The Work Response to a Guaranteed Income: A Survey of Experimental Evidence

*Gary Burtless**

Presidents and policy analysts are periodically seized with a passion to reform the nation's welfare system. This passion occasionally results in a serious proposal for thorough reform, such as President Nixon's Family Assistance Plan, President Carter's Program for Better Jobs and Income, or President Reagan's New Federalism. The only reform proposal that has received experimental scrutiny, however, is a suggestion advanced by academic economists—the negative income tax or guaranteed annual income plan. While popular among economists, the negative income tax proposal has never attracted much enthusiasm—or even attention—among politicians and voters. Fortunately, the findings from the negative income tax experiments are relevant to a wide variety of reform proposals, including the plans suggested by recent Presidents. Experimental results were used, in fact, to predict the behavioral consequences of both the Nixon and the Carter reform proposals. This essay summarizes the labor supply findings from the four negative income tax experiments and considers their implications for reforming the American welfare system.

It is useful at the start to distinguish among three different kinds of labor supply estimates that have been produced by the experiments. The first was obtained by measuring the simple difference between the work effort of people who were assigned to experimental negative income tax plans and that of people who were assigned to the control group. Those in the control group were not eligible to receive payments

*Senior Fellow, The Brookings Institution. David Betson, Robert Moffitt, and Philip Robins provided useful suggestions in the preparation of this paper. In addition to the two discussants, David Greenberg, Robert Haveman, Robert Reischauer, and Alice Rivlin gave helpful comments on an earlier version. The opinions are the author's own and should not be ascribed to the staff or trustees of The Brookings Institution.

and so were presumably unaffected by the experiment. The labor supply difference between these two groups is ordinarily expressed as a reduction in average hours of work per week or as a percentage change in comparison to the average hours worked by members of the control group.

A second type of estimate is produced by using structural models of work effort response. Structural models yield a decomposition of the overall work reduction into a change that is due to the net wage or tax change, on the one hand, and one that is due to the increase in family income, on the other. These two separate effects are referred to as the substitution and income effects, respectively. Economists usually prefer this type of measure of response for two reasons. It permits the results of the experiments to be directly compared with labor supply findings from nonexperimental studies. And it allows analysts to generalize the findings from the experiments to a much wider population than the one enrolled in the experiment and to estimate the effects of a broader range of plans than the ones actually tested in the experiment. The latter advantage is particularly important from the point of view of evaluating realistic reform proposals, since no plan suggested by congressmen or Presidents has borne much resemblance to the ones tested in the experiments.

The third kind of estimate of response is generated using microeconomic simulation. This type of estimate is simply a generalization of the results from an experiment to the national population. Using estimates of the income and substitution effects obtained in an experiment and a microeconomic census file representing all U.S. households, economists have predicted the response of low-income workers to alternative income maintenance tax plans and summed these responses to produce an estimate of the effect on national labor supply.

From a scientific viewpoint the most reliable estimate of work reduction is the simple difference in labor supply between members of the treatment and control groups. This is the measure of response that the experiments were specifically designed to produce and it is the one that has been most widely reported in the popular press. It is inherently more difficult to decompose the overall response into income and substitution effects, although in this respect the experiments possess substantial advantages over nonexperimental sources of data. The experimentally based simulations of national response are more problematical. National simulations are based on a specific (and perhaps erroneous) decomposition of the experimental response into income and substitution effects and on detailed assumptions about the responses of subpopulations that were unrepresented or poorly represented in the experiments.

Unfortunately, from the perspective of their policy usefulness, the

Table 1
Description of the Negative Income Tax Experiments

	Characteristics of the sample				Characteristics of the plans			
Experiment	Sample Size	Family Composition	Race	Income Truncation[a]	Duration	Range of Guarantee[a]	Range of Tax rates (percent)	Range of Breakeven[a]
New Jersey (1968-1972)	1,357	Husband-wife (100%)	White (32%) Black (37%) Hispanic (31%)	150	3 years	50 to 125	30 to 70	100 to 250
Rural (1969-1973)	809	Husband-wife (85%) Single female parent (15%)	White (65%) Black (35%)	150	3 years	50 to 100	30 to 70	100 to 250
Gary (1971-1974)	1,780	Husband-wife (41%) Single female parent (59%)	Black (100%)	None[b]	3 years	77 and 101	40 and 60	128 to 253
Seattle-Denver (1971-1982)	4,800	Husband-wife (61%) Single female parent (39%)	White (39%) Black (43%) Hispanic (18%)	325	3 years (71%) 5 years (25%) 20 years (4%)	92 to 135	50 and 70 $70-.0025Y^c$ $80-.0025Y^c$	140 to 300

[a]Measured as a percent of the poverty line. Breakeven is the income level at which the negative income tax payment is reduced to zero. Partial reimbursement of income and payroll taxes was phased out at higher income levels.

[b]The Gary sample was initially restricted to families with incomes below 240% of the poverty level, but a small sample with incomes above this limit was subsequently enrolled to minimize truncation bias.

[c]Declining marginal tax rate plans. Y is family income, implying that the marginal tax rate declined by 2.5 percentage points with every $1,000 increase in income.

Sources: Committee on Finance, U.S. Senate (1978), p. 316, Kehrer (1977), and Robins (1985).

three kinds of estimates of response would rank in the reverse order. The average difference in labor supply between treatment and control groups within a particular experiment may be suggestive, but it is not especially helpful for predicting the effect of a realistic welfare reform plan on a representative population. The most useful and meaningful estimates of response are ones that reflect the response of a nationally representative sample to a plausible program of reform. Unfortunately, such estimates are inherently the least reliable.

The remainder of this paper considers, in turn, the three kinds of estimates just described and their major limitations. The paper concludes with a discussion of the implications of the estimates for welfare policymaking.

Simple Estimates of Response

The negative income tax experiments produced a large number of estimates of average response to the tested plans. These estimates naturally vary across the four experiments, since the experiments tested different plans on different populations. Table 1 describes some of the main features of the samples and negative income tax plans tested in the experiments.

The samples varied tremendously in the different experiments. The first experiment, in New Jersey and Pennsylvania, enrolled low-income black, white, and Hispanic residents of declining urban areas. All of the enrolled families originally contained both husband and wife. The Rural experiment, which was conducted in Iowa and North Carolina, contained low-income rural families. Although most of the families contained both a husband and a wife, a small number of single-parent families were also enrolled. The two later experiments, in Gary, Indiana, and in Seattle and Denver, enrolled higher income samples drawn from low-income census tracts in large midwestern and western central cities. The samples in these experiments were purposefully drawn to represent single-parent as well as two-parent families. The Gary experiment was restricted to black families, while the Seattle-Denver experiment included large samples of white and Chicano, as well as black, families. Clearly, the differences in the samples are important enough so that significant differences might be expected in the average response even if each of the experiments had tested an identical set of plans.

The tested negative income tax plans were not identical, however. On average, the New Jersey, Rural, and Gary experiments tested less generous plans than the ones tried in Seattle and Denver. That is, the Seattle-Denver plans offered more generous payments to families without other income and provided payments to families at higher in-

Table 2
Changes in Hours and Earnings in Four Negative Income Tax Experiments
(Percentage changes in parentheses)

Experiment	Husbands		Wives		Single female heads of families	
	Hours per year	Annual earnings[a]	Hours per year	Annual earnings[a]	Hours per year	Annual earnings[a]
New Jersey						
White	−99 (−5.6)	+10 (+0.1)	−73 (−30.6)	−420 (−33.2)	—	—
Black	+36 (+2.3)	+1,180 (+9.3)	−5 (−2.2)	+110 (+7.8)	—	—
Hispanic	−10 (−0.7)	+800 (+6.4)	−99 (−55.4)	−560 (−54.7)	—	—
All[b]	−21 (−1.2)	+690 (+5.3)	−56 (−24.6)	−270 (−21.4)	—	—
Rural[c]						
White	+40 (+1.8)	−590 (−4.8)	−88 (−21.1)	−170 (−12.1)	—	—
Black	−152 (−8.0)	−630 (−6.8)	−268 (−31.3)	−1,360 (−41.6)	—	—
All	−56 (−2.8)	−610 (−5.7)	−178 (−27.9)	−770 (−32.8)	—	—
Gary						
Black	−114 (−6.5)	−830 (−5.0)	+14 (+5.0)	+160 (+10.5)	−112 (−30.0)	−280 (−13.9)
Seattle-Denver						
White	−144 (−7.6)	−1,310 (−7.5)	−107 (−17.1)	−590 (−16.5)	−85 (−8.6)	−900 (−13.9)
Black	−169 (−9.5)	−930 (−5.9)	−153 (−16.0)	−860 (−15.6)	−180 (−16.6)	−980 (−14.0)
Hispanic	−231 (−11.5)	−510 (−3.0)	−147 (−28.7)	−800 (−32.5)	−202 (−20.4)	−1,380 (−22.3)
All	−164 (−8.8)	−1,070 (−6.4)	−128 (−17.9)	−710 (−17.6)	−144 (−14.0)	−1,000 (−14.9)
3-year Sample	−133 (−7.1)	−810 (−4.8)	−101 (−14.2)	−580 (−14.4)	−134 (−13.0)	−940 (−14.1)
Weighted average[d]	−119 (−7.0)	−650 (−4.0)	−93 (−17.0)	−480 (−16.0)	−133 (−17.0)	−760 (−15.0)

[a]Annual earnings changes are measured in 1985 dollars. Earnings estimates reported in original reports were converted using the personal consumption expenditure deflator.

[b]Results for overall New Jersey response are obtained by weighting responses of separate racial groups. Racial weights are reported in Table 1.

[c]Results for Rural response are obtained by weighting of separately reported responses for white wage earners in Iowa (25% of sample), white wage earners in North Carolina (25%), and black wage earners in North Carolina (50%).

[d]Separate responses are weighted using reported estimation samples in four experiments. For husbands and wives, New Jersey = 0.20; Rural = 0.07; Gary = 0.17; Seattle-Denver = 0.56. For female heads, Gary = 0.34, and Seattle-Denver = 0.66.

Sources: New Jersey: Rees (1974), pp. 174-75; Rural: U.S. Department of Health, Education and Welfare (1976), pp. 23 and 29; Gary: Moffitt (1979), p. 482, and Greenberg, Moffitt, and Friedman (1981), p. 586; Seattle-Denver: SRI International (1983), pp. 120-22 (second experimental year results) and Robins and West (1980b), pp. 16, 19, 22, and 59-67.

come levels. Thus, other things equal, we would expect the Seattle-Denver plans to induce a larger response. Other things were not equal, however. I have already mentioned differences in the income distributions of the four samples. In addition, the nonexperimental welfare benefits available to members of the control group differed across the experiments. Local labor market conditions also differed. There is thus no reason to expect that the average response to the income maintenance plans would be identical across experiments.

Table 2 shows the average work effort and earnings reductions within various subsamples of the four experiments. The estimates are taken from the final reports of each of the experiments. Analysts essentially estimated a statistical model of the following type:

$$(1) \qquad Y = \alpha + \beta T + \gamma Z + \delta X + \epsilon,$$

where Y is the dependent variable of interest (either hours of work or earnings), T is a treatment dummy variable that takes the value one for people assigned to any of the negative income tax plans and zero for members of the control group, Z is a vector of variables originally used to stratify the sample in the experimental design (for example, pre-experimental income level), and X is a set of personal characteristics believed to affect the dependent variable (age, educational attainment, place of residence, and so forth). The treatment effect is β, and it captures the average effect of treatment on an average member of the sample assigned to negative income tax plans.[1]

Most but not all of the entries in table 2 are negative, implying that the negative income tax plans caused reductions in work effort and earnings for most subsamples enrolled in the experiments. All of the entries for the Seattle-Denver experiment are negative and, especially for men, are often larger than corresponding entries from the other three experiments. Virtually all of the Seattle-Denver estimates are significantly different from zero at the 95 percent confidence level, whereas estimates from the other experiments are frequently insignificant. There are two explanations for this pattern. As mentioned above, the average generosity of the Seattle-Denver plans was greater than that of the plans tested in the other experiments, causing a larger response, and the sample enrolled in Seattle-Denver was much larger, yielding a smaller standard error around the point estimate of response.

The bottom row in table 2 shows average hours and earnings reductions in all of the experiments for husbands and wives in two-parent families and female heads of single-parent families. Husbands reduced their reported work effort by approximately 7 percent, while wives and female heads reduced reported hours by 17 percent. The greater responsiveness of women than of men is consistent with the relative labor supply elasticities reported in nonexperimental studies. Analysts of the Gary and Seattle-Denver experiments concluded that most of the hours reduction was caused by shorter durations of employment and by longer durations of unemployment and labor force withdrawal among people enrolled in the negative income tax plans.[2] There was only a comparatively small effect on the weekly hours of those remaining at work.

On balance, the proportional reductions in earnings were quite close to the reductions in hours. Although the earnings reductions might appear to be relatively modest, they are sizable when compared with the negative income tax payment received by a typical family. In the Seattle-Denver experiment, for example, eligible two-parent families received transfer payments that were $2,700 larger than the nonexperimental payments sent to members of the control group.[3] The combined earnings reduction of husbands and wives in the Seattle-Denver treatment group was almost $1,800, or approximately two-thirds of the net experimental payment. The average tax rate of the Seattle-Denver plans was about 50 percent, implying that the $1,800 earnings reduction caused payments to be $900 above what they would have been in the absence of a work effort response. Thus, one-third of the net transfer cost of the Seattle-Denver plan was due to the reductions in reported earnings among participants. Another way to interpret the same set of figures is to say that the experiment spent nearly $2,700 on transfers and succeeded in raising the incomes of two-parent families by only $900.[4] Even if the earnings reductions are taken to be modest, it is reasonable to ask whether most taxpayers would be willing to spend $3 in order to raise the incomes of poor, two-parent families by only $1.

Several analysts have found evidence that at least part of the employment and earnings reduction reported in the experiments was spurious. Recipients of negative income tax payments had a clear incentive to underreport their employment and earnings, because to do so permitted them to receive a larger payment than the one to which they were legally entitled. Wage earners enrolled in the control group did not face this kind of misreporting incentive.

It is possible to analyze this issue with sufficiently accurate employment and earnings data which are not subject to reporting bias. The employment and earnings records of the unemployment insurance system provide one source of such data. The effects of underreporting

were systematically examined using these data in two of the experiments, Gary and Seattle-Denver. In both experiments underreporting was found to bias the estimates of employment and earnings response. The bias in the Gary experiment was large enough so that the entire earnings response and much of the apparent employment effect of the experiment disappeared.[5] In the case of the Seattle-Denver experiment the bias was somewhat smaller and less reliably estimated. Husbands and women heading single-parent families misreported their employment and earnings infrequently enough so that the response estimates reported in table 2 are probably only slightly overstated. On the other hand, the responses of wives and other secondary earners are greatly overstated.[6] The earnings reduction of wives, for example, virtually disappears when the response estimate is based on presumably accurate data from the unemployment insurance system. Of course, even if misreporting bias causes an exaggeration of the efficiency loss from a negative income tax, there is no reason for complacency about the earnings reductions reported in table 2. An earnings reduction caused by underreporting is just as costly to taxpayers as a reduction caused by a genuine reduction in work effort.

Offsetting the bias from misreporting is the effect of the limited duration of the experiments. There are at least two reasons to believe that a limited duration income maintenance program will elicit a smaller response than a permanent program that offers the same income guarantee and tax rate. The first is that workers may need time to respond to the incentives embedded in an income maintenance plan. If they are given only three years to respond, as they were in the experiments, their eventual response might not be fully observed. A second reason to expect a small response is that the income effect produced by a limited-duration program is by definition less than the income effect produced by an otherwise equivalent program which is expected to be permanent. A $1,000-per-year payment should cause a larger effect if it is to last indefinitely than if it is to continue only three years, unless the worker applies an extremely high discount rate to future income. On the other hand, because the experiments were temporary they essentially offered a sale on leisure, which participants were forced to take advantage of within a concentrated period. This encouraged greater responsiveness than would have been observed in a permanent program.

The Seattle-Denver experiment is the only one that permits us to examine the effects of limited duration in a reasonably satisfactory way. About 30 percent of the eligible sample in that experiment was enrolled for five years, while the remainder was enrolled for three years.[7] The pattern of response of families in both the five-year and three-year groups suggests that workers were somewhat slow in reacting to the negative income tax disincentives. Robins and West (1980b, p. 36)

estimate that 90 percent of the full response would only be observed after 2.4 years in the case of husbands, after 3.6 years in the case of wives, and after 4.5 years in the case of single women with dependent children.

The same authors also find that the average response of husbands and wives (though not of single women) in the five-year treatment group was substantially greater than that in the three-year group, even when the responses of the two groups are measured at the same point in time (for example, two years after enrollment). The maximum response of husbands in the five-year group occurred in the third and fourth years of the experiment, when the hours reduction was 13 percent and the earnings reduction approximately 12 percent (Robins and West, 1980b, p. 23). These reductions are about twice the magnitude of responses in the three-year group during the second year. The maximum response for wives in the five-year group occurred in the fourth and fifth years, when the hours reduction was 27 percent and the earnings reduction 26 percent (Robins and West, 1980b, p. 25). For single women heading families, the maximum response occurred in the fifth year, when the hours reduction was 32 percent and the earnings reduction about 35 percent (Robins and West, 1980b, p. 27).

Clearly, the sluggishness of the labor supply response and the attenuated response to a shorter duration plan cause the long-term impact of a permanent negative income tax to be substantially understated by the mid-experimental responses of families assigned to three-year plans. It should be stressed, however, that this conclusion is valid only for the relatively generous, low-tax plans tested in Seattle-Denver. Plans with high tax rates might have elicited a different pattern of response. Burtless and Greenberg (1982) found evidence that participants in the experiment, particularly women, reacted more strongly to the tax rates than they would have in a program of permanent duration. Taking advantage of the sale on leisure, participants in the three-year plans were significantly more responsive to the tax than were participants in the five-year plans. If a high-tax, low-guarantee plan had been tested, it is conceivable that the overall response of the three-year treatment group would have been larger, not smaller, than that of the five-year group.

The implications of table 2 may be summarized briefly. The four negative income tax experiments caused moderate to large proportional reductions in work effort. As expected, the proportional response was greater among women than men. The absolute reductions were largest in the Seattle-Denver experiment, which offered the most generous plans, and were smaller and less precisely estimated in the experiments testing plans with a lower income guarantee and breakeven point. The work effort reductions were overstated due to misreporting bias but understated because of the limited duration of the experiments, par-

ticularly in the case of high-guarantee, low-tax-rate plans. On balance, the experiments probably underestimated the permanent response to a negative income tax program with a generous guarantee (equal, for example, to the poverty line) and a relatively low tax rate (equal to or below 50 percent). It is less certain that the effect of low-guarantee, high-tax negative income tax plans would be understated in a short-duration experiment.

Even if we had perfect confidence in our estimates of average response, it is not clear how they could be used to predict the consequences of reform plans in which policymakers are actually interested. The Seattle-Denver experiment produced the most precise results and the ones that have been subject to the most thorough sensitivity analysis. But those results were obtained in an experiment in which the average negative income tax plan provided a guaranteed income of about 115 percent of the poverty line and taxed earned income at a marginal rate of approximately 50 percent. No feasible welfare reform plan could offer universal benefits this generous. The maximum combined benefit from aid to families with dependent children (AFDC) and food stamps is now only 73 percent of the poverty line in the median state and is equal to the poverty line in only the most generous state. The gross income limit for receiving benefits from AFDC is less than twice the poverty line in all but two states and below 1.5 times the poverty line in 37 states.[8] By contrast, over half of the families in the Seattle-Denver experiment were enrolled in plans with an income breakeven above twice the poverty level, and 86 percent were enrolled in plans with a breakeven above 1.5 times the poverty line.[9] Thus, not only were transfers quite generous in the Seattle-Denver experiment, they were available to families well up in the income distribution. For these reasons, the average recorded response to the Seattle-Denver plans does not provide a useful approximation of the expected response to plausible programs of welfare reform.

Structural Models of Response

Because estimates of the average negative income tax response are difficult to use, economists analyzing the experiments have sought to obtain structural estimates of response. We can distinguish between two broad classes of structural models. The first emphasizes the response within an experiment to the separate negative income tax plan characteristics—the income guarantee and the tax rate. The second emphasizes the individual-level response to unearned income and net wage levels more generally, or to changes in these two variables induced by the experiment.

At first blush, the first type of estimate might appear to be the easiest to produce and then apply in predicting the effects of income maintenance alternatives. To estimate the separate responses to income guarantees and tax rates, the analyst simply estimates equation (1) but replaces the variable T with a set of variables that reflect the level of the experimental guarantee, the marginal tax rate, and possibly some interaction between these two program features. For example, if the experiment tested a low, moderate, and high guarantee and a low, moderate, and high tax, with complete interaction of the guarantee and tax, a simple way to represent the treatment is:

$$(2) \qquad H = \alpha + \beta_1 T + \beta_2 (\Delta G) + \beta_3 (\Delta t) + \gamma Z + \delta X + \epsilon,$$

where T is again a variable for assignment to experimental treatment, ΔG is the dollar difference between the individual's assigned guarantee and the lowest guarantee tested, and Δt is the percentage-point difference between the assigned tax rate and the lowest tax rate tested. (ΔG and Δt both take the value zero for control observations and individuals assigned to the low-guarantee, low-tax plan.) With this specification, β_1 is interpreted as the average effect of the low-guarantee, low-tax program, β_2 is the average effect of a one-dollar rise in guarantee, and β_3 is the effect of a one-point rise in the tax rate. More complicated interaction effects of the guarantee and tax can also be specified. The expected effects of an alternative negative income tax plan can be predicted using the estimates of β_1, β_2, and β_3 and suitably defining T, ΔG, and Δt to represent accurately the alternative plan.

This approach to estimation, although straightforward, is less useful than it first appears. The samples enrolled in the experiments were not nationally representative, so the estimates of β_1, β_2, and β_3 will not necessarily be valid if applied to a wider population. For example, the average effect of a low-guarantee, low-tax plan, β_1, depends on the generosity of the welfare system against which it is compared. The plan might be expected to cause little work reduction in a state like Washington, where AFDC benefits are high, but significant reductions in Indiana, where the maximum payment for such aid is extremely low. Since the experiments did not enroll samples that faced a representative set of state welfare programs, it is not clear how estimates of β_1, β_2, and β_3 can be used to predict work effort responses in states where no experiment was conducted.

A more subtle problem arises because of the sampling design used to assign families to negative income tax treatments. The experiments did not use simple random assignment. The potential sample in each experiment was divided into subsamples defined by a set of stratifying variables. One important stratifying variable was preexperimental income level. The Seattle-Denver experiment, for example, divided the

sample into seven preexperimental income classes. Families within each income class were randomly assigned to one of the tested plans or to control status. The proportion assigned to a specific plan was not identical in each income class, however. In order to increase the number of families that could be enrolled given a fixed budget constraint, the experiments enrolled a higher proportion of low-income families into the least generous plans and a higher proportion of high-income families into the most generous plans. The income distribution of families assigned to the most generous plans was consequently not the same as that of families assigned to the least generous plans. This implies that the differences in average work effort response to two different plans in the same experiment may be due to differences in the composition of the samples assigned to the plans as well as to genuine differences in response induced by the plans.[10] This problem could be avoided in estimation by fully interacting the negative income tax plan parameters with the stratifying variables, but such a procedure is extremely cumbersome and yields statistically imprecise results.[11] No published study from the experiments relies on this approach. When analysts have estimated structural models of response to the income guarantee and tax rate, they have not estimated all of the interaction terms that would permit us to disentangle the effect of the sampling plan from that of the treatments themselves.

In a second approach to estimation, economists have specified labor supply models quite similar to those estimated with nonexperimental data. A model of this type was estimated by Keeley et al. (1978b) using data from the Seattle-Denver experiment:

$$(3) \qquad H = \alpha + \beta_1 [\Delta w|\bar{H}_p] + \beta_2 [\Delta Y|\bar{H}_p] + \gamma Z + \delta X + \epsilon,$$

where Δw is the change in the after-tax wage rate caused by an individual's assigned negative income tax plan and ΔY is the change in after-tax income. Both Δw and ΔY are computed at the individual's preexperimental level of work effort, H_p. Obviously, Δw and ΔY will vary widely even among individuals facing the same negative income tax plan. Under the usual assumptions we would expect β_1 to be positive and β_2 to be negative. That is, increases in the net wage, holding income constant, should cause individual labor supply to rise, while increases in income, holding the net wage constant, should reduce labor supply. The Keeley et al. specification is similar to one estimated with nonexperimental data and proposed by Ashenfelter and Heckman (1974). Like Keeley et al., Burtless and Hausman (1978) estimated a model that could be applied as easily to nonexperimental as to experimental data:

$$(4) \qquad \log (H) = \alpha + \beta_1 w + \beta_2 N + \delta X + \epsilon,$$

where w is the after-tax wage rate and N is the virtual income intercept

Table 3
Estimates of Substitution and Income Elasticities from Experimental and Nonexperimental Studies

Subjects		Uncompensated Substitution Elasticity[a]	Compensated Substitution Elasticity[b]	Total Income Elasticity
Men				
Negative Income Tax	[N = 21][c]	.0043 (.098)	.0795 (.068)	−.0757 (.093)
Weighted Negative Income Tax		−.0223	.0902	−.1139
Nonexperimental	[N = 26][c]	−.1045 (.178)	.2842 (.415)	−.3873 (.339)
Women				
Negative Income Tax Wives	[N = 20][c]	−.0420 (.368)	.1105 (.237)	−.1515 (.214)
Weighted Negative Income Tax		.0659	.1783	−.1115
Negative Income Tax Wives[d]	[N = 14][c]	.0957 (.225)	.1907 (.154)	−.0957 (.146)
Weighted Negative Income Tax		.1730	.2425	−.0696
Negative Income Tax Female Heads	[N = 11][c]	−.0373 (.123)	.1346 (.070)	−.1709 (.085)
Weighted Negative Income Tax		−.0426	.1355	−.1774
Nonexperimental	[N = 48][c]	1.9919 (3.162)	2.0248 (3.154)	−.0331 (.423)
Nonexperimental[e]	[N = 38][c]	1.3553 (1.319)	1.3661 (1.229)	−.0113 (.463)

Numbers in parentheses are standard deviations.

[a] $\left(\frac{\Delta H}{\Delta W(1-t)}\right) \div \left(\frac{H}{W(1-t)}\right)$

[b] $\left(\frac{\Delta H}{\Delta W(1-t)}\,|\bar{U}\right) \div \left(\frac{H}{W(1-t)}\right)$

[c] Number of separate estimates of response used to compute the reported elasticity.

[d] Excludes estimates from the New Jersey experiment.

[e] Excludes five estimates showing the highest compensated substitution elasticity and five estimates showing the lowest compensated substitution elasticity.

Sources: See text.

measured at an individual's desired hours of work.[12] For low-wage workers, it is reasonable to expect β_1 to be positive and β_2 to be negative.

A cynic might ask why it is necessary to invest $100 million collecting experimental data when analysts then estimate models that could as easily be estimated using nonexperimental data. While the question is a legitimate one, it has a straightforward answer. The experimental variation in tax rates and income guarantees produces a large amount of essentially random variation in Δw and ΔY and in w and N, the critical variables in equations (3) and (4). The variation is not totally random, of course, because these variables are correlated with a worker's gross wage rate and may be correlated with preexperimental work effort and other confounding variables through the effects of the negative income tax plan assignment procedure, discussed above.[13] But in spite of this correlation, random assignment of workers to widely differing negative income tax plans assures us that a greater fraction of the variation in Δw, ΔY, w, and N will be independent of observed and unobserved variables that affect H.[14] From a statistical standpoint, this should increase our confidence in the resulting coefficient estimates.

A large number of structural models have been estimated using data from the negative income tax experiments, particularly the Seattle-Denver experiment. Moffit and Kehrer (1981, pp. 138-42) and Robins (1985, p. 578) have reported individual and average estimates of income and substitution effects obtained in each of the negative income tax experiments. Table 3 presents a summary of income and substitution elasticities, averaged across the four experiments. These estimates of response in the New Jersey, Rural, and Gary experiments are based upon corrected estimates of elasticities reported by Moffitt and Kehrer (1981). Estimates for the Seattle-Denver experiment are based on the simple average of elasticities reported in seven separate studies using data from that experiment.[15] Table 3 shows the average experimental estimates of the (uncompensated) net wage elasticity, the compensated substitution elasticity, and the total income elasticity for husbands, wives, and female household heads. The substitution elasticities are useful in indicating the slope of the labor supply function and the rough magnitude of efficiency losses arising from imposition of higher tax rates on low-wage workers. The total income elasticity shows the percentage by which work effort falls with a one percentage point rise in income that is not accompanied by a change in the net wage rate.

Average elasticity estimates for the experiments were computed in two different ways. First, the simple arithmetic average of all of the estimates from separate studies of labor supply response was calculated. The top row in the table, for example, shows the simple average of 21 separate estimates of the labor supply elasticity for husbands enrolled in the experiments.[16] In parentheses below these elasticity averages the

table shows the standard deviation of the different point estimates of response around the average estimate. A more defensible way to compute the mean response is to account for the relative size of the samples used to obtain different elasticity estimates. It seems reasonable, for example, to attach a lower weight to the estimated elasticity within a sample of 200 New Jersey Hispanics than we attach to the response of 2,200 white, black, and Hispanic husbands in Seattle and Denver. A weighted estimate of the average elasticity was derived using a two-step procedure. First the average elasticity within each of the four negative income tax experiments was computed, and then the weighted average elasticity was calculated by suitably weighting the measured responses in the four experiments. (Weights are reported in a footnote to table 2. Where necessary the responses of separate racial groups within each experiment were weighted. See the weights reported in table 1.)

It might be argued that the estimates of response from the individual studies should be weighted by the quality of the research methodology rather than the size of the estimation sample. This is the implicit strategy of Borjas and Heckman (1978) in an early survey of the nonexperimental labor supply literature. Such a survey would yield more interesting and precise results than those reported here. However, it would also require thorough justification of the weights attached to the various studies. I will leave that exercise to others.

It is useful to compare the estimates obtained in the experiments with labor supply estimates reported in the nonexperimental literature. Killingsworth (1983) has provided an informative survey of elasticity estimates obtained in nonexperimental studies. Table 3 contains my computations of the average and standard deviation of elasticity estimates reported in 26 nonexperimental studies of U.S. prime-aged men and 48 studies of U.S. women.[17] Because the range of estimates for women was so large, average female elasticities were computed excluding the five studies with the highest and the five studies with the lowest estimates of compensated substitution elasticity.

The labor supply functions estimated with experimental data appear to be comparatively inelastic. For example, the uncompensated labor supply function of low-wage men is essentially vertical. A change in the net wage, holding nonwage income constant, has virtually no effect on annual male work effort. Even if we consider estimates of the uncompensated elasticity one standard deviation from the mean estimate, the elasticity appears to be quite moderate.

The uncompensated substitution elasticity of wives is less reliably estimated. Although the average estimated elasticity is only −0.04, this average is sensitive to the method of weighting. When the several elasticity estimates are weighted according to the size of the estimation samples, the mean elasticity rises to +0.07. Much of the uncertainty

arises because of the lack of robustness of estimates of wives' supply elasticities in the New Jersey experiment. When the New Jersey estimates are excluded, the mean uncompensated elasticity rises to 0.10–0.17, depending on how the remaining estimates are weighted. Note that the standard deviation around the unweighted average falls by more than one-third when New Jersey estimates are excluded. The relatively large dispersion in estimates of the labor supply of wives was caused by the income truncation imposed on samples enrolled in the experiments. Since the samples were restricted to very low-income families, they contained an abnormally small percentage of working wives. In the experiments with the lowest income limits (New Jersey and Iowa-North Carolina), the elasticity estimates were sensitive to the work effort changes of only a handful of women. The elasticity estimates for men and for women heading single-parent families seem to fall in a much narrower range than do the estimates for wives.

In comparison to the estimates from the nonexperimental literature, the elasticity estimates from the experiments tend to be much smaller in absolute value. This tendency is most pronounced with respect to the compensated substitution and the income elasticities estimates for men and, even more strikingly, for the uncompensated and compensated substitution elasticities for women. Whereas most nonexperimental estimates show a strongly positive uncompensated supply function for women, the experiments found only weakly positive or even backward-bending supply functions. The mean experimental estimates of the income elasticity for men and women are in the range -0.07 to -0.18. These estimates are below the average nonexperimental estimates in the case of men but above the average nonexperimental estimates for women.

On balance, the experimental estimates imply a smaller responsiveness to negative income tax disincentives than do most nonexperimental estimates. This conclusion was also reached by Moffitt and Kehrer (1981) in a survey of the earlier results from the negative income tax experiments. The average estimates of the compensated substitution elasticity from the experiments are uniformly lower than the average elasticities estimated in the nonexperimental literature. Since the economic efficiency costs of a particular tax or transfer plan are proportional to the compensated substitution effect, it follows that efficiency loss from a negative income tax was found to be smaller in the experiments than would have been predicted on the basis of the average elasticity estimated in nonexperimental studies.

Interestingly, the experimental estimates fall in a far narrower range than the nonexperimental estimates, though the experimental elasticities were estimated using four independent samples and a wide range of econometric models. The smaller dispersion in estimates is

obvious from a comparison of the standard deviations around the mean experimental and nonexperimental point estimates. The greater robustness of the experimental estimates is presumably due to the large amount of experimentally induced random variation in net wages and nonwage income levels. This random variation reduces the effect of specification error on parameter estimates and thus minimizes the effect of using alternative econometric models. Even though the average experimental and nonexperimental elasticity estimates in table 3 are sometimes far apart, the range of experimental estimates falls well within the range observed in the nonexperimental literature. Note, for example, that the average point estimate of response in the experiments is always within one standard deviation of the corresponding point estimate from nonexperimental studies. This is, of course, primarily due to the fact that the standard deviation of nonexperimental estimates is so large. The experiments thus appear to have achieved their major goal. They have substantially reduced our uncertainty about the size of work effort reductions in response to wage rate and income changes.

Implications for Welfare Reform

The labor supply estimates reported in the previous section can be used to analyze a variety of issues about welfare reform. The most important issues concern the net budgetary costs and work effort effects of particular proposals for reform. To predict the detailed effects of a reform it is necessary to incorporate estimates from a structural labor supply model into a microsimulation model. In comparison to the large number of studies of experimental labor supply response, there have been only few studies attempting to generalize the findings from the experiments to the U.S. population. Predictions of the nationwide response to a negative income tax are rare because they are costly to obtain.

The first requirement for decent prediction is a reliable source of information about a nationally representative sample of low-income families. Most sources of data, such as Census public-use tapes or the Current Population Survey, are expensive to use. A second requirement for prediction is a computer program that can accurately define or predict both the pre-reform and post-reform situations of individuals represented in the Census file. Certain pre-reform characteristics of individuals, such as employment status, weekly hours of work, annual earnings, and unearned income, may be directly reported in the file. Other characteristics, such as taxes paid, potential welfare benefits, and marginal tax rates, must be predicted on the basis of published tax and welfare schedules and sophisticated imputation procedures. Because

the United States contains 51 separate political jurisdictions with unique income tax schedules and welfare formulas, the burden of imputation is formidable. Using labor supply estimates from the negative income tax experiments (or some other source), the analyst must finally predict the amount of work effort change that will occur as a result of a reform in the transfer formula and calculate the budgetary cost of the reform, taking account of the labor supply response. Given the size of the computational burden, it is not surprising that microsimulation is seldom performed.

Table 4 shows predictions of the work effort effects and budgetary costs of four different negative income tax plans. The predictions are based on microsimulations performed by SRI International and Mathematica using estimates of work effort response from the Seattle-Denver experiment. The table shows the results of two separate simulations of response to each of the plans. The first simulation used population information covering the year 1974 and estimates of labor supply response reported in Keeley et al. (1978b). The second study used population information for 1975 and estimates of labor supply response reported in the final Seattle-Denver report (SRI International, 1983). Note that neither the baseline year nor the assumed labor supply parameters were the same in the two simulations. (Cost estimates are converted to 1985 dollars, however.) In addition, other details of the simulations differed, although the significance of these differences is difficult to interpret.[18]

The four negative income tax plans examined in the table offer two basic payment levels and two tax schedules. The lower income guarantee is 75 percent of the poverty level while the higher guarantee is one-third higher, or 100 percent of the poverty line (approximately $11,000 per year for a family of four in 1985). The plans are assumed to replace the present public assistance and food stamp programs. The lower guarantee is slightly more generous than the combined guarantees of AFDC and food stamps in a state offering the median aid benefit. However, states offer a wide range of basic aid plus food stamp payment levels, ranging from less than half to slightly more than the poverty line.[19] The two tax rates examined are 50 percent and 70 percent. By comparison, in the case of AFDC, the statutory tax rate on earnings is now 100 percent, though the statutory rate in the mid-1970s was only 67 percent. It should be stressed that effective rates have always been below statutory rates. The effective tax rate for AFDC might currently approach 70 percent, but in the mid-1970s it was as low as 30 percent (Fraker et al., 1985). The combined AFDC and food stamp effective tax rate in the mid-1970s was thus below 50 percent. In each of the negative income tax plans examined, the tax rate on unearned income is 100 percent. Positive income and payroll taxes are fully reimbursed for

families with gross income below the negative income tax breakeven point. This reimbursement implies that a 100-percent-of-poverty-line guarantee assures all families of a net income equal to at least the poverty line.

Table 4
Labor Supply and Budgetary Implications of Four Negative Income Tax Plans

Negative Income Tax Plan	(1) Work Effort Change: Among Recipients	(2) Work Effort Change: In Entire Population	(3) Percent Receiving Benefits[a]	(4) Net Additional Cost[b]	(5) Population Earnings Reduction[b]	(6) (5) ÷ (4)
75% Poverty Line Guarantee/50% Tax Rate						
Husband-Wife	− 9.5%	−1.4%	.19	$15.5	$ 9.0	.58
Female Heads	− 6.7	−2.4	.61	.8	.4	.50
Total			.24	16.3	9.4	.58
Alternative Estimate						
Husband-Wife	− 6.5	− .8	.17	11.5	5.1	.44
Female Heads	7.9	9.0	.57	−4.8	−3.0	.62
Total			.22	6.7	2.1	.31
75% Poverty Line Guarantee/70% Tax Rate						
Husband-Wife	−15.8%	− .5%	.07	$ 5.5	$ 2.2	.40
Female Heads	− 9.3	−1.2	.51	−1.0	.0	—
Total			.12	4.5	2.2	.49
Alternative Estimate						
Husband-Wife	− 8.0	.0	.06	1.2	−.7	—
Female Heads	5.2	11.5	.43	−6.5	−3.7	.57
Total			.10	−5.3	−4.4	.83
100% Poverty Line Guarantee/50% Tax Rate						
Husband-Wife	−10.0%	−3.5%	.39	$51.9	$27.1	.52
Female Heads	−12.0	−7.1	.73	9.2	1.8	.20
Total			.43	61.1	28.9	.47
Alternative Estimate						
Husband-Wife	− 9.8	−3.4	.39	51.4	26.7	.52
Female Heads	− 2.2	1.5	.71	4.1	− .6	—
Total			.43	55.5	26.1	.47
100% Poverty Line Guarantee/70% Tax Rate						
Husband-Wife	−20.6%	−1.5%	.15	$19.6	$ 8.6	.44
Female Heads	−14.9	−5.3	.61	6.1	1.0	.16
Total			.20	25.7	9.6	.37
Alternative Estimate						
Husband-Wife	−10.7	− .9	.14	14.8	5.2	.35
Female Heads	− 4.4	5.4	.57	.6	− 1.8	—
Total			.19	15.4	3.4	.22

[a]Percent of families in relevant population receiving negative income tax payments.

[b]Measured in billions of 1985 dollars. A negative sign indicates a net cost saving or net earnings increase. Estimated earnings reduction excludes the response of families who are nonrecipients before and after the reform.

Sources: Keeley et al. (1978a and 1978b). Alternative estimate from SRI International (1983).

Analysts performing the simulations assumed that the eligible population contained non-aged husband-wife families and female-headed families with children. Aged and single-person families were excluded from the simulation. The negative income tax represents a substantially different kind of reform for the three groups that would be eligible for payments. For single-parent families, the negative income tax would simply replace AFDC and food stamps, both of which are already received by a high proportion of single mothers with low incomes. For many single-parent families, the negative income tax payment might even be lower than the welfare benefit that it replaces. Two-parent families with children would be more generously treated under a negative income tax than they are under the current welfare system. These families are eligible to receive AFDC in only about half the states, and even in those states the program is less generous to two-parent families than it is to single-parent families. Childless husband-wife families would be treated far more generously under a negative income tax than they are under the current system. Such families are currently eligible to receive only food stamps and general assistance. General assistance is typically far less generous than AFDC.

The first column in table 4 shows the predicted reduction in annual hours of work among recipients of negative income tax payments. In two-parent families the work reduction under all four plans is moderately large, ranging from 6.5 percent to as much as 20.6 percent, depending on the characteristics of the plan and the details of the simulation. These estimates reflect the combined responses of both husbands and wives to the negative income tax incentives. In the first simulation there is a tendency for the percentage reduction in hours to rise strongly with increases in the guarantee and tax rate. The second simulation shows the same pattern, but it is much weaker. Note that the second simulation shows smaller work effort reductions than the first, particularly for plans with a higher marginal tax rate. In spite of their differences, both simulations show work effort reductions among husbands and wives receiving the negative income tax payments, with fairly large percentage reductions under the two plans that provide a poverty-line guarantee.

The two simulation programs do not conform in their predictions of the response among single-parent families. One simulation shows moderate to substantial hours reductions while the other shows only small reductions or even labor supply increases. It is unlikely that the inconsistencies are due to the differing labor supply parameters used.[20] They are probably caused by differences in the base year used and the assumed level of pre-reform welfare benefits. (The latter difference presumably has only a small effect in the case of husband-wife families because these families are typically ineligible for welfare benefits under the current system.) The striking differences in the predicted single-

parent responses to the same negative income tax are disturbing. The differences imply that work effort estimates are sensitive to alternative techniques in simulation as well as to varying assumptions about income and substitution effects.

The second column in the table shows the predicted population response to the negative income tax plans. These predictions include work effort changes among nonrecipients as well as recipients of payments. The numbers in the column show the percentage changes in population hours of work. (Note that the percentage change in population earnings will be much smaller than the percentage hours reductions because negative income tax recipients, who account for the work reductions, have lower wage rates than nonrecipients.) Among husband-wife families, the population response is always much smaller in percentage terms than the response among recipients. The reason is obvious in view of the participation rates reported in column (3). Only a fraction of the population receives negative income tax payments, so most husband-wife families will be unaffected by welfare reform. (Neither of the simulations includes a tax increase on nonrecipients to finance the added transfer payments.) Note that the husband-wife population response rises with increases in the guarantee level but *declines* with increases in the marginal tax rate. That is, a 70 percent tax rate causes less overall work reduction than a 50 percent tax rate. The explanation for this apparently perverse result is that the participation rate in a high-tax program will be lower than in a low-tax program that has the same income guarantee. As the tax rate rises, the income cutoff point for receipt of benefits falls. Fewer families will have incomes low enough to qualify for payments, so fewer will be affected by the work disincentives implicit in the transfer formula. The estimates in the table show a conflict between the goal of providing work incentives to transfer recipients and that of providing incentives to the population as a whole.[21] Recipients can be encouraged to work through a reduction in the tax rate, but such a reduction will increase the number of recipients and hence reduce aggregate work incentives.

The trade-off between work incentives for recipients and for the population as a whole is also evident in the case of single mothers. Both simulations show that aggregate work effort is greater under a high-tax plan than under a low-tax plan with the same guarantee.[22] Both simulations also show that work effort among recipients is lower under the high-tax plan than under the low-tax plan. The two simulations do not agree, however, in predicting the sign of the overall response to a negative income tax plan among single mothers. The first simulation implies that all four negative income tax plans, including the least generous, would reduce work effort. The second implies that the plans, including even the most generous, would cause an increase in labor sup-

ply among single parents. The discrepancy is due to different assumptions about the generosity of the existing welfare system. The second simulation is based on the assumption that the current system is relatively generous, so introduction of a negative income tax would reduce benefits for a substantial fraction of current welfare recipients.[23]

The budgetary implications of the four negative income tax plans are shown in column (4). The most interesting estimates are the ones for the two plans that offer income guarantees equal to the poverty line. By definition these plans eliminate poverty among husband-wife and single-parent families. The more generous plan would cost $56 billion to $61 billion more than the current welfare and food stamp programs, or approximately 1.5 percent of GNP.[24] The less generous, high-tax plan would cost $15 billion to $26 billion more, or 0.4 to 0.6 percent of GNP. How one views these estimates depends on one's attitude toward redistribution. A person favorably inclined toward redistribution might regard the less expensive high-guarantee plan as a bargain: poverty is eliminated among families containing children, and at modest cost. Federal taxes would have to rise 2 to 4 percent to finance the plan, however, so taxpayers less favorably inclined toward redistribution would have ample grounds to oppose the reform, especially for husband-wife families.

The last two columns provide evidence that might dissuade even advocates of redistribution from suggesting a universal negative income tax. Column (5) shows the earnings reductions in response to introduction of a negative income tax. Negative values are reported in a few cases, implying that a negative income tax would actually increase aggregate earnings. But most of the entries are positive, suggesting that earnings reductions would offset at least part of the income gains to the poor produced by a negative income tax. Column (6) shows the size of the earnings change as a fraction of the net additional transfer cost of the program. The fraction is especially high in the case of two-parent families. The first simulation implies that the earnings reduction would represent 40 to 58 percent of the added transfer costs of the program for two-parent families. The second simulation implies earnings reductions ranging from 35 to 52 percent of net program costs, except in the case of the least generous program, where there is a slight earnings gain.

Husbands and wives in families receiving benefits obviously "consume" a high percentage of their benefits in the form of additional leisure or other nonmarket uses of time. While the consumption of additional leisure increases the happiness of recipient families, it simultaneously raises the cost of payments to taxpayer donors and offsets a large part of the intended redistributional impact of the payments. Even more important to some taxpayers, it raises the dependence of poor two-parent families on government transfers.

The trade-off between earnings reductions and added transfer costs is more favorable in the case of single mothers. Only the two plans with a poverty-line guarantee involve substantial added costs to taxpayers. One of the simulations shows that under these plans earnings would fall by 16 to 20 percent of additional transfer costs, while the second shows that single mothers' earnings would actually rise as a result of introduction of a poverty-line guarantee. Though I am skeptical of the second set of predictions, it seems likely that the earnings response of single mothers would be less costly to taxpayers than the response in two-parent families. This is suggested by the actual pattern of response in the Seattle-Denver experiment. During the second year of that experiment, the earnings reduction among single mothers was 39 percent of the average negative income tax payment to one-parent families, while the combined husband and wife earnings reduction was 68 percent of the average payment to two-parent families.[25] Given the same payment, the net income gain to a single-parent family would be greater than the income gain in a two-parent family. A negative income tax thus represents a more attractive reform alternative for single-parent than for two-parent families.

One of the main obstacles to improving the generosity of means-tested transfers is the knowledge that more generous benefits will reduce the earnings and self-support of the poor. The simulation results reported in table 4 suggest that this concern is reasonable for two-parent families, but is less valid in the case of single-parent families. Even though the predicted work effort reduction among husbands and wives is small, the implied reduction in earnings is a large percentage of additional transfer benefits. Using Arthur Okun's analogy, it is obvious that a negative income tax does not provide a leakproof redistributive bucket.

The bucket is nonetheless more leakproof than sometimes suggested in the nonexperimental literature. Edgar K. Browning and William R. Johnson (1984) have recently argued, for example, that the disposable money income of the top three income quintiles is depressed by $9.51 for each one-dollar increase in money income successfully transferred to the lowest two quintiles. It is depressed by this large amount because transfer recipients reduce their work effort, thus increasing the amount of money that must be transferred to raise their net incomes by one dollar. In addition, Browning and Johnson's simulations show substantial work effort reductions among taxpayers who are faced with higher tax rates as a result of the increased transfers.

Findings from the experiments suggest that the cost of redistributing one dollar to the poor must be far less than $9.51. For example, estimates in table 4 of the cost and earnings impact of the most generous negative income tax plan imply that it would cost approximately $1.89 to transfer

an added dollar to the poor.[26] This estimate ignores the labor-supply response of taxpayers who must pay $1.89 in added taxes. If the net income of these taxpayers falls by $9.51, it must be the case that their net earnings fall by $7.62 (=9.51-1.89) in response to the higher tax rate. Using the assumptions of Browning and Johnson, this implies that gross earnings fall by at least $12.70.[27] The labor-supply response parameters estimated in the experiments appear inconsistent with the prediction that annual earnings of taxpayers would decline by $12.70 in response to a rise in net tax liabilities of only $1.89.

The experimental elasticity estimates reported in table 3 are in fact consistent with a slight rise in taxpayers' earnings, because the income effect of higher taxes should more than offset the substitution effect for most high-income families.[28] This is confirmed in the only microsimulation study that uses experimental labor supply parameters to predict the responses of both transfer recipients and taxpayers to the introduction of a negative income tax. In that simulation study, Betson, Greenberg, and Kasten (1982) find that the combined labor supply responses of transfer recipients and taxpayers actually cause national earnings to rise after introduction of a negative income tax. That is, the earnings gains of taxpayers more than offset the earnings reductions of transfer recipients.[29] If this conclusion is valid, the experimental results imply that the disposable money income of the top three income quintiles will fall by less than $1.89 for each one-dollar increase in money income successfully transferred to the working-age poor.[30] This estimate is, of course, far below the estimate reported by Browning and Johnson, who based their study on nonexperimental labor supply elasticities. The experimental results thus imply substantially lower costs to taxpayers of income redistribution.

Conclusions

The negative income tax plans tested in the experiments were expected to reduce work effort among participants, and they did so. The work reductions were probably smaller than most opponents of a negative income tax had feared, but larger than advocates had hoped. In comparison to predictions of work effort response based on prior nonexperimental research, the actual response to the tested plans was small. But the response was negative even among women previously receiving public welfare, with all of its attendant work disincentives. The estimates of income and substitution elasticities obtained in the experiments fall well within the very broad range of estimates obtained in nonexperimental studies. Moreover, the experimental estimates appear to be far more robust. That is, they fall within a narrow range even when

estimated using different samples and alternative econometric models. With the exception of the income elasticity estimated for women, the average experimental elasticities are lower in absolute value than corresponding nonexperimental estimates. In particular, the compensated substitution elasticity is only a fraction of the average elasticity estimated in nonexperimental studies, implying that the efficiency losses for redistribution to the able-bodied poor would be lower than could be predicted from the average nonexperimental estimates of response.

It has been argued, by Anderson (1978) and Murray (1984) among others, that the findings of the experiments greatly understate the long-run response to a negative income tax. While there is some evidence from the experiments themselves that the long-run impact is indeed understated, the evidence is neither as strong nor as unambiguous as these critics argue. The permanent income effect of negative income tax payments was almost certainly underestimated in the experiments, but the substitution effect of the tax rates was probably overstated, at least among wives. While it is true that participants in the experiments may not have had time to fully adjust their labor supply to its long-run equilibrium value, it is equally true that the experiments did not observe the long-run response of employers to a smaller supply of low-wage labor.[31] Moreover, at least part of the apparent labor supply response in the experiments is known to have been a reporting phenomenon rather than a true reduction in work effort. (It is arguable whether the protections against income misreporting in a national program would be greater or less than those available in an experiment.) Given these potentially offsetting biases, the long-run impact of a modest negative income tax is probably understated by no more than one-third by simple extrapolation of the experimental results.

The estimates obtained in the experiments have a number of implications for reform of the welfare system, especially reform that raises the generosity of benefits. The findings suggest that benefit increases would cause only moderate reductions in aggregate hours of work and even smaller reductions in aggregate earnings. But even if the overall work reduction is small, the resulting earnings loss among recipient breadwinners would represent a large fraction of the higher payments sent out to low-income families. Earnings reductions would therefore offset a substantial part of the income gain from more generous transfers.

The arithmetic of reform is especially melancholy in the case of two-parent families, where earnings reductions might represent 50 to 60 percent of the added cost of new transfers. A simple and moderately generous negative income tax appears to be far more feasible for single-parent families. The earnings response of single mothers is small or even slightly positive. The experiments thus provide some support for offer-

ing generous benefits only to families whose earnings are less responsive to work disincentives. George Akerlof (1978) has argued, for example, that high-benefit, low-tax transfer formulas should be made available to only the least responsive families so that benefits can be more generously provided to those in greatest need. The results of the experiments support Akerlof's argument that the trade-off between higher benefits and lower work effort would be less painful under a system of separate transfer formulas for one- and two-parent families. (As Akerlof also points out, this is a fair description of the current welfare system.) Unfortunately, such a system provides clear incentives for families to change their composition in order to become eligible for the more generous transfer formula.

The findings from the experiments also point up a conflict between creating work incentives for transfer recipients and for the population as a whole. If a major goal of a transfer formula is to provide work incentives for recipients, the findings imply that relatively low tax rates are desirable. If the goal is to reduce disincentives for the entire population, a much higher tax rate is preferable because it minimizes the size of the population subject to work disincentives. This trade-off is clearest in the case of husband-wife families, where reductions in the marginal tax rate (given a fixed and plausible guarantee level) cause rapid increases in the population eligible to receive benefits. For single-parent families the trade-off is less clear since so many single mothers are eligible to receive payments, even at low guarantee levels. Hence, reductions in the marginal tax rate do not cause such rapid increases in the proportion of one-parent families eligible to receive payments.

If the experiments have inspired pessimism about our ability to reduce poverty through a system of pure cash transfers, they have also stimulated an examination of alternatives to a negative income tax. One way to minimize the adverse earnings effects of generous transfers is to require recipients to work. The Carter administration proposed to do this through a program of guaranteed public sector jobs for welfare recipients who were expected to work. Recipients refusing to work would have been denied benefits under the more generous transfer formula and forced to rely on benefits computed under a less generous formula. While the Carter proposal would have reduced or even eliminated the adverse earnings impact of more generous transfers, it would have involved substantial additional costs in order to finance the guaranteed jobs program. Some of these costs can be avoided under workfare, which essentially requires welfare recipients to work but does not pay them anything in addition to their current welfare grant if they do so. Recipients who decline work can have their grants reduced or eliminated. The negative income tax experiments obviously shed little if any light on the effects of this kind of work requirement.

Wage subsidies and earnings subsidies represent an alternative approach to redistribution. A worker eligible for a wage subsidy receives a transfer payment that grows rather than declines as hours of work rise. Not only does the program redistribute income to the poor, but it offers larger transfers to breadwinners who work longer hours. The labor supply response to wage subsidies is thus assumed to reinforce rather than offset the direct redistributive effect of the transfer payments. The response estimates obtained in the negative income tax experiments can be used to predict the effects of wage subsidy plans as well as negative income tax plans. The elasticity estimates reported in table 3 do not appear especially encouraging for a wage or earnings subsidy scheme. The labor supply functions estimated in the experiments are vertical or backward-bending. Much of the response to negative income tax payments was caused by a reliably estimated income effect. Any wage or earnings subsidy thus has the potential to encourage work reductions among those breadwinners who would receive the largest subsidies, that is, those now working the longest hours. In a simulation study of the impact of wage-rate subsidy schemes based on labor supply estimates from the Seattle-Denver experiment, analysts have found that subsidy plans actually reduce hours and earnings in recipient families (Betson and Bishop, 1982). Contrary to the expectations of subsidy advocates, the work response to wage subsidies—like the response to negative income tax payments—tends to offset the direct redistributive impact of the transfers.

The experiments have confirmed that good deeds are not costless. Income redistribution to the poor has an efficiency price. The price is far lower than pessimists predicted, but it certainly exceeds zero. The reaction of policymakers and policy analysts to this set of findings is interesting. They seem far more impressed by our certainty that the efficiency price of redistribution is positive than they are by the equally persuasive evidence that the price is small.

[1]The results in table 2 were not based on an identical statistical specification across experiments, nor were the estimation samples selected with identical criteria. The estimates reflect the responses to negative income tax plans in the middle two years for the New Jersey experiment, in the entire three years of the Rural and Gary experiments, and in the second (or middle) year of the Seattle-Denver experiment. For estimates based on a similar model and set of sample selection criteria, see Robins (1985), who reports very similar results. I slightly prefer the results reported here because they reflect the judgments of analysts who were most familiar with data from the individual experiments.

[2]See Moffitt (1979, p. 479) and Robins and West (1980b, pp. 23, 25, and 27).

[3]SRI International (1983) p. 177.

[4]The after-tax income of eligible families was raised by somewhat more than $900. The estimated reduction in gross earnings is $1,800 but the implied reduction in net earnings is probably 10 to 20 percent below that figure.

[5]Greenberg, Moffitt, and Friedman (1981, p. 586).

[6]Greenberg and Halsey (1983, pp. 400-05). In an unpublished analysis of underreporting based upon earnings records from the Social Security Administration rather than the unemployment insurance system, SRI obtained similar results. Underreporting of income to the experiment caused a very slight overstatement of the true earnings reduction among husbands and single mothers and a somewhat larger overstatement of the reduction among wives in two-parent households. Some observers argue that the experiments' experience with income misreporting is not relevant in a fully operational national program, since a national program would have access to employer-reported earnings information, such as that available to the Social Security Administration. While it is possible to use Social Security and unemployment insurance administrative records to verify the earnings reductions estimated from interview data, it would be impractical to rely on these same administrative records to compute monthly negative income tax payments. The Social Security Administration and state unemployment agencies obtain individual earnings records only with a lag, which can range up to 18 months. This is clearly too long to permit the timely calculation of negative income tax benefits. Hence, any practical system of monthly (or bimonthly) transfer payments must rely on self-reported earnings information, at least to some degree. For that reason, the experimental findings on income underreporting are applicable to a wide range of feasible welfare reform plans.

[7]A very small number of families was enrolled for 20 years, but this sample is probably too small to yield useful results.

[8]Committee on Ways and Means (1986), pp. 373-74.

[9]Office of Income Security Policy (1983), p. 6.

[10]To illustrate the problem, consider the earnings reduction among Seattle-Denver husbands during the third experimental year. Men assigned to the lowest guarantee/50% tax plan reduced their earnings by an average of $962, while men assigned to the highest guarantee/50% tax plan reduced their earnings by only $592. An explanation for this perplexing pattern of response is provided by the sample assignment plan. Whereas 96% of men in the less generous plan had preexperimental income below $7,000, only 26% of men in the more generous plan had income below that level. If we estimate the effect of both plans separately for each preexperimental income level, we can compute what the expected responses would be in two samples with an identical income distribution. Suppose we consider a sample that has the income distribution of the combined samples assigned to the two negative income tax plans just mentioned. The expected response to the low guarantee/50% tax plan is an increase in earnings equal to $753 per year, while the expected response to the high guarantee/50% tax plan is an earnings reduction of $1,994. Both predictions are extremely imprecise because of the small number of men within particular income classes assigned to one or another of the plans. Clearly, the sampling plan had an enormous impact on the pattern and precision of estimated responses to the two plans.

[11]Results from this procedure are statistically imprecise because there are only a few observations in each cell when all conceivable interaction effects are estimated.

[12]A worker typically faces a segmented linear rather than a strictly linear budget constraint defining the trade-off between leisure and consumption. Each linear segment is defined by a slope (equal to the net or after-tax wage rate) and an intercept term referred to as "virtual income." If a worker faced a strictly linear budget constraint, the intercept would be equivalent to the amount of nonwage income to which the worker is entitled at zero hours of work.

[13]By definition, Δw and ΔY are directly correlated with preexperimental work effort since they are defined at the preexperimental level of hours. The correlation is nonetheless smaller than it would be in nonexperimental data.

[14]Strictly speaking, it would not concern us if Δw, ΔY, w, and N are correlated with observed variables so long as those variables are included in the estimation equation. As a practical matter, however, a high correlation between, say, w and X makes it difficult to estimate precisely the separate effects of w and X on H.

[15]The studies are Keeley et al. (1978b), Keeley and Robins (1980), Robins and West (1980a), Burtless and Greenberg (1982), Johnson and Pencavel (1982), SRI International (1983), and Johnson and Pencavel (1984). Labor supply elasticities for most of these studies are reported in Keeley (1981), pp. 159-67.

[16]The 21 estimates were not obtained in 21 different studies. Several studies reported separate labor supply estimates for different racial groups. For example, both New Jersey and Rural experimental studies often reported separate labor supply parameters for different racial groups.

[17]I include all elasticities reported by Killingsworth (1983) on pp. 119-122 and pp. 193-197 from U.S. studies where it is possible to compute them. Some individual studies provide several estimates of labor supply response; each response estimate is included with equal weight.

[18]For example, the first simulation considered the response of household heads aged 18 to 58, while the second considered responses of household heads between 16 and 65. The second simulation also used a significantly different method of imputing transfer benefits, which had important consequences for defining the pre-reform situation of low-income families (see below). Standard errors of the simulated national labor supply responses are reported in SRI International (1983), p. 181.

[19]See Committee on Ways and Means, (1986), pp. 370-75.

[20]The labor supply elasticities assumed in the two simulations do not differ very much for female heads. The uncompensated and compensated substitution elasticities and the total income elasticity were -0.03, 0.13, and -0.15, respectively, in the first simulation; they were -0.04, 0.17, and -0.22 in the second. For husbands, in the first simulation the elasticities were 0.02, 0.10, and -0.08; in the second they were -0.13, 0.09, and -0.22. For wives, in the first simulation the elasticities were 0.00, 0.22, and -0.22; in the second they were -0.11, 0.20, and -0.31.

[21]See also Levy (1979) and Moffitt (1985) for a discussion of this issue.

[22]This corresponds to Levy's (1979) findings with respect to AFDC but contradicts Moffitt's (1985) simulation of the effect of a pure negative income tax using nonexperimental labor supply elasticities.

[23]In fact, a majority of current single-parent welfare recipients is predicted to be worse off under three of the four plans examined. Even the most generous plan—offering a 100-percent-of-poverty-line guarantee and 50 percent tax rate—is predicted to make more than one-third of current welfare recipients worse off. See SRI International (1983), p. 189.

[24]These statements may understate the cost of a poverty-line income guarantee in the mid 1980s. The simulations are based on population responses in the mid 1970s when the employment rate of married and single mothers was somewhat lower. Since the labor supply response of women accounts for an important share of the net cost of a more generous program, the budgetary impact of a negative income tax could be higher in the 1980s.

[25]SRI International (1983), pp. 117 and 144.

[26]Table 4 contains four estimates of the net cost of guaranteeing a poverty-line income, two based on an assumed tax rate of 50 percent and two based on a tax rate of 70 percent. Column 6 shows the ratio of earnings reductions to net additional budget outlays. The highest reported ratio for the poverty-line plans is 0.47. This implies that \$1.00 in additional transfer benefits causes a \$0.47 reduction in earnings, suggesting that net income is only \$0.53 (or \$1.00 − \$0.47) higher than it would be without the additional transfers. By implication, taxpayers must spend \$1.89 to raise the net incomes of the poor by \$1.00.

[27]The marginal tax rate in the top three income quintiles is estimated to be about 40 percent (see Browning and Johnson, 1984, p. 184). With this tax rate, a \$12.70 decline in gross wages yields a \$7.62 decline in net wages.

[28]Oddly, Browning and Johnson argue that their simulation predictions are consistent with labor supply elasticities estimated in the experiments (Browning and Johnson, 1984, pp. 190-91). In fact, Browning and Johnson's assumed labor supply elasticities (p. 188) differ markedly from the experimental elasticities reported in table 3. The discrepancies are especially notable in the lowest income quintile.

[29]The results of this simulation are described in Betson, Greenberg, and Kasten (1982), p. 200. For a related discussion, see Betson and Greenberg (1986). We should be cautious in accepting simulations of the taxpayer response to tax increases that are based on response parameters obtained in the negative income tax experiments. The experiments enrolled low-income families; most taxes are paid by middle- and high-income families.

[30]We should carefully distinguish between the earnings effects of a tax increase and the welfare or economic efficiency effects. Even though the gross earnings of taxpayers might rise as a result of a tax increase, the welfare of such taxpayers must decline by at least as much as the added revenue raised by the tax. Thus, even if the net income of taxpayers falls by less than $1.89, the welfare of taxpayers must fall by more than $1.89. Depending on the size of the compensated substitution effect and existing marginal tax rate, the welfare loss could substantially exceed $1.89.

[31]In the long run, for example, wage offers by employers might be higher or the unemployment rate among nonrecipients of a negative income tax might be lower. The latter effect would occur if negative income tax recipients and nonrecipients are in competition for a limited number of jobs.

References

Akerlof, George A. "The Economics of 'Tagging'," *The American Economic Review,* 68 (March 1978), pp. 8-19.

Anderson, Martin. *Welfare,* Stanford, CA: Hoover Institution Press, 1978.

Ashenfelter, Orley, and James Heckman. "The Estimation of Income and Substitution Effects in a Model of Family Labor Supply," *Econometrica,* 42 (January 1974), pp. 73-85.

Betson, David M., and John H. Bishop. "Wage Incentive and Distributional Effects," in R.H. Haveman and J.L. Palmer, eds., *Jobs for Disadvantaged Workers: The Economics of Employment Subsidies,* Washington, DC: The Brookings Institution, 1982, pp. 187-209.

Betson, David M., and David Greenberg. "Labor Supply and Tax Rates: Comment," *The American Economic Review,* 76 (June 1986), pp. 551-56.

Betson, David M., David Greenberg, and Richard Kasten. "A Simulation Analysis of the Efficiency and Distributional Effects of Alternative Program Structures: The NIT Versus the CIT," in I. Garfinkel, ed., *Income-Tested Transfer Programs: The Case For and Against,* New York: Academic Press, 1982, pp. 175-203.

Borjas, George, and James Heckman. "Labor Supply Elasticities and Public Policy," *Proceedings of the 31st Annual Meeting of the Industrial Relations Research Association,* 1978.

Browning, Edgar K., and William R. Johnson. "The Trade-Off Between Equality and Efficiency," *Journal of Political Economy,* 92 (April 1984), pp. 175-203.

Burtless, Gary, and David Greenberg. "Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments," *The American Economic Review,* 72 (June 1982), pp. 488-97.

Burtless, Gary, and Jerry A. Hausman. "The Effect of Taxation on Labor Supply: Evaluating the Gary NIT Experiment," *Journal of Political Economy,* 86 (December 1978), pp. 1103-30.

Committee on Finance, U.S. Senate. *Welfare Research and Experimentation.* Hearings of November 15-17, 1978. Washington, DC: U.S. Government Printing Office, 1978.

Committee on Ways and Means, U.S. House of Representatives. *Background Material and Data on Programs within the Jurisdiction of the Committee on Ways and Means: 1986 Edition,* Washington, DC: U.S. Government Printing Office, 1986.

Fraker, Thomas, Robert Moffitt, and Douglas Wolf. "Effective Tax Rates and Guarantees in the AFDC Program, 1967-82," *The Journal of Human Resources,* 20 (Spring 1985), pp. 264-77.

Greenberg, David, and Harlan Halsey. "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment," *Journal of Labor Economics,* 1 (October 1983), pp. 380-407.

Greenberg, David, Robert Moffitt, and John Friedmann. "Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment," *The Review of Economics and Statistics,* 63 (November 1981), pp. 581-89.

Johnson, Terry R. and John H. Pencavel. "Forecasting the Effects of a NIT Program," *Industrial and Labor Relations Review,* 35 (January 1982), pp. 221-234.

_____."Dynamic Hours of Work Functions for Husbands, Wives, and Single Females," *Econometrica,* 52 (March 1984), pp. 363-89.

Keeley, Michael C. *Labor Supply and Public Policy: A Critical Review,* New York: Academic Press, 1981.

Keeley, Michael C., and Philip K. Robins. "The Design of Social Experiments: A Critique of the Conlisk-Watts Assignment Model," in R. Ehrenberg, ed., *Research in Labor Economics, Vol. 2,* Greenwich, CT: JAI Press, 1980.

Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, and Richard W. West. "The Labor Supply Effects and Costs of Alternative NIT Programs," *The Journal of Human Resources,* 13 (Winter 1978a), pp. 3-36.

_____. "The Estimation of Labor Supply Models Using Experimental Data," *The American Economic Review,* 68 (December 1978b), pp. 873-87.

Kehrer, Kenneth C. *The Gary Income Maintenance Experiment: Summary of Initial Findings,* (mimeo) Gary, IN: Indiana University, 1977.

Killingsworth, Mark R. *Labor Supply,* New York: Cambridge University Press, 1983.

Levy, Frank. "The Labor Supply of Female Heads, or AFDC Work Incentives Don't Work Too Well," *The Journal of Human Resources,* 14 (Winter 1979), pp. 76-97.

Moffitt, Robert A. "The Labor Supply Response in the Gary Experiment," *The Journal of Human Resources,* 14 (Fall 1979), pp. 477-87.

_____. "A Problem with the Negative Income Tax," *Economics Letters,* 17 (1985), pp. 261-65.

Moffitt, Robert A., and Kenneth C. Kehrer. "The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments," in R.G. Ehrenberg, ed., *Research in Labor Economics, Vol. 4,* Greenwich, CT: JAI Press, 1981, pp. 103-50.

Murray, Charles. *Losing Ground: American Social Policy 1950-1980,* New York: Basic Books, 1984.

Office of Income Security Policy, U.S. Department of Health and Human Services. *Overview of the Seattle-Denver Income Maintenance Experiment Final Report,* Washington, DC: U.S. Government Printing Office, 1983.

Rees, Albert. "An Overview of the Labor-Supply Results," *The Journal of Human Resources,* 9 (Spring 1974), pp. 158-180.

Robins, Philip K. "A Comparison of the Labor Supply Findings from the Four NIT Experiments," *The Journal of Human Resources,* 20 (Fall 1985), pp. 567-582.

Robins, Philip K., and Richard W. West. "Labor Supply Response to the Seattle and Denver Income Maintenance Experiments: Alternative Estimates of a Structural Model," (mimeo) Menlo Park, CA: SRI International, January 1980a.

_____. "Labor Supply Response to the Seattle and Denver Income Maintenance Experiments: Analysis of Six Years of Data," (mimeo) Menlo Park, CA: SRI International, May 1980b.

SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1: Design and Results,* Washington, DC: U.S. Government Printing Office, 1983.

U.S. Department of Health, Education, and Welfare. *Summary Report: Rural Income Maintenance Experiment,* Washington, DC: U.S. Government Printing Office, 1976.

Discussion

*Orley C. Ashenfelter**

Having been commissioned to write a paper similar to the one by Gary Burtless at a much earlier state in the development of the negative income tax experiments (Ashenfelter 1978), I had hoped to see a major effort to address some of the puzzles that were evident to any serious scientist examining the early results of those experiments. I am afraid that Burtless has passed over all of these basic issues in his apparent determination to reach strong and definite conclusions about public policy. The result is that Burtless's paper is at best an incomplete catalogue of the research that has already been done with the negative income tax experiment data. At worst it leaves the impression that many of the important reasons for experimentation have now disappeared. Quite to the contrary, I believe most of the important research with social experiments of this type remains to be done. Careful analysis of the data already available and the design and implementation of new and better experiments could have enormous payoffs for our understanding of the effects of public policies on the poor *and* on our understanding of behavior in the labor market.

In order to demonstrate the veracity of my assertion in the limited space available, I will simply take up the two most important issues that troubled me in my review of the rural negative income tax experiment a decade ago. These issues are, as it turns out, of fundamental importance for the interpretation of the results of a negative income tax experiment, and they seem to remain as unresolved now as they were a decade ago.

First, what is the size of the effect of a negative income tax on hours worked? Burtless produces a handy table 2 that, at first blush, provides

*Professor of Economics and Director of the Industrial Relations Section, Princeton University.

the answer to this question for the programs detailed in table 1. Unfortunately, the data in table 2 are taken from the statements by program participants to the survey research houses responsible for data collection in these experiments. A key point about a negative income tax program, however, is that, like a positive income tax, it sets up an incentive for workers to underreport their incomes. The more they can reduce income reported to the experimenter, the greater will be their transfer payments.

In the reports of the New Jersey and rural negative income tax experiments, underreporting was little discussed. As Burtless states, there is some research on this issue in the Gary and Seattle-Denver experiments that indicates that income underreporting is a major (and perhaps the only) cause of the observed decline in earnings in both of these experiments. Of course, the design of these experiments did not incorporate the likelihood that income underreporting would be a serious problem, so the way it is studied is indirect. In particular, earnings from government administrative records are used to measure "true" earnings and then these are compared against the survey data.

The conclusion that Burtless draws from his appraisal of the studies of underreporting is that, "even if misreporting bias causes an exaggeration of the efficiency loss from a negative income tax . . . an earnings reduction caused by underreporting is just as costly to taxpayers as a reduction caused by a genuine reduction in work effort." Although Burtless is only adopting the same conclusion as many others, it seems to me to be in serious error. After all, a genuine negative income tax program *will* operate from government administrative reports on income. Thus, payments in a genuine negative income tax program would be based on the "true" records used by the experimenter here to establish the extent of underreporting. Unless participants actually did change their labor supply behavior or found a way to misreport their income to government officials, it is possible that the additional program costs of a genuine negative income tax scheme attributable to reductions in work effort might be very small. Who is to say whether there would be any labor supply response, further income underreporting, or neither, if an experiment with conventional administrative procedures were implemented? Only an experiment fully informed at the design stage about the possibility for income underreporting, and that tested for its effect, would shed any light on this critical issue. Sadly, the design of none of these experiments was so informed.

A second important issue revolves around the determination of precisely why a labor supply response is produced by a negative income tax experiment. To economists there are effects associated with (a) the size of the tax rate in the program and (b) the generosity of the program. Sorting out these effects is an issue of high priority if the results of the

experiments are to be used to predict the expected response to a program not yet tested. No less importantly, for an economist, the incentive effects of a negative income tax program must operate through variations in the tax rate and generosity of the program, if we are to put much faith in the conventional models of labor supply often used to analyze these issues.

The reports of the results of all these experiments rarely, if ever, provide simple, nonparametric two-way contrasts of labor supply behavior by experimentals and controls. Most analysts estimate parametric models (of the form (4) in Burtless's paper) before providing any tabulation of nonparametric results. A partial exception is the final report on the Seattle-Denver income maintenance experiments (1983). A key finding there is that simple two-way contrasts show no clear evidence that higher tax rates are associated with higher labor supply responses than lower tax rates. Furthermore, no clear relationship was found between program generosity and labor supply responses. My guess is that, at best, the reported magnitudes of income and substitution effects in Burtless's table 3 are based on parametric models so weakly related to the data available that most of the results mainly reflect prior views of the experimenter, and not the actual data. At the very least we are owed some notion of the extent to which the data discipline these results, rather than the prior views of those who calculated them. I find it quite surprising that, a decade after this research was begun, it is still difficult to find out precisely what it is that nonparametric models fit to the basic experimental data reveal, if anything, about the nature of income and substitution effects on labor supply.

References

Ashenfelter, Orley C. "The Labor Supply Response of Wage Earners," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment*, Washington, DC: The Brookings Institution, 1978, pp. 109-48.

SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1: Design and Results*, Washington, DC: U.S. Government Printing Office, 1983.

Discussion

*Robert E. Hall**

No topic could be further removed from discussion in Washington today than a guaranteed income for all Americans, financed by a steep tax on the first few thousand dollars of income. Instead, the whole thrust of policy has been toward tightly limited categorical benefits financed by low marginal rates on all earnings. Hence, the experiments discussed by Burtless and the other authors at this conference cannot be seen now as bearing on policy choices. Rather, they provide data points for scientific investigations of the responses of families to changing economic incentives. I agree strongly with the basic theme of the Burtless paper that the main focus of research should be the incorporation of experimental data into structural labor supply estimation, and not the evaluation of the effects of the particular plans that were the subjects of the experiments.

Burtless in his paper notes the bias toward a finding of high elasticities of labor supply in the experimental data because it was in the interest of the subjects to understate their earnings in order to enlarge their payments. He reviews the attempts that have been made to measure the bias by measuring earnings from extrinsic data. In some cases, such as the Gary experiment, most of the observed decline in hours of work appears to be underreporting. In the Seattle-Denver experiment, primary earners did not underreport but secondary earners did. The reader is left with some unresolved questions: Why go on to use the data that are contaminated by known underreporting later in the paper? Why is underreporting rampant in some instances yet absent in others, where the incentive is just as strong? As the paper stands, it ap-

*Professor of Economics and Senior Fellow, Hoover Institution, Stanford University.

pears that labor supply responses should be studied directly with the extrinsic data, ignoring the reports of the subjects themselves, or at least that studies should be confined to those cases where the problem of underreporting is known to be mild.

The experimental data dramatically improve the variation in the right-hand variables in labor supply estimation. Moreover, thanks to random assignment, the variation is fully exogenous. Hence, both the bias and the randomness of estimated labor supply elasticities are smaller with experimental data than with survey data. In this respect, the scientific value of the experiments has been enormous.

Before looking at the labor supply findings, Burtless considers the biases that arise from the temporary nature of the experiments. He notes that adjustment costs and temporary income subsidies cause the experimental data to understate the long-run effect on labor supply, but that intertemporal substitution causes the data to overstate the long-run effect. His conclusion is that the net effect is an understatement of the response, but I see this as an unsettled issue.

For men, Burtless observes that econometric work has almost universally found that both the substitution and income responses of labor supply are substantially smaller in the experimental data than in survey data. That observation confirms the misgivings that veterans of labor supply estimation in survey data have always had—wages and preferences favoring work are positively correlated in the population. The cross-sectional labor supply function has a positive wage elasticity even if the labor supply function of each individual has zero elasticity. The comparison of two men, one earning $10 per hour and the other $5, shows the former working more hours than the latter. Conclusions about the labor supply functions of either of the men are hard to reach. On the other hand, in the experimental data, we can study a man earning $10 per hour before the experiment, who starts paying a 50 percent tax and hence faces a decline in his wage to $5 per hour. His decline in hours of work is unambiguously a measure of his labor supply elasticity.

For women, the results collected by Burtless show much smaller substitution responses in the experimental data than in the survey data, by an order of magnitude. The high substitution elasticities found in survey data for women are apparently the result of an even higher correlation between wages and preferences favoring work than is the case for men. However, the income responses in the experimental data are larger than those found in survey data, the opposite of what is found for men.

Burtless goes on to apply the labor supply findings to evaluate the effects of possible negative income tax programs for the U.S. economy. As I mentioned at the outset, this exercise is of relatively minor importance, since no plan of this type has any chance of active consideration,

but still it is an interesting way to draw out the implications of the labor supply findings. One of the interesting things we learn as part of the exercise is that the biggest uncertainty about the effect of a move to a negative income tax as a replacement for state-administered welfare programs is the economic characterization of those programs, not the elasticities of labor supply. From the point of view of his table 4, it is just as important to carry out research on benefit levels and implicit tax rates for the existing systems in 51 states as it is to process data from the experiments.

In table 4 and earlier in the paper, Burtless invites evaluation of negative income tax programs in terms of the ratio of earnings reductions to "costs." I find this type of calculation a mystery. A negative income tax is a lump-sum benefit (a demo-grant) paid to every family, financed in part by a tax at a high rate on the first few thousand dollars of earnings of all workers and in part by the general tax system. The cost in terms of resources—government purchases of goods and services—is zero. We could also talk about the cost in the sense of the deadweight burden of the tax, but this is not what Burtless does. Yet another sense of the cost would be the total amount of the lump-sum benefits paid to all families. Again, this is not what he considers. Rather, he makes an economically arbitrary distinction between the revenue raised by the new tax on earnings and the revenue from the existing tax. The "cost" is the difference between the lump-sum benefits and the revenue from the new part of the tax. I cannot see any economic sense in which this is a cost.

Burtless seeks some kind of normalization of the aggregate earnings reduction so that it can be expressed as a percent rather than a total dollar amount. However, his choice of normalization, the "cost," is small, because most negative income tax plans generate most of the revenue needed to finance their lump-sum benefits from their own taxes. Hence his normalized earnings effects are very large. A much superior normalization, in my view, is simply the total amount of earnings. In other words, the percent reduction in earnings is the best normalized way to express the magnitude of the earnings reduction.

An important finding of Burtless's study and many earlier ones is that there is a positive relation between the tax rate and total work effort, even though each worker's labor supply function has a negative relation between his tax rate and his work effort. The reason is that a higher tax rate means that a smaller fraction of workers are subject to the tax. This finding was the explicit rationalization for welfare changes introduced in the early 1980s, when implicit tax rates for the welfare system were raised dramatically.

Burtless notes but does not stress the cruel dilemma of income supplements—under a straight negative income tax, most of the benefits

go to two-parent families, yet correcting this inequality strongly subsidizes the splitting up of families. Since the conference failed to resolve the central question of the impact of welfare and negative income tax incentives on family splitting, it is hard to know how to balance the two goals of helping the neediest most and providing incentives for intact families.

In this paper, Burtless has done a commendable job in bringing together the results of a huge body of research and reducing it to its essential elements.