. Although the basic effect of the experiment is small, the statistical reliability of the estimate is sufficiently high to make it implausible that there is actually no effect at all. Statistically insignificant regression coefficients in the elaborate models of other studies have left this question somewhat up in the air. The fact that the effect is statistically significant does not mean that it is large, only that the experiment was sufficiently well designed to detect a small effect.

A major theme of the paper is that permanent unobserved determinants of labor supply vary among families. Econometrically, these induce a positive correlation between the average hours of work of the subjects before and during the experiment. The presence of this correlation suggests an important respect in which the design of the experiment could have been improved. As conducted, the experiment failed in two ways to take full advantage of the opportunity to let the subjects serve as their own controls. First, too few observations were made of the subjects' preexperimental hours of work. Additional quarters of data would have provided a better baseline from which to measure the response of each individual. Second, if half of the subjects had received the negative income tax first and had been interviewed for a number of quarters after its termination, it would have been possible to dispense with the controls altogether. A proposed revision of the experiment following both of these suggestions could have gathered the same amount of information with only a little over one-sixth the number of subjects.

A major surprise in studies of the data from the experiment is the apparent finding of large differences in the responses of the three ethnic groups distinguished in it: whites, blacks, and the Spanish speaking. Other investigators seem to take this finding at face value, concluding that there are important ethnic differences in labor supply functions. I question this conclusion on the grounds that attrition in the black and Spanish-speaking samples is so large as to make the data unusable without a thorough analysis of the interaction between the response to the negative income tax and the departure of families from the sample. Among the experimental subjects, 46 percent of the black and 56 percent of the Spanish-speaking families were excluded from the category of "continuous husband-wife sample" that is used extensively by others. Problems in the data may well explain their puzzling findings of large ethnic differences.

Theory

A discussion of the negative income tax experiment must begin with a fairly complete development of the economic theory of the response to a negative income tax. The paper by Albert Rees and Harold W. Watts has provided a good graphical discussion of the theory. Here—and in much greater detail in appendix A to this paper—I develop an analytical version of the same theory in which an important issue can be treated explicitly: families must decide whether or not to be covered by the negative income tax. They may escape its tax and forgo the income guarantee simply by earning more than the breakeven level of income. The theory must pay close attention to the endogeneity of participation in the negative income tax. Within the experiment, 13 percent of the families who were offered the generous plan paying a guarantee of 125 percent of the poverty level and taxing earnings at 50 percent opted against the plan by earning more than the breakeven level.[1] Practically all families, 94 percent, opted against the least generous, or 50–50, plan. Slightly more than one-quarter of the families who were offered the 100–50 plan decided against it. Surprisingly, the writings by the economists associated with the experiment omit any theoretical treatment of the issue of the choice that families make: whether to earn above or below the breakeven level and thus whether or not to be covered by the negative income tax.[2]

Closely related to the issue of the endogeneity of participation in the negative income tax is the need to treat the diversity of responses to the negative income tax and other economic opportunities among an apparently homogeneous population of families. Virtually all discussions of the response to the negative income tax are framed in terms of a representative family and suggest that the average response in the population is the same as the response of that family. But the negative income tax has a differential appeal to families whose ability or propensity to work is small. In many cases the representative or median family will not respond at all to a negative income tax, but an important minority will decide to earn less, so the average response is negative. In appendix A, I deal with this issue by formulating a model in which an unobserved variable indexes attitudes and

1. See Henry J. Aaron, "Cautionary Notes on the Experiment," in this volume.

2. As Zvi Griliches remarked at the conference, it is misleading to talk as if all of the experimental subjects were paying the tax and receiving the guarantee when in many cases only a small fraction were covered by the negative income tax.

abilities for work. The distribution of the variable among the members of the population implies a distribution of hours of work when no negative income tax is available and permits a prediction of the distribution of responses to a negative income tax when it becomes available. I discuss two models within this framework. In model 1, families differ in the fraction of their total resources they devote to purchases of goods. Some families work a good deal, purchase a large volume of goods, and spend relatively little time at home. These families have little or no response to the negative income tax. Others work less, purchase less, and spend more time at home. In model 1, these families have a large response to the negative income tax.

The parameters of model 1 are fitted informally to data on average hours of work (about thirty-five a week) and to the dispersion of hours of work (a standard deviation of about eighteen a week). For a family with a wage of $2.00 an hour that is offered a negative income tax with a guarantee of $80 a week and a tax rate of 50 percent, it makes a number of predictions. First, all families opt to participate in the negative income tax; not even the hardest-working family finds it desirable to earn above the break-even level of $160 a week, or eighty hours of work. Second, only 45 percent of the families continue working after the negative income tax becomes available. Because they pay the tax of 50 percent on their earnings and have higher incomes nonetheless, they reduce their work by almost twenty-four hours per week. The remaining families stop work altogether, representing an average reduction of about twenty-two hours of work a week—those who stop work are also those who worked less than average before the negative income tax. The average reduction within the whole population is 22.4 hours a week. Model 1 confirms the worst fears of the opponents of the negative income tax: the program causes wholesale withdrawal from the labor force as a large fraction of the population retires and lives on the guaranteed income alone.

In my view, however, the experiment has conclusively laid to rest any fears of a response of hours of work to the negative income tax of the magnitude predicted by model 1. The defect of model 1 is its assumption that families with relatively low hours of work in the absence of the negative income tax are those which spend a large fraction of their resources on time at home. When the negative income tax expands total resources, they spend the same fraction of the increase on time at home and consequently work considerably less. Model 1 assumes that the labor supply functions of a fraction of the population are highly income elastic, which must be false in view of the results of the experiment.

Model 2 embodies an alternative view of diversity among families. It assumes that they differ in the number of hours a week over which they have discretion in choosing between work in the market and work and other activities at home. All families are assumed to have a relatively low propensity to spend increases in resources on additional time at home. Model 2 gives about the same distribution of hours of work among families as model 1 but predicts a much smaller response to a negative income tax. Families with wages of $2.00 an hour opt unanimously in favor of a plan with a guarantee of $80 a week and a tax rate of 50 percent, and all of them continue to work. Average hours of work fall by 2.4 when the negative income tax becomes available, a reduction that seems consistent with the findings of the experiment. According to model 2, only 84 percent of families with wages of $3.00 an hour would opt in favor of the same negative income tax plan. Average hours of work would fall by only 1.6 a week. A program whose guarantee was only $40 a week would attract just 54 percent of families with wages of $2.00 an hour. Those opting for the program would reduce their hours of work by 1.2 a week, so the average would fall by half that amount, 0.6 hour a week.

Model 2 seems to give a reasonable prediction of the distribution of responses to the negative income tax, as well as a reasonable prediction of the average response. Future research will attempt to fit an econometric model along the lines of model 2 to the data from the experiment. For now, the main econometric contribution of this study of the theory of the response to the negative income tax is emphasis on the importance of the distribution of unobserved determinants of the labor supply of the families in the experiment.

Predicting the Effect of a Negative Income Tax from Cross-sectional Studies

Interest in the potential impact of a negative income tax has stimulated considerable research on labor supply functions over the past few years. In this section I will look briefly at the predictions of a cross-sectional regression study of hours of work in the light of the findings of the experiment.[3] Although I will use my own results, I should emphasize that most of the studies are in substantial agreement about the response of hours of work to changes in wages and income.[4] I repeat my point, however, that in principle regression studies are incapable of predicting the effect of the negative income tax because they are unable to deal with the opportunity to opt out of the program. The best they can do is to predict the effect of a mandatory tax and transfer program with the same parameters as the negative income tax. They will invariably overstate the effect of the program, and the overstatement will be most severe for families whose alternative sources of income are large enough that only a fraction decides to participate.

Table 1 presents estimates of the effect of a program with a guarantee of $80 per week and a tax rate of 50 percent. The first two rows give estimates of the response of hours of work to wages and income, derived from my earlier study—which involved putting them on a weekly rather than annual basis and converting from income per adult to family income. It is important to understand the measure of income used in the study: full-time income = unearned income + (38.5 · husband's wage) + (38.5 · wife's wage).[5]

The family considered in Table 1 has a full-time income of $154 per week arising solely from the value of the family's time. The negative income tax imposes a tax of 50 percent, or $77 per week, on the family and also gives it $80 in transfers. Measured by full-time income, the family is ahead by only $3 per week. In fact, the negative income tax has an income effect much larger than this, because the tax actually applies only to work in the market; time spent at home escapes taxation. In the husband's labor supply function, there is a cross-income effect from the implicit lowering of the wife's wage caused by the tax: that is, tax rate · wage · (38.5 − wife's actual hours). This appears in row 4 of the table and is the main ingredient in the total income effect in row 5. A similar though much smaller adjustment is made for wives.

Husbands show no strong evidence in favor of a positive wage elasticity in my study, so the effect of the negative income tax operates entirely

3. Sandra Christensen currently is preparing a more complete survey of cross-sectional evidence on the impact of a negative income tax.

4. See Robert E. Hall, "Wages, Income, and Hours of Work in the U.S. Labor Force," in Glen G. Cain and Harold Watts (eds.), *Income Maintenance and Labor Supply: Econometric Studies* (Markham, 1973), pp. 102–62.

5. The logic of the use of this measure of income is discussed in ibid., pp. 104–07.

Table 1. Effects of a 50 Percent Negative Income Tax on Income and Hours of Work of Husbands and Wives, by Color, from Cross-sectional Regression[a]

Effect	Husbands		Wives	
	White	*Black*	*White*	*Black*
1. Wage effect (hours/week per dollar/hour)[b]	0	0	12.7	11.5
2. Income effect (hours/week per dollar/week)[b]	0.110	0.045	0.075	0.115
3. Change in full-time income (dollars per week)[c]	3	3	3	3
4. Adjustment for cross-income effect (dollars per week)[d]	23	21	0	5
5. Adjusted change in income (dollars per week)[e]	26	24	3	8
6. Income effect of the program (hours per week)[f]	2.9	1.1	0.2	0.9
7. Wage effect of the program (hours per week)[g]	0	0	9.5	8.6
8. Total effect (hours per week)[h]	2.9	1.1	9.7	9.5

a. Assumptions: husband's and wife's wages, $2.50 and $1.50 per hour, respectively, with no other income; husband's hours, 38.9 for whites, 34.1 for blacks; wife's hours, 7.6 for whites, 10.8 for blacks; negative income tax guarantee, $80 per week; tax rate, 50 percent.
b. Derived from Robert E. Hall, "Wages, Income, and Hours of Work in the U.S. Labor Force," in Cain and Watts (eds.), *Income Maintenance and Labor Supply*, Table 3.5, pts. A–D, pp. 133–34.
c. $\$80 - 0.5[(38.5 \cdot 2.50) + (38.5 \cdot 1.50)]$.
d. Spouse's wage times (38.5 minus spouse's hours) times 0.5.
e. Row 3 plus row 4.
f. Row 2 times row 5.
g. Row 1 times 0.5 times wage.
h. Row 6 plus row 7.

through the income effect. The effect for whites is just under three hours a week and just over one hour for blacks. Both figures are larger than most estimates derived from the experiment, although both sets of evidence agree that whites are more likely to reduce their hours of work in response to the negative income tax than are blacks. In view of the tendency for regression results to overstate the response, I find no reason to think that the cross-sectional and experimental results are in conflict.

For wives, the regression prediction is unreasonable on its face, implying negative hours for whites and essentially complete withdrawal for blacks. The regression model in my study—and in all others known to me—fails to take account of nonnegativity by attenuating the response to wage reductions as they gradually push a larger fraction out of the labor force altogether. The data from the experiment are much more suitable for this type of analysis than those obtained from surveys, since they deal with individuals facing much lower effective wages.

Econometric Appraisal of the Effect of the Negative Income Tax within the Experiment

Econometric methods can answer two somewhat distinct questions about the experiment. Did the program have any effect at all? If so, precisely how large was it and how did it vary among different groups in the sample? Most of the attention of the investigation associated with the experiment has been devoted to the second question, in part, apparently, because it never occurred to them—or to anyone else—that the first question was a live issue. The models and methods used have been sophisticated and complex. In my view, however, they are not well suited to answering the first question. As all experienced users of statistical methods are aware, the more complex the model, the greater is the dispersion in the estimates of its parameters, and the more likely is the acceptance of the null hypothesis that any one of them is zero. In statistical terms, complexity reduces the power of some tests. I think that Harold Watts and his coauthors mislead their readers slightly in summarizing their results by stating: "Overall, it would not require a determined skeptic to claim that there is no evidence of any disincentive at all for husbands."[6] In this section I will argue that, on the contrary, there are small but clearly positive reductions in average hours of work among those offered the experimental negative income tax. The methods I use will be fairly simple and directed toward the first of my two questions.

I will begin with the following numbers (which are calculated from all data reported in the "Vargn" file for the experiment), representing average weekly hours of white husbands:

	Experimental group	*Control group*
Before experiment	34.1	34.8
During experiment	31.8	34.4

From these numbers I will try to derive a simple estimate of the average effect of the program on the labor supply of the experimental subjects.

6. Harold W. Watts and others, "The Labor-Supply Response of Husbands," *Journal of Human Resources*, vol. 9 (Spring 1974), p. 199.

Statistical analysis requires the specification of the expected values of the observed statistics in terms of unknown parameters and a specification of their variances and covariances. For now, I will take the simplest possible model of the expected values:

$$E(H_{S,B}) = E(H_{C,B}) = E(H_{C,D}) = \mu, \text{ and}$$
$$E(H_{S,D}) = \mu + \beta,$$

where H is average hours; S refers to the experimental group and C to the control group; and B refers to the observations before the experiment took place and D to those made during the experiment.

The statistical problem, then, is to find an estimator b of β, the average effect of the program. The basic question is whether to compare the hours of the subjects during the experiment to their own hours before the experiment or to the hours of the controls during the experiment. There is also a subsidiary question of whether or how to use the data from the controls before the experiment. One polar choice is to measure the effect by the change in the hours of the subjects only; this yields what I will call the difference-in-means-of-subjects estimator, or *DMS*:

$$b_{DMS} = H_{S,D} - H_{S,B}$$
$$= -2.3 \text{ hours per week.}$$

This estimator is unbiased.[7] The other polar choice is to use data from the experimental period only, giving the difference-in-means-during-the-experiment estimator, or *DMD*:

$$b_{DMD} = H_{S,D} - H_{C,D}$$
$$= -2.6 \text{ hours per week.}$$

Again, this estimator is unbiased.[8]

Finally, a compromise, the equal-weighting estimator, *EW*, gives a weight of one-third to each of the three observations on groups not subject to the program:

$$b_{EW} = H_{S,D} - \tfrac{1}{3}(H_{S,B} + H_{C,B} + H_{C,D})$$
$$= -2.6 \text{ hours per week.}$$

This estimator also is unbiased.[9]

Which estimator is best? The usual answer is the one with minimum variance. Derivation of the variances of these estimators requires some preliminary examination of the stochastic character of the underlying data.

7. $E(b_{DMS}) = \mu + \beta - \mu = \beta.$
8. $E(b_{DMD}) = \mu + \beta - \mu = \beta.$
9. $E(b_{EW}) = \mu + \beta - \tfrac{1}{3}(\mu + \mu + \mu) = \beta.$

The theory of family labor supply presented earlier in this paper suggested that the wide dispersion in hours of work among apparently homogeneous families was the consequence of the dispersion of an underlying more or less permanent unobserved index of tastes, attitudes, and abilities among the families. In practical work, this model of labor supply must be amended by adding a transitory component representing the effects of short-run fluctuations in employment opportunities, especially overtime and layoffs, and in the ability to work. A simple model then has the form:

$$H_{it} = \mu + \eta_i + \epsilon_{it},$$

where H_{it} is hours of work of family i in time t; μ is the average over all families; η_i is a permanent unobserved shift for family i with mean zero across all families (in the models of appendix A, η_i is a multiple of θ, shifted downward to have mean zero); and ϵ_{it} is the transitory shift at time t. I make the simple stochastic assumptions that the permanent and transitory components are uncorrelated and that the transitory component does not exhibit serial correlation.[10] Further, I assume that families are stochastically independent of each other. Of these assumptions, the only one open to serious question is the lack of serial correlation of the transitory component. I shall present evidence on this point shortly.

The stochastic model just outlined is familiar to all students of panel data on individuals. It has been used extensively in analyzing data from the negative income tax experiment by Harold Watts, although he uses an elaborate regression specification in place of my μ.

In appendix B to this paper, I derive the variances and covariances of the four means implied by this stochastic model. I calculate the variances of the three estimators—*DMS*, *DMD*, and *EW*—and then derive the minimum variance or efficient estimator, called *GLS*, and calculate its variance. In all of these calculations, a critical parameter, λ, appears; it is defined as the fraction of the total variance in the data contributed by the permanent component. If λ is close to zero, the efficient estimator is similar to *DMD*—most information comes from the comparison of subjects and controls during the experiment. But if λ is positive, the situation is rather different. The efficient estimator takes maximal advantage of the correlation between the average hours of the subjects before and during the experiment by

10. The assumptions are as follows:

$$\text{var}(\eta_i) = \delta^2, \text{ and } \text{var}(\epsilon_{it}) = \sigma^2;$$

$$\text{cov}(\eta_i, \epsilon_{it}) = 0, \text{ for all } t, \text{ and } \text{cov}(\epsilon_{it}, \epsilon_{it}) = 0, \text{ if } t \neq t'.$$

making the subjects serve as their own controls. When λ is much above 0.25, the efficient estimator is more like *DMS*. Data from the controls receive little weight, and indeed the question arises whether the experiment should have had controls at all.

Evidence on Variances and Covariances of Hours of Work in the Experiment

The next step is to establish the relevance of the stochastic model and to estimate the value of λ for the data at hand. Table 2 presents estimates of the variances and covariances for the whole sample. The model predicts equal covariances for all lags, but it is no surprise that the covariances tend to be smaller for longer lags. Still, the evidence in favor of an important permanent component is strong. For both men and women, the covariance of hours in adjacent quarters is substantially less than the variance—about one-half for men and two-thirds for women. The covariance for hours two quarters apart drops by much less. The covariance seems to approach constant levels of 110 for men and 60 for women, and falls some more at the longest lag. It is reasonable to approximate this variance-covariance

Table 2. Variances and Covariances of Hours at Work of Husbands and Wives[a]

Description	*Husbands*	*Wives*
Variance	330	143
Covariance[b]		
Lag, j		
1	174	100
2	145	89
3	138	82
4	132	75
5	122	71
6	117	68
7	109	63
8	112	61
9	104	59
10	99	56
11	93	52
12	72	41

Source: Derived from data in "Vargn" files of the New Jersey Graduated Work Incentive Experiment. See text and appendix B of this paper for the method of derivation.

a. Includes all individuals who were husbands and wives in the first interview.

b. Covariance of H_t and H_{t-j}, where H_t is hours of work in time t and j is quarterly lag.

structure with the simple permanent transitory model with $\delta^2 = \text{cov}(H_t, H_{t-6})$ and $\sigma^2 = \text{var}(H_t) - \delta^2$. This gives $\lambda = 117/330 = 0.35$ for men and $\lambda = 68/143 = 0.48$ for women.

Estimates and Standard Errors for White Husbands

The following tabulation presents, for white husbands, the three alternative estimates already discussed (*DMS*, *DMD*, and *EW*), the efficient estimate (*GLS*), and their standard errors (in parentheses):

	Weight for preexperimental data		
Estimator	*Experimental group* (ω)	*Control group* (γ)	*Estimated effect (hours per week)* (b)
DMS	1	...	−2.3 (1.0)
DMD	0	0	−2.6 (1.1)
EW	0.33	0.50	−2.6 (0.9)
GLS	0.56	0.08	−2.4 (0.8)

In all cases, the standard errors are calculated from formulas in appendix B, taking account of the permanent component of the variance. Only in the case of the *DMD* estimator is the standard error the same as the usual standard error of the difference of two uncorrelated means.

All four estimators agree in rejecting the null hypothesis of no disincentive effect of the experimental negative income tax. It is essentially impossible that random variation in hours of work could account for the observed reduction in hours of the experimental subjects.

The standard errors give a ranking of the four estimators. Poorest is the comparison between subjects and controls during the experiment (*DMD*), and third is the comparison of subjects before and during the experiment (*DMS*). If there were no permanent effect, *DMD* would have been substantially better than *DMS* because there are twelve times as many observations on controls during the experiment as there are on subjects before the experiment. *DMS* would have had a crushing advantage over *DMD* if the design of the experiment had divided the thirteen observations more equally between those before and those during the experiment. The second-best estimator is *EW*: it is advantageous to give some weight to the controls. Best of all is the *GLS* estimator, which improves on *EW* by assigning more weight to the preliminary data for the subjects and less to the preliminary data for the controls.

A Model with Changes in Hours over Time

A major objection to the results just presented is their dependence on the assumption that nothing changed the mean hours of work of the subjects except the negative income tax. The conclusion that data on controls are not as useful as data on subjects derives from the lack of a need for information about the probable behavior of the subjects during the experiment. In this section I reconsider the previous results under a model with a time effect, χ:

$$E(H_{S,B}) = E(H_{C,B}) = \mu;$$
$$E(H_{C,D}) = \mu + \chi; \text{ and}$$
$$E(H_{S,D}) = \mu + \chi + \beta.$$

The general class of unbiased estimators in this model is:

$$b = H_{S,D} - \omega H_{S,B} - (H_{C,D} - \omega H_{C,B}).$$

Note that $H_{C,B}$ entered b with a negative sign in the efficient estimator without a time effort, but enters here with a positive sign. Further, it is no longer possible to reduce the variance introduced by $H_{C,D}$ by multiplying it by a weight less than unity. The best that can be done is to subtract a fraction of $H_{C,B}$, which is positively correlated with $H_{C,D}$.

The *DMD* estimator is a member of this class, with $\omega = 0$. If there is a permanent component in the stochastic element, however, a more efficient estimator uses a positive ω to reduce the variance of $H_{S,D} - \omega H_{S,B}$ and of $H_{C,D} - \omega H_{C,B}$.[11] With a single observation before the onset of the program, as in the experiment, the optimal procedure is to take quasi-differences with a weight for the earlier observation equal to the correlation, λ, of hours of work in different periods. In a regression framework, this corresponds to including initial hours of work as a righthand variable, a procedure used extensively by the investigators associated with the experiment.

The data for white husbands previously studied shows no evidence of a time effect. On the other hand, the data on the average weekly hours of

11. The optimal ω can be shown to be equal to:

$$\omega = \frac{\lambda}{\lambda + \frac{1-\lambda}{T_B}} [= \lambda, \text{ if } T_B = 1],$$

where T_B is the number of observations obtained before the experiment.

their wives suggests rather strongly that they increased during the experiment:

	Experimental group	*Control group*
Before experiment	3.8	2.7
During experiment	3.9	4.5

Why should there be a time effect? First, the labor force participation rate of women in the entire U.S. economy was rising fairly rapidly during the experiment. Second, the subjects and controls were chosen deliberately because they were the poorer members of a much larger population. Some families were included because of the temporary withdrawal of the wife from the labor market. Both controls and subjects tend to work more as the experiment progresses and they return to their normal patterns of work. The disincentive effect of the negative income tax tends to counterbalance this increase, so the experimental effect must be inferred indirectly from the data on both subjects and controls.

Results for white wives, assuming time effect, for the two estimators that are unbiased in the presence of a time effect are shown below, where *DMD* represents differences in means of subjects and controls, experimental period, and *GLST*, efficient weights subject to constraint of unbiasedness with time effect, and the figures in parentheses are standard errors:

Estimator	*Weight for preexperimental data* (ω)	*Estimated effect* (b)
DMD	0	−0.5
		(0.8)
GLST	0.47	−1.5
		(0.6)

The simple comparison of subjects and controls (the *DMD* estimator) gives an effect in the expected direction, but of small magnitude. The estimator that makes optimal use of the sample data, *GLST*, not only has a smaller standard error, but estimates a larger effect. The null hypothesis of no response to the negative income tax is rejected by the *GLST* results at any reasonable level of significance.

Among white families, the simple methods of this paper demonstrate statistically unambiguous evidence of small disincentive effects of the negative income tax on the labor supply of both husbands and wives. The fact that the effects are statistically significant does not mean that they are large, only that the design of the experiment and the variance-components estimators measure the effects with sufficient precision to detect even a small effect.

Black and Spanish-speaking Families

The results for black and Spanish-speaking families are more difficult to interpret. The relevant means of the average weekly hours of work for black and Spanish-speaking husbands and wives—tabulated from all data reported in the experiment's "Vargn" file—are shown below:

	Experimental group		*Control group*	
	Before	*During*	*Before*	*During*
Black husbands	31.8	31.2	31.9	28.5
Black wives	5.6	6.0	6.2	6.3
Spanish-speaking husbands	32.9	31.7	36.7	32.9
Spanish-speaking wives	3.3	2.0	3.1	4.4

Taken at face value, the data for husbands in both groups suggest, paradoxically, that labor supply hardly changed for those offered the negative income tax program, whereas it decreased substantially for the control group. The results for wives are more reasonable: black wives show no effect and Spanish-speaking wives in the control group show an effect of about the same magnitude as white wives. In the experimental group, however, a considerable decline occurred, whereas white wives changed hardly at all. I have serious reservations, however, about the validity of the data for the black and Spanish-speaking families. In spite of intense effort, the administrators of the experiment found great difficulty in collecting data from these ethnic groups. The following data on the percentages of families lacking continuous structure and data throughout the experimental period illustrate the problem:[12]

	Experimental group	*Control group*
White	22	34
Black	46	45
Spanish speaking	56	61
Total group	41	45

12. Computed from Harold W. Watts, Dale J. Poirier, and Charles Mallar, "Concepts Used in the Central Analysis and Their Measurement," in Harold W. Watts and Albert Rees (eds.), *Final Report of the New Jersey Graduated Work Incentive Experiment* (University of Wisconsin—Madison, Institute for Research on Poverty, and Mathematica, 1974), vol. 1., sec. 1, pt. B, chap. 1.

In the critical experimental group, just over one-fifth of the white families failed to report fully usable data, either because the husband or wife left the household or because interviewing stopped before the end of the experiment. This level of attrition is troublesome but probably not capable of obscuring the basic effect of the program. But for black families, nearly one-half of the data are not fully usable, and for Spanish-speaking families, more than one-half of the data are absent or contaminated. The situation is no better for the less important control group: an astonishing 61 percent of the data for the Spanish speakers are not fully usable.

The high rate of attrition in the sample and its implications have not escaped the attention of the proprietors of the experiment. Jon K. Peck[13] has studied the relation between family characteristics at the first interview and subsequent attrition and confirms the differential among ethnic groups reported above. The analyses of the impact of the negative income tax on labor supply in the other chapters of the final report seem to ignore the implications of the high rates of attrition. In particular, I suspect that biases on this account are an important part of the explanation of the paradoxical response of black husbands reported by Harold Watts and associates.[14] Nowhere do Watts and his coauthors mention what a small fraction they are able to use of the potential number of observations for black families, although they confess some mystification at the behavior of the sharply truncated black control group.

I do not know why attrition should cause the hours of the experimental group to rise over time and the hours of the control group to fall, but I am reluctant to calculate estimates of the effect of the negative income tax when the surviving part of the sample involves so much self-selection. A thorough treatment of the relation between attrition and response to the negative income tax is needed.

Benefits of Additional Observations before Onset of the Program

Permanent unobserved differences make the experimental subjects their own best controls. As the data for white husbands below illustrate, with the same effort in collecting data the experiment could have reduced the variance of almost all of the estimators by delaying the onset of the program by five quarters:

13. Jon K. Peck, "The Problem of Attrition," in ibid., vol. 2, pt. C, chap. 1.
14. Watts and others, "The Labor-Supply Response of Husbands," pp. 181–200.

	Variance	
Estimator	*As conducted* ($T_B = 1, T_D = 12$)	*Proposed modification* ($T_B = 6, T_D = 7$)
DMS	0.54	0.17
DMD	0.75	0.83
EW	0.45	0.41
GLS	0.33	0.16
GLST	0.48	0.30

Particularly noteworthy is the low variance of the *DMS* estimator in the modified design. If time effects are ignored, the controls make almost no contribution to reducing the variance of b; *GLS*, which weights the controls optimally, has a variance only slightly below that of *DMS*. The experiment could have been run with half the data collected if it had been designed to avoid the need to estimate a time effect. It seems to me that this could have been done easily by dividing the experimental subjects into two groups, one of which would have six quarters of control observation followed by six quarters of the program, and the other group the reverse. Then the time effects in the *DMS* estimator would be canceled and it would be unbiased. The variance of b estimated by *DMS* in the modified experiment would be 0.17, compared with the variance of the optimal estimator for the experiment as conducted, 0.48. It is a remarkable implication of these calculations that the same amount of information about the overall effect of the program that is available from the existing body of data for some 1,200 families could have been gathered by observing fewer than 200 families.

Regression Analysis and More Powerful Tests of the Hypothesis of No Experimental Effect

The defect of the methods considered in this paper is their use of a fixed parameter, β, to represent the effect of the program on the hours of work of the subjects. In fact, the subjects received benefits according to a variety of guarantee levels and tax rates and differed in predictable ways in their likely response to the program. This information can be used to advantage in statistical analysis of the data from the experiment. A slight extension of my previous model is:

$$E(H_{S,D} \mid w, Y_0, \tau, B_0) = \mu(w, Y_0) + \rho\beta(w, Y_0, \tau, B_0),$$

where $\mu(w, Y_0)$ is the expected labor supply in the absence of the program for a family with wage w and income Y_0; $\beta(w, Y_0, \tau, B_0)$ is the expected effect of the program; and ρ is a parameter to be estimated, interpreted as the fraction of the expected effect that is actually observed. Harvey Rosen[15] has studied the effect of the personal income tax on the hours of work of married women within this framework and has found that ρ is close to unity.

Harold Watts has attempted something rather similar in defining his θ variable, which takes on large values for susceptible families and small values for those who are unlikely to respond to the program. Watts's procedure is ad hoc, but a variable in the same spirit can be calculated from the theory of the first section of this paper. It is exactly the $E(H_N) - E(H_P)$ function defined, for example, for model 2 in Table 4. Under this definition of $\beta(w, Y_0, \tau, B_0)$, tests of the null hypothesis $\rho = 0$ are more powerful tests of whether the negative income tax has any effect at all, and tests of $\rho = 1$ test whether it has the magnitude predicted by economic theory.

Conclusions for Future Research

A close look at the data from the experiment has resolved some of its mysteries. There is, after all, a measurable decline in the hours of work of white families when they are offered a negative income tax. With an appropriate statistical technique, the magnitude of the decline can be estimated with reasonable precision. For blacks and Spanish speakers, the data will not support the same kind of analysis because attrition in the samples was so high. Several lessons for future research and experimental design have emerged from my study.

First, statistical work needs to take full account of the permanent and transitory components of individual behavior.

Second, when an explicit model of labor supply is fitted to data from the experiment, it should recognize the endogeneity of participation in the negative income tax. This requires an explicit model of the diversity of preferences.

Third, future experiments should gather as much data as possible on their subjects before and after the negative income tax or other treatment is administered. The most efficient estimators of the effect of the experiment let the subjects serve as their own controls.

15. Harvey S. Rosen, *The Impact of U.S. Tax Laws on the Labor Supply of Married Women* (Ph.D. dissertation, Harvard University, 1974).

Fourth, it is worthwhile to devote a great deal of effort to keep in touch with all experimental families. An experiment with 100 subjects and 10 percent attrition is more useful than one with 1,000 subjects and 50 percent attrition.

Appendix A: Theory of the Response to a Negative Income Tax

The theory needs to deal with three complications: the diversity of attitudes about work and consumption within the population; the opportunity to choose whether or not to be covered by the program; and the fact that hours of work cannot be negative. The analysis of this section is the working out of a simple story about family decisions. The family starts by formulating a work-consumption plan on the assumption that it is not covered by the program. Then it formulates an alternative plan on the assumption that it pays the tax and receives the transfer from the program. Finally, it decides whether or not to participate by choosing the plan that yields the greater satisfaction. There is a strong presumption that the second plan, which assumes participation, will involve less market work than the first plan. The program reduces the incentive to work by taxing earnings and stimulates the use of time at home by providing income to purchase goods. I distinguish three kinds of response to the program:

—The family would not work in the absence of the program. Then it is certain to participate and will not work under the program either.

—The family would work if it did not participate, but would not work if covered by the tax and transfer. Then the program causes families to withdraw from the labor force if by spending all of their time at home and using the transfer to buy goods they achieve a higher level of satisfaction than by spending less time at home and using earnings to buy goods.

—The family works whether or not it is covered. It decides in favor of the program if the combination of more time at home and smaller purchases of goods—financed by the earnings after the tax and the transfer income—is more satisfying than the alternative. If the program makes it possible to reduce work and yet spend more on goods, then it is certain that the family will participate. The effect of the program is to reduce the hours of work of those families who opt to participate.

My next step is to provide a general analytical apparatus to describe family decisions. The basic tool is the indirect utility function, which measures the level of satisfaction achieved by the family when it faces a reward, w, for each hour of work, and has income, Y, from sources other than earnings. Call it $g(w, Y)$. Then the level of satisfaction (u) conditional on not participating is:

$$u_N = g(w, Y_0),$$

where Y_0 is the family's unearned income from interest, pensions, and the like. The alternative of participating yields:

$$u_P = g[(1 - \tau)w, B_0 + (1 - \tau)Y_0],$$

where τ is the tax rate imposed by the program and B_0 is its transfer component, or income guarantee. A family opts for the program if u_P exceeds u_N, but otherwise not. For future use I note here that the labor supply function can be extracted from the indirect utility function by differentiating it:

$$H_N = \frac{\partial g(w, Y_0)/\partial w}{\partial g(w, Y_0)/\partial Y_0}, \text{ and}$$

$$H_P = \frac{\partial g[(1 - \tau)w, B_0 + (1 - \tau)Y_0]/\partial w}{\partial g/\partial Y_0}.$$

The actual labor supply depends on the family's decision about the program:

$$H = H_N \text{ if } u_N > u_P, \text{ or } H = H_P \text{ if } u_P > u_N.$$

So far I have dealt explicitly with the decision about participation and implicitly with the nonnegativity of labor supply—the indirect utility function takes account of the family's inability to reduce hours of work below zero. The problem of nonnegativity will arise explicitly when I assign a particular functional form to the indirect utility function.

Next I want to deal with the issue of diversity among families. Unexplained variation in hours of work is quite large even when an extensive set of explanatory variables is available. For example, even the best regression for hours of work has a residual standard error of about half the average level of hours. The casual development of a theory of the representative family, together with an unexplained additive random deviation, would not be suitable for the range of issues raised in the negative income tax experiment. Rather, the existence of diversity in the determinants of hours of work ought to be built into the theory from the start.

I propose a theory along the following lines. Suppose θ is an index of family attitudes about work and abilities to work. Everything else held constant, families with low values of θ choose to work more than those with high values. The distribution of θ among the population accounts for the apparent randomness in hours of work. In the absence of the negative income tax, there will be a critical value of θ, say θ_N, such that families with θ exceeding θ_N will not work at all. The probability that a family drawn at random from the population will work is just the fraction having θ below θ_N. Note that θ_N depends on the observable characteristics of the family, especially its wage and nonlabor income. Similarly, among families participating in the program there is a threshold, θ_P, beyond which they will not work. Because I assume that labor supply is inversely related to θ, it is evident that θ_P is less than θ_N—some families would choose to work if the program were not available but would choose not to work if they participated. No family would do the opposite; working under the program but not working otherwise.

There is another threshold, say θ^*, dividing families who opt in favor of the program from those who do not. Again, from the assumption that families with low values of θ favor work, those whose θ is less than θ^* will remain off the program, whereas those above θ^* will take advantage of it. A negative income tax is inherently more attractive to families that are less willing or less able to work. In terms of an indirect utility function now considered as a function of θ as well as w and Y, say $g(w, Y, \theta)$, the threshold θ^* is defined succinctly by:

$$g(w, Y_0, \theta^*) = g[(1 - \tau)w, B_0 + (1 - \tau)Y_0, \theta^*].$$

Under the assumption that the program is always attractive to nonworkers, θ^* must be less than or equal to θ_N. The three groups I distinguished earlier now can be restated as:

Group 1, $\theta > \theta_N$,

Group 2, $\theta_P < \theta < \theta_N$, and

Group 3, $\theta < \theta_P$.

The fraction of the population falling in each of the three groups is the fraction having values of θ in each of the corresponding intervals. Two distinct cases need to be considered in deriving the response of the three groups to the negative income tax. Case 1, in which θ^* is less than θ_P (some participants work), and case 2, in which θ^* is greater than θ_P (no participants work). The responses then are:

	Case 1 $\theta^* < \theta_P$	*Case 2* $\theta_P < \theta^* < \theta_N$
Group 1 ($\theta > \theta_N$)	All participants, none work	All participants, none work
Group 2 ($\theta_P < \theta < \theta_N$)	All participants, none work	Those above θ^* participate and do not work; those below θ^* work and do not participate
Group 3 ($\theta < \theta_P$)	All work; those above θ^* participate; those below do not	All work; none participate

Labor supply functions can be derived immediately from this matrix: expected hours of work in absence of program equals [fraction of population in group 2] · [average hours in group 2] + [fraction of population in group 3] · [average hours in group 3]; expected hours of work when program is available, in case 1 ($\theta^* < \theta_P$), equal [fraction of population in group 3 with $\theta < \theta^*$] · [average hours of these nonparticipants] + [fraction of population in group 3 with $\theta > \theta^*$] · [average hours of these participants], and, in case 2 ($\theta_P < \theta^* < \theta_N$), equal [fraction of population in group 2 with θ less than θ^*] · [average hours of these] + [fraction of population in group 3] · [average hours in group 3].

The difference between these is the expected effect of the program: expected reduction in hours of work, in case 1, equals [fraction of population in group 2] · [average hours in group 2 when the program is not available] + [fraction of population in group 3 with $\theta > \theta^*$] · [average reduction in hours of these participants]; and, in case 2, equals [fraction of population in group 2 with θ above θ^*] · [average hours of these when program is not available].

In case 1, the program causes some families to withdraw from the labor market (those in group 2) and others to reduce their hours (part of group 3). In case 2, on the other hand, withdrawal is the only effect. The magnitude of either effect depends on the distribution of the population among the groups as well as on the response within each group. I emphasize that no simple application of conventional labor supply equations can answer the questions of the expected impact of a negative income tax.

Within this general framework, a number of steps are required to create an operational model:

—First, specify decisions in terms of an indirect utility function $g(w, Y, \theta)$.

Identify θ as a parameter of attitudes and tastes. Be certain that $g(w, Y, \theta)$ correctly incorporates the constraint that hours of work cannot be negative.

—Second, derive the labor supply functions $L_N(w, Y, \theta)$ and $L_P[(1-\tau)w, B_0 + (1-\tau)Y, \theta]$ conditional on not participating and on participating, respectively. Derive the thresholds θ_N and θ_P from them.

—Third, derive the participation threshold θ^* from:

$$g(w, Y, \theta^*) = g[(1-\tau)w, B_0 + (1-\tau)Y_0, \theta^*].$$

—Fourth, specify the distribution $f(\theta)$ of θ within the population.

—Fifth, calculate expected labor supply in the absence of the program as:

$$E(H_N) = \int_0^{\theta_N} L_N(w, Y, \theta) f(\theta)\, d\theta.$$

—Sixth, calculate expected labor supply in the presence of the program as:

$$\text{(Case 1)}\quad E(H_P) = \int_0^{\theta^*} L_N(w, Y_0, \theta) f(\theta)\, d\theta + \int_{\theta^*}^{\theta_P} L_P[(1-\tau)w, B_0 + (1-\tau)Y_0, \theta] f(\theta)\, d\theta.$$

$$\text{(Case 2)}\quad E(H_P) = \int_0^{\theta^*} L_N(w, Y_0, \theta) f(\theta)\, d\theta.$$

—Seventh, calculate expected impact of the program as $E(H_N) - E(H_P)$.

First Application

My first attempt to create a model was unsuccessful, but its failure was instructive, so I will describe it briefly. The model starts from the assumption that both the income and wage elasticities of the demand function for nonwork are unity. That is, if $\bar{H}$ is the total amount of time to be divided between work (H) and nonwork $(\bar{H} - H)$, then:

$$\bar{H} - H = \theta \frac{R}{w},$$

where R is total family resources (what Gary Becker calls full income)[16]

16. Gary S. Becker, "A Theory of the Allocation of Time," *Economic Journal*, vol. 75 (September 1965), pp. 493–517.

and θ is the fraction of total resources allocated to nonwork. In this model, variations in this budget share account for the dispersion in hours of work among an apparently homogeneous group of families. Now total family resources are the sum of the total value of its time, $w\bar{H}$, and its nonlabor income Y_0:

$$R = w\bar{H} + Y_0.$$

Thus its labor supply function is:

$$H = \bar{H} - \theta \frac{Y_0 + w\bar{H}}{w}$$

if positive; otherwise:

$$H = 0.$$

It is easy to derive the thresholds θ_N and θ_P:

$$\theta_N = \frac{w\bar{H}}{Y_0 + w\bar{H}} \text{ and}$$

$$\theta_P = \frac{(1-\tau)w\bar{H}}{B_0 + (1-\tau)(Y_0 + w\bar{H})}.$$

Note that both lie between 0 and 1 and that θ_N exceeds θ_P, as required. The indirect utility function is:

$$g(w, Y_0, \theta) = \theta^\theta(1-\theta)^{1-\theta}\frac{Y_0 + w\bar{H}}{w\theta} \quad \text{if} \quad \frac{Y_0}{w} \leqq \frac{1-\theta}{\theta}\bar{H}$$

$$= \bar{H}^\theta Y_0^{1-\theta} \quad \text{if} \quad \frac{Y_0}{w} \geqq \frac{1-\theta}{\theta}\bar{H}.$$

The kink is caused by the constraint of nonnegativity. Satisfaction does not depend on the wage rate for families with values of θ sufficiently large that they do not work at all. In case 1:

$$\theta^* = \frac{\log\left[\dfrac{Y_0 + w\bar{H}}{B_0 + (1-\tau)(w\bar{H} + Y_0)}\right]}{-\log(1-\tau)}, \text{ if positive, and 0 if not.}$$

This threshold behaves in the expected way, rising with w and Y_0 (the plan is less attractive to families with more resources of their own), rising with τ (higher tax rates are less attractive) and falling with B_0 (higher guarantees are more attractive). For high enough B_0, θ^* would become negative by the first formula. This occurs when:

$$Y_0 + w\bar{H} \leqq B_0 + (1 - \tau)(w\bar{H} + Y_0)$$

or

$$B_0 \geqq \tau(w\bar{H} + Y_0).$$

That is, the lump sum B_0 exceeds the amount of tax paid even for a family that devotes all of its time to market work. In this case all families will opt for the program. Taking $\theta^* = 0$ in this case will account for it correctly.

For this model, case 2 is irrelevant for all reasonable values of the parameters, so I omit its consideration.

The next step is to find a distribution that will account for the dispersion observed in hours of work. The beta density, $f(\theta) = 6\theta(1 - \theta)$, does well in fitting the observed standard error of hours, although it does not account for the tendency for hours to bunch around forty a week. I doubt that this is important here. Note that the model does predict that a positive fraction of the sample—those with θ below the appropriate threshold—will have zero hours of work. In this respect, the model is in the Tobit family of probability models.[17]

Putting all these assumptions into the recipe at the end of the previous section, I get:

$$E(H_N) = 6\int_0^{\theta_N}\left(\bar{H} - \theta\frac{Y_0 + w\bar{H}}{w}\right)\theta(1 - \theta)\,d\theta$$

and

$$E(H_P) = 6\int_0^{\theta^*}\left(\bar{H} - \theta\frac{Y_0 + w\bar{H}}{w}\right)\theta(1 - \theta)\,d\theta$$

$$+ 6\int_{\theta^*}^{\theta_P}\left(\bar{H} - \theta\frac{B_0 + (1 + \tau)(Y_0 + w\bar{H})}{(1 - \tau)w}\right)\theta(1 - \theta)\,d\theta.$$

It is easy but not illuminating to carry out these integrations. I will pass directly to the results shown in Table 3. The first three columns compare families facing wages of \$1.50, \$2.00, and \$3.00 an hour, holding constant the other determinants of labor supply. The first row of calculated responses shows the expected or average labor supply in the absence of the negative income tax. Both its level and its response to the wage rate are in

17. James Tobin, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, vol. 26 (January 1958), pp. 24–36.

Table 3. Response to a Negative Income Tax, Model 1: Effect on Hours of Work of Variations in Labor Supply Determinants

Item	Combinations of parameters							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labor supply determinants								
Wage (dollars per hour) (w)	1.50	2.00	3.00	2.00	2.00	2.00	2.00	2.00
Nonlabor income (dollars per week) (Y_0)	10	10	10	50	10	10	10	10
Guarantee (dollars per week) (B_0)	80	80	80	80	40	120	80	80
Tax rate (percent) (τ)	50	50	50	50	50	50	30	70
Attitude thresholds								
Switch on to program (θ^*)	0	0	0.25	0.15	0.42	0	0	0.19
No work, participants (θ_P)	0.40	0.47	0.57	0.42	0.63	0.38	0.55	0.35
No work, nonparticipants (θ_N)	0.92	0.94	0.96	0.75	0.94	0.94	0.94	0.94
Calculated responses								
Expected hours, no program [$E(H_N)$]	34.2	35.0	35.8	26.4	35.0	35.0	35.0	35.0
Expected hours, with program [$E(H_P)$]	9.5	12.6	18.8	10.8	24.1	8.6	16.3	9.1
Effect of program (hours) [$E(H_N) - E(H_P)$]	24.7	22.4	17.0	15.6	10.9	26.4	18.7	25.9
Percent of families opting for program	100	100	84	94	63	100	100	91
Percent of families continuing to work[a]	35	45	45	31	31	32	57	19
Average reduction in hours of those continuing to work	27.0	23.6	22.2	7.5	6.4	28.8	19.3	36.5
Percent of families stopping work[a]	63	54	39	47	31	67	42	71
Average reduction in hours of those stopping work	24.1	21.9	18.3	17.3	4.5	25.8	18.4	26.7

Source: Model 1, discussed in the text.
a. Percent of families who continue to work plus percent of families stopping work is less than 100 because some families do not work in the absence of the negative income tax.

accord with evidence from empirical studies of hours of work of groups who are not eligible for income supplements. Even though the model embodies a rather wage-elastic demand function for nonwork, the derived wage elasticity of average labor supply is very low. Further, the derived standard error of hours of work, eighteen hours a week, is close to that found in the data from the experiment and in other studies of weekly hours. Judged by its prediction of average hours, the standard deviation of hours and the response of average hours to wages model 1 seems quite satisfactory.

But the third row of calculated responses shows that the model greatly overstates the response to the negative income tax. In the benchmark case—w = \$2.00, Y_0 = \$10.00, B_0 = \$80.00, and τ = 50 percent—all families opt in favor of the program. A little under half of them (45 percent) work but reduce their hours by an average of 23.6 relative to hours of work in the absence of the program. The rest (54 percent) would work an average of 21.9 hours in the absence of the program but do not work at all in its presence. The average reduction is 22.4 hours a week, far in excess of anything suggested by the results of the experiment for a plan with these parameters.

The responses to changes in nonlabor income shown in column 4 give an indication of the trouble. Families with \$40 more a week in nonlabor income reduce their labor supply from 35.0 to 26.4 hours a week. Model 1 clearly has an excessive response of hours of work to increases in income. A glance at the labor supply function, $H = \bar{H} - \theta(Y_0 + w\bar{H})/w$, shows why. Families with large values of θ not only supply smaller amounts of labor to the market but also are more responsive to changes in Y_0. These families reduce their supply by a large amount under the negative income tax. But the results of the experiment are inconsistent with large responses in even a small minority of the population. Model 1 cannot be made to give correct predictions by reducing the dispersion of θ, because its predicted dispersion in hours of work would then be too low. It is clear that model 1 is unsatisfactory because its explanation of diversity of hours of work is incorrect. Families that supply few hours cannot be as income elastic as model 1 assumes.

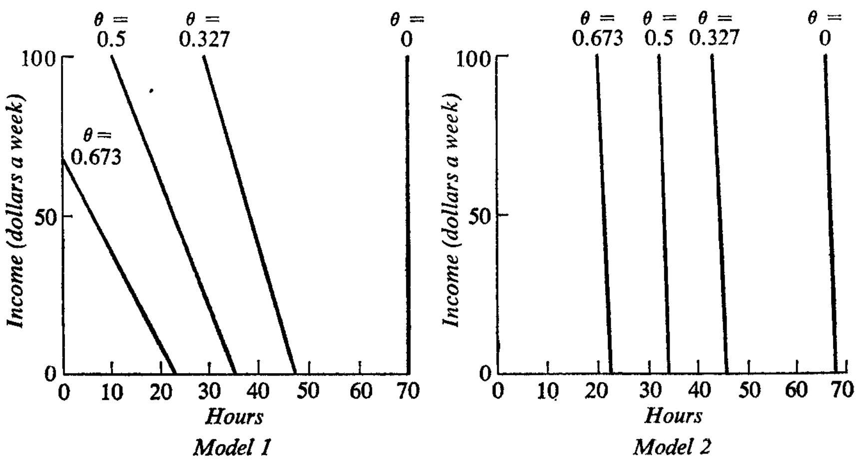
Second Application

Model 2 overcomes this problem by explaining diversity in hours of work through differences in the maximum number of hours available for work. In common with model 1, it assumes unitary income and wage elasticities of the underlying demand function for nonwork. Its labor supply function for a family of type θ is:

$$H = (1 - \theta)\bar{H} - \alpha \frac{Y_0 + w(1 - \theta)\bar{H}}{w},$$

where $(1 - \theta)\bar{H}$ is the amount of time that the family has available to allocate between work on the market and work and other activities at home. Figure 1 shows the difference between the two models.

Figure 1. Labor Supply as a Function of Nonwage Income in Models 1 and 2



Note: $\theta = 0.327$ is the first quartile of the distribution of θ; $\theta = 0.5$ is the median; and $\theta = 0.673$ is the third quartile.

The points of zero labor supply for model 2 are:

$$\theta_N = 1 - \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{Y_0}{w\bar{H}}\right)$$

and

$$\theta_P = 1 - \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{B_0 + (1-\tau)Y_0}{(1-\tau)w\bar{H}}\right).$$

The crossover point dividing families who opt for the program from those remaining off it is:

$$\theta^* = 1 - \left(\frac{1}{w\bar{H}}\right)\left(\frac{B_0}{(1-\tau)^\alpha - (1-\tau)} - Y_0\right).$$

In realistic cases, this formula needs no modification on account of nonnegativity. Table 4 reports the response of hours of work to variations in wages and income and in the guarantee and tax rates of the negative income tax. These responses are much closer to those found in the experiment, even though models 1 and 2 share the same unitary wage and income elasticities of demand for time at home. As in model 1, the ordinary labor supply function, $E(H_N)$, is not very wage elastic. The average family raises its hours of work by only 0.1 a week as the wage rises from \$1.50 to \$3.00. On the other hand, neither is model 2 very income elastic: as income rises

Table 4. Response to a Negative Income Tax, Model 2: Effect on Hours of Work of Variations in Labor Supply Determinants

Item	Combinations of parameters							
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Labor supply determinants								
Wage (dollars per hour) (w)	1.50	2.00	3.00	2.00	2.00	2.00	2.00	2.00
Nonlabor income (dollars per week) (Y_0)	10	10	10	50	10	10	10	10
Guarantee (dollars per week) (B_0)	80	80	80	80	40	120	80	80
Tax rate (percent) (τ)	50	50	50	50	50	50	30	70
Attitude thresholds								
Switch on to program (θ^*)	0	0	0.25	0.17	0.48	0	0	0.21
No work, participants (θ_P)	0.95	0.96	0.97	0.95	0.98	0.94	0.97	0.94
No work, nonparticipants (θ_N)	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
Calculated responses								
Expected hours, no program [$E(H_N)$]	33.7	33.8	33.8	33.2	33.8	33.8	33.8	33.8
Expected hours, with program [$E(H_P)$]	30.6	31.4	32.5	31.0	33.2	30.2	32.1	30.3
Effect of program (hours) [$E(H_N) - E(H_P)$]	3.1	2.4	1.3	2.2	0.6	3.6	1.7	3.5
Percent of families opting for program	100	100	84	93	54	100	100	89
Percent of families continuing to work	99	100	84	92	54	99	100	87
Average reduction in hours of those continuing to work	3.2	2.4	1.6	2.4	1.2	3.6	1.7	4.0
Percent of families stopping work	1	0	0	1	0	1	0	1
Average reduction in hours of those stopping work	2.1	1.6	1.0	1.4	0.8	2.4	1.1	2.6

Source: Model 2, discussed in the text.

from \$10 to \$50 a week, average hours in the absence of the negative income tax fall by only 0.6, in contrast to 8.6 hours in model 1. The weaker effect of income on labor supply in model 2 makes it predict much smaller responses to the negative income tax. In the benchmark case—w = \$2.00, Y_0 = \$10, B_0 = \$80, and τ = 50 percent—average hours fall by 2.4, from 33.8 to 31.4, when the plan becomes available. All families opt in favor of the plan, and none of them stop work. In the high-wage group, the average effect is smaller, only 1.3 hours, partly because 16 percent of the families opt against the program and do not reduce their hours at all, and partly because the response of those under the negative income tax is smaller. The difference in the responses of high- and low-income families is similar.

Model 2 predicts a particularly small response, 0.6 hour on the average, to a plan with a guarantee level of only $40 a week. Again, part of the population decides against the plan (46 percent), and the remainder reduce their hours by a relatively small amount (1.2 hours). On the other hand, the plan with a high guarantee achieves universal acceptance and even drives 1 percent of the families out of the labor market altogether. Finally, the last two columns of Table 4 show how the response to the negative income tax depends on the tax rate. The low tax rate, 30 percent, corresponds to a more generous plan yet causes a reduction of 1.7 hours a week, compared with 2.4 hours for the benchmark rate of 50 percent. The high tax rate of 70 percent causes 9 percent of the families to opt against the program, so that although those on the program reduce their hours of work by 4.0 hours a week, average hours in the population fall by only 3.5 when the plan becomes available. The unattractiveness of high tax rates tends to counterbalance their discentive effects on labor supply.

Both model 1 and model 2 would fit fairly well the earlier cross-sectional data on the mean and dispersion of hours of work. Among low-wage families, the defective income elasticity of model 1 would not cause trouble because so few of the families have significant nonwage income. The experiment has provided the first opportunity to observe the response of hours of work to large amounts of income. The small size of the response requires the use of a theoretical model of labor supply that does not attribute large income responses even to a minority of families. Model 2 is such a model and demonstrates that the findings of the experiment are not contradictory to the economic theory of the household.

Appendix B: Derivation of Variances and Covariances of Means and Variances of Estimators

The first step is to derive the variances and covariances of the four basic statistics implied by the variance-components model. I define:

$\lambda = \dfrac{\delta^2}{\delta^2 + \sigma^2}$, the fraction of the total variance contributed by the permanent component;

$N_S = 225$, the number of experimental subjects reporting valid data;

$N_C = 200$, the number of controls reporting valid data;

$T_B = 1$, the number of observations before the experiment; and
$T_D = 12$, the number of observations during the experiment.

Then it is not hard to show, first:

$$\mathrm{var}(H_{S,B}) = \frac{\delta^2 + \sigma^2}{N_S}\left(\lambda + \frac{1 - \lambda}{T_B}\right);$$

and, second:

$$\mathrm{cov}(H_{S,B}, H_{S,D}) = \frac{\delta^2 + \sigma^2}{N_S}\lambda.$$

That is, the statistics $H_{S,B}$ and $H_{S,D}$ are positively correlated because they share the common variation arising from permanent unobserved differences among families. As λ, the fraction arising from the permanent component, becomes larger, so does the correlation. A well-chosen estimator will take advantage of this positive correlation to reduce the variance of the estimated effect of the experiment.

Similar calculations show that:

$$\mathrm{var}(H_{S,D}) = \frac{\delta^2 + \sigma^2}{N_S}\left(\lambda + \frac{1 - \lambda}{T_D}\right), \text{ and}$$

$$\mathrm{var}(H_{C,B}) = \frac{\delta^2 + \sigma^2}{N_C}\left(\lambda + \frac{1 - \lambda}{T_B}\right);$$

$$\mathrm{cov}(H_{C,B}, H_{C,D}) = \frac{\delta^2 + \sigma^2}{N_C}\lambda, \text{ and}$$

$$\mathrm{var}(H_{C,D}) = \frac{\delta^2 + \sigma^2}{N_C}\left(\lambda + \frac{1 - \lambda}{T_D}\right).$$

From this information about the stochastic relation among the means, I can calculate the variances of the various estimators:

$$\begin{aligned}\mathrm{var}(b_{DMS}) &= \mathrm{var}(H_{S,D}) + \mathrm{var}(H_{S,B}) - 2\,\mathrm{cov}(H_{S,D}, H_{S,B}) \\ &= \frac{\delta^2 + \sigma^2}{N_S}\left(\lambda + \frac{1 - \lambda}{T_D} + \lambda + \frac{1 - \lambda}{T_B} - 2\lambda\right) \\ &= \frac{\delta^2 + \sigma^2}{N_S}\left((1 - \lambda)\frac{1}{T_D} + \frac{1}{T_B}\right).\end{aligned}$$

Note the cancellation of the two λ terms from the variances by the λ from the covariance. The variance of b_{DMS} arises solely from the variances of the transitory components, which contribute the terms $(1 - \lambda)/T_B$ and $(1 - \lambda)/T_D$ to the variances. These terms are not canceled by the covariance.

If both T_B and T_D are large, the variance of b_{DMS} can be made very small even in an experiment with a limited number of subjects.

The estimator b_{DMD} fares less well with this stochastic specification because it fails to cancel the variance from permanent components through the use of data from before the experiment:

$$\begin{aligned}\mathrm{var}(b_{DMD}) &= \mathrm{var}(H_{S,D}) + \mathrm{var}(H_{C,D}) \\ &= (\delta^2 + \sigma^2)\left(\frac{1}{N_S} + \frac{1}{N_C}\right)\left(\lambda + \frac{1-\lambda}{T_D}\right).\end{aligned}$$

Similar calculations apply to the compromise estimator b_{EW}, but the resulting formula is too complicated to justify discussing it here.

None of the three estimators discussed so far is optimal in the sense of minimizing variance. The general class of linear unbiased estimators can be written as:

$$b = H_{S,D} - \omega H_{S,B} - (1-\omega)[(1-\gamma)H_{C,D} + \gamma H_{C,B}].$$

The parameter ω represents the major decision about the weight for the preexperiment data for the subjects against the weight for the controls, both before and during the experiment. It will turn out that the more important the permanent component of the difference among families, the larger will be ω. The parameter γ represents the subsidiary decision of how to distribute the total weight assigned to the controls between the data for before and during the experiment. For a given fixed choice of ω and γ, the variance of b is:

$$\begin{aligned}\mathrm{var}(b) = {} & \mathrm{var}(H_{S,D}) + \omega^2\,\mathrm{var}(H_{S,B}) - 2\omega\,\mathrm{cov}(H_{S,D}, H_{S,B}) \\ & + (1-\omega)^2[(1-\gamma)^2\,\mathrm{var}(H_{C,D}) + \gamma^2\,\mathrm{var}(H_{C,B}) \\ & + 2\gamma(1-\gamma)\,\mathrm{cov}(H_{C,D}, H_{C,B})].\end{aligned}$$

It is not difficult to show that the optimal weight to assign to the data for the controls before the experiment is: $\gamma = T_B/(T_D + T_B)$. Given optimal weighting of the controls, some additional algebra shows that the optimal weight for the data for the subjects before the experiment is:

$$\omega = \frac{\dfrac{\lambda}{N_S} + \dfrac{1}{N_C}\left(\lambda + \dfrac{1-\lambda}{T_B + T_D}\right)}{\dfrac{1}{N_S}\left(\lambda + \dfrac{1-\lambda}{T_B}\right) + \dfrac{1}{N_C}\left(\lambda + \dfrac{1-\lambda}{T_B + T_D}\right)},$$

if $\lambda = 1$, $\omega = 1$, and the best estimator is the *DMS* estimator. High values of λ are associated with high ω weights for the preexperiment data for subjects. As long as λ is positive, ω also approaches unity as T_B, the number of observations before the experiment began, becomes large. On the other hand, if there are no permanent differences ($\lambda = 0$) and all of the samples are of equal size ($N_S = N_C$ and $T_B = T_D$), then $\omega = 1/3$, $\gamma = 1/2$, and $b = H_{S,D} - 1/3[H_{S,B} + H_{C,D} + H_{C,B}]$, the *EW* estimator discussed earlier. Finally, if $\lambda = 0$ and $T_B = 1$, ω approaches zero as T_D becomes large: in the absence of a permanent error component, the *DMD* estimator is close to minimal variance where the number of periods during the experiment is large relative to the number before the experiment. If, however, λ is positive, the efficient estimator gives weight to the preliminary data $H_{S,B}$ no matter how large T_D may be. Data on the experimental subjects before the experiment are extremely useful in reducing the variance of the estimated effect of the program within the experiment.

Comment by Zvi Griliches

Hall's paper outlines a theoretical model with individual differences in behavior built into it; recomputes the statistical significance of the mean effect of the experiment; and comments on the desirable design of such experiments. In my comments, I shall try to follow this outline.

The Theoretical Model

There are two reasons for worrying about individual behavioral differences in such data. First, and most important, they dominate the data. Despite a large number of variables included in the analysis, at the individual level the fits are poor and the bulk of the observed variance remains unexplained. Tacking on a routine additive disturbance to the estimated equation does not do justice to the problem. Second, and quite important for the econometric analysis of the results, a model is needed that would explain participation rates in the negative income tax experiment. One cannot simply regress labor supply behavior on the guaranteed income levels and tax rates because for a significant fraction of the subjects the tax rate is not binding if they are above the breakeven point. Moreover, this switching point is endogenous and has to be built into the model explicitly—unless the experiment had been so designed that participation rates were always 100 percent.

Hall presents two versions of a simple model that tries to capture the phenomenon of variations across individuals. In the first, the random variable is directly the relative preference for leisure. In the second, which gives a bit more realistic results, the individual variability is introduced through the variability of available or uncommitted hours. It is assumed that the week is shorter for some people than others, that people have different amounts of available hours to allocate between leisure and work.

The econometrics of the second model would be relatively simple if the effective tax rate were not endogenous to the experiment. The labor supply equation can be written as $H = A_i - \alpha y_0/w$, where H is hours worked, y_0 is nonlabor income, w is the after-tax wage rate, α is a parameter fixed across the whole population, and A_i is an unobservable but estimable individual parameter connected to the distribution parameter as follows: $A_i = (1 - \alpha)(1 - \theta_i)\bar{H}$, where $\bar{H}$ is the ceiling value of total available hours. Econometric problems arise because the actual effective tax rate is endogenous and because H cannot take on negative values, converting this into a Tobit type problem. The latter point is discussed ably and in some detail by Hall.

I have two points to make about this model, one constructive and one critical.

First, this model is reminiscent of the ones actually estimated by Watts and his associates, who also allow, in effect, separate constant terms for each individual. But Hall's model is more powerful and hence also easier to refute because it predicts nonparticipation rates, making them conditional on the true A_i (or θ_i). Because such an equation could be estimated for the entire working sample, one could get estimates of α and the individual A_i. One could then insert these estimated values into the participation formula and see whether in fact they explain who does or does not participate in the program. Alternatively, one could estimate the hours equation and participation equation jointly, since the same parameters appear in both.

Second, I am not certain that I like the assumed form of the labor supply function for a given θ_i. Note that in both models the response of hours worked to wages depends on the availability of nonlabor income. Zero response to the wage rate occurs in both models 1 and 2 if nonlabor income is zero, irrespective of the level of earnings. Similarly, the elasticities of hours worked with respect to nonlabor income and with respect to the wage rate are opposite in sign but equal, and their absolute magnitude depends on the ratio of nonlabor income to labor earnings, implying lower elasticities as earnings increase and as nonlabor income decreases. The first may be reasonable if an asymptote in hours worked is approached, but in the second, the strong dependence of the labor supply elasticity on the level of nonlabor income seems too specific and may lead to perverse results. (Nonlabor income is far from a homogeneous entity and may be related to the θ_is.)

In the next section Hall shows, in spite of the caveats in the text, that the cross-sectional surveys do in fact predict an average response for men on the order of magnitude observed in the experiment, supporting both the earlier estimates and the experiment. On the other hand, there is a substantial discrepancy for women, which in fact is predicted by the model outlined by Hall.

Statistical Significance of the Experiment

The third section of Hall's text ingeniously recomputes the mean experimental response and the associated standard errors and shows that in fact there was a statistically significant average response in the experiment. I have several comments and queries about it.

First, should not an attempt be made to reconcile these results with those of Watts and others to determine the sources of the difference?

Second, given the rather rich econometric theory of the earlier sections, I am a bit disappointed that it is not carried over into the estimation of the mean effect. In particular, I would have liked to see an estimate of the effect of the experiment on the participators alone: that is, those families that actually have been affected by the experiment and are taking money from it, and excluding those who found themselves above the breakeven point. To do so, however, would have required the estimation of the complete model outlined by Hall earlier, since participation itself is endogenous to the model. My guess is that Hall's mean effect estimates—as well as those of Watts—are underestimates of the "effective effects."

Third, the estimated effects are not so small after all. For white males a reduction of 2.6 hours a week represents a 7 percent reduction in total labor supply. For participators, for those who actually were affected by the experiment, the effect probably is double that figure—which is not insignificant.

Fourth, I am not so certain what this means, however, because the actual amount of money transferred to the subjects was rather small. If I read the numbers correctly in Rees and Watts, of the experimental families whom they analyzed, about half got less than $15 a week, and fewer than a hundred families received an average payment of as much as $45 a week in payments. That is not very generous, and it may not be so surprising that larger effects were not observed.

Design of the Experiment

The problem of attrition in the experiment clearly is serious and requires further study. Hall suggests that more data should have been collected before the onset of the experiment. Let me expand this to suggest that more data should have been collected after the experiment. From the point of view of Hall's model, whether the additional observations are from $t - h$ or $t + h$ does not really matter. It would have been relatively cheap, I think, to have kept interviewing the same people for another year after the completion of the actual experiment, and perhaps we should return and resurvey them again.

Let me end with a few scattered comments. More analyses of these data certainly are needed. They should be analyzed in relation to actual benefits received, not merely plans under which enrolled. Something more than simply averaging could be done with the time structure of the model. Wages, hours, and husband and wife labor force behavior should be analyzed jointly, not one at a time. In retrospect, it would be desirable, I think, to have a postmortem analysis of the original design of the experiment. What were the particular places at which it went wrong?

Finally, we should thank the designers and executors of the experiment for having gone through with it and shown us that it can be done—that behavior can be studied in this way. With all its imperfections, I think that the experiment is something we can all be proud of. Next time we might do it better, but there might not have been a next time and we might not have known how to do it better without the care and effort invested in this endeavor.

Comment by Jacob Mincer

It may be nice to sit on top of a pyramid, but the view of the ground below is rather distant, and it might actually be difficult to get down at all. This is the feeling I get in attempting to survey the findings of the experiment through the successive layers of past progress reports, the thirty papers that make up the final report, and the papers prepared for this conference. Although I read only a small fraction of all this material, I take the liberty to comment on the results of the experiment, as well as to remark on Robert Hall's excellent paper.

Design and Data Problems

There is some danger that the analyses of the experiment's findings might be colored by attitudes toward the need for welfare reform and toward the desirability of experimental research in the social sciences, including economics. Let me confess that my attitudes about these matters are quite positive, although I insist that they do not at all inhibit my critical comments. It is much easier, of course, for an onlooker to be detached than for those present at the creation. So it is possible that the perceived political urgency of the experimental program led to a less than perfect design: too little time and attention to preexperimental observation, too short a duration of the experiment, and some problems with data quality. Perhaps the realization of some of these shortcomings did not require an actual experience with the experiment. The experimental design, according to Hall, could have been improved by a factor of 6 to 1 at no additional cost, if half of the sample were studied preexperimentally half the time, the other half post-experimentally half the time. This, of course, would have cut the effective duration of the experiment in half, but Hall considered only the statistical efficiency of face value estimators.

One troublesome problem is the inadequate quality of the earnings and wage-rate data. The interesting questions on the possible effect of the experiment on wage rates, which Rees enumerates, are largely unanswerable because of the differentially inaccurate understanding of the concept by the experimental and control groups. This loss of opportunity to gather much cleaner data on wage rates than the usual surveys provide is bothersome, because it will hamper possible reanalyses or other kinds of analyses of the data generated by the experiment.

The deterioration of data, mainly by attrition, appears to be so massive that it leads Hall to discard the findings on blacks and the Spanish speaking, reducing the sample size to one-third. Multiply this by the 6-to-1 estimate of loss in efficiency due to design and the staggering loss in potential information is more than 90 percent.

Undaunted by the more than decimated efficiency of the experiment, Hall argues that the experiment provided a good deal of new evidence about the labor supply response of poor families to a negative income tax. This is probably true, if what we wanted to learn is the short-run response of a particular part of the population to a temporary flow of cash grants. I say temporary cash grants, without adding the proviso "work-conditioned receipts," because no evidence was produced on effects of differential marginal tax rates.

It is unfortunate that 0 percent and 100 percent marginal tax plans were not studied. If the evidence is to be taken at face value, the experimental results are either silent on this important question or they do not disparage categorical 100 percent tax welfare systems. Aside from the nominal high tax feature (100 percent before 1967, 66⅔ percent thereafter), much of the opposition to AFDC (for which intact families are not eligible) is based on its tax on family integrity and its consequent plausible association with family disintegration. Regrettably, though perhaps inevitably, the experiment was not designed to study this problem. Yet the important labor supply effects probably are related to family dissolution, but they escape the observers who focus on the behavior of the deserted mothers.

Of course, the question of how well actual work behavior and actual family dissolution can be inferred from reported behavior is partly a question of law enforcement, for which the experiment was not designed and certainly not suitable.

Other Biases

Aside from questions of design, attrition, and varying quality of data, the basic problem in interpreting the experimental results is the short duration of the experiment. The purpose of the experiment, as defined by its managers, was to find out what would be the responses of the families of the working poor to a legislated, permanent negative income tax program, which would replace the current system. Do we get the answer to this question from the three-year experiment? Hall's answer is not explicitly positive, though perhaps implicitly, when he finds that after some adjustments made on his own nonexperimental cross-sectional estimates of labor supply parameters, the latter become consistent with experimental effects for men, at least within sampling error. It is not clear whether the same conclusion would obtain if Hall's recipe for adjustments were used on other available nonexperimental estimates. Although I would agree that these are not of uniform quality and I do not object to Hall's healthy respect for his own work, this exercise is not likely to serve as a final answer.

If consistency with nonexperimental results were the measure of validity of the experimental results, there would be no need for experiments. Short of this happy state, we must judge the validity of the experimental results by a number of indications emerging from the experiment. Among these there are suggestions of perceptions, misperceptions, and selectivity biases on the part of the experimental subjects that militate against an unambiguous reading of the experimental results as a prognostication of the effects of a permanent program. I proceed to a discussion of doubts about the validity of such prognostication.

SELECTIVITY PROBLEMS. The managers of the experiment found it very difficult to locate intact working families who were poor by the standards of the experiment. Consequently, the selected families have the unusual composition of above-average numbers of children, below-average labor force attachment of secondary workers, and temporarily low incomes. The labor force attachment of male heads in such families is particularly strong, and the mothers have strong incentives to drop out of market work when they can afford it. The prior existence of an AFDC program for which intact families were not eligible—not to be confused with the AFDC-UP program instituted later—suggests an additional selection bias toward family stability and strong work motivation of the fathers in the families eligible for the experiment.

In addition, the truncation of low current family income three months before the experimental period predicts a subsequent regression toward the mean and consequently increases in income during the experiment. This bias is mitigated to the extent that it is shared by the controls, but the attrition and replacement of controls makes this escape doubtful. If the black and Spanish-speaking sample comes from a New Jersey population with higher permanent income than the whites in Scranton, and the major attrition was in the New Jersey controls, the differences in results are better understood as statistical artifacts than as ethnic mystique.

SHORT SPAN OF EXPERIMENT. "Nothing is as permanent as the temporary," a French proverb, may well apply to politics but not to the consumption function—and even less so, I would judge, to labor supply functions. Cain notes that male heads of families "have a firm attachment to a job that requires them to work a fixed number of hours" a week on a full-time basis, and to expect them to make major changes in their schedules, knowing that at the end of three years they will have to return to their old schedules, is not realistic. By contrast, the options of women to work intermittently and less than full time permit wives "to show a greater responsiveness to the experimental incentives."[18]

The response of husbands to a permanent program probably is understated—but by how much? It is the assessment of magnitudes, not just direction of responses, that was the purpose of the experiment. The response of wives may be overstated, though this is not clear, given the small number of working wives and the truncation bias that may have prevailed despite the controls.

It is rather clear from the consumption data presented by Metcalf that the households were acutely aware of the transitoriness of the grants they were receiving. The marginal propensity to consume food out of the experimental payments was only one-half to one-third as large as the marginal propensity to consume out of normal income. The experimental payments were viewed the same way as transitory components of income, whereas public assistance subsidies were treated as permanent. The same and perhaps an even stronger conclusion emerges from the data on purchases of appliances. Their timing was rather quick in response to the payments and more sensitive to the subsidies than to other variable components of income. This is to say that the windfall nature of the experimental payments was clearer than that of the other sources of income.

INTRUSION OF STATE WELFARE PROGRAMS. The introduction of an AFDC-UP program after the start of the experiment obscured the comparisons between control and experimental families, probably reducing the differentials. It seems to me that the appearance of these state programs may have presented an opportunity to contrast effects of what are viewed as permanent programs with those of the temporary experimental program. It is shown by Garfinkel that the lower the guarantee and the higher the marginal tax rate, the greater the withdrawal of experimental families into state welfare.[19] And, of course, many more control than experimental families went on welfare. Moreover, families on welfare tended to work and earn substantially less than those not on welfare. According to Garfinkel's Table 3, heads of families who went on welfare for less than half the period—that is, six quarters or less—worked five hours a week less, and those on welfare for more than half the period worked nine hours less than those who did not go on welfare. Compare this with the experimental

18. Glen G. Cain and others, "The Labor-Supply Response of Married Women, Husbands Present," *Journal of Human Resources*, vol. 9 (Spring 1974), p. 203.

19. Garfinkel, "The Effects of Welfare," Table 3.

differential of a reduction of only one or two hours a week. But we are not told the work behavior of the same families before the AFDC-UP was instituted. The matter deserves closer attention.

PERCEPTION OF THE "RULES OF THE GAME." It is my understanding that after about a year in the rural experiment—and I recall a similar progress report on the urban experiment—25 percent of the families thought that their subsidies would be reduced if they quit or lost a job. I am not sure whether this response represented ignorance of the rules of the game on the part of participants or a conviction prevailing despite the overt rules that proper response to the experimental payments must be in the socially appropriate direction that the experiment set out to demonstrate. Whichever the interpretation, ignorance or attitude, the effects were to diminish the observed disincentive effects.

Some Conclusions

First, at best, we have learned the short-run responses of a particular population group to short-run negative income tax type programs. I find it impossible to extrapolate from these findings to predict general and long-term effects of a bona fide and rather sweeping institutional change. Second, urgent policy research is likely to be more vulnerable to haste, impatience, pressure, and Hawthorne effects than are basic research studies. And, unfortunately, policy research is more likely to be funded. Thus, purely scientific progress in experimental economics is likely to be slow unless grantors of funds recover their respect for basic research. I should add, however, that despite the potentially thin line that existed in the New Jersey project between a political demonstration project and a scientific experiment, the work appears to have been conducted with impeccable standards of scientific objectivity.

Where do we go from here? As far as reanalyses of the generated data are concerned, it would be useful to follow up the families if they can be located. Closer attention to age and to job mobility may shed more light on some of the questions that were posed in the experiment. To a person of sixty-two years, a three-year program is more permanent than it is to a person of thirty. The turnover observed in the group probably is not aimless, and if so, the brevity of the experiment is the more serious a problem.

I hope these comments concerning the interpretation of the outcome of the New Jersey experiment are not taken as an expression of doubt about the potential of experimental economics. I have no doubt about the promise of this tool, which is new in the social sciences but has proven its power in the traditional sciences, including agriculture and medicine. Of course, the feasibility and payoff must differ depending on the research question. It does seem to me that the problem of gauging long-run macrosocietal adjustments from necessarily limited samples and periods is inherently difficult. Despite these reservations, I believe the experiment is to be applauded as a milestone in the history of economic research.

An offprint from

Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment

Editors: JOSEPH A. PECHMAN
P. MICHAEL TIMPANE

1775 Massachusetts Avenue, N.W.,
Washington, D.C. 20036

The Brookings Institution
Washington, D.C.