Volume Title: Annals of Economic and Social Measurement, Volume 5, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-4

Publication Date: October 1976

Chapter Title: The Evaluation of Results from Truncated Samples: The New Jersey Income Maintenance Experiment

Chapter Author: Jerry A. Hausman, David A. Wise

Chapter URL: http://www.nber.org/chapters/c10489

Chapter pages in book: (p. 421 - 445)

Annals of Economic and Social Measurement, 5/4, 1976

THE EVALUATION OF RESULTS FROM TRUNCATED SAMPLES: THE NEW JERSEY INCOME MAINTENANCE EXPERIMENT

BY JERRY A. HAUSMAN AND DAVID A. WISE*

Estimates of the effects of tax and income guarantee values on hours worked by white males in the New Jersey income maintenance experiment are presented after developing a procedure to take explicit account of the "truncating" sampling procedure used to select participants in the experiment. The estimated effects of an income maintenance scheme like that imposed by the experiment are substantially larger than those obtained by other investigators.

Two models are developed. The first deals with one endogenous variable, annual earnings. Our estimates reveal a negative experimental effect of about 6 percent on earnings. They also lead to an estimate of the proportion of variation in income due to "permanent" factors of about 86 percent. This means that even if one were to use only experimental data, the truncation in the first period would lead to parameter estimates with large bias. The second model decomposes earnings into two endogenous variables—wages and hours worked. Our estimates reveal an elasticity of hours worked with respect to the wage rate, or the tax rate, of about 14 percent, and with respect to non-wage income, or the income guarantee, about 2 percent. Because of the truncation, other investigators who did not correct for it often found a negative wage coefficient in equations similar to ours. These coefficients although small, suggest that for persons who elect to participate in an income maintenance scheme the effect on hours worked could be substantial, possibly as high as 16 or 17 percent. It is of interest that the results were surprisingly close to those obtained using pre-experimental observations only.

The oft-touted power of controlled experiments derives from their theoretical ability to isolate the effects of specific actions, treatments, or more general policies. This theoretical ability is based on the assumption of careful randomization. Randomization, however, may be difficult to realize in practice. Retreat from such an optimal state may result from consideration of cost, convenience, technical expertise, legal constraints, ethical bounds, or any number of other compelling reasons. In some cases, such factors lead to rather well defined deviations from random selection and assignment. The primary limit on randomization in the New Jersey Negative Income Tax Experiment, for example, was the restriction of participation in the experiment to families who earned less than one-and-one-half times the poverty level. That is, participants had to have been below this earnings limit in the year just before the experiment began. Although not a hindrance to some uses of the experimental data, this "truncation" does hamper their applicability for others. For example, any uses of the New Jersey data that treat earnings or components of earnings—hours and wages—as endogenous variables, are affected by the truncation. This is true of the other income maintenance experiments as well.

* The research reported herein was performed pursuant to a grant from the Department of Health, Education, and Welfare, Washington, D.C. The opinions and conclusions expressed herein are solely those of the authors, and should not be construed as representing the opinion or policy of any agency of the United States government.

The authors were supported by the above grant through *MATHEMATICA*, to whose staff we owe a considerable debt for assistance in the work. In particular, we thank Francis Drevers for helping to bring together the data used in the study. Several members of the staff, including R. Hollister, C. Metcalf, R. Moffitt, and K. Kehrer, provided helpful comments. Neil Goldman at Harvard provided valuable aid in checking our mathematical derivations. Wendy Gelberg at Harvard drew the graphs and typed the manuscript.

The problem and a method of "correcting" for it were discussed by Hausman and Wise in an earlier paper [1975]. A straightforward maximum likelihood procedure was proposed. The approach suggested in that paper, however, is strictly applicable only to observations that serve as a basis for truncation. In the New Jersey experiment, for example, persons were selected for the experiment on the basis of observations pertaining to the year prior to the experiment. This information was used to "truncate" the sample. Our earlier procedure is directly applicable to data for this year. We would like a procedure for dealing with information collected during the course of the experiment. The experiment lasted for three years. Once selected for the experiment, there was no further restriction on earnings. Family income could have been three times the poverty level in any of the three years of the experiment, for example, as long as it was less than one-and-one-half times the poverty level the year prior to the experiment. In fact, because earnings are so highly correlated from one year to the next, the effect of the truncation was almost as strong for pre-experimental as for experimental data. This means that it is not possible to avoid the truncation problem by using only experimental data. We propose in this paper a method for treating experimental data, given the procedure used to select participants. As in our previous paper, we will "carry along" an empirical example to demonstrate the technique.

The *first section* deals with the case of a single endogenous variable observed before the experiment began as well as during the experiment. Our example will be annual earnings. We are able to estimate, in particular, the average effect of the "treatment" on earnings. The *second section* extends the methodology to a simultaneous equation situation with two endogenous variables observed before and during the course of the experiment. These variables are the components of earnings, the hourly wage and the number of hours worked. This example is particularly relevant to the New Jersey and other income maintenance experiments. The primary goal of all of them is to assess the impact on hours worked or labor supply of various negative income tax schemes. Other methodologies (e.g., Hall [1975]) allow estimation of total experimental effects on hours worked, but do not provide reasonable estimates of income and substitution effects. Both are inputs into reliable estimates of the results and cost of negative income tax plans. To separate the experimental effect into income and substitution components requires a structural model. Such separation is important for evaluating a broad range of possible plans. The experiment allowed for eight possible income guarantee—marginal tax rate combinations. There are, of course, many other possibilities. One might suppose that relevant parameters could simply be estimated from non-experimental observations, and they have been; but under existing tax laws, no observations of low income families would be observed with the high marginal tax rates (as high as 70 percent) imposed in the experiment. The *third section* contains a short summary. Finally, computational considerations are discussed in an *appendix*.

1. A Single Endogenous Variable: Earnings

To aid in exposition, we will begin by describing the situation where the analysis is restricted to pre-experimental data. These paragraphs and graphs are borrowed from our previous work. We will then extend this heuristic discussion to the case where both pre-experimental and experimental observations are considered. This latter situation is then described in a more formal manner.

Assume that in the population the relationship between earnings and exogenous variables is of the form

$$Y_i = X_i\beta + \varepsilon_i, \tag{1.1}$$

where Y is earnings; X is a vector of exogenous variables including education, intelligence, etc. . . . ; i indexes individuals; β is a vector of parameters; and ε is a disturbance term with expected value zero and variance σ^2 for each individual. Thus Y_i is distributed normally with mean $X_i\beta$ and variance σ^2, $N(X_i\beta, \sigma^2)$.

The sample we have is not, however, randomly drawn from the population, nor from some segment of the population defined by values of the exogenous variables, and therefore is not representative of it. Families were selected from an otherwise eligible population in four cities in New Jersey and Pennsylvania. Those families who were subsequently found to have incomes during the year preceding the experiment, greater than one and one-half times the 1967 poverty line were eliminated from the study. The poverty line is dependent on family size; therefore the "cut-off" point varied from family to family. The truncation thus takes the form,

$$Y_i < L_i \tag{1.2}$$

where L_i depends on family size. [The reader will note that equation (1) pertains to individual income, while the truncation is based on family income. For the time being, we will act as if the two were the same and return to the problem when discussing the empirical example.] If we substitute for Y_i, the final selection criterion for families considered for inclusion in the experiment could be stated as follows:

$$\begin{aligned} Y_i &= X_i\beta + \varepsilon_i \leq L_i, \text{ included} \\ Y_i &= X_i\beta + \varepsilon_i > L_i, \text{ excluded} \end{aligned} \tag{1.3}$$

where Y_i pertains to earnings during the year prior to the experiment. This formulation affords an explicit comparison with the "Tobit" situation. We discussed it in our earlier paper, but it may warrant re-emphasis here. In the Tobit case L_i is equal to some L for all i (although this is not logically necessary), and Y_i would be *equal* to L for Y_i greater than L. Here we can think of a measuring device that misses all observations above L, rather than assigning them the value L. Both statistical models, however, are members of a wider class of models associated with truncated distributions.

To fix ideas, consider the following graph, where the solid line indicates the "average" relation between education and earnings and the dots represent the distribution of earnings around this mean for selected values of education. Assume that family size is the same for each observation. All individuals with earnings above a given level L, indicated by the horizontal line would be eliminated from the experiment. In estimating the effect of education on earnings, using pre-experimental data, we would observe only the points below the limit (circled) and would thus tend to underestimate the effect of education using ordinary least squares. In other words, the sample selection procedure introduces correlation between right-hand variables and the error term, which we know leads to biased parameter estimates. The estimated regression line is dashed in the graph. From the graph, it can be seen that the magnitude of the bias depends on L, β, σ^2, and the values of X_i.

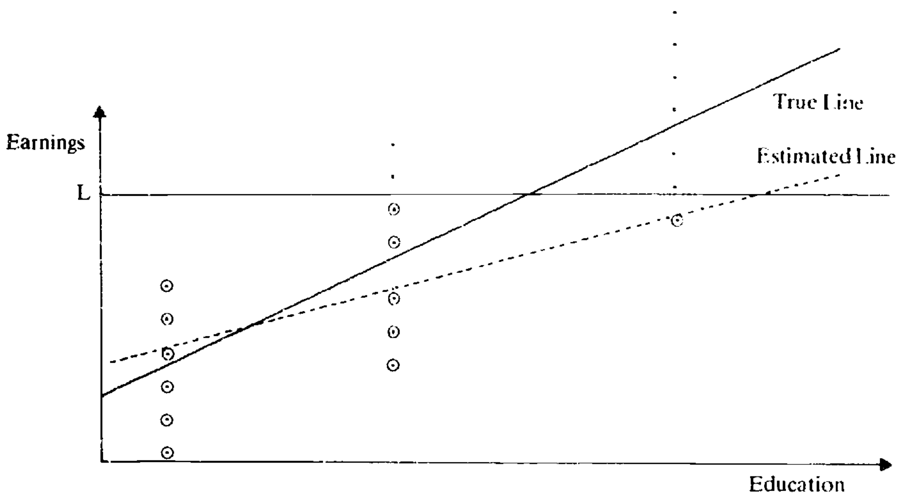

Figure 1

Thus for any given value of education (or in general X), the observed distribution of earnings during the year prior to the experiment can be thought of as truncated at L, where the "extent" of the truncation depends on the level of education. Graphically, for a given X_i the distribution of Y_i may look as follows, with the right-hand tail eliminated.

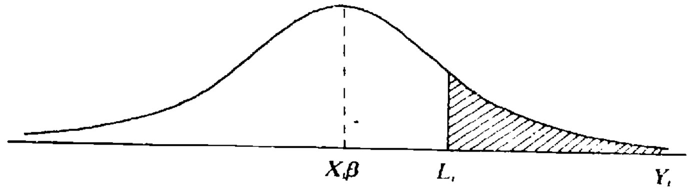

Figure 2

Now assume that we observe earnings prior to *and* during the experiment. We would like to use both. Let prior observations be indexed by 1 and those during the experiment by 2. Then

(1.4)
$$Y_{1i} = X_{1i}\beta + \varepsilon_{1i}, \quad \text{and}$$
$$Y_{2i} = X_{2i}\beta + \varepsilon_{2i},$$

where ε_1 and ε_2 are jointly normal with zero expected values. Observations are available, however, only for persons for whom Y_{1i} was less than or equal to L_i. For given X_{1i} and X_{2i}, the joint distribution of Y_{1i} and Y_{2i} may look like that in Figure 3. The ellipses represent equal probability contours and $X_1\beta$ and $X_2\beta$ are the means in the population of Y_1 and Y_2 respectively. The truncation on earnings in period 1 precludes observation of points to the right of L_i. The idea is analogous to that depicted in Figure 2 except that we now need to consider the bivariate distribution of Y_1 and Y_2.

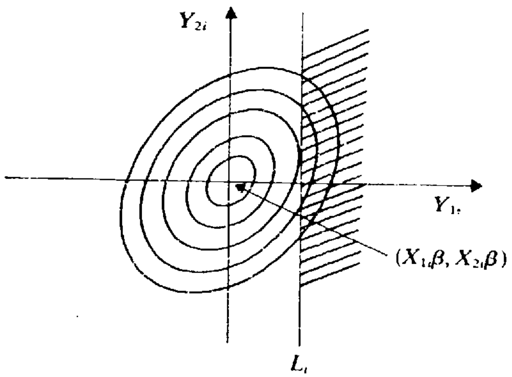

Figure 3

More formally, we need to find the joint distribution of Y_1 and Y_2 for values of Y_1 less than or equal to L. Let f be their joint density function. Assume that in the population, given X_{1i} and X_{2i}, Y_{1i} and Y_{2i} are jointly normal with mean vector $(X_{1i}\beta, X_{2i}\beta)$ and covariance matrix Σ given by,

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{bmatrix}. \tag{1.5}$$

Then,

$$f(Y_{1i}, Y_{2i}) = \begin{cases} 0 \quad \text{if } Y_{1i} > L_i, \quad \text{and} \\ \dfrac{\tilde{\phi}(Y_{1i}, Y_{2i})}{\Pr(Y_{1i} \leq L_i)} \quad \text{if } Y_{1i} \leq L_i. \end{cases} \tag{1.6}$$

In this formulation, $\tilde{\phi}(Y_{1i}, Y_{2i})$ is a bivariate normal density function with the mean vector and the covariance matrix shown above. Note that the probability that Y_{1i} is less than or equal to L_i can be written as $\Phi[(L_i - X_{1i}\beta)/\sigma]$. We can now write the likelihood function for N observations on Y_1 and Y_2 as,

$$L = \prod_{i=1}^{N} f(Y_{1i}, Y_{2i}) = \prod_{i=1}^{N} \frac{\tilde{\phi}(Y_{1i}, Y_{2i})}{\Phi[(L_i - X_{1i}\beta)/\sigma]}. \tag{1.7}$$

By writing $\tilde{\phi}(Y_{1i}, Y_{2i})$ as the product of the marginal distribution of Y_{1i} and the conditional distribution of Y_{2i} given Y_{1i}, we can replace it with,

$$(1.8)\quad \tilde{\phi}(Y_{1i}, Y_{2i}) = \phi\left(\frac{Y_{1i}-X_{1i}\beta}{\sigma}\right)\phi\left(\frac{Y_{2i}-\rho Y_{1i}-(X_{2i}-\rho X_{1i})\beta}{\sigma\sqrt{1-\rho^2}}\right)\left(\frac{1}{\sigma}\right)\left(\frac{1}{\sigma\sqrt{1-\rho^2}}\right),$$

where $\phi(\cdot)$ represents a unit normal density function and ρ is the correlation coefficient between Y_1 and Y_2.

We can write the log-likelihood function using (1.7) as,

$$(1.9)\quad \mathscr{L} = \frac{N}{2}\ln D - \frac{1}{2}\sum_{i=1}^{N}[(Y_{1i}-X_{1i}\beta,\ Y_{2i}-X_{2i}\beta)\Sigma^{-1}(Y_{1i}-X_{1i}\beta,\ Y_{2i}-X_{2i}\beta)'] - \sum_{i=1}^{N}\ln\Phi[(L_i-X_i\beta)/\sigma],$$

where D is the determinant of Σ^{-1}, or, taking advantage of the relationship (1.8), as

$$(1.10)\quad \mathscr{L} = -N\ln 2\pi - N\ln(\sigma^2\sqrt{1-\rho^2}) - \tfrac{1}{2}\sum_{i=1}^{N}[(Y_{1i}-X_{1i}\beta)/\sigma]^2 - \sum_{i=1}^{N}\ln\Phi[(L_i-X_{1i}\beta)/\sigma] - \tfrac{1}{2}\sum_{i=1}^{N}\left[\left((Y_{2i}-\rho Y_{1i})-(X_{2i}-\rho X_{1i})\beta\right)\Big/(\sigma\sqrt{1-\rho^2})\right]^2.$$

This is the form that we will use for estimation.

We will find the values of β, σ^2, and ρ that maximize this function for a given specification of (1.4). (The maximization technique is discussed in the appendix.) Recall that ρ is an estimate of the proportion of the variation in Y, given X, that results from "permanent" versus "transitory" components of the random term ε. If we write ε_{it} as $\varepsilon_{it} = u_i + \eta_{it}$, where t indexes time, and u is an individual specific effect: and we assume that $E(u_i\eta_{it}) = 0$ and $E(\eta_{it}\eta_{it'}) = 0$ for $t \neq t'$, then ρ is given by,

$$\rho = \sigma_u^2/(\sigma_u^2+\sigma_\eta^2).$$

An Example: One Endogenous Variable

We will present an example analogous to the one used in our previous work. That example is extended to handle observations for two time periods. In addition, while the data for the earlier example covered both whites and nonwhites, the data used in this case will be restricted to whites.[1]

The technique is demonstrated with empirical estimates of the effect of education and "intelligence," as well as the experimental "treatment," on the

[1] The attrition rate for blacks was considerably higher than for whites. See Peck in vol. 2, pt. c, chap. 1 of Watts and Rees [1974].

earnings of workers in "low-level" jobs with relatively little education. The mean level of education of persons in the sample is 8.76 years. They are distributed among occupations (census classifications) as follows: 2.4 percent are professional, technical, and kindred workers, or managers, officials and proprietors; 6.1 percent are clerical or sales workers; 63.1 percent craftsmen, foremen, operatives, and kindred workers; 10.8 percent private household or service workers; and 17.6 percent are laborers. The experimental treatment consisted of marginal tax rates and income guarantees assigned to the "experimental" group, but not to "controls." There were eight different tax rate guarantee level combinations. In this example we will test only for an "experimental effect," and will not distinguish between the different combinations. In the simultaneous equation example below, however, we will take account of differences across individuals in marginal tax rates and income guarantees or non-wage income.

The sample is comprised of male heads of households who participated in the negative income tax experiment, and for whom relevant information was available. The truncation point for each observation was taken to be one and one-half times the appropriate poverty level, less any family income other than the earnings of the male head. If the male earned more than this, given other family income, the observation would not appear in the sample.

The variables used are defined as follows:

Earnings in period 1: gross annual wage income during the year prior to the experiment. (Average reported earnings per week times the number of weeks worked.) Depending on the city of residence, this year ended between August 1968 and August 1969.

Earnings in period 2: average of weekly earnings over twelve weeks (one in each quarter) during the experiment, multiplied by the number of weeks worked in the previous period.

Education: education in years.

"*IQ*": the number of correct answers, out of 50 questions, on the Ammons and Ammons Quick Test.

Age: age in years.

Experience: age, minus the age at which the individual reported obtaining his first full-time job.

Union: takes the value one if the individual reported being a member of a union, zero otherwise.

Training: months of training of a vocational nature.

Illness: takes the value one if the individual reported that he had an illness that limited his working.

Time: takes the value one for the experimental period and zero for the pre-experimental period.

Experimental effect: takes the value one in period 2 if the individual is in the experimental group (not a control), and zero otherwise.

The following variables are not used in the single endogenous variable example, but are used in the simultaneous equation example below. For completeness, they are defined here.

Hours in period 1: earnings during the year prior to the experiment, divided by an estimated wage. The wage was estimated by the average of earnings, divided by hours worked during the weeks prior to the second through fifth quarterly interviews. Only those weeks for which positive earnings were reported were used in the average.

Hours in period 2: average of hours worked over twelve weeks (one in each quarter) during the experiment, multiplied by the number of weeks worked in the previous period.

Wage in period 1: average of earnings divided by hours worked during the weeks prior to the second through fifth quarterly interviews. Only those weeks for which positive earnings were reported were used in the average.

Wage in period 2: same as above except that data for twelve quarters were used.

Family size: the number of persons in the family.

Coefficient estimates are presented in Table 1. The dependent variable is the logarithm of earnings. The independent variables are entered linearly (not in logarithm form). The maximum likelihood procedure estimates β, σ^2, and ρ in equation (1.10), where the logarithm of earnings is Y.

Comparable results pertaining to the pre-experimental period only were discussed in our previous work. The primary result was that least squares coefficient estimates were found to be strongly biased toward zero in comparison with consistent estimates analogous to these. Here we will draw attention only to those estimates that pertain to the results of the experiment. The estimated experimental effect on earnings is 6 percent and negative. It is interesting to note that the average change in the logarithm of earnings of the controls was 0.190 and for the experimental group, 0.134.[2] The difference of 0.056 is quite close to the difference estimated as the experimental effect, 0.058, in the "behavioral" equation. This suggests that the two groups were in fact selected randomly with respect to the variables in the equation. This is, of course, what the experimental randomization was designed to do.

The variation in earnings, given the independent variables controlled for, due to "permanent" versus "transitory" factors, is estimated by $\hat{\rho}$ at almost 86 percent. This strongly suggests that the bias of least squares estimates found in results based on the pre-experimental data would persist even if experimental data only were used.

Finally, the positive time effect of about 3 percent may be thought of as representing the effect of inflation, as well as other factors which may have influenced the trend in earnings of both controls and experimentals from the first period to the second.

2. Two Endogenous Variables: Wages and Hours

The primary goal of the New Jersey experiment—and the goal of others, some still in process—was to determine the effect of "income maintenance" schemes on labor supply. Thus we would like to isolate the effect of the experimental programs on labor supply, while at the same time taking account of the

[2] The average over both groups of the logarithm of earnings in the year prior to the experiment was 8.275. It was 8.431 for the experimental period.

TABLE 1

ONE ENDOGENOUS VARIABLE, RESULTS*

Variable	Estimated Coefficient (Standard Error)
Constant	8.5778 (0.0343)
Education	0.0343 (0.0198)
IQ	0.0042 (0.0043)
Training	0.0132 (0.0109)
Union	0.3248 (0.1843)
Illness	−0.5497 (0.2240)
Age < 35	0.0209 (0.0197)
Age 35 to 45	0.0064 (0.0058)
Age > 45	−0.0317 (0.0444)
Time	0.0274 (0.0185)
Experimental Effect	−0.0579 (0.0305)

Number of Individuals = 276
Number of Experimentals = 156
Number of Controls = 120
Number of Observations = 276 · 2 = 552

S.E.E. = $\hat{\sigma}$ = 0.589
Correlation Coefficient $\hat{\rho}$ = 0.859

* The three age variables beginning with the variable corresponding to age <35 are defined as follows:

$$A_1 = \begin{cases} \text{Age, if Age} \le 35 \\ 35, \text{ if Age} > 35 \end{cases}$$

$$A_2 = \begin{cases} 0, \text{ if Age} \le 35 \\ \text{Age} - 35, \text{ if } 35 < \text{Age} \le 45 \\ 45 - 35, \text{ if Age} > 45 \end{cases}$$

$$A_3 = \begin{cases} 0, \text{ if Age} \le 45 \\ \text{Age} - 45, \text{ if Age} > 45 \end{cases}$$

truncation introduced by the selection procedure. In addition, even if we were only interested in using the data to investigate the effect of "academic variables" on earnings, the above approach would have at least two shortcomings.

It obscures the process by which earnings are generated; they result from a choice of hours of work made by the individual together with the hourly wage that he commands in the market. And, when investigating the relationship between personal attributes and "productivity" what we really would like to know is the wage per unit of time that an individual commands in the market, his "marginal product." This relationship is partly hidden when we look only at annual earnings. In addition, the variance of the error term in earnings, the product of hours and a wage rate, is larger than that of a wage equation. Thus the accuracy with which we can estimate the effect of I.Q., for example, should be greater if we break the relationship into its component parts. It would be possible in general to consider a wage equation separately, although at some expense in efficiency, given the simultaneous nature of the wage-hour relationship. But in our particular case the truncation point is based on annual earnings, so that if we consider hourly wage we must also consider hours worked.

Recall that $Y = H \cdot W$, where H is hours of work and W is the hourly wage. Thus $\ln Y = \ln H + \ln W$. Assume that in the population $\ln W$ and $\ln H$ are jointly distributed with,

$$\begin{aligned} \ln W_{1i} &= X_{1i}\delta + \varepsilon_{1i} \\ \ln H_{1i} &= \ln W_{1i}\beta + Z_{1i}\alpha + \eta_{1i} \\ \ln W_{2i} &= X_{2i}\delta + \varepsilon_{2i} \\ \ln H_{2i} &= \ln W_{2i}\beta + Z_{2i}\alpha + \eta_{2i}, \end{aligned} \tag{2.1}$$

where X and Z are vectors of exogenous variables, δ and α vectors of parameters, and β is a scalar parameter. We will let the hours equation depend on the wage rate net of taxes in practice. To simplify exposition, however, we will proceed with this model for the time being and make the appropriate alterations below.

We assume that ε_1, η_1, ε_2, and η_2 are jointly normal with covariance matrix given by,

$$\Sigma = \begin{array}{c} \begin{array}{cccc} \varepsilon_1 & \eta_1 & \varepsilon_2 & \eta_2 \end{array} \\ \left[\begin{array}{cc|cc} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\eta} & \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\eta_2} \\ \sigma_{\varepsilon\eta} & \sigma_{\eta\eta} & \sigma_{\varepsilon_1\eta_2} & \sigma_{\eta_1\eta_2} \\ \hline \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\eta_2} & \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon_1\eta_2} & \sigma_{\eta_1\eta_2} & \sigma_{\varepsilon\eta} & \sigma_{\eta\eta} \end{array}\right] \end{array} \begin{array}{c} \\ \varepsilon_1 \\ \eta_1 \\ \varepsilon_2 \\ \eta_2 \end{array} . \tag{2.2}$$

Note that variances and covariances in the two periods are assumed equal (the upper left and lower right two-by-two matrices). We also assume that the covariance between ε_1 and η_2 is equal to the covariance between ε_2 and η_1.

For some purposes, it is informative to think of the random terms in (2.1), and the corresponding variances and covariances, as having both "permanent" and "transitory" components. The relevant components of variance can in fact be identified, given estimates of the parameters in (2.1). To see this, let

$$\begin{aligned} \varepsilon_{1i} &= u_i + e_{1i} \\ \eta_{1i} &= v_i + n_{1i} \\ \varepsilon_{2i} &= u_i + e_{2i} \\ \eta_{2i} &= v_i + n_{2i} \end{aligned} \tag{2.3}$$

where $\sigma^2_{e_1}=\sigma^2_{e_2}$, $\sigma^2_{n_1}=\sigma^2_{n_2}$, and $\sigma_{ue}=\sigma_{vn}=\sigma_{e_1e_2}=\sigma_{n_1e_2}=0$. Then,

$$
\begin{aligned}
\sigma_{\varepsilon\varepsilon} &= \sigma^2_u + \sigma^2_e \\
\sigma_{\eta\eta} &= \sigma^2_v+\sigma^2_n \\
\sigma_{\varepsilon\eta} &= \sigma_{uv}+\sigma_{en} \\
\sigma_{\varepsilon_1\varepsilon_2} &= \sigma^2_u \\
\sigma_{\eta_1\eta_2} &= \sigma^2_v \\
\sigma_{\varepsilon_1\eta_2} &= \sigma_{uv},
\end{aligned}
\tag{2.4}
$$

and we see that, given the above restrictions, the parameters of the variance components formulation are identified.

To estimate the parameter of (2.1) and (2.2), we will use the reduced form of (2.1), given by:

$$
\begin{aligned}
\ln W_1 &= X_1\delta+\varepsilon_1 \\
\ln H_1 &= X_1\delta\cdot\beta+Z_1\alpha+\varepsilon_1\beta+\eta_1 \\
\ln W_2 &= X_2\delta+\varepsilon_2 \\
\ln H_2 &= X_2\delta\cdot\beta+Z_2\alpha+\varepsilon_2\beta+\eta_2.
\end{aligned}
\tag{2.5}
$$

The reduced form covariance matrix is given by,

$$
\Omega=\left[\begin{array}{cc|cc}
\omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} \\
\omega_{12} & \omega_{22} & \omega_{14} & \omega_{24} \\
\hline
\omega_{13} & \omega_{14} & \omega_{11} & \omega_{12} \\
\omega_{14} & \omega_{24} & \omega_{12} & \omega_{22}
\end{array}\right]=\left[\begin{array}{c|c}\Omega_1 & \Omega_2 \\ \hline \Omega_2 & \Omega_1\end{array}\right],
\tag{2.6}
$$

where,

$$
\Omega_1=\begin{bmatrix}\sigma^2_\varepsilon & \sigma^2_\varepsilon\beta+\sigma_{\varepsilon\eta} \\ \sigma^2_\varepsilon\beta+\sigma_{\varepsilon\eta} & \sigma^2_\varepsilon\beta+\sigma_{\varepsilon\eta}2\beta+\sigma^2_\eta\end{bmatrix},
$$

and

$$
\Omega_2=\begin{bmatrix}\sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\varepsilon_2}\beta+\sigma_{\varepsilon_1\eta_2} \\ \sigma_{\varepsilon_1\varepsilon_2}\beta+\sigma_{\varepsilon_1\eta_2} & \sigma_{\varepsilon_1\varepsilon_2}\beta^2+\sigma_{\varepsilon_1\eta_2}2\beta+\sigma_{\eta_1\eta_2}\end{bmatrix}.
$$

Note that there are only six unknown parameters in the covariance matrix, giving it a rather simple structure.

The joint density function for the logarithms of wages and hours in the two periods, analogous to the development in section 1, is given by,

$$
f(\cdot)=\begin{cases} 0, \text{ if } \ln W_{1i}+\ln H_{1i}>\ln L_i, \text{ and} \\ \dfrac{\tilde{\phi}(\ln W_{1i}, \ln H_{1i}; \ln W_{2i}, \ln H_{2i})}{\Pr(\ln W_{1i}+\ln H_{1i}\le \ln L_i)}, \text{if } \ln W_{1i}+\ln H_{1i}\le L_i. \end{cases}
\tag{2.7}
$$

Here, $\tilde{\phi}$ is a multivariate normal density function with mean vector given by the non-random terms in equations (2.5) and covariance matrix Ω, shown in (2.6). To evaluate the denominator in (2.7), recall that $\ln W_{1i}$ and $\ln H_{1i}$ are distributed bivariate normal with mean vector $(X_1\delta, X_1\delta\beta + Z_1\alpha)$ and covariance matrix Ω_1. Then, the distribution of $\ln W_{1i} + \ln H_{1i}$ is given by

$$\ln W_{1i} + \ln H_{1i} \sim N(X_1\delta + X_1\delta\beta + Z_1\alpha\,;\, \omega_{11} + \omega_{22} + 2\omega_{12}). \tag{2.8}$$

The likelihood function for N observations is thus,

$$\mathscr{L} = \prod_{i=1}^{N} \tilde{\phi}_i(\cdot)/\Phi[d_i], \tag{2.9}$$

where

$$\tilde{\phi}_i(\cdot) = \tilde{\phi}(\ln W_{1i}, \ln H_{1i}; \ln W_{2i}, \ln H_{2i}),$$

$$d_i = (\ln L_i - (X_{1i}\delta + X_{1i}\delta\beta + Z_{1i}\alpha))/\sqrt{\omega_{11} + \omega_{22} + 2\omega_{12}},$$

and the log-likelihood function by,

$$\mathscr{L} = \sum_{i=1}^{N} \ln \tilde{\phi}_i(\cdot) - \sum_{i=1}^{N} \ln \Phi[d_i]. \tag{2.10}$$

If we let,

$$V_{11i} = \ln W_{1i} - X_{1i}\delta,$$

$$V_{12i} = \ln H_{1i} - X_{1i}\delta\beta - Z_{1i}\alpha,$$

$$V_{21i} = \ln W_{2i} - X_{2i}\delta,$$

$$V_{22i} = \ln H_{2i} - X_{2i}\delta\beta - Z_{2i}\alpha,$$

and

$$V_i = (V_{11i}, V_{12i}, V_{21i}, V_{22i}),$$

the log-likelihood function is given by

$$\mathscr{L} = \frac{N}{2}\ln D - \tfrac{1}{2}\sum_{i=1}^{N} V_i\Omega^{-1}V_i' - \sum_{i=1}^{N} \ln \Phi[d_i]. \tag{2.11}$$

If, as in section 1, we take advantage of the fact that $\tilde{\phi}_i(\cdot)$ can be written as the product of marginal and conditional density functions—both bivariate in this case, we can write (2.11) as,

$$\begin{aligned}\mathscr{L} = {} & \frac{N}{2}\ln(\operatorname{Det}\Omega_1^{-1}) + \frac{N}{2}\ln(\operatorname{Det} B^{-1}) \\ & - \tfrac{1}{2}\sum_{i=1}^{N} V_{1i}\Omega_1^{-1}V_{1i'} - \sum_{i=1}^{N} \ln \Phi[d_i] \\ & - \tfrac{1}{2}\sum_{i=1}^{N} (V_{2i} - CV_{1i'})B^{-1}(V_{2i} - CV_{1i'})',\end{aligned} \tag{2.12}$$

where $B = \Omega_1 - \Omega_2\Omega_1^{-1}\Omega_2$, $C = \Omega_2\Omega_1^{-1}$, $V_{1i} = (V_{11i}, V_{12i})$, and $V_{2i} = (V_{21i}, V_{22i})$. We have chosen to use this form for estimation of the parameters of the model—δ,

β, α, and the six parameters in the covariance matrix (2.2). The precise procedure used is discussed more fully in the appendix. We proceed now to an empirical example.

An Example: Two Endogenous Variables

Before presenting empirical estimates, we need to digress somewhat to describe more precisely the "world" that generated our data, and the concomitant problems of estimation. The discussion is directed in particular to the specification of the hours worked equations in (2.5). We will consider first the specification of these equations for "non-experimentals" (all participants before the experiment began—period 1—and controls during the experiment—period 2), and related estimation problems. The specification suggested for this case will then be extended to include the "experimental" group.

Consider the graph in Figure 4, where the solid line represents an "after-tax" budget constraint and the dotted line the "pre-tax" constraint. Assume that, faced with the after-tax budget constraint, an individual chooses to work h hours. Assume that this point represents a tangency between the budget constraint and an indifference curve. Then this individual would work the same number of hours if faced with the linear budget constraint represented by the tangent to the true constraint at $H = h$. This "as if" constraint may be completely described by two values: its slope, $\tilde{W}$, and $\tilde{Y}_0$, "adjusted" non-wage income.[3] The point H^* represents the number of hours the individual can work before the marginal tax rate becomes greater than zero. Non-wage income $\tilde{Y}_0$ is, for our purposes, all family income other than wage income of the male head of the household.

The hours equations in (2.1) can now be specified as,

$$
\begin{aligned}
(2.13)\qquad \ln H_{1i} &= \ln [W_{1i}(1-t_{1i})]\beta + \ln \tilde{Y}_{01i}\alpha_1 + Z_{1i}\alpha + \eta_{1i}\\
\ln H_{2i} &= \ln [W_{2i}(1-t_{2i})]\beta + \ln \tilde{Y}_{02i}\alpha_2 + Z_{2i}\alpha + \eta_{2i},
\end{aligned}
$$

where t_i is the marginal tax rate "faced" by the ith individual. The value $\ln \tilde{Y}_0$ can be thought of as one of the elements of the vector Z in (2.1). The reduced form equations in (2.5) now become,

$$
\begin{aligned}
(2.14)\qquad \ln H_{1i} &= [X_{1i}\delta + \ln (1-t_{1i})]\beta + \ln \tilde{Y}_{01i}\alpha_1 + Z_{1i}\alpha + \varepsilon_{1i}\beta + \eta_{1i}\\
\ln H_{2i} &= [X_{2i}\delta + \ln (1-t_{2i})]\beta + \ln \tilde{Y}_{02i}\alpha_1 + Z_{2i}\alpha + \varepsilon_{2i}\beta + \eta_{2i}.
\end{aligned}
$$

Note that $\tilde{Y}_0$ is taken, at this point, to be $\tilde{Y}_0 = HW + Y_0 - T - H\tilde{W} = HWt + Y_0 - T$, where T is total taxes paid. We see that both $X\delta + \ln(1-t) = \ln \tilde{W}$ and $\tilde{Y}_0$ depend on hours worked and are thus correlated with the disturbance terms in the hours equations. We know which tax rate an individual faces for any given year, ex post—and for experimentals, whether or not they were "on" the experiment as described below—but it is endogenous, in that it depends on endogenous variables. To circumvent this difficulty we will evaluate t and $\tilde{Y}_0$ at

[3] So far, this is a variant of a procedure used by Hall [1973]. His method, however, does not take account of the "endogeneity" of both $\tilde{W}$ and $\tilde{Y}_0$. Our method for handling this is taken up shortly.

the same number of hours for all persons.[4] We have chosen 1,500 hours: this is about the average number of hours worked per year by participants in the experiment.[5]

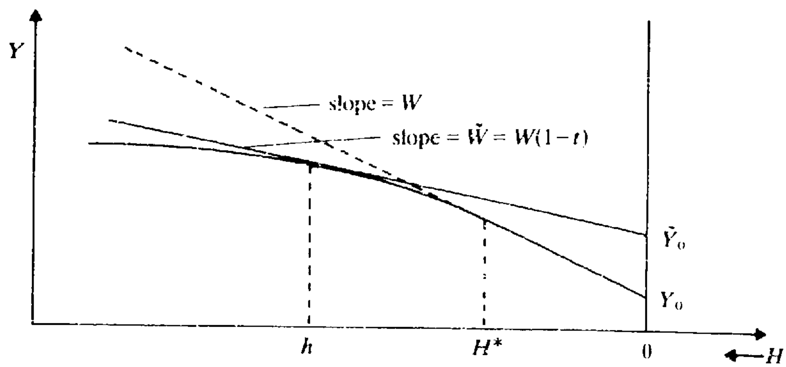

Figure 4

To assign comparable values of $\tilde{W}$ and $\tilde{Y}_0$ to the experimental group, we need to consider the nature of the experimental "treatment." Each family in the experimental group was assigned a tax rate. call it τ, and a guaranteed income G. Family income would be at least G. even if no family members worked. Associated with each family is a "break-even" point. It is defined as the number of hours that would lead to income under the experiment equal to income that would have been earned had the family not been in the experimental group. We let this value be H^{**}. There is a kink in the budget constraint at this point.[6] Beyond H^{**} the individual no longer faces the treatment tax rate; he is faced with the same rate as a non-experimental participant with his characteristics. The idea is demonstrated graphically in Figure 5. The guarantee is for the family and all family income is taxed at the rate τ. Because we are considering the labor supply of the

[4] We have discovered that a comparable procedure was followed by Rosen [1974]. This procedure is. of course, rather ad hoc. and is limited by the fact that an endogenous variable. the observed wage rate, is used in the calculations. In planned subsequent work we will be more systematic in taking account of the "on"-"off" decision.

[5] For example. consider the "as-if" tax rate of a person whose wage is W and whose non-wage income is \$1,125 and who has 5 dependents. He pays (in 1970) no federal income tax on income up to \$1,100+5(\$625) = \$4,125. He then pays a marginal rate of about 18 percent up to \$10.000. (Note that persons with income less than \$10,000 who take standard deductions do *not* face a simple progressive marginal tax rate, which for higher income families started at 14 percent in 1970. The 18 percent figure is taken to represent the average for this group, if taxes are paid at all. Because all persons in our sample have low incomes, we will assume this marginal rate for everyone. after standard deductions are taken.) He will also pay social security tax at the rate of 4.8 percent up to \$7,800 (1970). (We assume this rate for all income of persons in our sample.) If the wage rate of this individual is less than \$2.00, his as-if tax rate is taken to be 0.048: if it is greater than \$2.00 it would be $0.180+0.048 = 0.228$. The "as-if" non-wage income would be $\tilde{Y}_0 = 1.500Wt + Y_0 - T$. where t is the tax rate calculated above and T is (1,500)(0.048)(W) + (federal income tax at $H = 1.500$). Federal income tax would be zero if W were less than \$2.00. otherwise it would be $(1,500 - H^*)Wt$, where H^* is (\$4,125 − \$1,125)/W. Social security tax would be (1.500)(0.048)(W).

[6] In fact, the procedures of the experiment took account of federal income tax in the determination of the break-even point, but not social security tax. This is presumably a minor discrepancy.

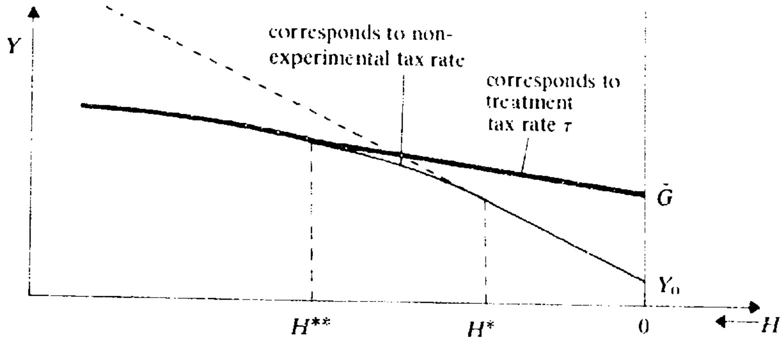

Figure 5

male head of the family we will define a $\tilde{G} = G - \tau Y_0 + Y_0$ where Y_0 is family income other than his wage income.

We will define an "as-if" tax rate and a non-wage income level for this group analogous to those for the non-experimental group. Again, we ask what marginal tax rate the individual would be facing if he worked 1,500 hours. This rate is the marginal tax rate assigned to him.[7] The value for $\tilde{Y}_0$ is defined as above. Note that $\tilde{Y}_0 = \tilde{G}$ if the individual is assigned that treatment tax rate. If H^{**} is greater than 1,500 hours, the experimental tax rate τ is assigned. If H^{**} is less than or equal to 1,500, the appropriate non-experimental rate is assigned.[8]

The results for two endogenous varaibles are shown in Table 2. Again we will draw attention only to those estimates that are of particular relevance to the goals

[7] In planned subsequent work we will be more systematic in predicting whether or not an individual will be "on" or "off" the experiment. The problem is similar to predicting, on general, the tax rate an individual will face.

[8] Algebraically, the procedure for non-experimentals and experimentals is as follows: For *non-experimentals*, calculate

$$H^* = (1{,}100 + F \cdot 625 - Y_0)/W,$$

where F is the family size. The tax rate is given by,

$$t = \begin{cases} 0.048 & \text{if } H^* > 1{,}500, \\ 0.048 + 0.18 & \text{if } H^* \leq 1{,}500, \end{cases}$$

where 0.048 is social security tax rate and 0.18 federal income tax rate. The "as-if" non-wage income is given by,

$$\hat{Y}_0 = 1{,}500Wt + Y_0 - T,$$

where

$$T = \begin{cases} 1{,}500W(0.048) & \text{if } H^* > 1{,}500, \\ 1{,}500W(0.048) + (1{,}500 - H^*)Wt & \text{if } H^* \leq 1{,}500. \end{cases}$$

For *experimentals*, H^{**} is given by,

$$H^{**} = \begin{cases} (G - Y_0 - tH^*W)/W(\tau - t) & \text{if } G + WH^*(1-\tau) > Y_0 + H^*W, \\ (\tilde{G} - Y_0)/W\tau & \text{if } \tilde{G} + WH^*(1-\tau) \leq Y_0 + H^*W. \end{cases}$$

Then, for $H^{**} < 1{,}500$, the tax rate is τ and $\hat{Y}_0 = \tilde{G}$. For $H^{**} \geq 1{,}500$, the tax rate and $\hat{Y}_0$ are the same as for a "like" non-experimental.

TABLE 2

TWO ENDOGENOUS VARIABLES, RESULTS*

Variable	Wage Equation Estimates (Standard Error)	Hours Equation Estimates (Standard Error)
Constant	0.8294	7.4620
	(0.3093)	(0.9897)
Education	0.0155	—
	(0.0119)	
IQ	0.0045	—
	(0.0023)	
Training	0.0021	—
	(0.0011)	
Union	0.2625	—
	(0.0342)	
Illness	−0.2614	−0.0397
	(0.0522)	(0.0248)
Age < 35	0.0110	0.0077
	(0.0086)	(0.0046)
Age 35 to 45	−0.0050	−0.0002
	(0.0074)	(0.0003)
Age > 45	−0.0047	−0.0024
	(0.0032)	(0.0018)
Family Size	—	0.0461
		(0.0286)
Time	0.0340	0.0475
	(0.0028)	(0.0274)
Log Wage	—	0.1401
		(0.0643)
Log Non-Wage Income	—	−0.0233
		(0.0066)

$$\hat{\Omega}_1 = \begin{bmatrix} 0.0683 & 0.0383 \\ & 0.4946 \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} 0.0050 & 0.0013 \\ & 0.1614 \end{bmatrix}$$

* See footnote to Table 1, p. 429.

of the experiment. The estimated coefficient on the logarithm of the wage is 0.140 and that on the logarithm of non-wage income, −0.022. That is, a 100 percent increase in the wage is estimated to increase hours worked by 14 percent, and a 100 percent increase in non-wage income, to decrease hours worked by 2.2 percent. It is worth pointing out that although the experimental selection procedures induced a negative correlation between wages and hours in this sample, taking explicit account of the truncation leads to a quite plausible coefficient on the wage rate in the hours equation. We note first that these numbers are rather close to those estimated from pre-experimental data (0.095 and −0.024, respectively) in our previous work. This is so, even though the estimation procedures used in the two cases are substantially different. In particular, variations in marginal tax rates were not of substantial importance with respect to pre-experimental data, and we did not take account of them.

Although these estimates may not seem very large, they do imply that for persons who "elect" to be "on" the experiment, the effects could be sizable. For some individuals, the experimental guarantee implies a sizable percentage increase in non-wage income, and the tax rate a substantial reduction in the marginal wage rate. For example, consider an individual whose non-wage income increased by 700 percent and whose marginal wage rate fell by 50 percent. Hours worked by him would be expected to fall by over 22 percent. Similar calculations for all individuals in our sample who were on the experiment (observed to be below the break-even point) suggest an estimated average decrease in hours worked of 16.1 percent. The comparable number for persons predicted to be below this breakeven point at 1,500 hours of work was 17.6 percent. In subsequent work we will take explicit account of individual decisions to be "on" or "off" the experiment.

The time parameters should be interpreted as representing both the influence of factors that affected experimentals and controls, as well as differences in the methods used to obtain wage and hours worked data in the two periods.

The specification discussed above does not distinguish between "experimentally induced" and other components of the wage rate and non-wage income. That is, the use of the net wage variable in the hours equation (2.14) constrains the coefficient, β, on the logarithm of the wage, $X\delta$, to be the same as that on the logarithm of the tax rate, $\ln(1-t)$; and, the single coefficient on non-wage income for the experimental group, $\tilde{Y}_0 = \tilde{G} = G - \tau Y_0 + Y_0$, does not distinguish between income fixed by the experiment, G, and other non-wage income, Y_0. To test for possible differences in response to these two components of the wage rate on non-wage income, we have separated the net wage into two variables $\ln W$ and $\ln(1-t)$, and non-wage income into two variables $\ln(G-\tau Y_0)$ and $\ln Y_0$. Separate coefficients were estimated for each. The estimated coefficients on the logarithms of the tax and the wage rate were 0.1126 and 0.1567 respectively. The coefficients on $\ln(G-\tau Y_0)$ and $\ln Y_0$ were -0.0161 and -0.0468. Using a likelihood ratio test, the hypothesis that the wage rate coefficient was equal to the tax coefficient and the coefficient on the experimental guarantee equal to other non-wage income could not be rejected, even at a significance level as high as 20 percent. (Twice the ratio of the likelihood is 3.2 and is distributed as χ^2 with 2 degrees of freedom.)

We also estimated a simple experimental effect in the hours equation, deleting the wage and non-wage income variables. The estimated coefficient on the dummy variable identifying the experimental group was -0.0375, with an asymptotic standard error of 0.0230. It is comparable to those obtained by others (e.g. Hall 1975 or Watts and Rees 1974). This estimate, of course, does not isolate the effect of the experimental treatment on non-wage income and the wage rate and thus does not deal with the biased estimates of these effects that the sample truncation induces.

Finally, we also estimated the hours equation specification in table 2 using standard two stage least squares. The estimated coefficient on the wage rate was -0.382 with an asymptotic standard error of 0.224. The negative coefficient results from the sampling procedure that tended to eliminate from the sample persons with both high wage rates and high hours worked. Or, persons with relatively high wage rates who responded by working relatively long hours tended to be eliminated from the sample.

3. Summary

We have presented a maximum likelihood procedure for estimating behavioral equations from observations over two time periods when the observations are selected by "truncating" on an endogenous variable during the first period. It is an extension of a procedure, developed in an earlier paper, that is strictly applicable only to data pertaining to the truncation period—the first time period in this case. In particular, the method of this paper has been used to analyze the effects of the New Jersey Negative Income Tax Experiment.

Two models were developed. The *first* deals with one endogenous variable. Our example was annual earnings. Our estimate revealed a negative experimental effect of about 6 percent on the earnings of white males. It also led to an estimate of the proportion of variation in income (given X) due to "permanent" factors of about 86 percent. This means that even if one were to use only experimental data, the truncation in the first period would lead to parameter estimates with large bias, as demonstrated in our previous work using pre-experimental data. The *second* model handled two endogenous variables—wages and hours worked. Our sample estimates revealed an elasticity of hours worked with respect to the wage rate of about 14 percent, and with respect to non-wage income, about 2 percent. (Note that because of the truncation, other investigators who did not correct for it often found a negative wage coefficient in equations similar to ours.) These coefficients, although small, suggest that for persons who elect to be "on" the experiment, the effect on hours worked may be substantial, even though it does not appear to be large on the average. It is also of interest that the results were surprisingly close to those obtained in our work using pre-experimental observations only.

Subsequent work will systematically treat the tax rate—or, the decision to be "on" or "off" the experiment—as endogenous. The procedure used here apparently worked well, but did not allow explicit prediction of the "on"–"off" decision for each individual. This is an important aspect of the evaluation of any income maintenance scheme of the type encountered in negative income tax experiments.

Massachusetts Institute of Technology
Harvard University

Appendix on Estimation

Parameter estimates for both the one- and two-endogenous variable models were obtained using a generalization of the Guass–Newton maximization algorithm suggested by Bernt, Hall, Hall, and Hausman [1974]. It uses only first derivatives. In both cases we maximized the likelihood function obtained after writing the appropriate multivariate normal density functions as the product of marginal and conditional density functions, as in equations (1.10) and (2.12). We had hoped that this transformation would simplify computation. However, after having derived first order conditions with and without the transformation, one approach did not seem to recommend itself over the other. First order conditions for the single endogenous variable case are straightforward, but those for the simultaneous case require somewhat more explanation.

We can write the first-order conditions for maximization of (1.10) as:

$$\text{(A.1)} \qquad \frac{\partial \mathscr{L}}{\partial \beta_j} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (Y_{1i} - X_{1i}\beta) X_{1ij} + \sum_{i=1}^{N} X_{1ij}\phi(d_i)/\sigma\Phi[d_i],$$

$$+ \frac{1}{\sigma^2(1-\rho^2)} \sum_{i=1}^{N} (\tilde{Y}_i - \tilde{X}_i\beta)\tilde{X}_{ij},$$

$$\frac{\partial \mathscr{L}}{\partial \sigma^2} = \frac{-N}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{N} (Y_{1i} - X_{1i}\beta)^2$$

$$+ \frac{1}{2\sigma^3} \sum_{i=1}^{N} (L_i - X_{1i}\beta)\phi(d_i)/\Phi[d_i]$$

$$+ \frac{1}{2\sigma^4(1-\rho^2)} \sum_{i=1}^{N} (\tilde{Y}_i - \tilde{X}_i\beta)^2,$$

$$\frac{\partial \mathscr{L}}{\partial \rho} = \frac{-\rho}{\sigma^2(1-\rho^2)^2} \sum_{i=1}^{N} (\tilde{Y}_i - \tilde{X}_i\beta)^2$$

$$+ \frac{1}{\sigma^2(1-\rho^2)} \sum_{i=1}^{N} (\tilde{Y}_i - \tilde{X}_i\beta)(Y_{1i} - X_{1i}\beta)$$

$$+ \frac{N\rho}{(1-\rho^2)},$$

where $d_i = (L_i - X_{1i}\beta)/\sigma$, $\tilde{Y}_i = Y_{2i} - \rho Y_{1i}$, $\tilde{X}_i = X_{2i} - \rho X_{1i}$.

Estimation of the parameters in the simultaneous equation case requires maximization of the likelihood function (2.12) with respect to *to* δ, β, α, and the six parameters in the covariance matrix Ω (2.6), three in Ω_1, and three in Ω_2. For convenience, equation (2.12) is reproduced here as,

$$\text{(A.2)} \qquad \mathscr{L} = \frac{N}{2} \ln (\text{Det}\, \Omega_1^{-1}) + \frac{N}{2} \ln (\text{Det}\, B^{-1})$$

$$- \tfrac{1}{2} \sum_{i=1}^{N} V_{1i}\Omega_1^{-1}V'_{1i} - \sum_{i=1}^{N} \ln \Phi[d_i]$$

$$- \tfrac{1}{2} \sum_{i=1}^{N} (V_{2i} - CV'_{1i})B^{-1}(V_{2i} - CV'_{1i})'.$$

recall the following definitions:

$$V_{11i} = \ln W_{1i} - X_{1i}\delta$$

$$V_{12i} = \ln H_{1i} - X_{1i}\delta\beta - Z_{1i}\alpha$$

$$V_{21i} = \ln W_{2i} - X_{2i}\delta$$

$$V_{22i} = \ln H_{2i} - X_{2i}\delta\beta - Z_{2i}\alpha$$

$$V_i = (V_{11i}, V_{12i}, V_{21i}, V_{22i})$$

$$V_{1i} = (V_{11i}, V_{12i})$$

$$V_{2i} = (V_{21i}, V_{22i})$$
$$d_i = (\ln L_i - (X_{1i}\delta + X_{1i}\delta\beta + Z_{1i}\alpha))/\sqrt{\omega_{11} + \omega_{22} + 2\omega_{12}}$$
$$C = \Omega_2\Omega_1^{-1}$$
$$B = \Omega_1 - \Omega_2\Omega_1^{-1}\Omega_2.$$

We see that elements of Ω show up in C and B^{-1}, as well as in Ω^{-1}. We need both to account for the relationships between these matrices and to insure that Ω_1, B, and Ω_2, will be positive definite, since they are covariance matrices. To do this we let:

$$\Omega_1^{-1} = \Gamma\Lambda\Gamma'$$
$$C = \Omega_2\Omega_1^{-1} = (\Gamma^{-1})'\Lambda\Gamma'$$
$$B^{-1} = (\Omega_1 - \Omega_2\Omega_1^{-1}\Omega_2)^{-1} = \Gamma(\Lambda^{-1} - \Lambda)^{-1}\Gamma',$$

where,

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix};$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

with λ_1 and λ_2 constrained to be greater than zero.

We maximize (A.2) with respect to δ, β, α, and the six elements of Γ and Λ. We then rely on the invariance theorem to recover the elements of Ω.

Before taking derivatives, we need to make a few calculations and some additional definitions. Note that:

$$\text{Det}\,\Gamma = \Gamma_{11}\Gamma_{22} - \Gamma_{12}\Gamma_{21},$$
$$\Omega_1^{-1} = \Gamma\Lambda\Gamma' = \begin{bmatrix} \lambda_1\Gamma_{11}^2 + \lambda_2\Gamma_{12}^2 & \lambda_1\Gamma_{11}\Gamma_{21} + \lambda_2\Gamma_{12}\Gamma_{22} \\ \lambda_1\Gamma_{11}\Gamma_{21} + \lambda_2\Gamma_{12}\Gamma_{22} & \lambda_1\Gamma_{21}^2 + \lambda_2\Gamma_{22}^2 \end{bmatrix} \equiv \begin{bmatrix} \omega^{11} & \omega^{12} \\ \omega^{12} & \omega^{22} \end{bmatrix},$$
$$\text{Det}\,\Omega_1^{-1} = \lambda_1\lambda_2\Gamma_{11}^2\Gamma_{22}^2 + \lambda_1\lambda_2\Gamma_{12}^2\Gamma_{21}^2 - 2\lambda_1\lambda_2\Gamma_{11}\Gamma_{21}\Gamma_{12}\Gamma_{22}$$
$$= \lambda_1\lambda_2(\text{Det}\,\Gamma)^2,$$
$$\Omega_1 = \frac{1}{\lambda_1\lambda_2(\text{Det}\,\Gamma)^2}\begin{bmatrix} \lambda_1\Gamma_{21}^2 + \lambda_2\Gamma_{22}^2 & -\lambda_1\Gamma_{11}\Gamma_{21} - \lambda_2\Gamma_{12}\Gamma_{22} \\ -\lambda_1\Gamma_{11}\Gamma_{21} - \lambda_2\Gamma_{12}\lambda_{22} & \lambda_1\Gamma_{11}^2 + \lambda_2\Gamma_{12}^2 \end{bmatrix}$$
$$\equiv \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix},$$
$$\sigma_c^2 = \omega_{11} + \omega_{22} + 2\omega_{12}$$
$$= \frac{1}{\lambda_1\lambda_2(\text{Det}\,\Gamma)^2}[(\lambda_1\Gamma_{21}^2 + \lambda_2\Gamma_{22}^2) + (\lambda\Gamma_{11}^2 + \lambda_2\Gamma_{12}^2)$$
$$-2(\lambda_1\Gamma_{11}\Gamma_{21} + \lambda_2\Gamma_{12}\Gamma_{22})],$$

$$B^{-1}=\Gamma(\Lambda^{-1}-\Lambda)^{-1}\Gamma'=\begin{bmatrix}\gamma_1\Gamma_{11}^2+\gamma_2\Gamma_{12}^2 & \gamma_1\Gamma_{11}\Gamma_{21}+\gamma_2\Gamma_{12}\Gamma_{22}\\ \gamma_1\Gamma_{11}\Gamma_{21}+\gamma_2\Gamma_{12}\Gamma_{22} & \gamma_1\Gamma_{21}^2+\gamma_2\Gamma_{22}^2\end{bmatrix}=\begin{bmatrix}b^{11} & b^{12}\\ b^{12} & b^{22}\end{bmatrix},$$

where $\gamma_1\equiv\lambda_1/(1-\lambda_1^2)$ and $\gamma_2\equiv\lambda_2/(1-\lambda_2^2)$,

$$\operatorname{Det}B^{-1}=\gamma_1\gamma_2\Gamma_{11}^2\Gamma_{22}^2+\gamma_1\gamma_2\Gamma_{12}^2\Gamma_{21}^2-2\gamma_1\gamma_2\Gamma_{11}\Gamma_{21}\Gamma_{12}\Gamma_{22}$$

$$=\gamma_1\gamma_2(\operatorname{Det}\Gamma)^2$$

$$C=\Omega_2\Omega_1^{-1}$$

$$=\begin{bmatrix}(\lambda_1\Gamma_{11}\Gamma_{22}-\lambda_2\Gamma_{12}\Gamma_{21})/\operatorname{Det}\Gamma & (\lambda_1\Gamma_{21}\Gamma_{22}-\lambda_2\Gamma_{21}\Gamma_{22})/\operatorname{Det}\Gamma\\ (-\lambda_1\Gamma_{11}\Gamma_{12}+\lambda_2\Gamma_{11}\Gamma_{12})/\operatorname{Det}\Gamma & (-\lambda_1\Gamma_{12}\Gamma_{21}+\lambda_2\Gamma_{11}\Gamma_{22})/\operatorname{Det}\Gamma\end{bmatrix}$$

$$\equiv\begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix}$$

$$CV_1'=\begin{bmatrix}C_{11}V_{11}+C_{12}V_{12}\\ C_{21}V_{11}+C_{22}V_{12}\end{bmatrix}\equiv\begin{bmatrix}S_{11}\\ S_{12}\end{bmatrix},$$

where the last term defines S_{11} and S_{12},

$$V_1\Omega_1^{-1}V_1'=\omega^{11}V_{11}^2+2\omega^{12}V_{12}V_{11}+\omega^{22}V_{12}^2$$

$$\begin{aligned}(V_2-CV_1')B^{-1}(V_2-CV_1')'&=b^{11}(V_{21}^2-2V_{21}S_{11}+S_{11}^2)\\&\quad+2b^{12}(V_{21}V_{22}-V_{21}S_{12}-S_{11}V_{22}+S_{11}S_{12})\\&\quad+b^{22}(V_{22}^2-2V_{22}S_{12}+S_{12}^2)\\&\equiv b^{11}f_1+2b^{12}f_2+b^{22}f_3,\end{aligned}$$

where the last term defines f_1, f_2, f_3. Finally, if we let $(\omega_{11}+\omega_{22}+2\omega_{12})^{1/2}\equiv\sigma_c$, the relevant first order conditions are given by:

$$\begin{aligned}\frac{\partial\mathscr{L}}{\partial\delta_j}&=\sum_{i=1}^{N}[\omega^{11}V_{11_i}+\omega^{12}V_{12_i}+\omega^{12}V_{11_i}\beta+\omega^{22}V_{12_i}\beta]X_{1_{ij}}\\&\quad+\sum_{i=1}^{N}\frac{\phi(d_i)(X_{1_{ij}}+X_{1_{ij}}\beta)}{\Phi[d_i]\sigma_c}\\&\quad+\sum_{i=1}^{N}b^{11}[V_{21_i}X_{2_{ij}}-V_{21}(C_{11}X_{1_{ij}}+C_{12}X_{1_{ij}}\beta)\\&\quad-S_{11}X_{2_{ij}}+S_{11}(C_{11}X_{1_{ij}}+C_{12}X_{1_{ij}}\beta)]\\&\quad+b^{12}[V_{21}X_{2_{ij}}\beta+V_{22}X_{2_{ij}}-V_{21}(C_{21}X_{1_{ij}}+C_{22}X_{1_{ij}}\beta)\\&\quad-S_{12}X_{2_{ij}}-S_{11}X_{2_{ij}}\beta-V_{22}(C_{11}X_{1_{ij}}+C_{12}X_{1_{ij}}\beta)\\&\quad+S_{11}(C_{21}X_{1_{ij}}+C_{22}X_{1_{ij}}\beta)+S_{12}(C_{11}X_{1_{ij}}+C_{12}X_{1_{ij}}\beta)]\\&\quad+b^{22}[V_{22}X_{2_{ij}}\beta-V_{22}(C_{21}X_{1_{ij}}+C_{22}X_{1_{ij}}\beta)\\&\quad-S_{12}X_{2_{ij}}\beta+S_{12}(C_{21}X_{1_{ij}}+C_{22}X_{1_{ij}}\beta)]\end{aligned}$$

$$\frac{\partial \mathscr{L}}{\partial \beta} = \sum_{i=1}^{N} [\omega^{12} V_{11} + \omega^{22} V_{12}] X_{1_i} \delta$$

$$+ \sum_{i=1}^{N} \frac{\phi(d_i)}{\Phi[d_i]} \frac{X_{1_i} \delta}{\sigma_c}$$

$$+ \sum_{i=1}^{N} \{b^{11}[-V_{21} C_{12} X_{1_i} \delta + S_{11} C_{12} X_{1_i} \delta]$$

$$+ b^{12}[V_{21} X_{2_i} \delta - V_{21} C_{22} X_{1_i} \delta - S_{11} X_{1_i} \delta$$

$$- V_{22} C_{12} X_{1_i} \delta + S_{11} C_{22} X_{1_i} \delta + S_{12} C_{12} X_{1_i} \delta]$$

$$+ b^{22}[V_{22} X_{2_i} \delta - V_{22} C_{22} X_{1_i} \delta - S_{12} X_{2_i} \delta + S_{12} C_{22} X_{1_i} \delta]\}$$

$$\frac{\partial \mathscr{L}}{\partial \alpha_k} = \sum_{i=1}^{N} [\omega^{12} V_{11} + \omega^{22} V_{12}] Z_{1_{ik}}$$

$$+ \sum_{i=1}^{N} \frac{\phi(d_i) Z_{1_{ik}}}{\Phi[d_i] \sigma_c}$$

$$+ \sum_{i=1}^{N} \{b^{11}[-V_{21} C_{12} Z_{1_{ik}} + S_{11} C_{12} Z_{1_{ik}}]$$

$$+ b^{12}[V_{21} Z_{2_{ik}} - V_{21} C_{22} Z_{1_{ik}} - S_{11} Z_{2_{ik}}$$

$$- V_{22} C_{12} Z_{1_{ik}} + S_{11} C_{22} Z_{1_{ik}} + S_{12} C_{12} Z_{1_{ik}}]$$

$$+ b^{22}[V_{22} Z_{2_{ik}} - V_{22} C_{22} Z_{1_{ik}} - S_{12} Z_{2_{ik}} + S_{12} C_{22} Z_{1_{ik}}]\}$$

$$\frac{\partial \mathscr{L}}{\partial \Gamma_{11}} = \frac{2N\Gamma_{22}}{\operatorname{Det} \Gamma} - \sum_{i=1}^{N} [\lambda_1 \Gamma_{11} V_{11}^2 + \lambda_1 \Gamma_{21} V_{12} V_{11}]$$

$$- \sum_{i=1}^{N} \left\{ b^{11} \left[(-V_{21} + S_{11}) \left(\frac{\lambda_1 \Gamma_{22} V_{11}}{\operatorname{Det} \Gamma} - \frac{\Gamma_{22} S_{11}}{\operatorname{Det} \Gamma} \right) \right] + f_1 \gamma_1 \Gamma_{11} \right.$$

$$+ b^{12} \left[(-V_{21} + S_{11}) \left(\frac{(\lambda_2 - \lambda_1) \Gamma_{12} V_{11}}{\operatorname{Det} \Gamma} + \frac{\lambda_2 \Gamma_{22} V_{12}}{\operatorname{Det} \Gamma} - \frac{\Gamma_{22} S_{12}}{\operatorname{Det} \Gamma} \right) \right.$$

$$\left. + (-V_{22} + S_{12}) \left(\frac{\lambda_1 \Gamma_{22} V_{11}}{\operatorname{Det} \Gamma} - \frac{\Gamma_{22} S_{11}}{\operatorname{Det} \Gamma} \right) \right] + f_2 \gamma_1 \Gamma_{21}$$

$$\left. + b^{22} \left[(-V_{22} + S_{12}) \left(\frac{(\lambda_2 - \lambda_1) \Gamma_{12} V_{11}}{\operatorname{Det} \Gamma} + \frac{\lambda_2 \Gamma_{22} V_{12}}{\operatorname{Det} \Gamma} - \frac{\Gamma_{22} S_{12}}{\operatorname{Det} \Gamma} \right) \right] \right\}$$

$$+ \sum_{i=1}^{N} \frac{\phi(d_i)}{\Phi[d_i]} \frac{d_i}{\sigma_c^2} \left[\frac{\Gamma_{11}}{\lambda_2 (\operatorname{Det} \Gamma)^2} \; \frac{-\Gamma_{21}}{\lambda_2 (\operatorname{Det} \Gamma)^2} \; \frac{-\sigma_c^2 \Gamma_{22}}{\operatorname{Det} \Gamma} \right],$$

$$
\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \Gamma_{22}} = {} & \frac{2N\Gamma_{11}}{\operatorname{Det}\Gamma} - \sum_{i=1}^{N} [\lambda_2\Gamma_{12}V_{12}V_{11} + \lambda_2\Gamma_{22}V_{12}^2] \\
& - \sum_{i=1}^{N} \left\{ b^{11}\left[(-V_{21}+S_{11})\left(\frac{\lambda_1\Gamma_{11}V_{11}}{\operatorname{Det}\Gamma} + \frac{(\lambda_1-\lambda_2)\Gamma_{21}V_{12}}{\operatorname{Det}\Gamma} - \frac{\Gamma_{11}S_{11}}{Det\,\Gamma}\right)\right] \right. \\
& + b^{12}\left[(-V_{21}+S_{11})\left(\frac{\lambda_2\Gamma_{11}V_{12}}{\operatorname{Det}\Gamma} - \frac{\Gamma_{11}S_{12}}{\operatorname{Det}\Gamma}\right)\right. \\
& \left. + (-V_{22}+S_{12})\left(\frac{\lambda_1\Gamma_{11}V_{11}}{\operatorname{Det}\Gamma} + \frac{(\lambda_1-\lambda_2)\Gamma_{21}V_{12}}{\operatorname{Det}\Gamma} - \frac{\Gamma_{11}S_{11}}{\operatorname{Det}\Gamma}\right)\right] + f_2\gamma_2\Gamma_{12} \\
& \left. + b^{22}\left[(-V_{22}+S_{12})\left(\frac{\lambda_2\Gamma_{11}V_{12}}{\operatorname{Det}\Gamma} - \frac{\Gamma_{11}S_{12}}{\operatorname{Det}\Gamma}\right)\right] + f_3\gamma_2\Gamma_{22} \right\} \\
& + \sum_{i=1}^{N} \frac{\phi(d_i)}{\Phi[d_i]} \frac{d_i}{\sigma_c^2} \left[\frac{\Gamma_{22}}{\lambda_1(\operatorname{Det}\Gamma)^2} - \frac{\Gamma_{12}}{\lambda_1(\operatorname{Det}\Gamma)^2} - \frac{\sigma_c^2\Gamma_{11}}{\operatorname{Det}\Gamma}\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \Gamma_{12}} = {} & -\frac{2N\Gamma_{21}}{\operatorname{Det}\Gamma} - \sum_{i=1}^{N} [\lambda_2\Gamma_{12}V_{11}^2 + \lambda_2\Gamma_{22}V_{12}V_{11}] \\
& - \sum_{i=1}^{N} \left\{ b^{11}\left[(-V_{21}+S_{11})\left(\frac{-\lambda_2\Gamma_{21}V_{11}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{21}S_{11}}{\operatorname{Det}\Gamma}\right)\right] + f_1\gamma_2\Gamma_{12} \right. \\
& + b^{12}\left[(-V_{21}+S_{11})\left(\frac{(-\lambda_1\Gamma_{11}+\lambda_2\Gamma_{11})V_{11}}{\operatorname{Det}\Gamma} - \frac{\lambda_1\Gamma_{21}V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{21}S_{12}}{\operatorname{Det}\Gamma}\right)\right. \\
& \left. + (-V_{22}+S_{12})\left(\frac{-\lambda_2\Gamma_{21}V_{11}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{21}S_{11}}{\operatorname{Det}\Gamma}\right)\right] + f_2\gamma_2\Gamma_{22} \\
& \left. + b^{22}\left[(-V_{22}+S_{12})\left(\frac{(-\lambda_1\Gamma_{11}+\lambda_2\Gamma_{11})V_{11}}{\operatorname{Det}\Gamma} - \frac{\lambda_1\Gamma_{21}V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{21}S_{12}}{\operatorname{Det}\Gamma}\right)\right]\right\} \\
& + \sum_{i=1}^{N} \frac{\phi(d_i)\, d_i}{\Phi[d_i]\sigma_c^2} \left[\frac{\Gamma_{12}}{\lambda_1(\operatorname{Det}\Gamma)^2} - \frac{\Gamma_{22}}{\lambda_1(\operatorname{Det}\Gamma)^2} + \frac{\sigma_c^2\Gamma_{21}}{\operatorname{Det}\Gamma}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \Gamma_{21}} = {} & -\frac{2N\Gamma_{12}}{\operatorname{Det}\Gamma} - \sum_{i=1}^{N} [\lambda_1\Gamma_{11}V_{12}V_{11} + \lambda_1\Gamma_{21}V_{12}^2] \\
& - \sum_{i=1}^{N} \left\{ b^{11}\left[(-V_{21}+S_{11})\left(\frac{-\lambda_2\Gamma_{12}V_{11}}{\operatorname{Det}\Gamma} + \frac{(\lambda_1\Gamma_{22}-\lambda_2\Gamma_{22})V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{12}S_{11}}{\operatorname{Det}\Gamma}\right)\right] \right. \\
& + b^{12}\left[(-V_{21}+S_{11})\left(\frac{-\lambda_1\Gamma_{12}V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{12}S_{12}}{\operatorname{Det}\Gamma}\right)\right. \\
& \left. + (-V_{22}+S_{12})\left(\frac{-\lambda_2\Gamma_{12}V_{11}}{\operatorname{Det}\Gamma} + \frac{(\lambda_1\Gamma_{22}-\lambda_2\Gamma_{22})V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{12}S_{11}}{\operatorname{Det}\Gamma}\right)\right] + f_2\gamma_1\Gamma_{11} \\
& \left. + b^{22}\left[(-V_{22}+S_{12})\left(\frac{-\lambda_1\Gamma_{12}V_{12}}{\operatorname{Det}\Gamma} + \frac{\Gamma_{12}S_{12}}{\operatorname{Det}\Gamma}\right)\right] + f_3\gamma_2\Gamma_{21} \right\} \\
& + \sum_{i=1}^{N} \frac{\phi(d_i)\, d_i}{\Phi[d_i]\sigma_c^2} \left[\frac{\Gamma_{21}}{\lambda_2(\operatorname{Det}\Gamma)^2} - \frac{\Gamma_{11}}{\lambda_2(\operatorname{Det}\Gamma)^2} + \frac{\sigma_c^2\Gamma_{12}}{\operatorname{Det}\Gamma}\right]
\end{aligned}
$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_1} = \frac{N}{2\lambda_1} + \frac{N}{2\gamma_1}\left(\frac{1+\lambda_1^2}{(1-\lambda_1^2)^2}\right)$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left[\Gamma_{11}^2 V_{11}^2 + 2\Gamma_{11}\Gamma_{21}V_{12}V_{11} + \Gamma_{21}^2 V_{12}^2\right]$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left\{b^{11}\left[2(-V_{21}+S_{11})\left(\frac{\Gamma_{11}\Gamma_{22}V_{11}}{\mathrm{Det}\,\Gamma}+\frac{\Gamma_{21}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]\right.$$

$$+f_1\Gamma_{11}^2\left(\frac{1+\lambda_1^2}{(1-\lambda_1^2)^2}\right)$$

$$+2b^{12}\left[(-V_{21}+S_{11})\left(\frac{-\Gamma_{11}\Gamma_{12}V_{11}}{\mathrm{Det}\,\Gamma}-\frac{\Gamma_{12}\Gamma_{21}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right.$$

$$\left.+(-V_{22}+S_{12})\left(\frac{\Gamma_{11}\Gamma_{22}V_{11}}{\mathrm{Det}\,\Gamma}+\frac{\Gamma_{21}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]$$

$$+2f_2\Gamma_{11}\Gamma_{21}\left(\frac{1+\lambda_1^2}{(1-\lambda_1^2)^2}\right)$$

$$+b^{22}\left[2(-V_{22}+S_{12})\left(\frac{-\Gamma_{11}\Gamma_{12}V_{11}}{\mathrm{Det}\,\Gamma}-\frac{\Gamma_{12}\Gamma_{21}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]$$

$$\left.+f_3\Gamma_{21}^2\left(\frac{1+\lambda_1^2}{(1-\lambda_1^2)^2}\right)\right\}$$

$$+\tfrac{1}{2}\sum_{i=1}^{N}\frac{\phi(d_i)\,d_i}{\Phi[d_i]\sigma_c^2}\left[\frac{-\Gamma_{22}^2}{\lambda_1^2(\mathrm{Det}\,\Gamma)^2}\,\frac{-\Gamma_{12}^2}{\lambda_1^2(\mathrm{Det}\,\Gamma)^2}+\frac{2\Gamma_{12}\Gamma_{22}}{\lambda_1^2(\mathrm{Det}\,\Gamma)^2}\right]$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_2} = \frac{N}{2\lambda_2} + \frac{N}{2\gamma_2}\left(\frac{1+\lambda_2^2}{(1-\lambda_2^2)^2}\right)$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left[\Gamma_{12}^2 V_{11}^2 + 2\Gamma_{12}\Gamma_{22}V_{12}V_{11} + \Gamma_{22}^2 V_{12}^2\right]$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left\{b^{11}\left[2(-V_{21}+S_{11})\left(\frac{-\Gamma_{12}\Gamma_{21}V_{11}}{\mathrm{Det}\,\Gamma}-\frac{\Gamma_{21}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]\right.$$

$$+f_1\Gamma_{12}^2\left(\frac{1+\lambda_2^2}{(1-\lambda_2^2)^2}\right)$$

$$+2b^{12}\left[(-V_{21}+S_{11})\left(\frac{\Gamma_{11}\Gamma_{12}V_{11}}{\mathrm{Det}\,\Gamma}+\frac{\Gamma_{11}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right.$$

$$\left.+(-V_{22}+S_{12})\left(\frac{-\Gamma_{12}\Gamma_{21}V_{11}}{\mathrm{Det}\,\Gamma}-\frac{\Gamma_{21}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]$$

$$+2f_2\Gamma_{12}\Gamma_{22}\left(\frac{1+\lambda_2^2}{(1-\lambda_2^2)^2}\right)$$

$$+b^{22}\left[2(-V_{22}+S_{12})\left(\frac{\Gamma_{11}\Gamma_{12}V_{11}}{\mathrm{Det}\,\Gamma}+\frac{\Gamma_{11}\Gamma_{22}V_{12}}{\mathrm{Det}\,\Gamma}\right)\right]$$

$$+f_3\Gamma_{22}^2\left(\frac{1+\lambda_2^2}{(1-\lambda_2^2)^2}\right)\bigg\}$$

$$+\frac{1}{2}\sum_{i=1}^{N}\frac{\phi(d_i)\,d_i}{\Phi[d_i]\sigma_\epsilon^2}\left[\frac{-\Gamma_{21}^2}{\lambda_2^2(\mathrm{Det}\,\Gamma)^2}-\frac{\Gamma_{11}^2}{\lambda_2^2(\mathrm{Det}\,\Gamma)^2}+\frac{2\Gamma_{11}\Gamma_{21}}{\lambda_2^2(\mathrm{Det}\,\Gamma)^2}\right].$$

References

Berndt, E. K., Hall, B. H., Hall. R. E., and Hausman, J. A., "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement.* 1974. pp. 653–665.

Hall, R. E., "The Effects of the Experimental Negative Income Tax on Labor Supply," in J. A. Pechman and P. M. Timpane (eds.), *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment,* The Brookings Institution, Washington, D.C., 1975.

———. "Wages, Income, and Hours of Work in the U.S. Labor Force," in G. Cain and H. Watts (eds.), *Income Maintenance and Labor Supply Studies.* Rand McNally. Chicago, 1973. pp. 102–162.

Hausman, J. A. and D. A. Wise, "Social Experimentation, Truncated Distributions, and Efficient Estimation," John Fitzgerald Kennedey School of Government, Discussion Paper No. 31D, Harvard University, June 1975, Forthcoming in *Econometrica.*

Rosen, H. S., "The Impact of U.S. Tax Laws on the Labor Supply of Married Women," unpub. diss. Harvard University, Department of Economics, March 1974.

Watts, H. W. and A. Rees, (eds.). *Final Report of the New Jersey Graduated Work Incentive Experiment,* Institute for Research on Poverty, University of Wisconsin (Madison) and MATHEMATICA, 1974.