153-73

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

AN APPROPRIATE ECONOMETRIC FRAMEWORK FOR ESTIMATING A LABOR-SUPPLY FUNCTION FROM THE SEO FILE

Dennis Aigner

UNIVERSITY OF WISCONSIN-MADISON

AN APPROPRIATE ECONOMETRIC FRAMEWORK FOR ESTIMATING A LABOR-SUPPLY FUNCTION FROM THE SEO FILE

Dennis Aigner

I am grateful to Glen Cain for comments on a previous draft of this paper, to Arthur Goldberger for helpful discussion in the formative stage of the research, and to Dale Poirier for programming assistance. The data were made available through grants from the Office of Economic Opportunity. Some of the work reported here was supported in part by funds granted to the Institute for Research on Poverty by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. The conclusions remain the author's responsibility.

ABSTRACT

This paper presents an econometric model for estimating a labor-supply function from the SEO file that explicitly accounts for a potentially spurious relationship between labor supply and the wage rate arising from certain measurement problems in the data. The proposed estimators possess optimal large sample properties. An example of application is included.

AN APPROPRIATE ECONOMETRIC FRAMEWORK FOR ESTIMATING A LABOR-SUPPLY FUNCTION FROM THE SEO FILE

1. INTRODUCTION

The wide variation in estimates of the substitution and income effects of labor supply obtained recently by Irwin Garfinkel, David H. Greenberg and Marvin Kosters, and E. D. Kalachek and Fredric Q. Raines using the 1967 Survey of Economic Opportunity (SEO) file surely must be disconcerting to those intending to use such estimates to evaluate the viability of a negative-income-tax program. There would be substantive problems involved in the use of such results even if they were to agree closely. These include the fact that valid differences of opinion exist about the choice of an appropriate universe on which to base a prediction of labor response to a negative-income-tax program and that the wage and (nonemployment) income behavior of SEO sample respondents is not being observed in the presence of such an income tax program. In this latter respect the prediction problem is similar to forecasting the demand for a nonexistent product.

In this paper we abstract from these issues and concentrate on the more mundane econometric question that arises if we take these various studies as essentially equivalent in their approach to the estimation problem. Then, the wide variation in results on comparable large samples is symptomatic of the fact that the estimation approach is suspect, in the sense that it is quite sensitive to what appear to be minor differences in model specification and data manipulation. A basic problem in using the SEO file is that an independent measurement on wages is not available. Rather, a wage measurement must be constructed from earnings and hours of work; this latter variable being a component of the measure of "labor supply." The authors who subsequently use this wage variable remark on how its use is likely to bias the reported least squares coefficient estimates for the labor-supply equation. But only Robert Hall in "Wages, Income and Hours of Work in the U.S. Labor Force;" Michael Boskin in "Income Maintenance Policy, Labor Supply and Income Redistribution;" and Sandra Christensen in "Income Maintenance and Labor Supply," attempt to alter the estimation procedure in order to obtain consistent estimates of labor-supply parameters.

The initial purpose in this paper is to criticize the Hall approach, as it is specifically applied by him, pointing out how it fails to produce consistent estimates of parameters. This is also done by Christensen, who subsequently produces consistent estimates via the instrumental variables technique applied to an equation with earnings as a dependent variable. As is the usual case with an instrumental variables estimator, though consistent if certain conditions are met, its asymptotic precision depends on the choice of instruments, and it cannot generally be taken to possess the efficiency property (even asymptotically). As Boskin applies Hall's approach, he also gets an instrumental variables estimator (though it differs from Christensen's), which is thereby subject to this same shortcoming. A brief discussion of these points is included in Section 2 of the present paper.

Following in the spirit of Hall's and Christensen's efforts, in Section 3 an econometric model is suggested that captures faithfully the structure of the problem and produces estimates with the guaranteed large-sample properties of consistency, efficiency, and normality. The model is a multivariate linear regression subject to nonlinear and linear constraints and requires more elaborate computations than does ordinary least squares on a linear model. Nonetheless, the required calculations are not onerous given today's array of available computing hardware. In presenting the model we do as the cited authors did and assume away questions of simultaneity in the labor-supply function. Moreover, we do not attempt to integrate the choice problem of work/no work into the model, a matter that is recognized as quite important by some.[1]

In Section 4 we present empirical results from an application of the model to data that are constructed in a fashion similar to that of Greenberg and Kosters in "Income Guarantees and the Working Poor: The Effect of Income Maintenance Programs on the Hours of Work of Male Family Heads." Our main findings corroborate an income coefficient of (essentially) zero, but with a positive substitution effect, yielding a compensated substitution elasticity that is close to zero in absolute value and positive. That the wage coefficient is positive we attribute to the model's explicit treatment of the errors-in-variables problem.

Additional discussion of the methodological aspects of the problem and some caveats are the province of the concluding section.

2. ECONOMETRIC ISSUES

There is substantial agreement among the cited studies on several matters. First, a linear model is the basic functional form used,[2] say

$$y_{1i} = \beta_1 w_i + \beta_2 s_i + \underline{X}_i' \underline{\beta}_3 + u_{1i} \qquad i=1,\ldots,n. \qquad (1)$$

where: y_{1i} is labor supply (hours worked annually) for the $i^{\underline{th}}$ individual;

w_i is the wage rate;

and s_i is his nonemployment income. $\underline{X}_i$ is an (r x 1) vector of certain demographic variables included as "controls" directly in the labor-supply function (including, perhaps, a variable that is always one, corresponding to an intercept term) and $\underline{\beta}_3$ is its corresponding (r x 1) parameter vector. Assumptions on u_{1i} include a mean of zero, constant variance, and independence across individuals.

Authors differ slightly on what and how variables should be included in $\underline{X}_i$. In Hall's approach the demographic variables come into equation (1) indirectly, as will be explained below. For Greenberg and Kosters a "preferences" variable is included that in their opinion helps isolate the income effect and allows freedom to combine universes--hence getting a broader-based labor-response prediction than is admitted by many of the other authors, who tend to estimate equation (1) for narrower subsamples of SEO respondents.

Likewise, Garfinkel and Hall discretize the wage and/or income variables in attempting to gain some freedom of functional form.[3] And, as mentioned previously, there is reason to view the prediction of labor supply as a sequential problem, involving at the outset a decision to work or to not work, followed by a decision as to the number hours of work given the affirmative action. Acceptance of this point of view leads Boskin, Kalachek and Raines, and Garfinkel, to estimate a probability function for the work/no work decision, and then to estimate the expected labor supply given labor-force participation. There are further differences of opinion on the construction of s_i.

What is agreed upon, using equation (1) as "the" model, is that w_i is not measured accurately. There is also a problem in measuring y_{1i}, but given the usual assumptions on this observation error it can be taken to be already represented in u_{1i}. Writing

$$y_{2i} = w_i + u_{2i} \qquad i=1,\ldots,n \qquad (2)$$

to represent the measured wage (y_{2i}) as a function of the true wage and an additive error, we assume u_{2i} to be independent of w_i, with a mean of zero. w_i and u_{2i} are supposed to have constant variances (σ_w^2 and $\sigma_{u_2}^2$, respectively) and to be independently distributed over individuals. If it is further assumed that $E(u_{1i}u_{2i}) = 0$, the effect of the existence of this observation error on least squares estimates of coefficients in equation (1) with the measured wage used as regressor in place of the unobservable w_i is to bias them toward zero.[4] When $E(u_{1i}u_{2i}) \neq 0$ an additive bias effect must also be considered, a case that will be of some interest later.

Hall's suggestion to eliminate errors-in-variables bias is to construct a predicted wage from least squares applied to a wage equation,

$$y_{2i} = \underline{Z}_i' \, \underline{\gamma} + \delta_i \qquad i=1,\ldots,n \qquad (3)$$

which has the measured wage represented by a linear regression in certain demographic variables, $\underline{Z}_i' = (z_{1i} \; . \; . \; z_{ki})$.[5] Again, one of the z's may be one for all observations, corresponding to an intercept term in equation (3). In this equation, δ_i is assumed to be a well-behaved disturbance, with a mean of zero and a constant variance, independently distributed over individuals, and uncorrelated with the variables in $\underline{Z}_i$. Moreover, the variables in $\underline{Z}_i$ are assumed to be uncorrelated with u_{1i}.

In the sample we will have

$$y_{2i} = \underline{Z}_i' \, \hat{\underline{\gamma}} + e_i = \hat{w}_i + e_i \qquad i=1,\ldots,n \qquad (4)$$

where $\hat{\underline{\gamma}}$ is the least squares estimate of the (k x 1) parameter vector $\underline{\gamma}$ and e_i is the corresponding calculated residual. By construction, $\Sigma \hat{w}_i e_i = 0$.

For Hall, $\underline{\beta}_3 = \underline{0}$ in equation (1), since he stratifies samples with regard to the variables in $\underline{X}_i$, so that upon implementation he estimates equation (1) by a least squares regression of y_{1i} on $\hat{w}_i$ and s_i. Given that equation (3) is fitted over the same sample represented in equation (1), these least squares estimates of β_1 and β_2 will be consistent, since $\hat{w}_i$ and s_i are uncorrelated with u_{1i}. Their asymptotic precision will depend on the particular choice of demographic variables in equation (3), however, which fact is underscored by interpreting the resulting estimates as instrumental variables estimates.[6] Boskin's work is apparently consistent with the analysis just given though in contrast to Hall's assumption that $\underline{\beta}_3 = \underline{0}$ he includes some of the variables in $\underline{Z}_i$ as $\underline{X}_i$ $(k > r)$ in (1).[7]

Hall's actual use of his suggested technique fails to produce consistent estimates of the labor-supply coefficients for two reasons. The first is that Hall tries to meld the labor-supply quantity and participation prediction problems by imputing a wage to persons in the samples who did not work. He fits a wage equation over wage earners on the basis of various demographic characteristics and then uses the least squares coefficients so obtained to predict a wage for each individual in the various labor-supply function he estimates. The specifications required for consistent parameter estimation outlined above are not met because the same sample is not used to fit both equations (3) and (1) in the sequence suggested by Hall.

The second problem with Hall's empirical work is that he uses the imputed wage to compute a labor-supply variable (the ratio of measured earnings to the imputed wage) rather than using as dependent variable the product of hours worked last week by weeks worked last year. As Christensen points out, a biasing factor is thereby introduced through the spurious relation between this particular measurement of y_{1i} and $\hat{w}_i$.[8]

Christensen's approach to these issues--in work that builds on Hall's--is noteworthy in two respects. First, she develops a theoretical basis for the labor-supply function from an intertemporally additive Stone-Geary utility function. Beyond deducing the anticipated signs for the substitution and income effects, the theory suggests the linear form and that variables should appear in a particular way. Christensen's labor-supply function is, in the notation of equation (1),

$$y_{1i} = \beta_1 + \beta_2 \left(\frac{s_i}{w_i}\right) + \left[\frac{\underline{X}_i}{w_i}\right]' \underline{\beta}_3 + u_{1i} \qquad i=1,\ldots,n \qquad (5)$$

where $[\underline{X}_i/w_i]'$ is a row vector with each variable in $\underline{X}_i$ divided by the wage rate, w_i. Multiplying through by w_i,

$$E_{1i} = \beta_1 w_i + \beta_2 s_i + \underline{X}_i' \underline{\beta}_3 + u_{1i} w_i \qquad i=1,\ldots,n \qquad (6)$$

which is Christensen's "earnings" (E_{1i}) equation. Under the assumptions associated with equation (1), conditional on w_i, s_i, and $\underline{X}_i$, $u_{1i}w_i$ has the requisite properties to guarantee the BLUE properties of least squares except constant variance. So, were least squares applied to equation (6), without regard to the difficulties in measuring w_i, we would expect the resulting estimates not to be efficient. Christensen considers the errors-of-observation problem in w_i and notes that Hall's procedure as applied by him will not produce consistent estimates even when "reported" labor supply is used as the dependent variable. She proposes to estimate equation (3) from a sample that has no overlap with the sample (or samples) of interest for the prediction of labor response. This forced independence makes possible a conventional application of the instrumental variables technique. If the wage rate used to fit equation (1) is earnings divided by hours worked, then the imputed wage from her version of equation (3) is used as its instrument and vice versa. Given the nonoverlapping of her two samples, these two sets of estimates will each be consistent. However, their asymptotic efficiency is in doubt not only because they are instrumental variables estimates but because she does not recognize the heteroscedasticity of the disturbance in equation (6) explicitly in the estimation procedure.

3. THE "APPROPRIATE" MODEL

In the spirit of Hall's approach to the estimation of labor-supply parameters, we have constructed an econometric model through which can be obtained maximum likelihood estimates of the substitution and income effects. Its essence is contained in three equations:

$$y_{1i} = \beta_1 w_i + \beta_2 s_i + \underline{X}_i' \underline{\beta}_3 + u_{1i}; \tag{1}$$

$$y_{2i} = w_i + u_{2i}; \tag{2}$$

and

$$w_i = \underline{Z}_i' \underline{\gamma} + \varepsilon_i \qquad i=1,\ldots,n \tag{7}$$

Other than being written for a single sample (for example, as if the model applied to the problem of estimating labor supply given labor-force participation), the only difference between this model and what we have represented as Hall's is that we write the true wage as a (stochastic) function of $\underline{Z}_i$. Perhaps a more faithful reproduction of Hall's specification would use

$$w_i = \underline{Z}_i' \underline{\gamma}, \tag{7a}$$

instead of equation (7), which says that the true wage is an exact function of $\underline{Z}_i$. All the disturbance terms are assumed to posses means of zero and constant variances, and be independently normally distributed over individuals. There is no particular a priori reason to impose $E(u_{1i}u_{2i}) = 0$, $E(u_{1i}\varepsilon_i) = 0$, and $E(u_{2i}\varepsilon_i) = 0$. We assume that $\underline{X}_i$ is a subvector of $\underline{Z}_i$ $(k > r)$. For notational convenience $\underline{Z}_i$ is partitioned into $\underline{Z}_i = (\underline{Z}_{1i}' \; \underline{Z}_{2i}')'$, where $\underline{Z}_{1i} = \underline{X}_i$, and, correspondingly, $\underline{\gamma} = (\underline{\gamma}_1' \; \underline{\gamma}_2')'$. Then, upon substitution of equation (7) into equations (1) and (2), the full reduced form of the model, written for all n observations, appears as

$$(\underline{y}_1 \; \underline{y}_2) = (\underline{Z}_1 \; \underline{Z}_2 \; \underline{s}) \begin{pmatrix} (\underline{\gamma}_1\beta_1 + \underline{\beta}_3) & \underline{\gamma}_1 \\ \underline{\gamma}_2\beta_1 & \underline{\gamma}_2 \\ \beta_2 & 0 \end{pmatrix} + (\underline{v}_1 \; \underline{v}_2). \tag{8}$$

In equation (8), $\underline{y}_1$ and $\underline{y}_2$ each have dimension (n x 1), and the submatrix $\underline{Z}_1$ has the vectors $\underline{Z}_{1i}'$ as its <u>rows</u>, with dimension (n x r). In the same way follow the definitions of $\underline{Z}_2$ [n x (k-r)] and $\underline{s}$ (n x 1). $(\underline{v}_1 \; \underline{v}_2)$ is an (n x 2) matrix of reduced form distrubances, defined by $\underline{v}_1 = (v_{11} \ldots v_{1n})'$, with $v_{1i} = u_{1i} + \beta_1\varepsilon_i$, and $\underline{v}_2 = (v_{21} \ldots v_{2n})'$, with $v_{2i} = u_{2i} + \varepsilon_i$. Equation (8) is in the form of a multivariate regression system subject to certain parameter constraints. It is clear that complete overlap of the demographic variables in equations (1) and (7) causes underidentification of the labor-supply parameters. For, in that case, the parameter matrix in equation (8) would appear as

$$\begin{pmatrix} (\underline{\gamma}\beta_1 + \underline{\beta}_3) & \underline{\gamma} \\ \beta_2 & 0 \end{pmatrix}, \tag{9}$$

leaving β_1 and $\underline{\beta}_3$ not identified in the reduced form.

If we relabel the parameter matrix in equation (8) as it is partitioned there by

$$\Pi = \begin{pmatrix} \underline{\pi}_{11} & \underline{\pi}_{12} \\ \underline{\pi}_{21} & \underline{\pi}_{22} \\ \underline{\pi}_{31} & \underline{\pi}_{32} \end{pmatrix} \tag{10}$$

the restrictions are (a) $\pi_{32} = 0$, and (b) $\underline{\pi}_{21} = \underline{\pi}_{22}\beta_1$. In effect, restriction (b) says that the submatrix $(\underline{\pi}_{21}\ \underline{\pi}_{22})$ cannot be of full (column) rank, but must satisfy the homogeneous equations $(\underline{\pi}_{21}\ \underline{\pi}_{22})\ \underline{\eta} = \underline{0}$, where $\underline{\eta}$ is an unknown (2 x 1) vector.

No further restrictions to be observed in the estimation procedure present themselves. Even if it is assumed that $E(u_{1i}u_{2i}) = E(u_{1i}\varepsilon_i) = E(u_{2i}\varepsilon_i) = 0$, the covariance matrix of reduced form disturbances is "full." If, however, (7a) is used in conjunction with $E(u_{1i}u_{2i}) = 0$, the reduced form covariance matrix will be diagonal, a restriction that then should be incorporated into the estimation procedure. We continue under the assumption that this covariance matrix is unrestricted, and that a multivariate normal distribution characterizes the joint density of $\underline{v}_1$, $\underline{v}_2$.

A computational technique that generates maximum likelihood estimates for $\underline{\gamma}_1$, $\underline{\gamma}_2$, $\underline{\beta}_3$, β_1, and β_2, in equation (8) is as follows. We transform equation (8) given a value for β_2 into

$$[(\underline{y}_1 - \beta_2\underline{s})\ \underline{y}_2] = (\underline{Z}_1\ \underline{Z}_2) \begin{pmatrix} \underline{\pi}_{11} & \underline{\pi}_{12} \\ \underline{\pi}_{21} & \underline{\pi}_{22} \end{pmatrix} + (\underline{v}_1\ \underline{v}_2). \tag{11}$$

The rank constraint on the $(\underline{\pi}_{21}\ \underline{\pi}_{22})$ submatrix of $\underline{\Pi}$ in equation (11) is the same constraint that appears in limited-information maximum likelihood (LIML) estimation.[9] What we are proposing is a search over values for β_2, for each using a LIML algorithm to estimate the $\underline{\Pi}$-matrix in equation (11). These estimates and the current value of β_2 would then be used to compute the maximum likelihood criterion function $|\underline{W}|$, where

$$\underline{W} = (\underline{Y} - \underline{Z}\ \hat{\underline{\Pi}})'(\underline{Y} - \underline{Z}\ \hat{\underline{\Pi}}), \tag{12}$$

with $\underline{Y} = [(\underline{y}_1 - \beta_2\underline{s})\ \underline{y}_2]$, $\underline{Z} = (\underline{Z}_1\ \underline{Z}_2)$, and $\hat{\underline{\Pi}}$ as it is estimated for equation (11). The maximum likelihood estimates are found when $|\underline{W}|$ is minimized. Since the search is only over a single parameter and $\underline{W}$ is (2 x 2), the computational burden of the procedure suggested is not great. The rewards from it are that the resulting estimates possess the desirable asymptotic properties of consistency and efficiency. Hypothesis testing is accomplished within the multivariate normal distribution, which holds asymptotically.

If, as may be the case in practice, the sample is not stratified according to demographic characteristics that are unique to the labor-supply function (in Hall's case it is, leaving only s_i as a variable unique to the labor-supply function), the search procedure must be expanded to cover the additional parameters involved.

4. EMPIRICAL RESULTS

Our choice of variables to specify equations (1) and (7) is as follows, which represents a compromise between the Hall work and that of Greenberg and Kosters. The sample used is Greenberg and Kosters' "working sample."[10]

y_{1i} = "labor supply," in annual hours, calculated as weeks worked in 1966 multiplied by hours worked or spent looking for work in the week preceding the interview,[11]

w_i = observed hourly wage, computed as last week's earnings divided by hours worked last week,

s_i = nonemployment income,[12]

$\underline{X}_i$ = intercept, age of head, race (white or nonwhite),

$\underline{Z}_i$ = intercept, age of head, race (white or nonwhite), years of education of head, SMSA (1-12), residence at age 16 (4-way dummy), union member? (yes or no), health (good, poor)

In addition, as an alternative to sample stratification, we included directly in the labor-supply equation variables for the number of adults in the family unit and for the number of children in the family unit. Effectively, then, two additional parameters (call them β_4 and β_5) are introduced into the search procedure (whereas in the model as set out in Section 2 there is but one parameter, β_2).

By way of comparison to the data bases and specifications of Hall and Greenberg and Kosters, our labor supply and wage variables are those used by Greenberg and Kosters. While the nonemployment income variable is quite similar to theirs also, it is not exactly the same. They include an "asset preferences" variable,[13] variables relating to earnings of the spouse and other family members (which we include directly in nonemployment income), and squared age in their labor-supply function, whereas we do not. Our model specification (which variables belong where) follows Hall's.

For the sample indicated the ML estimates obtained are as follows:

$\beta_1 = 52.21$

$\beta_2 = .004025$

$$\underline{\beta}_3 = \begin{cases} 2,131 \\ -140.2 \\ -.9128 \end{cases}$$

$$\beta_4 = .9101$$

$$\beta_5 = -14.96$$

$$\underline{\Upsilon}_1 = \begin{cases} 1.718 \\ -1.099 \\ .003690 \end{cases}$$

$$\underline{\Upsilon}_2 = \begin{cases} .1506 \\ .05087 \\ -.07803 \\ -.04961 \\ -.4209 \end{cases}$$

Since the sample is quite large and because the estimates we compute are already "optimal," for the main purpose at hand--namely, comparisons with results of other studies based on essentially the same sample--reporting of standard errors seems unnecessary.[14]

The wage (β_1) and income (β_2) coefficients, of course, are of primary interest. Greenberg and Kosters, for example, find a small, positive income coefficient (.03) in a labor-supply regression that excludes their preferences variable and other demographic controls.[15] In that same regression, the wage coefficient is negative (-78.55). Unfortunately, they do not report results for a model that includes demographic factors but excludes the preferences variable. Were the models otherwise completely comparable, the traditional errors-in-variables bias analysis would suggest a relative bias toward zero for the Greenberg and Kosters estimated wage coefficient. If a direct comparison of +52.21 and -78.55 is of any value, therefore, it must be to indicate the magnitude of the biasing factor due to the particular construction of the wage rate used. It is important to note, however, that if $E(u_{1i}u_{2i}) \neq 0$, two sources of errors-in-variables bias occur, one relative, the other additive. To take the simple two-variable textbook model (written in our notation) as an illustrative case, when w_i is uncorrelated with u_{1i} and u_{2i}, but $E(u_{1i}u_{2i}) \neq 0$, we find:

$$\text{plim } \hat{\beta}_1 = \frac{\beta_1}{1+\sigma_w^2/\sigma_{u_2}^2} + \frac{E(u_{1i}u_{2i})}{\sigma_w^2 + \sigma_{u_2}^2}, \qquad (13)$$

where $\hat{\beta}_1$ is the least squares estimate of β_1 from a regression of y_{1i} (labor supply) on y_{2i} (the erroneously measured wage). Aside from the bias toward zero in $\hat{\beta}_1$ (as indicated from the first term), there is an additive bias that pushes either upward or downward depending on the sign of $E(u_{1i}u_{2i})$.

$E(u_{1i}u_{2i}) \neq 0$ subsumes within it the situation we face here, where y_{1i} and w_i (as measured) are seen to contain a potentially "spurious" element of correlation, and is allowed for explicitly in our "appropriate" model. If β_1 is, in fact, a positive parameter, $E(u_{1i}u_{2i})$ must be negative in order to get a negative wage coefficient, which is a consistent interpretation of our findings in light of the anticipated negative direction of the "spurious" relationship between y_{2i} and y_{1i}.

In their main results based on a linear model, Greenberg and Kosters report a negative income coefficient (-.069) and a somewhat larger (absolute) wage coefficient (-103.2).[16] These convert to an income-compensated wage elasticity of +.064.[17] In our results, because the income coefficient is small and positive, it has a small negative effect (compensation). Our income-compensated substitution elasticity is, coincidentally, +.059, very close to theirs considering quantitative differences in the two coefficients involved. For us, almost all the response is from "price" effects, with very little effect of income compensation. For them, the income term dominates in what is, at best, a backward-bending supply function.

5. CONCLUSIONS

Our main purpose in this paper has been to identify and treat an econometric problem rather than to present an exhaustive set of results to complement and/or substitute for work already reported in other studies. This latter task we leave to those most qualified to effectively utilize the data and interpret their subsequent findings. What we have accomplished is an explicit treatment of the errors-in-variables problem arising in this specific setting, as applied to the SEO file and its peculiar characteristics. An indication of the magnitude of the biasing effect in others' work comes as a by-product of the analysis.

Other approaches to conquer the errors-in-variables bias problem are possible, but ours can be argued to be "optimal" in a meaningful sense. Moreover, any treatment other than an explicit analytical one can only open the door to additional difficulties that may cloud the main issue. For example, given the construction of the measured wage series (y_{2i}), various other measurements of labor supply might be used that do not have the obvious potential correlation with y_{2i} that our version does. One such alternative would multiply weeks worked or spent looking for work last year by 20 hours or 40 hours ("part-time" or "full-time"), this latter factor determined by whether the particular observation (person) reportedly worked fewer than or more than some "critical" number of hours last week (for example, 35). Because the sample is dominated by persons who claim they spent 51 weeks working or looking for work, however, the resulting dependent variable becomes an essentially dichotomous one. Such a "grouping" technique is a reasonable approach for reducing the errors-in-variables problem. But the resulting model is that of a "linear probability function," which is agreed by most econometricians to be a clear example of functional misspecification. So, while the least squares coefficient estimates so derived may not be subject to major criticism from the standpoint of errors-in-variables bias, functional misspecification causes them still to be regarded with extreme caution.

Although our empirical results--presumably obtained using the best "single equation" econometric framework available for the data--are quite favorable, one must question whether the assumption of an exogenous wage is really justified and hence whether the equation we have estimated is the supply schedule. The idea that this segment of the work force faces a given set of wages and merely supplies hours at those wages is surely somewhat naive, as is a simultaneous equations model that presumes observations are being generated by a simple underlying behavioral process. The movement to an econometric model which incorporates the notion that jobs and people are "matched" as the underlying process generating observations on hours worked and wages is discussed by Sherwin Rosen.[18] Within that framework one may have the potential to test the validity of an assumed exogenous wage, whereas here we use that assumption as a maintained hypothesis.

NOTES

[1]See Michael Boskin, "Income Maintenance Policy, Labor Supply and Income Redistribution," (Stanford: Stanford University, 1971). Research Center in Economic Growth Memo No. 111, and Irwin Garfinkel, "On Estimating the Labor Supply Effects of a Negative Income Tax," (Madison: Institute for Research on Poverty, 1971). Discussion Paper No. 101.

[2]Some prefer fitting the equation in "elasticity" form, so that (1) is log-linear. While little of what we say in this Section is affected by this matter of "taste," the analysis of bias due to observation error and our subsequent presentation of an appropriate econometric framework does not carry over directly to the log-linear case.

[3]David H. Greenberg and Marvin Kosters in "Income Guarantees and the Working Poor: The Effect of Income Maintenance Programs on the Hours of Work of Male Family Heads" (Santa Monica, Cal.: RAND Corporation, 1971), also report on results with explicit nonlinear forms for (1).

[4]For a two-variable model this effect is well-known. Its magnitude depends on the relative sizes of σ_w^2 and $\sigma_{u_2}^2$. The net effect on <u>all</u> coefficient estimates in a multiple regression with only one independent variable subject to measurement error is also downward in general, as Sandra S. Christensen illustrates in "Income Maintenance and the Labor Supply," (Madison: University of Wisconsin, 1971). Ph.D. Dissertation, pp. 53-59.

[5]In actuality, Hall fits an analysis of variance using the logarithm of y_{2i} as dependent variable. To analyze that specification directly would involve more sophisticated analysis than is necessary to make the essential points. So, I have chosen to misrepresent him.

[6]Ignoring the presence of s_i just to enable a simple demonstration of this statement, β_1 would thereby be estimated by $\Sigma y_{1i}\hat{w}_i/\Sigma\hat{w}_i^2$. Using $\hat{w}_i$ as an instrumental variable for the measured wage, the estimate of β_1 would be calculated as $\Sigma y_{1i}\hat{w}_i/\Sigma y_{2i}\hat{w}_i$. Since $\Sigma y_{2i}\hat{w}_i = \Sigma(\hat{w}_i + e_i)\hat{w}_i = \Sigma\hat{w}_i^2$ (using the fact that $\Sigma\hat{w}_i e_i = 0$ <u>in the sample</u>), the two estimates are identical.

[7]Although it is appealing to include control variables directly in the labor-supply function, with Hall's approach (as used by Boskin) the effect of

this overlap is to induce a "multicollinearity" problem, since $\hat{w}_i$ is a linear function of the variables in $\underline{Z}_i$, a subset of which form $\underline{X}_i$. When there is exact overlap ($\underline{Z}_i = \underline{X}_i$), of course, multicollinearity is "perfect," and unique least squares estimates of coefficients in equation (1) cannot be obtained. Boskin, "Income Maintenance Policy, Labor Supply and Income Distribution," p. 7, n. 4 also comments on this matter as one that involves identification of the labor-supply function, the proper interpretation in our "appropriate" model, which is described in the next section.

[8]Christensen, "Income Maintenance and the Labor Supply," pp. 15-17.

[9]See Arthur S. Goldberger, "Criteria and Constraints in Multivariate Regression" (Madison: Social Systems Research Institute, University of Wisconsin, 1970) Workshop Paper EME 7026; Arthur S. Goldberger and Ingram Oklin, "A Minimum-Distance Interpretation of Limited Information Estimation," Economica Vol. 39 (May 1971):635-39; and Arnold Zellner, "Estimation of Regression Relationships Containing Unobservable Variables," International Economic Review Vol. 11 (October 1970):441-54, which contain the necessary background material to support the claims made here for the estimation procedure. There also are contained descriptions of LIML in the context of Zellner's "seemingly unrelated regressions."

[10]Greenberg and Kosters, "Income Guarantees and the Working Poor: The Effect of Income Maintenance Programs on the Hours of Work of Male Family Heads," (Santa Monica, Cal.: RAND Corporation, 1971), p. 83. Report R-579-OEO.

[11]Ibid., p. 87.

[12]Our treatment of nonemployment income is as follows: to "unearned" income [Robert Hall, Wages, Income and Hours of Work in the U.S. Labor Force" (Cambridge: Massachusetts Institute of Technology, Department of Economics 1970), p. 27] we add imputed income from net home equity, net equity in vehicles, and net business or farm equity, and subtract an imputed value of other debts. All imputation is done using an interest rate of 8 percent. Finally, the resulting nonemployment income variable is deflated by the appropriate SMSA price deflator [Hall, p. 127].

[13]Greenberg and Kosters, "Income Guarantees and the Working Poor:" pp. 19-22.

[14]Moreover, owing to the nature of the algorithm used to calculate estimates, one must be satisfied with confidence regions generated from contours of the likelihood function rather than standard errors *per se*.

[15]Greenberg and Kosters, "Income Guarantees and the Working Poor:" p. 23.

[16]Ibid., p. 27.

[17]Ibid., p. 21.

[18]Sherwin Rosen, "A Theory of Hedonic Prices" (Rochester, N.Y.: University of Rochester, Department of Economics, 1972). No. 72-8.

REFERENCES

1. Boskin, Michael. "Income Maintenance Policy, Labor Supply and Income Redistribution." Stanford: Stanford University, May 1971. Research Memo. No. 111, Research Center in Economic Growth.

2. Christensen, Sandra S. "Income Maintenance and the Labor Supply." Madison: University of Wisconsin, 1971. Ph.D. Dissertation.

3. Garfinkel, Irwin. "On Estimating the Labor Supply Effects of a Negative Income Tax." Madison: Institute for Research on Poverty, University of Wisconsin, 1971. Discussion Paper 101-71.

4. Goldberger, Arthur S. "Criteria and Constraints in Multivariate Regression." Madison: Social Systems Research Institute, University of Wisconsin, July 1970. Workshop Paper EME 7026.

5. ____________________, and Oklin, Ingram. "A Minimum-Distance Interpretation of Limited-Information Estimation." Econometrica Vol. 39 (May 1971): 635-39.

6. Greenberg, David H. and Kosters, Marvin. "Income Guarantees and the Working Poor: The Effect of Income Maintenance Programs on the Hours of Work of Male Family Heads." Santa Mcnica, Cal.: RAND Corporation, December 1970. Report R-579-OEO.

7. Hall, Robert. "Wages, Income and Hours of Work in the U.S. Labor Force." Cambridge: Massachusetts Institute of Technology, Department of Economics, August 1970. Working Paper.

8. Kalachek, E. D., and Raines, Fredric Q. "Labor Supply of Low Income Workers." President's Commission on Income Maintenance Programs Technical Studies (1970), pp. 159-86.

9. Rosen, Sherwin. "A Theory of Hedonic Prices." Rochester: University of Rochester, Department of Economics, 1972. No. 72-8.

10. Zellner, Arnold. "Estimation of Regression Relationships Containing Unobservable Variables." International Economic Review Vol. 11 (October 1970):441-54.