ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions

Author(s): Thomas A. Mroz
Source: *Econometrica*, Vol. 55, No. 4 (Jul., 1987), pp. 765-799
Published by: The Econometric Society
Stable URL: https://www.jstor.org/stable/1911029
Accessed: 26-11-2019 18:58 UTC

Econometrica, Vol. 55, No. 4 (July, 1987), 765-799

THE SENSITIVITY OF AN EMPIRICAL MODEL OF MARRIED WOMEN'S HOURS OF WORK TO ECONOMIC AND STATISTICAL ASSUMPTIONS

BY THOMAS A. MROZ[1]

This study undertakes a systematic analysis of several theoretic and statistical assumptions used in many empirical models of female labor supply. Using a single data set (PSID 1975 labor supply data) we are able to replicate most of the range of estimated income and substitution effects found in previous studies in this field. We undertake extensive specification tests and find that most of this range should be rejected due to statistical and model misspecifications. The two most important assumptions appear to be (i) the Tobit assumption used to control for self-selection into the labor force and (ii) exogeneity assumptions on the wife's wage rate and her labor market experience. The Tobit models exaggerate both the income and wage effects. The exogeneity assumptions induce an upwards bias in the estimated wage effect; the bias due to the exogeneity assumption on the wife's labor market experience, however, substantially diminishes when one controls for self-selection into the labor force through the use of unrestricted generalized Tobit procedures. An examination of the maintained assumptions in previous studies further supports these results. These inferences suggest that the small responses to variations in wage rates and nonwife income found here provide a more accurate description of the behavioral responses of working married women than those found in most previous studies.

KEYWORDS: Female labor supply, specification tests, sample selection biases, taxes and labor supply.

EVERYONE FAMILIAR with the past ten years' research on empirical models of female labor supply is aware of the wide range of estimated income and substitution effects. Many studies and review articles have used economic and statistical arguments to explain some of this across study variation, and a few, such as Davanzo, DeTray, and Greenberg (1973), Heckman (1980), Borjas (1980), and Cogan (1981) have tried to test explicitly for the consequences of several economic and statistical misspecifications. Most empirical studies address some subset of these possible misspecifications, but the overlap of these studies is not sufficient for one to reach any firm conclusions about the practical importance of these considerations. Questions relating to the consequences of measurement error, sample selection bias, and the inclusion of taxes, to name only a few, remain unanswered. This study attempts a systematic analysis of many of the theoretical and statistical issues raised in previous studies of female labor supply. By using a single data set and by addressing these issues one at a time, it is possible to control for many of the methodological differences across studies. Consequently, many of the results reported here should serve as an extremely useful resource for future studies in this field.

[1] I would like to thank Angus Deaton, Mark Gritz, Jim Heckman, Joe Hotz, Harry Paarsch, John Pencavel, the referees, and especially Tom MaCurdy for many valuable comments. Support for this project was provided by NICHD (Grant #HD14256 to ERC/NORC), the Alfred P. Sloan Foundation, and the Hoover Institution.

Table I illustrates the range of estimates found in fairly comparable studies of married women's labor supply functions published during the last decade.[2] In an attempt to control for some sources of variation, we exclude studies restricted to low income samples and those convoluting the labor force participation decision with the hours of work decision from this table.[3] In addition, the table contains only studies using a measure of annual hours of work as the dependent variable. Each study presented here has stressed different theoretical and methodological issues. Some control for taxes, others control for wage rate endogeneity, and several take into account the issue of self-selection into the labor force.

Many reviews, such as Cain and Watts (1973), DaVanzo, DeTray, and Greenberg (1973), Garfinkel (1973), Borjas and Heckman (1978), Cogan (1980b; 1981),

TABLE I

ESTIMATES OF MARRIED WOMEN'S LABOR SUPPLY RESPONSES FROM PREVIOUS STUDIES

Evaluated at:

Wife's Wage:	$4.50	Husband's Wage:	$7.00
Wife's Hours:	1500	Husband's Hours:	2000

Household nonlabor income:	$1000
Labor income marginal tax rate:	0.339
Nonlabor income marginal tax rate:	0.280

		Wage Effect	Income Effect (per $1000)
1. Boskin (1973)	Instrumental Variables	29	−16.9
2. Cogan (1980a)	Tobit	865	−32.3
3. Cogan (1980a)	Instrumental Variables	349	−11.7
4. Cogan (1980b)	Tobit	632	−22.8
5. Cogan (1980b)	Fixed Costs	196	−8.5
6. Cogan (1981)	Fixed Costs	269	−22.4
7. Greenhalgh (1980)	Instrumental Variables	213	−65.6
8. Hausman (1981)	Convex Budget Set	328	−125.0
9. Hausman (1981)	Non-convex Budget Set	335	−118.0
10. Hausman (1981)	Fixed Costs	305	−113.0
11. Heckman (1976)	Tobit	1462	−73.4
12. Heckman (1976)	Generalized Tobit	−499	51.0
13. Heckman (1980)	Generalized Tobit	1401	−18.7
14. Layard et al. (1980)	Tobit	128	−118.2
15. Layard et al. (1980)	Instrumental Variables	22	−11.8
16. Leuthold (1978)	1967 Estimates	14	−3.0
17. Leuthold (1978)	1969 Estimates	45	−7.1
18. Leuthold (1978)	1971 Estimates	33	5.8
19. Nakamura and Nakamura (1981)		−16	−15.0
20. Schultz (1980)	Tobit	123	−67.0
21. Schultz (1980)	Instrumental Variables	−26	−1.9

Note: See Mroz (1984) for information on how these wage and income effects are calculated. All effects evaluated in terms of 1975 dollars.

[2] These wage and income effects are all evaluated at the point indicated at the top of Table I. The criteria used to translate the estimates from the various studies can be found in Mroz (1984). The relative ranking of the estimates and the general orders of magnitude in their differences change very little for variations in the point of evaluation.

[3] This latter restriction excludes, among others, Hall (1973) and Masters and Garfinkel (1977) who assign zero hours of work to nonworkers and include the nonworkers in a regression with workers.

and Moffitt and Kehrer (1981), discuss possible sources for disparate estimates of labor supply parameters. In Table II we classify the studies in Table I by several potential sources of discrepancy. A careful examination of these tables together can suggest several explanations for the disparate estimates, but in order to make any definitive statements, a detailed empirical analysis that varies a number of economic and statistical assumptions one at a time is needed.

We undertake such a sensitivity analysis by focusing on a simple model of the labor supply behavior of married women.[4] In this model, the husband's behavior is considered exogenous, and a woman's labor supply is given by

$$h_i = a_0 + a_1 \ln (w_{fi}) + a_2 Y_i + a_3' Z_i + e_i, \tag{1}$$

where h_i is the ith woman's hours of work during a given year, w_{fi} is a measure of her wage rate, Y_i is a measure of other income received by the household, Z_i is a set of control variables, e_i is a stochastic disturbance, and a_0, a_1, a_2, and a_3 are the parameters of the labor supply function. The vector Z_i includes the wife's age, her years of schooling, the number of children less than six years old in the household, and the number of children between the ages of five and nineteen. When taxes are introduced into this model, a linearization of the budget set is used. The marginal after-tax wage rates replace the gross wages, and the intercept of the linearized budget constraint at zero hours of work replaces the income term.[5]

Several factors determined this choice of functional form. Most importantly, these or similar models are those most frequently found in the literature. Other models derived explicitly from a specification of the preference function could be analyzed, but the introduction of a new or rarely encountered functional form would introduce an additional source of discrepancies with previous studies.[6] Furthermore, the linearity in the parameters allows for relatively simple estimation schemes and makes possible extensive testing of the hypotheses under consideration.

In this paper we examine three methodological considerations: exogeneity assumptions, statistical control for self-selection into the labor force, and the impact of controlling for taxes. In the first category, we test for the exogeneity of wage rates, nonwife income, children in the household, and the wife's labor market experience. In the second category, we test for the significance of the "Tobit" assumption and the presence of sample selection biases under a variety of distributional assumptions. We use statistical specification tests to uncover the importance of the methodological assumptions in the first two categories. The importance of accounting for taxes cannot be captured through parameter constraints and, as a result, we cannot explicitly test for the appropriate model in

[4] Mroz (1984) contains a similar sensitivity analysis for a household model of married women's hours of work.

[5] Specifically, let w be the pre-tax wage rate, Y_N be the family's after tax income, h be the observed choice of hours of work, and τ be the marginal tax rate. The wage measure used is $\ln[(1-\tau)w]$, and the income measure is the intercept of the linearized budget constraint, $Y_V = Y_N - (1-\tau)wh$. Hausman (1981) calls this income measure the wife's virtual income.

[6] See Stern (1986) for the specification of the indirect utility function which yields the labor supply function used in this paper.

TABLE II

METHODOLOGICAL CLASSIFICATION OF PREVIOUS STUDIES

	Data Set	Single Worker Model	Exogenous Wage	Exogenous Experience	No Sample Selection Controls	Tobit Restrictions	No Taxes	Exogenous Taxes
1. Boskin (1973)	SEO, 1967				×			×
2. Cogan (1980a)	NLS, 1966	×		×		×	×	
3. Cogan (1980a)	NLS, 1966	×		×	×		×	
4. Cogan (1980b)	PSID, 1975	×		×		×	×	
5. Cogan (1980b)	PSID, 1975	×		×			×	
6. Cogan (1981)	PSID, 1975	×		×			×	
7. Greenhalgh (1980)	GHS, 1971	×		×	×			×
8. Hausman (1981)	PSID, 1975	×	×	?		×		
9. Hausman (1981)	PSID, 1975	×	×	?		×		
10. Hausman (1981)	PSID, 1975	×	×	?				
11. Heckman (1976)	NLS, 1967			×		×	×	
12. Heckman (1976)	NLS, 1967			×			×	
13. Heckman (1980)	NLS, 1967						×	
14. Layard, et al. (1980)	GHS, 1974			×		×		×
15. Layard, et al. (1980)	GHS, 1974			×				×
16. Leuthold (1978)	NLS, 1967		×		×			×
17. Leuthold (1978)	NLS, 1969		×		×			×
18. Leuthold (1978)	NLS, 1971		×		×			×
19. Nakumura and Nakumura (1981)	1970 U.S. Census	×						
20. Schultz (1980)	SEO, 1966					×		
21. Schultz (1980)	SEO, 1966				×			

this framework. Instead we report the estimates for the different specifications and examine the point estimates to see whether or not the choice of economic model is important in reconciling the differences in the previous studies.

The data for this analysis come from the University of Michigan Panel Study of Income Dynamics (hereafter PSID) for the year 1975 (interview year 1976). Although this year was atypical of most of the 1970's, only in 1976 did PSID directly interview the wives in the households. During all other years the head of the household's interview supplied information about the wife's labor market experiences during the previous year. One suspects that the own reporting is more accurate, and it is for this reason that many recent studies of married women's labor supply have used these data.

Our sample consists of 753 married white women between the ages of 30 and 60 in 1975, with 428 working at some time during the year. This sample size is smaller than most used in the studies reported in Table I. The dependent variable, the wife's annual hours of work, is the product of the number of weeks the wife worked for money in 1975 and the average number of hours of work per week during the weeks she worked. The measure of the wage rate is the average hourly earnings, defined by dividing the total labor income of the wife in 1975 by the above measure of her hours of work. The nonwife income is defined as the household's total money minus the wife's labor income. The sample characteristics

TABLE III

MEANS OF THE DATA

(Standard Deviation in Parentheses.)

Variable name	Full Sample	Working Women
Wife's age	42.5	42.0
	(8.1)	(7.7)
Wife's education	12.3	12.7
	(2.3)	(2.3)
Children less than 6	0.24	0.14
	(0.52)	(0.39)
Children between 6 and 18	1.35	1.35
	(1.32)	(1.32)
Husband's age	45.1	44.6
	(8.1)	(8.0)
Husband's education	12.5	12.6
	(3.0)	(3.0)
Wife's wage (in dollars)	—	4.18
(average hourly earnings)		(3.31)
Husband's wage	7.48	7.23
(average hourly earnings)	(4.23)	(3.57)
Household nonlabor income ($1000)	3.76	3.39
	(5.90)	(6.07)
Wife's hours of work	740.6	1302.9
	(871.3)	(776.3)
Husband's hours of work	2267.3	2233.5
	(595.6)	(582.9)
Number of Observations	753	428

are presented in Table III and a detailed description of the data set construction can be found in Appendix 1.

This paper contains three sections. The first examines the statistical assumptions in the basic model. The second section presents the results with the controls for taxes. The final section summarizes the main conclusions and uses the results of the empirical analysis to shed some light on the empirical discrepancies found in previous studies in this field.

1. THE BASIC LABOR SUPPLY MODEL

1.1. *Choice of Baseline Specification*

The estimates presented in Table IV demonstrate the sensitivity of the wage and income coefficients to minor variations in the variables used to instrument the wage rate.[7] In this table (and Tables VI, VII, and VIII) we use the subsample of working women to calculate the estimates, and our estimation procedures (ordinary least squares and two stage least squares) do not control for self-selection into the labor force. Although many recent studies of female labor supply have stressed the importance of controlling for self-selection, we find

TABLE IV

CHOICE OF BASELINE SPECIFICATION

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *I*. Estimation Method: Two-Stage Least Squares.)

	$\ln(w_w)$	Nonwife Income/1000	Young Children	Older Children	Addtitional Instruments and Comments (R^2: R^2 in reduced form wage equation)
1.	−17	−4.2	−342	−115	$\ln(w_w)$, OLS, $R^2 = 1.0$
	(81)	(3.1)	(131)	(29)	
2.	1282	−8.3	−235	−60	E, $R^2 = .17$
	(461)	(4.6)	(182)	(49)	
3.	831	−7.0	−271	−78	E, $F2$, $R^2 = .18$
	(312)	(3.8)	(155)	(39)	
4.	672	−6.4	−283	−85	E, $F3$, $R^2 = .21$
	(217)	(3.6)	(147)	(36)	
5.	482	−5.8	−300	−93	E, $F3$, $H3$, $R^2 = .23$
	(171)	(3.4)	(138)	(33)	
6.	638	−6.3	−287	−87	E, $F4$, $R^2 = .22$
	(197)	(3.5)	(145)	(35)	
7.	−182	−3.7	−356	−122	$F2$, $R^2 = .15$
	(355)	(3.5)	(138)	(33)	
8.	46	−4.4	−337	−112	$F3$, $R^2 = .18$
	(220)	(3.3)	(131)	(30)	
9.	−30	−4.2	−338	−113	$F3$, $H3$, $R^2 = .20$
	(174)	(3.3)	(129)	(29)	
10.	129	−4.7	−330	−108	$F4$, $R^2 = .19$
	(201)	(3.2)	(130)	(30)	

[7] All standard errors reported in this study are corrected for arbitrary forms of heteroscedasticity. None of the tables report the intercept or the coefficients on age and education in the labor supply equation. See Mroz (1984, Appendix 2), for several estimates of the complete labor supply function and reduced form wage equation.

these simple, "first generation," labor supply models an informative starting point for this analysis. In Section 1.6 we examine the empirical consequences of failing to account for the possible sample selection bias.

The first row of Table IV contains the ordinary least squares estimates. Rows two through six use as instruments for the wife's wage rate the wife's reported labor market experience,[8] this variable squared, and several different polynomials in the wife's and husband's age and education. Rows seven through ten use the identical age-education polynomials as instruments, but do not contain measures of the wife's labor market experience. In order to translate the ln(wage) coefficients to uncompensated wage effects, we use the same point of evaluation as used in Table I, namely, \$4.50. Dividing the ln(wage) coefficients by this wage rate gives a range of estimated uncompensated effects of −40 to 280 hours per year per dollar increase in the wage rate.[9]

At the top of each table is a list of the assumptions maintained throughout the table. In Table IV, for example, each set of estimates uses two-stage least squares as the estimation method and the wife's background variables, the children variables, and the nonwife income as instruments. The definitions of the instrumental variable sets are in Table V. The rightmost column in the table lists any

TABLE V

DEFINITIONS OF INSTRUMENTAL VARIABLES

Each set of instrumental variables contains a constant, the wife's age and the wife's education.

B:	Background variables: County unemployment rate, SMSA dummy, number of years of schooling of wife's mother, number of years of schooling of wife's father.
C:	Children variables: Number of children less than six in the household, number of children between the ages of five and nineteen.
E:	Wife experience variables: Number of years the wife worked since age eighteen and this variable squared.
$F2$:	Quadratic terms in the wife's age and education: Age, age squared, education, education squared, age times education.
$F3$:	Cubic terms in wife's age and education: Variables in $F2$ plus age^3, education^3, age^{2}education, education^{2}age.
$F4$:	Quartic terms in wife's age and education: Variables in $F3$ plus age^4, education^4, age^{2}education^2, age^{3}education, education^{3}age.
$H2$:	Quadratic terms in husband's age and education: Analogous to terms in $F2$, plus number of years of schooling of husband's mother and number of years of schooling of husband's father.
$H3$:	Cubic terms in husband's age and education: Variables in $H2$ plus analogous terms to the variables in $F3$.
$H4$:	Quartic terms in husband's age and education: Variables in $H3$ plus analogous terms to $F4$.
HW:	The logarithm of the husband's average hourly earnings.
I:	Nonwife income.
$\ln(w_w)$:	Wife's 1975 average hourly earnings.
$\ln(\text{wage}_{1976})$:	Wife's wage on job at time of the 1976 interview.
NL:	Household nonlabor income.

[8] The measure of labor market experience used in this study is the number of years the woman worked for money since her eighteenth birthday.

[9] All conversions in this study are evaluated at the same point as used in Table I.

additional instrumental variables for that row and other deviations from the common assumptions. Specification three, for instance, includes the wife's labor market experience variables and the fully interacted quadratic terms in her age and education in the instrument set.

The primary use of Table IV is to aid in selecting a baseline set of instrumental variables. For this purpose we use two tests: goodness of fit tests for the reduced form wage equation and tests of overidentifying restrictions. The goodness of fit tests are standard F tests, and the tests for overidentifying restrictions are a variant of Basmann's (1960) test.[10] Using a five per cent level of significance, the best wage equation, in the sense of the simplest model which is not rejected in favor of the model containing the next higher order terms in the age-education polynomials, contains the cubic terms. This result holds for the wage specifications with the wife's labor force experience variables (specification 4), those without the experience variables (specification 8), and the specifications with the husband variables (specifications 5 and 9).[11] When the wife's labor force experience is not in the instrument set (rows 7 through 10) none of the overidentifying restrictions is rejected at the five per cent level. There is, however, some evidence of invalid overidentifying restrictions when the wife's labor market experience variables are used to help identify the wage effect. On the basis of these tests, we choose the specifications with the third order terms in the instrument set as the baseline specifications.[12]

1.2. *Testing Model Specifications*

The wide range of estimated wage effects found in Table IV suggests that assumptions concerning the sets of instruments used to estimate the model can have considerable impact upon the estimated structural parameters. Notice that estimates using the set of instrumental variables with the wife's market experience

[10] To test for the overidentifying restrictions, the model is estimated by two-stage least squares, and a new variable is constructed as the difference between the hours of work variable and the estimated wage coefficient times the ln(wage) variable. This new variable is regressed against all the variables in the set of instruments (including the exogenous variables in the labor supply equation). A likelihood ratio test at the five per cent level (assuming homoscedastic normal disturbances) is used to test for the inclusion of the overidentifying instruments in this auxiliary regression. Note that in these tests we have assumed that the ln(wage) variable is the only right-hand side endogenous variable.

[11] Specifically, the reduced form wage equations implied by rows 2 and 3 are rejected in favor of row 4. The quartic terms in row 6 do not improve the fit of the wage equation with the cubic terms (row 4). Similarly, row 7 is rejected in favor of row 8, and row 10 does not increase the explanatory power of the wage equation in row 8. Although not reported here, the tests conducted by adding higher order polynomials in the husband's age and education at the same time that the wife's polynomials are increased imply that the cubic instrumental variable sets are the preferred specifications (rows 5 and 9).

[12] The use of the husband variables in the wife's reduced form wage equation is not typical of most analyses of married women's labor supply. They are introduced into this table to demonstrate that the basic models are not sensitive to their inclusion. The specifications with the husband variables are used here to increase the efficiency of the estimates when the exogeneity assumptions on children and nonwife income are tested.

(rows 2–6) yield larger wage responses than the rows without this set of instruments (rows 7–10). This suggests a possible specification error. To test for such errors, we apply variants of the specification tests proposed by Durbin (1954), Wu (1973), Hausman (1978), and White (1982b).

The motivation for these tests arises from the observation that two estimators, both consistent under some null hypothesis, should yield similar sets of estimates of the structural parameters. For example, consider the following linear regression model

$$y_i = b'X_i + g'Y_i + v_i, \tag{2}$$

where X_i is a set of assumed exogenous variables, Y_i is a set of variables possibly correlated with the disturbance v_i, and b and g are the parameters of this structural relationship. Define X^* as the set of maintained exogenous variables excluded from the structural equation. Under the maintained assumptions, applying the instrumental variables estimator to equation (2) using as instruments $Z_1 = \{X, X^*\}$ yields a consistent estimate $\hat{\theta}_0 = (\hat{b}, \hat{g})$ of the true parameter vector $\theta_0 = (b_0, g_0)$.

Now consider a set of variables W whose inclusion in the instrumental variables set may yield more efficient estimates of (b, g). (This set W may contain variables in Y.) Suppose one suspects that some elements in W may be correlated with the disturbance v in equation (2); i.e., W contains invalid instrumental variables. A test of this hypothesis has the following form:

H_0: W is exogenous with respect to the disturbance in equation (2);

H_1: W is not exogenous.

Define another instrumental variable set $Z_2 = \{X, X^*, W\}$. Estimating equation (2) by instrumental variables using Z_2 as instruments will yield an estimate $\hat{\theta}_1 = (\bar{b}, \bar{g})$ of the parameters $\theta_1 = (b_1, g_1)$. Under the null hypothesis, θ_1 is identical to θ_0 (see Sargan (1959) and Amemiya (1974)). If W is not exogenous, then θ_1 need not equal θ_0. Consequently, the hypothesis can be stated more specifically as:

H_0: $\theta_1 = \theta_0$;

H_1: $\theta_1 \neq \theta_0$.

If the value of θ_0 were known, such a hypothesis test would be straightforward; here only estimates of θ_0 and θ_1 are available. But under the maintained assumptions, if H_0 is true, then plim $(\hat{\theta}_1 - \hat{\theta}_0) = 0$.

Suppose, for the moment, that a procedure for obtaining a consistent estimate of the covariance matrix of $(\hat{\theta}_1, \hat{\theta}_0)$ is available and that these estimates have asymptotically a normal distribution. The test of the null hypothesis can be described as follows: First, estimate the model with the instrumental variable set Z_2 (i.e., under the null hypothesis). Next, estimate the model with the instrumental variable set Z_1 (under the alternative hypothesis). Construct an estimate of the covariance matrix of $(\hat{\theta}_1, \hat{\theta}_0)$ and use a standard χ^2 test to test whether $(\hat{\theta}_1 - \hat{\theta}_0)$ is significantly different from the zero vector. If $(\hat{\theta}_1 - \hat{\theta}_0)$ is significantly different from zero, then one rejects the null hypothesis that the instrumental variable set W is exogenous.

Because the estimator under the null hypothesis need not be efficient, one cannot use Hausman's (1978) formulae for the covariance matrix of the differences between the two sets of estimates. White (1982a) provides formulae for the covariance matrix of $(\hat{\theta}_1, \hat{\theta}_0)$ when quasi-maximum likelihood procedures are used to estimate θ_1 and θ_0. He shows that $(\hat{\theta}_1, \hat{\theta}_0)$ have an asymptotic normal distribution. In addition, the construction of the covariance matrix does not rely on either the normality or homoscedasticity of the disturbances, and it takes into account the correlation between these two sets of estimates. The actual formulae used to construct the covariance matrix of $(\hat{\theta}_1, \hat{\theta}_0)$ are described in Appendix 2.

1.3. *Endogeneity of Wage Measures and Labor Market Experience Without Controls for Self-Selection Biases*

In Table VI we present several sets of instrumental variables estimates for the wife's labor supply equation as well as tests of the equality of coefficients across these specifications. A comparison of the first two specifications suggests that the exogeneity assumption on the average hourly earnings is not unreasonable. This is a surprising result, for the average hourly earnings is defined by dividing the wife's labor earnings by the dependent variable; any measurement error in the hours of work measure should introduce a spurious negative correlation between this wage measure and the dependent variable. We shall return shortly to this observation and offer an alternative explanation for the similarity between these estimates. The second and third specifications differ only by the inclusion of the labor force experience variables in the reduced form wage equation. The difference in the point estimates on the wage coefficients is 627, and the asymptotic normal statistic for the equality of the wage coefficients takes the value 3.0; we conclude that the wife's labor market experience is an invalid instrumental variable. This result is in accord with Heckman's (1980) rejection of the exogeneity of the wife's labor force experience in the labor supply equation.

The comparison of specifications two and three belabors the obvious point that women who have worked many years in the past tend to have higher wages and work more in the present. Intuitively, the difference in the number of years worked between two women (identical in all other observed exogenous characteristics) reflects a systematic difference in the unobservables influencing their labor supplies (e.g., "tastes for work"). This makes the women's labor market experience endogenous to the labor supply function given in equation (1). As a result, the correlation between the predicted wage rate and a woman's hours of work obtained by predicting the wage with her previous labor market experience does *not* correspond to the economic notion of an uncompensated wage effect. The appropriate conceptual experiment requires an exogenous change in the wage rate (e.g., holding "tastes" constant), and the above test suggests that the variations across women in previous work experience do not satisfy the requisite *ceteris paribus* assumptions.[13]

The unexpected similarity between specifications one and two can be explained by a mixture of two opposing effects: endogeneity of the actual wage rate and the spurious correlation between hours and wages resulting from defining wages as average hourly earnings. Specifications four through six in Table VI present the results leading to this conclusion. In the 1976 interview the PSID asked each woman the wage rate on her current job. This measure should be an ideal instrument to control for the average hourly earnings measurement error.[14] This reported wage rate, however, is not available for women who are not at work at the time of the interview. Only 326 of the 428 women who worked in 1975 have this measure of the wage rate, and this may introduce an additional sample selection bias.[15]

TABLE VI

ENDOGENEITY OF WAGES

(Standard Errors in Parentheses. Common Instrument Sets: B, C, I, $F3$. Estimation Method: Two-Stage Least Squares.)

Specification	$\ln(w_w)$	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
1.	−17	−4.2	−342	−115	$\ln(w_w)$, OLS
	(81)	(3.1)	(131)	(29)	
2.	46	−4.4	−337	−112	
	(220)	(3.3)	(131)	(30)	
3.	672	−6.4	−283	−85	E
	(217)	(3.6)	(147)	(36)	
4.	−32	−4.4	−275	−98	$\ln(w_w)$, OLS, Restricted subsample
	(126)	(3.4)	(168)	(31)	
5.	348	−5.9	−301	−88	$\ln(\text{wage}_{1976})$, Restricted subsample
	(133)	(3.6)	(171)	(31)	
6.	−244	−3.6	−260	−103	Restricted subsample
	(260)	(3.6)	(171)	(31)	
Differences between Specifications					
3 and 2	627	−2.0	54	27	Reject spec. 3 (i.e., reject exogenous labor market experience)
	(209)	(1.8)	(69)	(19)	
5 and 4	380	−1.5	−26	10	Reject spec. 4 (i.e., reject exogenous average hourly earnings)
	(142)	(1.1)	(41)	(11)	
6 and 5	−592	2.4	41	−15	Reject spec. 5 (i.e., reject exogenous 1976 reported wage)
	(247)	(1.7)	(64)	(18)	

[13] Note that the loss in precision by excluding the wife's labor force experience from the instrument list is trivial. The only effect of including these variables is to increase the size of the wage coefficient.

[14] The R^2 for the regression of 1975 average hourly earnings on the 1976 reported wage is 0.42, and the coefficient on the reported wage is 0.91.

[15] Controlling for the bivariate sample selection rule, the woman must work at some time during 1975 and be at work at the time of the 1976 interview, yields similar results to those reported below. Note that this wage measure is not observed for 427 women. It is, therefore, infeasible to use this variable in the sample selection function and the construction of the conditional means. This may explain why we reject the exogeneity of this wage measure even after controlling for sample selection biases.

Specification four reports the ordinary least squares estimates over the restricted subsample. The point estimates are quite close to those in the unrestricted sample. Specification five uses the 1976 wage measure as an instrument for the average hourly earnings. The point estimate increases by 380, and this increase is significant at the five per cent level. We interpret this as evidence of a substantial amount of measurement error in the average hourly earnings.[16]

The large estimated wage effect from specification five is at odds with the estimates in specification two. Specification six relates the exogeneity assumption on the 1976 wage measure for the restricted subsample, and we find a significant difference between the specifications with and without this measure. As in the case of the wife's labor market experience, the 1976 wage rate is correlated with unobservables in the labor supply equation, and the variations in this measure do not capture an exogenous change in the 1975 wage rate. The comparisons between specifications four through six strongly support the hypothesis that two sources of endogeneity are working through the average hourly earnings variable. In this sample these effects tend to cancel. The standard errors of the ordinary least squares estimates, however, suggest a greater degree of precision of the estimates than may be warranted. In other samples or with nonlinear specifications one should not expect these two effects to cancel.

1.4. *Endogeneity of Children and Nonwife Income*

Several previous studies have discussed the endogeneity of variables measuring previous life-cycle decisions in a static model of labor supply. Greenberg and Kosters (1973) and Smith (1980), for example, discuss the endogeneity of unearned income resulting from persistent unobserved taste components. Schultz (1980) questions the interpretation of labor supply estimates containing family composition variables which are themselves the result of previous household decisions, and Hotz (1980) explicitly incorporates family formation into a life-cycle model of female labor supply. As a first step towards assessing the importance of the potential biases introduced by these variables, we use the tests for exogeneity described above.

Table VII presents estimates of the female labor supply function under various assumptions about the endogeneity of the nonwife income and the wife's labor force experience. In *none* of the specifications do we reject the exogeneity assumption on the nonwife income, and only when experience is treated as exogenous do the point estimates change appreciably. From this table there is no evidence that the nonwife income variable is endogenous.[17]

From the estimates in Table VIII we do not reject the exogeneity assumptions for the children variables in the labor supply equation. The point estimates of

[16] If the unobserved determinants of the woman's 1975 labor supply have a large impact on her wage rate in 1976, then one might argue that this is not a test for measurement error but rather a test for endogeneity of the 1976 wage rate.

[17] When the nonwife income is treated as endogenous (specifications one and three in Table VII), we again reject the exogeneity assumption on the wife's labor force experience variables. We reject both the equality of the wage and nonwife income coefficients.

TABLE VII

ENDOGENEITY OF NONWIFE INCOME

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *F*3, *H*3. Estimation Method: Two-Stage Least Squares.)

Specification	$\ln(w_u)$	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
1.	−29	−3.1	−341	−117	
	(174)	(8.7)	(130)	(31)	
2.	−29	−4.2	−334	−116	*I*
	(174)	(3.3)	(130)	(30)	
3.	405	−18.6	−316	−84	*E*
	(174)	(8.1)	(318)	(33)	
4.	482	−5.8	−300	−93	*E*, *I*
	(171)	(3.4)	(138)	(33)	
Differences between Specifications					
2 and 1	0	−1.1	2	1	Fail to reject spec. 2
	(5)	(8.1)	(6)	(8)	(exogenous nonwife income)
4 and 3	78	12.7	16	10	Fail to reject spec. 4
	(57)	(7.7)	(21)	(8)	(exogenous nonwife income)
4 and 1	512	2.8	−42	23	Reject spec. 4 (Joint exogeneity
	(148)	(8.3)	(56)	(17)	of nonwife income and labor market experience)

TABLE VIII

ENDOGENEITY OF CHILDREN AND NONWIFE INCOME

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *F*3, *H*3. Estimation Method: Two-Stage Least Squares.)

Specification	$\ln(w_u)$	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
1.	−30	−4.2	−344	−116	*C*, *I*
	(174)	(3.3)	(130)	(30)	
2.	−52	−5.4	−298	14	*I*
	(189)	(3.6)	(380)	(85)	
3.	−29	−3.1	−341	−117	*C*
	(174)	(8.7)	(130)	(31)	
4.	−53	−3.7	−278	14	
	(190)	(9.2)	(385)	(84)	
5.	482	−5.8	−301	−93	*C*, *E*, *I*
	(171)	(3.4)	(138)	(33)	
6.	442	−5.3	−966	−176	*E*, *I*
	(175)	(3.7)	(343)	(87)	
Differences between Specifications					
1 and 2	22	1.7	−45	−129	Fail to reject spec. 1
	(57)	(1.2)	(344)	(79)	(exogenous children)
3 and 4	24	0.7	−63	−131	Fail to reject spec. 3
	(58)	(2.3)	(351)	(78)	(exogenous children)
5 and 6	40	−0.6	663	83	Fail to reject spec. 5
	(82)	(1.8)	(357)	(81)	(exogenous children)

the coefficients on these variables change appreciably (but not significantly) only when the wife's labor market experience is included in the instrument list and the children variables are excluded from the instrument list. Given the previous rejection of the wife's labor market experience as a valid instrument, such a result is not unexpected and does not indicate any evidence for the endogeneity of children.

1.5. *Estimating and Testing Model Specifications with Controls for Self-Selection Biases*

All of the estimates in the previous sections were constructed from a subsample of working women without any controls for self-selection into the labor force. This subsample contains only fifty-eight per cent of the women in our sample and, as is well known, estimates derived from self-selected samples may be biased due to correlations between the independent variables and the stochastic disturbance induced by the sample selection rule. Indeed, most of the studies of female labor supply over the past ten years have focused on statistical controls for these sample selection biases.

In this study we examine several methods to correct for sample selection biases. All of these methods can be described in a common statistical framework. Define d_i as a dummy variable equal to 1 if the ith woman works and zero otherwise. Let Z_1 be the set of all exogenous variables in the model, and consider the latent variable $I_i = f'Z_{1i} + u_{Ii}$. We define the labor force participation function as $d_i = 1$ if and only if $I_i > 0$, and $d_i = 0$ otherwise. In general, one can consider I_i as an unobserved measure of the difference in utility between working and not working; an individual works only if the utility of working (receiving labor income and forgoing home time) is greater than the utility from not working. After making a distributional assumption on the disturbances u_{Ii}, one can estimate the parameters f up to a common scale factor by binary choice methods.

Now consider the reduced form specifications for the hours of work function and the log-wage function:[18]

$$h_i = a_0 + a_1(c'Z_{1i}) + a_2 Y_i + a_3' Z_i + u_{1i}, \tag{3}$$

$$\ln(w_{fi}) = c'Z_1 fi_i + u_{2i}, \tag{4}$$

where $u_{1i} = a_1 u_{2i} + e_{1i}$. We make the following restrictions on the disturbances in these reduced form equations:

Assumption 1: $u_{ji} = \rho_j u_{Ii} + v_{ji}^*$, *for* $j = 1, 2$.

Assumption 2: $E(v_{ji}^*) = 0$, *and the* v_{ji}^* *are independent of* u_{Ii} *and* Z_{1i}.

The Tobit model used by Heckman (1974) and considered by Schultz (1980) and Cogan (1980a, 1980b, 1981) can be represented by the following restrictions on this model.

[18] For ease of presentation, we treat the elements of Z_i and Y_i as exogenous. Conceptually there is no difficulty in allowing these variables to be endogenous. All that is required is to introduce their reduced form specifications and substitute these reduced forms into equation (3). The required assumptions are analogous to those introduced on the log-wage reduced form disturbance.

RESTRICTION 1: $\rho_1 = 1$, $\mathrm{Var}(v_{1i}^*) = 0$.

RESTRICTION 2: The parameters f incorporate the reduced form hours equation parameter restrictions up to a constant of proportionality equal to $[\mathrm{Var}\,(u_{1i})]^{1/2}$.

RESTRICTION 3: u_{1i} and u_{2i} are distributed joint normal.

In this study we estimate the Tobit model by the method of maximum likelihood. Heckman (1974) describes the construction of the likelihood function for the random variables h_i and $\ln(w_{fi})$. We also consider a conditional Tobit specification. The likelihood function in this formulation is constructed by conditioning the standard Tobit likelihood function on the event that the woman works. This model is also known as the truncated normal regression model. It is estimated only over the subsample of working women.

The other estimators in this study fall in the class of generalized Tobit estimators. These estimators relax the restrictions of the constrained Tobit model. We estimate these models with a multi-stage procedure. This procedure is simpler to estimate than maximum likelihood and imposes fewer distributional assumptions.[19] The reduced form labor supply function conditional upon the wife working is defined by

$$h_i = a_0 + a_1(c'Z_{1i}) + a_2 Y_i + a_3' Z_i + \rho_1 E(u_{Ii} | I_i > 0) + v_{1i} \tag{5}$$

where $v_{1i} = \rho_1[u_{Ii} - E(u_{Ii} | I_i > 0)] + v_{1i}^*$. For the subsample of working women, v_{1i} has mean zero and is uncorrelated with the independent variables in the reduced form specification. Similarly, the reduced form wage equation is given by

$$\ln(w_{fi}) = c'Z_{1i} + \rho_2 E(u_{Ii} \,|\, I_i > 0) + v_{2i}, \tag{6}$$

where $v_{2i} = \rho_2[u_{Ii} - E(u_{Ii} \,|\, I_i > 0)] + v_{2i}^*$. The new disturbance in this equation also has desirable properties. After estimating the parameters of the labor force participation function, one can construct consistent estimates of the conditional expected values. These estimates replace the conditional expected values in equation (6), and least squares applied to this equation yields consistent estimates of the wage parameters. The estimated wage parameters, $\hat{c}$, are used to construct a consistent estimate of $c'Z_{1i}$. These estimates replace the $c'Z_{1i}$ in equation (5) and the consistent estimates of the conditional means replace the conditional mean. Least squares applied to this modified equation (5) yields consistent estimates of the labor supply parameters.[20] Lee (1982) provides formulae for the asymptotic variance-covariance matrix of these estimates.

[19] These multi-stage procedures relax the assumption that the v_{ji}^* are normally distributed, as would be necessary in the maximum likelihood framework.

[20] The actual procedure used here was first to obtain the estimates of the conditional expected value. The structural labor supply function given in equation (1) was modified to include the estimated conditional mean and was estimated by two-stage least squares. By following this procedure, the coefficient on the conditional mean estimates the correlation of the structural residual with the sample selection disturbance, $\rho_1 + a_1\rho_2$, instead of ρ_1.

The estimates of the labor supply parameters may depend upon the distribution assumed for the disturbance in the labor force participation function. One important aspect of this study concerns the sensitivity of the estimated structural responses to this assumption. In order to assess the importance of the distributional assumption, we use a variety of distributions with differing degrees of skewness and kurtosis. The most common assumption in the literature is the normality assumption. As is well known, the conditional expected value of this distribution is given by

$$E_n(u_{Ii} \mid I_i > 0) = n(-f'Z_{1i})/N(-f'Z_{1i}),$$

where $n(\cdot)$ is the standard normal density and $N(\cdot)$ is the standard normal cumulative. To allow for more kurtosis, we chose the logistic distribution. Its conditional expected value is given by

$$E_1(u_{Ii} \mid I_i > 0) = [1+\exp(-f'Z_{1i})] \ln [1+\exp(-f'Z_{1i})] + f'Z_{1i} \exp(-f'Z_{1i}).$$

We attempted several different distributional assumptions to allow for skewness in the disturbance. First, we estimated a binary choice model based on the assumption of a singly truncated bivariate normal. This distribution function is defined by specifying a bivariate normal distribution function, $n_2(u_{Ii}, u_i^*, \rho)$, where ρ is the correlation coefficient; this bivariate distribution is then conditioned on the event $u_i^* > \mu$. For this distribution, the density function of u_{Ii} is given by

$$\frac{\int_\mu^\infty n_2(u_{I_i}, u^*, \rho)\,du^*}{\int_\mu^\infty n(u^*)\,du^*},$$

where both μ and ρ are parameters to be estimated. This univariate distribution allows for either positive or negative skewness and contains the normal distribution as a special case.

In several of the models analyzed using this distribution, the estimated value of ρ was arbitrarily close to 1. The limiting distribution in this case is a singly truncated normal distribution. Imposing this limiting distribution yielded exact predictions of several women's labor force participation, and we deemed it an inappropriate statistical model. The shape of the estimated singly truncated normal distribution suggested that a skewed distribution might be appropriate. Thus, we used a log-normal distribution with conditional mean given by

$$E_{ln}(u_{Ii} \mid I_i > 0) = \exp(.5\sigma^2) \{N[\sigma - \ln(-f'Z_{1i})/\sigma]/N[-\ln(-f'Z_{1i})/\sigma] - 1\},$$

where $N[\cdot]$ is the cumulative standard normal.

Testing exogeneity assumptions in these models that control for sample selection biases is analogous to the testing in the simple instrumental variables models. We form two different sets of independent variables, Z_1 and Z_2. As before, Z_1

is a proper subset of Z_2. We estimate the above models first using the set Z_1 and then with the set of variables Z_2, and we test for the equality of the coefficients of the labor supply equation in these two specifications. A rejection of the equality of these coefficients is considered a rejection of the exogeneity of the variables in the set Z_2 not included in the set Z_1. The construction of the variance-covariance matrix of these correlated estimates is quite complex, and requires accounting for the cross-specification covariance of the binary choice estimated parameters as well as the correlation of the structural parameter estimates with the binary choice estimates. The formulae for these asymptotic covariance matrices appear in Appendix 2. These formulae correct for the heteroscedasticity induced by the conditional mean adjustments as well as the estimation error from the multiple stage estimation method.

The exogeneity tests presented in the following two sections rely heavily upon the distributional assumption in the labor force participation function. The most restrictive assumption is that the reduced form disturbances in this function, under both the null and alternative hypotheses, fall in the same class. Specifically, suppose that the reduced form of the supposed endogenous instruments is given by $W_i = \delta' Z_{1i} + \eta_i$. The assumption maintained under these tests is that the disturbances u_{Ii} and $u_{Ii} + f'_w \eta_i$, where f_w are the coefficients on the potential endogenous variables in the participation function, fall in the same class of distributions. Under the normality assumption, one set of sufficient conditions for the constancy of the distribution function is that both u_{Ii} and η_i are homoscedastic normal random variables. Under the lognormal and logistic assumptions, there are no simple characterizations to guarantee this assumption.[21]

1.6. *Comparison of Sample Selection Models*

In Table IX we report the estimates of the labor supply function under different controls for self-selection into the labor force. The comparison of the Tobit estimates (specification one) and the conditional Tobit estimates (specification two) provides a specification test of the sort proposed by Hausman (1978). The comparison of specifications one and three yields a test of the Tobit restrictions for the case of normal censoring. In both instances we reject the Tobit specification.[22] This result is in accord with Cogan's (1981) examination of the Tobit model.[23] Specifications three, four, and five use alternative distributional assumptions on the disturbances in the labor force participation function; the estimated labor supply parameters are remarkably similar across these different specifications. A comparison of these specifications to the two-stage least squares estimates without sample selection controls provides a test for the importance of

[21] In work in progress, Mroz (1986), we have relaxed this assumption. For the case of u_{Ii} normally distributed (and no additional assumptions on the distribution of $u_{Ii} + f'\eta_i$), we find the same results for the test of the exogeneity of the wife's labor force experience as reported in Section 1.7.

[22] See Appendix 2 for the construction of the across specification variance-covariance matrices when one estimator is maximum likelihood and the other is a multistage estimator.

[23] Cogan's test, however, treated the wife's labor force experience as exogenous.

TABLE IX

SELF-SELECTION MODELS: WITHOUT WIFE'S EXPERIENCE

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *F*3, *I*. Estimation Method: See Comments.)

Specification	ln (w_w)	Nonwife Income/1000	Young Children	Older Children	$\sigma^2_{u_1}$	Additional Instruments and Comments
1.	261	−22.9	−1035	−97	1.6×10^6	Tobit
	(357)	(5.1)	(140)	(48)		
2.	157	−7.2	−561	−166	8.1×10^5	Conditional Tobit
	(395)	(6.5)	(145)	(55)		
3.	64	−1.0	−183	−106	—	Multistage estimates
	(227)	(9.2)	(408)	(34)		Normal distribution
4.	110	−0.8	−183	−103	—	Multistage estimates
	(223)	(10.2)	(418)	(36)		Lognormal distribution
5.	66	−1.1	−184	−105	—	Multistage estimates
	(226)	(8.7)	(385)	(34)		Logit distribution
6.	46	−4.4	−337	−112	—	Two-stage Least
	(220)	(3.3)	(131)	(30)		Squares
Differences between Specifications						
1 and 2	−106	15.7	474	70	7.8×10^5	Reject spec. 1 (Tobit
	(403)	(6.6)	(158)	(56)	(1.4×10^5)	spec. restrictions)
1 and 3	−197	21.8	853	−9	—	Reject spec. 1 (Tobit
	(342)	(10.0)	(413)	(46)		spec. restrictions)
6 and 3	−19	−3.4	−153	−7	—	Fail to reject spec. 6
	(84)	(8.8)	(375)	(14)		(no self-selection bias)
6 and 4	−64	−3.6	−153	−9	—	Fail to reject spec. 6
	(85)	(9.7)	(382)	(18)		(no self-selection bias)
6 and 5	−20	−3.4	−153	−7	—	Fail to reject spec. 6
	(74)	(8.2)	(352)	(14)		(no self-selection bias)

these controls. For none of these three distributions is there any evidence that the failure to control for self-selection yields biased results. Table X repeats the specifications in Table IX when the wife's labor force experience variables are considered exogenous. The Tobit specification is again rejected. We do, however, find considerable evidence of self-selection biases from the generalized Tobit models with experience included in the set of independent variables.

1.7. *Exogeneity Tests with Self-Selection Controls*

Very few of the earlier tests of exogeneity change when we control for self-selection into the labor force. In general, the magnitudes of the differences across various specifications do not change. The standard errors of the estimates do tend to increase, resulting in the failure to reject several of the null hypotheses. Due to the rejection of the Tobit models and the similarity of the estimates for the three different distributional assumptions in the generalized Tobit models, we carry out these tests with the more conventional generalized Tobit model under the normality assumption. The principal findings are discussed below.

TABLE X

SELF-SELECTION MODELS: WITH WIFE'S EXPERIENCE

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *E*, *F*3, *I*. Estimation Method: See Comments.)

Specification	ln (w_u)	Nonwife Income/1000	Young Children	Older Children	σ^2_{u1}	Additional Instruments and Comments
1.	4097	−37.6	−849	59	1.3×10^6	Tobit
	(1094)	(17.9)	(354)	(135)	(8.2×10^4)	
2.	1903	−13.0	−403	−87	7.1×10^5	Conditional Tobit
	(502)	(10.9)	(207)	(78)	(5.5×10^4)	
3.	122	+3.9	+53	−87	—	Multistage estimates
	(225)	(4.5)	(173)	(34)		Normal distribution
4.	161	+3.9	+65	−80	—	Multistage estimates
	(226)	(4.7)	(179)	(34)		Lognormal distribution
5.	157	+3.5	+38	−86	—	Multistage estimates
	(217)	(4.5)	(171)	(34)		Logit distribution
6.	672	−6.4	−283	−85	—	Two-stage Least
	(217)	(3.6)	(147)	(36)		Squares
Differences between Specifications						
1 and 2	2194	24.6	446	−145	-5.3×10^5	Reject spec. 1 (Tobit
	(788)	(12.8)	(237)	(95)	(1.0×10^5)	spec. restrictions)
1 and 3	3976	−41.4	−905	146	—	Reject spec. 1 (Tobit
	(1120)	(18.2)	(375)	(132)		spec. restrictions)
6 and 3	551	10.2	337	2	—	Reject specification 6
	(218)	(3.7)	(128)	(24)		(no self-selection bias)
6 and 4	511	10.4	349	5	—	Reject specification 6
	(209)	(3.9)	(132)	(22)		(no self-selection bias)
6 and 5	516	10.0	321	2	—	Reject specification 6
	(209)	(3.7)	(121)	(22)		(no self-selection bias)

TABLE XI

TESTING EXOGENEITY OF EXPERIENCE WITH CENSORING EXPERIENCE

(Standard Errors in Parentheses. Estimates from Tables IX and X.)

	ln (w_u)	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
	Difference between Specifications in Tables 8 and 9				
1.	3837	14.7	184	155	Tobit
	(1153)	(16.1)	(319)	(116)	Reject exogenous experience
2.	1748	−5.8	156	78	Conditional Tobit
	(579)	(7.1)	(161)	(63)	Reject exogenous experience
3.	57	4.8	237	18	Normal distribution multistage estimates
	(163)	(8.9)	(383)	(24)	Fail to reject exogenous experience
4.	51	4.8	249	24	Log normal distribution multistage
	(171)	(9.9)	(390)	(26)	estimates. Fail to reject exogenous
					experience
5.	91	4.6	222	18	Logit distribution multistage estimates.
	(100)	(8.4)	(360)	(24)	Fail to reject exogenous experience

The only major change occurs in the test for the exogeneity of the wife's labor force experience. From Table XI we see that only in the Tobit and conditional Tobit models do we reject the exogeneity of these variables. For the three generalized Tobit models we do not reject the exogeneity assumption on the wife's labor force experience.[24] The income effects and the impact of young children both become appreciably more positive. These changes, however, are not significant.

This finding conflicts with Heckman's (1980) rejection of the exogeneity of previous labor market experience in the labor supply equation.[25] In order to examine this discrepancy, we attempted to replicate Heckman's estimation and testing procedures on our data set. After controlling for self-selection into the labor force we were still unable to reject the hypothesis of exogenous wife's experience. In several specifications we found that controlling for the possible endogeneity of the wife's labor force experience did increase the estimated wage effect. However, this is not a robust result, and in none of the specifications examined were the estimates precise enough to reject the hypothesis of zero wage effects.

The large impact of the exogeneity assumption on experience in the Tobit models appears to arise from the fact that previous labor market experience is an excellent predictor of whether or not the woman is in the labor force during 1975. In these Tobit models, the impact of experience on participation can only take place through the wage effect in the labor supply equation. As a test of this restriction, we included the experience variables as separate regressors in the "structual" labor supply equation as well as using them in the reduced form wage equation. In this reformulation of the statistical model, the estimated wage coefficient fell from 4097 (1094) to 67 (300) and the income coefficient rose from −37.6 (17.9) to −9.2 (4.9) (standard errors in parentheses). These dramatic changes strongly support the conclusion that labor market experience "explains" participation and hours of work over and above its impact on a woman's wage rate.[26]

[24] It appears that the conditional mean constructed using the wife's experience measure is controlling for the invalid "overidentifying restriction" noted in Section 1.1. See footnote 15 for a further discussion of the interaction of "endogenous" variables, the sample selection controls, and the disturbance in the labor supply equation.

[25] Heckman's (1980) test for endogeneity of experience controls for sample selection induced biases, and he rejects the hypothesis of exogenous labor force experience in a model without the Tobit restrictions.

[26] Another possible source of the discrepancy between the Tobit models and the generalized Tobit models arises from the fact that the Tobit models estimate jointly the wage and hours of work coefficients. In order to examine this possibility we modified the generalized Tobit estimation procedure to estimate simultaneously the hours and wage coefficient. First we estimated the reduced form labor force participation function and constructed the conditional expected values. Second, we used a nonlinear least squares procedure to estimate jointly the parameters in equations (5) and (6). We found (i) that the estimated wage and income effects are not sensitive to whether we used the instrumental variables procedures or the joint estimation procedures and (ii) that the models with exogenous experience (and conditional mean adjustments based upon exogenous experience) are not sensitive to either the choice of estimation procedure or to the inclusion of the experience variables in the "structural" labor supply equation. These results suggest that the interaction of exogenous experience and the Tobit models reflects the power of previous labor market experience to predict current labor force status.

Table XII contains the estimates of the labor supply parameters under various exogeneity assumption on the nonwife income, the children variables, and the wife's labor force experience.[27] As before, we do not reject the exogeneity of either the nonwife income or the children variables. There does appear to be some interaction between the wife's labor force experience variables and the

TABLE XII

EXOGENEITY OF NONWIFE INCOME, CHILDREN, AND LABOR FORCE EXPERIENCE

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *F*3, *H*3. Estimation Method: Multistage Conditional Mean Adjustments.)

Specification	$\ln(w_w)$	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
1.	−67	−0.4	−184	−109	*C*, *I*
	(192)	(6.2)	(278)	(32)	
2.	−70	−2.4	−209	−2	*I*
	(209)	(7.6)	(413)	(77)	
3.	−87	+0.8	−158	−111	*C*
	(195)	(10.2)	(270)	(33)	
4.	−120	+1.6	−135	−10	
	(225)	(11.5)	(428)	(97)	
5.	3	+3.7	+33	−89	*E*, *C*, *I*
	(200)	(4.4)	(173)	(34)	
6.	48	+3.2	−268	−81	*E*, *I*
	(203)	(4.6)	(418)	(83)	
7.	−14	0.0	+23	−88	*E*, *C*
	(201)	(10.2)	(179)	(36)	
8.	34	−1.2	−301	−80	*E*
	(203)	(10.4)	(421)	(83)	
Differences between specifications without wife's labor force experience					
1 and 2	2	2.0	26	−107	Fail to reject specification 1
	(136)	(6.5)	(405)	(98)	(exogeneity of children)
1 and 3	20	−1.2	−26	2	Fail to reject specification 1
	(119)	(10.0)	(209)	(15)	(exogeneity of nonwife income)
1 and 4	52	−2.0	−48	−100	Fail to reject specification 1
	(156)	(11.3)	(426)	(93)	(joint exogeneity of children and nonwife income)
Differences between specifications with wife's labor force experience					
5 and 6	−45	+0.6	+301	−9	Fail to reject specification 5
	(170)	(3.5)	(390)	(79)	(exogeneity of children)
5 and 7	17	3.6	+9	−1	Fail to reject specification 5
	(160)	(9.9)	(117)	(22)	(exogeneity of nonwife income)
5 and 8	−31	4.9	+334	−9	Fail to reject specification 5
	(168)	(10.1)	(393)	(78)	(joint exogeneity of children and nonwife income)

[27] As in the earlier analysis, we reject the hypotheses of no measurement error in the average hourly earnings and the exogeneity of the 1976 wage rate. These tests are complicated by the fact that two sample selection rules determine the subsample of women needed to perform these tests. The woman had to work sometime during 1975 and had to be working at the time of the 1976 interview. Estimation of a bivariate selection model yielded estimates with unacceptable statistical properties of the estimates. Consequently, we used the univariate index function framework to model the joint event.

number of young children in the household: when experience is treated as exogenous, relaxing the exogeneity assumption on children makes the estimated impact of young children more negative. The sample used here, however, is not large enough to uncover any statistically significant differences in the estimates.

2. CONTROLLING FOR VARYING MARGINAL TAX RATES

In this section we examine how the estimated labor supply parameters change when we take into account the fact that the relevant economic wage measure is the after-tax marginal wage rate. To explore this issue, we use Hall's (1973) linearization of the budget set. The marginal wage rate replaces the wage measure, and the virtual income, defined as the intercept of the linearized budget set at zero hours of work, replaces the nonwife income.[28] With varying marginal tax rates, the marginal wage rates and virtual income explicitly depend upon the chosen hours of work. Hence, unlike Hall, we do not treat the marginal tax rate and virtual income as exogenous.

The estimation procedures used here are similar to those used in the first two sections. In the models without controls for self-selection into the labor force, these procedures allow one to estimate the relevant economic parameters. When modeling the joint labor force participation decision and the hours of work decision, however, it is difficult to justify the simple conditional mean adjustments. In this case the relevant economic decision depends upon the parameters of the tax system, and the statistical model must be able to evaluate the woman's preferences for work at all points along the budget set rather than just about zero hours of work. Such models are difficult to estimate and require exceptionally strong assumptions on the tax structure and the stochastic disturbances in the model. Hausman (1981) provides one model of this type to control for the effect of taxes on labor supply. Given these potential problems, the conditional mean adjustments presented here should be considered a first order approximation.

In Table XIII we present estimates of the single worker model with and without adjustment for taxes. Controlling for taxes appears to affect only the estimated wage coefficient. When taxes are taken into account, the estimated log-wage coefficient falls in each instance. The magnitude of this change is at most 33 hours, which translates to a change in the uncompensated wage effect of less than 10 hours per dollar increase in the before tax wage rate. All of the estimated wage coefficients with controls for taxes lie within one-fifth of one standard deviation of the estimates without taxes. In the light of the other possible sources of bias examined in the previous sections, the influence of taxes on the estimates of the labor supply parameters appears to be at most a second order effect.

[28] This study only controls for federal income and social security taxes. To compute the taxes paid and the marginal tax rates, we use the standard deductions from the 1975 tax tables. In the single worker model, the virtual income is defined under the assumption that the husband does not vary his hours of work in response to a change in the wife's hours of work. This would only be valid if the husband's hours of work were perfectly inelastic to changes in wages and income.

TABLE XIII

TAXES IN THE SINGLE WORKER MODEL

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *F*3, *H*3, *I*, *NL*, *HW*. Estimation Method: Two-Stage Least Squares.)

Specification	ln (w_w)	Nonwife Income/1000	Young Children	Older Children	Additional Instruments and Comments
1.	−21	−4.2	−342	−116	No taxes
	(175)	(3.2)	(130)	(30)	
2.	−37	−5.8	−342	−114	Taxes
	(176)	(4.4)	(130)	(29)	
3.	−78	0.0	−171	−110	No taxes; normal conditional
	(195)	(6.1)	(260)	(32)	mean adjustment
4.	−98	−1.0	−164	−109	Taxes; normal conditional
	(195)	(7.2)	(260)	(31)	mean adjustment
5.	−32	+4.3	+41	−91	*E*, No taxes; normal conditional
	(201)	(4.5)	(173)	(34)	mean adjustment
6.	−65	+5.1	+47	−92	*E*, Taxes; normal conditional
	(202)	(5.2)	(174)	(34)	mean adjustment

Rosen (1976) carries out a test of whether or not individuals take taxes into account in their labor supply decisions. One can rewrite equation (1) as

$$(7)\qquad h_i = a_0 + a_1 \ln[W_{f1}(1-\tau_i)] + a_i^* \ln(1-\tau_i)$$
$$+ a_2 Y_{vi} + a_2^* Y_i + a_3' Z_i + e_i,$$

where τ_i is the wife's marginal tax rate and Y_{vi} is the wife's virtual income. The hypothesis that married women optimally take taxes into consideration can be examined by testing

$$H_0:\quad a_1^* = a_2^* = 0;$$

$$H_1:\quad a_1^* \neq 0 \quad \text{or} \quad a_2^* \neq 0.$$

One could also consider an alternative null hypothesis, namely, that the individual ignores taxes when making labor supply decisions. The test in this instance would be

$$H_0^*:\quad a_1^* = -a_1 \quad \text{and} \quad a_2 = 0;$$

$$H_1^*:\quad a_1^* \neq -a_1 \quad \text{or} \quad a_2 \neq 0.$$

Table XIV contains the estimates of equation (7). From these estimates we cannot reject either the hypothesis that the women optimally take taxes into account or the hypothesis that the women fail to take taxes into consideration. The failure to reject in these instances is due to our inability to estimate precisely

TABLE XIV

EXAMINING RESPONSES TO VARYING MARGINAL TAX RATES

(Standard Errors in Parentheses. Common Instrument Sets: *B*, *C*, *F*3, *H*3, *I*, *HW*. Estimation Method: Multistage Normal Conditional Mean Adjustment.)

Specification Coefficient:	$\ln[w_w(1-\tau)]$ a_1	$\ln[1-\tau]$ a_1^*	Virtual Nonwife Income/1000 a_2	Nonwife Income/1000 a_2^*	Additional Instruments and Comments
1.	−119	−1644	+2.8	−17.8	
	(186)	(1779)	(94)	(72)	
2.	−75	−2507	−42.2	+12.7	*E*
	(194)	(1968)	(102)	(76)	

Specification	Tests of hypotheses (2 degrees of freedom, critical, 95% $\chi^2 = 5.99$)	
	H_0: $a_1^* = a_2^* = 0$	H_0^*: $a_1^* = -a_1$ and $a_2 = 0$
	H_1: $a_1^* \neq 0$ or $a_2^* \neq 0$	H_1^*: $a_1^* \neq -a_1$ or $a_2 \neq 0$
	(H_0: optimally take taxes into account)	(H_0: Ignore taxes)
1.	$\chi^2_{2\mathrm{df}} = 1.66$	$\chi^2_{2\mathrm{df}} = 0.91$
2.	$\chi^2_{2\mathrm{df}} = 1.73$	$\chi^2_{2\mathrm{df}} = 1.04$

the coefficients on $\ln(1-\tau_i)$ and the income terms due to the high correlation of our prediction of $\ln(1-\tau_i)$ and the nonwife income variables.

There is, in addition, a third hypothesis that is consistent with the estimates presented in Table XIV. Pechman and Okner (1974) have argued that although the federal income tax code implies a nonproportional tax rate, the actual incidence of taxation appears to reflect a proportional tax system.[29] Under this hypothesis, $a_1 = -a_1^*$ and $a_2 = 0$. The intercept a_0 would include $a_1 \ln(1-\bar{\tau})$, where $\bar{\tau}$ is the constant marginal tax rate and the coefficient a_2^* would be the product of the income effect and $(1-\bar{\tau})$. Consequently, our failure to reject the hypothesis that individuals fail to take taxes into account in their labor supply decisions is empirically indistinguishable from the hypothesis that all individuals face the same proportional tax rate.[30] The power of this hypothesis test, however, is quite low.

The failure to reject the null hypothesis that individuals face a constant marginal tax rate has one important implication in this analysis. As discussed earlier, our controls for self-selection into the labor force are only an approximation to the correct controls in the presence of varying marginal tax rates. Under this null hypothesis, however, our controls are exactly those required to undertake an analysis such as in Hausman (1981). The failure to reject this null hypothesis, then, implies that the procedures we use are not inconsistent with the observed labor supply data.

[29] Specifically, their simulation methods show that for a wide range of income that the average ratio of tax burden to income is approximately a constant. There is, however, a fairly substantial variation within income categories around the mean.

[30] One objection to this conclusion is that Pechman and Okner (1974) do find substantial variations around the mean tax rate. This objection would also apply to studies that use tax tables and standard deductions to define the marginal tax rate, which is how we derived the marginal tax rate used in Tables XIII and XIV.

The result conflicts with Rosen's (1976) finding that taxes have a significant impact on married women's labor supply.[31] A main difference between our analysis and Rosen's lies in the specification of the functional form of the labor supply function; we use a semilogarithmic form while Rosen uses a linear form. The semilogarithmic form implies that the uncompensated wage effect diminishes as wages increase, while the linear specification obviously implies a constant uncompensated effect. The only way a linear specification such as the one used by Rosen could capture a diminishing uncompensated effect would be for the interaction of the marginal tax rate (which depends upon the woman's wage) and the wage to have a nonzero coefficient. The difference between our results and Rosen's could be due to a difference in the functional form of the labor supply equation. A more detailed examination of this issue is certainly an important topic for future research.

3. SUMMARY AND SYNTHESIS

In the previous sections we have shown how the imposition of various economic and statistical assumptions can influence the estimates of the married women's labor supply function. These findings serve two purposes. First, they should be a valuable resource for the formulation of models of married women's labor supply in the future. Second, they allow one to reconcile the wide range of estimated labor supply effects in previous studies. Several of the results in this study, such as the rejection of the Tobit assumptions and the rejection of the exogeneity of the wife's labor force experience, have been previously documented and used to explain some of the across study variation in the estimates of behavioral labor supply responses. The framework used here, however, allows one to examine explicitly the consequences of each economic and statistial specification, and it provides the most comprehensive attempt to reconcile the wide range of estimated wage and income effects found in previous studies.

The most important set of findings in this paper pertains to the treatment of the average hourly earnings measure of the wage rate. Like DaVanzo, DeTray, and Greenberg (1973) and Borjas (1980), we find substantial measurement error in the average hourly earnings, and this measurement error is negatively correlated with the woman's annual hours of work. We examine one alternative measure of the woman's wage rate, namely, her wage rate on her current (time of interview) job, and it appears to be endogenous to the labor supply equation. The use of this wage rate as an instrumental variable to control for measurement error in the average hourly earnings induces a positive bias in the estimated wage effect. This suggests that the average hourly earnings measure combines measurement error and wage rate endogeneity, resulting in ordinary least squares point estimates that are not significantly different from the two stage least squares estimates

[31] Hausman (1981) imposes the budget constraint obtained from the tax tables with standard deductions. His analysis does not contain a test of whether taxes are important or whether the Pechman and Okner (1974) observation is consistent with the data.

which control for these two sources of potential bias. In addition, we explore the use of the woman's previous labor market experience as an instrument for the average hourly earnings. Without controlling for self-selection into the labor force, we find it to be endogenous to the labor supply function, resulting in significant overestimates of the wage effect.

When one treats the woman's labor market experience as endogenous, there is no evidence that the failure to control for self-selection into the labor force results in biased estimates of the labor supply parameters. As in Cogan's (1980b, 1981) analyses of fixed costs and labor supply, the imposition of the Tobit constraints to control for these possible self-selection biases leads to significant overestimates of the magnitudes of the wage effect, the income effect, and the impact of young children. In other words, the hours of work decisions made when the woman is in the labor force appear quite distinct from her labor force participation decision.

With conditional mean controls for self-selection into the labor force, we do not reject the exogeneity assumption on the woman's labor market experience. This result conflicts with Heckman's (1980) rejection of the exogeneity of previous labor market experience. When we impose this exogeneity assumption, however, we do find substantial evidence of self-selection biases. The inclusion of the woman's labor market experience in the set of instrumental variables does little to reduce the standard errors of the estimated wage and income coefficients, and it has a large, although not statistically significant, impact on the estimate of the effect of young children. This measure appears to be an instrument which does little to increase the accuracy of the estimates while complicating the required statistical model.

The children variables and the household's nonwife income do not appear to be endogenous to the woman's labor supply function. The corrections for varying marginal tax rates trivially reduce the estimates of the wage effect and have a small and inconsistent impact on the estimates of the income effects. These tax-corrected estimates lie within one-fifth of one standard error of the estimates that do not take taxes into account, and this suggests that taxation is of second-order importance in explaining the across study variation of the estimates of the married women's labor supply function.[32]

On the basis of these results, we can exclude as invalid several of the sets of estimates found in Tables VI–XIII. What is surprising, however, is the narrow range of estimates of the income and substitution effects found in those specifications that we fail to reject.[33] The maximum point estimate of the log wage coefficient in the 27 unique sets of estimates which pass the specification tests is 161. This translates to a 40 hour per year uncompensated effect when

[32] In Mroz (1984), the estimates from the household model imply the same behavioral responses as those from the single worker model presented here and yield the same results in the tests for exogeneity and self-selection biases.

[33] The set of specifications we fail to reject is Table VI: Specification 2 and 6; Table VII: Specification 1 and 2; Table VIII: Specification 1–4; Table IX: Specifications 3–6; Table X: Specifications 3–5; Table XII: Specifications 1–8; Table XIII: Specification 1–6.

evaluated at the sample's mean wage. The maximum upper bound of the 27 ninety-five per cent confidence intervals is an uncompensated wage effect of 150 hours per year per dollar increase in the wage rate. For the income effects, the largest point estimate implies a 6 hour reduction in annual hours for a $1000 increase in the family's nonwife income. The maximum bound of the 27 ninety-five per cent confidence intervals for this income effect is a 22 hour per year reduction in the wife's labor supply. These estimates are small and precise. They also suggest that the range of estimates from previous studies presented in Table I is quite misleading.

The many sets of estimates and specification tests presented in this study are not independent. One should view the 27 sets of estimates referred to in the preceding paragraph more as a single experiment which exhibits only minor sensitivities to local variations of assumptions than as corroborating experiments. In order to augment this evidence, a possible cross-validation of these results can be obtained by asking how well the empirically important theoretical and methodological criteria uncovered in this study explain the range of estimates found in Table I. On the one hand, an extrapolation from one study to another or one data set to another is, of course, fraught with unobservable sources of error and should be viewed with some caution. On the other hand, the simple and consistent explanations we offer for why these studies disagree with our estimates suggest that such an extrapolation is a valuable tool for reconciling the wide range of estimates.

This study indicates that specifications using the wife's labor market experience as an instrumental variable without controls for self-selection into the labor force and specifications using the Tobit model to control for self-selection biases yield large and biased estimates of the wage and income effects. As predicted, the studies using Tobit models or the wife's labor market experience to instrument the wage rate without controls for self-selection into the labor force (Cogan (1980a)-Tobit; Cogan (1980b)-Tobit; Heckman (1976)-Tobit; Layard, Barton, and Zabalza (1980)-Tobit; and Schultz (1980)-Tobit[34]) yield large estimates for both the income effect and the uncompensated wage effect.

Cogan's fixed cost models (1980b, 1981) allow for nonconvexities in the labor force participation function by relaxing the Tobit model restrictions. His estimation procedure controls for sample selection biases and he treats a woman's labor market experience as exogenous. According to the results presented above, such procedures should not yield biased results. However, even though in one of these studies he uses a data set derived from the same source as this study (PSID, 1975 labor supply data), his estimated wage and income effects are somewhat larger

[34] Schultz does not use the wife's actual labor market experience in the reduced form wage equation. The comparison of the Tobit estimates with and without the wife's experience variable found in Table XI suggests that this is why his estimates of the wage effect are much smaller than the other Tobit estimates in Table I. Layard, Barton, and Zabalza do use the wife's labor market experience to predict wage rates. An examination of their reduced form wage equation, however, reveals that the experience variables have almost no impact on the wage rate. This most likely explains why their wage effects are much smaller than those from the other studies using the wife's labor force experience to predict the wage rate.

in magnitude than those found in our analysis. Here we explore several possible reasons for Cogan's results.

First, Cogan uses many fewer overidentifying restrictions to estimate his model than are used in the results reported here. When we impose the assumption of an exactly identified model (leaving the structural equation unchanged and exactly identifying the wage effect through the log-experience measure used by Cogan), we are able to obtain larger estimates of the wage effect quite close to Cogan's point estimates. The multiple stage estimation procedures we use, however, yield large standard errors which include our range of point estimates in a ninety-five per cent confidence interval. When we relax the exactly identifying restriction by using the log-experience variable and its square as identifying variables for the wage effects, the point estimate of the log-wage coefficient falls from 1100 to −120. Thus Cogan's model appears to be quite sensitive to his exact specification. These results, in conjunction with Heckman's (1980) rejection of the exogeneity of experience, do bring to question the power of our test for the exogeneity of the wife's previous labor market experience with the controls for sample selection biases.

A second possible source of discrepancy stems from Cogan's model of the sample selection rule. Cogan imposes proportionality restrictions between the coefficients on the identifying variables in the wage equation and the coefficients on those variables in the labor force participation function. This approach partially constrains the impact of experience on labor force participation to operate through the wage effect. Although Cogan's assumptions are not as restrictive as those in the Tobit model, the extreme sensitivity of the Tobit estimates to the identifying restriction on the experience variables (discussed in Section 1.7) suggests a possible source of specification error. The generalized Tobit models used in this study do not impose these restrictions, and an investigation of these assumptions would be a useful extension of this study.

Only two studies in Table I treat the woman's wage rate as exogenous, Leuthold (1978) and Hausman (1981). Leuthold's wage measure is the woman's wage rate on her current or previous job; this is not an average hourly earnings measure. These variables are not available in the PSID, making any simple comparison difficult. In addition, her controls for taxes do not correspond to those used in this study,[35] and she treats both the marginal tax rate and disposable income as exogenous. Her estimated wage effect does not fall outside the range of estimates we fail to reject, but given her treatment of taxes, it is not obvious that she measures the same labor supply effects as this study.[36]

Because of methodological differences, Hausman's (1981) estimates are the least comparable to those in this analysis. Although his hours of work function is linear in wages and income, the complex switching regression framework required to take account of the discontinuous and possibly nonconvex after-tax

[35] Leuthold uses the marginal after-tax wage rate and the after-tax disposable income rather than a measure of the virtual income from the linearized budget set.

[36] Another possible explanation for why she obtains such small wage coefficients is classical measurement error in her wage measure.

budget constraint makes any comparisons based on the previous sections merely speculative. Additionally, he does not report the instrumental variables used to predict the wages for nonworkers, making it impossible to evaluate his exogeneity assumptions. We do note, however, that the assumptions in his convex and nonconvex models closely correspond to Cogan's assumptions, and this may be one reason why he obtains such large estimates of the income and wage effects. A more detailed examination of Hausman's models is required before one can conclude that it is only his more exacting controls for taxes that generate the large behavioral responses found in his analyses.

The only estimates not yet discussed are those of Boskin (1973), Heckman (1976)-generalized Tobit, Heckman (1980), Nakamura and Nakamura (1981), and Schultz (1980)-instrumental variables. Of these five sets of estimates, only the two by Heckman do not correspond to the estimates reported here, even though he controlled for self-selection bias in both sets of estimates and also controlled for endogenous experience in his 1980 estimates. Heckman's (1976) generalized Tobit procedure yields a large negative uncompensated wage effect and a large positive income effect. When we used procedures similar to Heckman's with our data set, we were unable to find such point estimates. Our estimates, however, had large standard errors, and we could not reject the hypothesis that our point estimates were equal to the point estimates reported by Heckman. Unfortunately these estimates by Heckman were only initial consistent estimates for a Tobit procedure, and he did not report enough information to calculate the standard errors of these estimates. Our results suggest that Heckman's results may be estimated imprecisely, but a more detailed examination with his data set will be needed in order to make any more definitive inferences.

Heckman's (1980) estimates imply a large positive wage response that is significantly different from those found in the specifications that we fail to reject on methodological grounds. He treats experience as endogenous and uses a generalized Tobit procedure to control for self-selection into the labor force. According to the methodological results reported here, we would not expect such procedures to yield large wage effects. When we used his procedures with our data set, we were unable to replicate his point estimates.[37] We did find certain choices of instrumental variables that yielded larger point estimates than those reported in the previous tables, but none of these were precisely estimated.

The procedure Heckman uses to obtain his estimates that control for both self-selection bias and endogeneity of wages, however, does not yield consistent estimates when one rejects the joint hypothesis of exogenous experience and no self-selection bias.[38] In order to demonstrate this, it is necessary to examine his procedures in detail. First, he treats experience as exogenous and uses a probit

[37] Unlike our results, Heckman fails to reject exogenous experience without controls for self-selection bias, and he also fails to reject the hypothesis that self-selection is not important when he treats experience as exogenous. Only when he tests the joint hypothesis does he reject the individual hypotheses. This suggests that there may be some fundamental difference between the two data sets.

[38] I would like to thank Tom MaCurdy, Jim Heckman, Joe Hotz, Ricardo Barros, and Colin Cameron for discussing this point with me.

procedure to estimate the labor force participation function. From these estimates that treat experience as exogenous he constructs the conditional expected value of the error term and includes this term as an explanatory variable in the labor supply equation. Second, he estimates the labor supply equation by instrumental variables, treating both the wife's experience and the conditional mean as endogenous variables. A comparison of the instrumental variables estimates with the ordinary least squares estimates does provide a test of the joint hypothesis of exogenous experience and no sample selection bias. After rejecting this hypothesis, the conditional mean is misspecified, and the instrumental variables procedure does not yield consistent estimates of the labor supply equation. Consequently, the results reported by Heckman are not consistent estimates of the behavioral relationship. His rejection of the joint hypothesis is valid, and this result suggests that there may be some fundamental differences between the two data sets.

The remaining three studies—Boskin (1973), Nakamura and Nakamura (1981), and Schultz (1980)-Instrumental Variables—treat the wife's wage as endogenous and do not include measures of the wife's labor market experience in the set of instrumental variables.[39] Their estimates of the labor supply parameters are close to those in our preferred specifications. Except for Layard, Barton, and Zabalza's instrumental variables estimates, these are the only estimates found in Table I in agreement with our preferred specifications.[40] The important methodological considerations uncovered by our statistical specification tests appear to single out studies with similar estimates and exclude only estimates differing from our preferred set. These inferences strongly support our methodological criteria, and suggest that the small income and wage effects found in this study provide a much more accurate picture of the behavioral responses of working women to variations in nonlabor income and wages than those found in most previous studies.[41]

The negative income tax experiments provide additional data sources for cross-validating the income and substitution effects found in this study. It has been argued that these experiments do permit one to examine more "exogenous" variations in wages and incomes than can be found in nonexperimental data. Indeed, the range of estimates reported by Moffitt and Kehrer (1981) for the estimates of the wage and income effects for married women from the experimental literature suggests smaller responses than one would infer from a casual examination of the estimates in Table I. Given the conceptual difficulties of

[39] Only Nakamura and Nakamura control for self-selection into the labor force.

[40] See the above note on Layard, Barton, and Zabalza's reduced form wage equation for a possible explanation for why their instrumental variables estimates with the wife's labor market experience variables and without controls for self-selection yield small estimates of the wage effect. Leuthold's estimates are also similar to our preferred estimates, but she does not control for possible measurement error in the wage rate.

[41] Our results do not imply that the labor force participation decision is insensitive to variations in wage rates and nonlabor income. In fact, the large estimates of the income and wage effects from the Tobit models suggests that the labor force participation decision may be quite responsive to changes in these economic factors.

analyzing the nonrandom treatment and control assignments in the experimental data and the small number of studies directly estimating behavioral responses, we are hesitant to use the experimental results as a strong cross-validation of our estimates. However, we do believe that a thorough examination of the experimental data could provide a strong test of the robustness of the results found in this study.

CONCLUSION

In this study we find that economic and statistical assumptions can have a substantial impact on the estimates of the behavioral labor supply parameters. The three most important assumptions are (i) the Tobit assumption to control for self-selection into the labor force, (ii) the exogeneity assumption on the wife's wage rate, and (iii) the use of the wife's labor market experience as an instrumental variable to control for the endogeneity of the wife's wage. The Tobit models exaggerate both the income and wage effects. The two exogeneity assumptions induce an upward bias in the estimated wage effect; the bias due to the exogeneity of the wife's labor market experience, however, greatly diminishes when one controls for self-selection into the labor force through the use of generalized Tobit procedures. After controlling for these specification problems, the estimates of the income and substitution effects are invariant to a wide range of assumptions. Among the potential specifications found to be unimportant are (i) exogeneity assumptions on the nonwife income and the number of children in the household, (ii) controls for nonproportional income taxes, and (iii) controls for self-selection into the labor force when experience is treated as endogenous.

The range of labor supply estimates that we fail to reject suggests that the labor supply behavior of working married women matches the estimated behavior of prime aged males. Such a conclusion conflicts with most commonly held beliefs about female labor supply. Killingsworth (1983, p. 432) recently summarized the current evidence for such beliefs, "... most of the available evidence suggests that female labor supply, measured as either labor force participation or hours of work, is considerably more wage and property income elastic than male labor supply." In this study, we are able to obtain large estimates of the income and wage coefficients. Our statistical tests, however, emphatically reject the economic and statistical assumptions needed to obtain these large wage and income effects.

The principal finding of this analysis is that economic factors such as wage rates, taxes, and nonlabor incomes have a small impact on the labor supply behavior of *working* married women. The array of estimates we have reported and our examination of the methodologies and estimates from previous studies clearly support this conclusion. Of course, this study and those reviewed here are all based upon a simple economic model of the behavior of married women. In order to interpret the estimated coefficients as behavioral labor supply responses, a woman must face a continuum of job offers at a constant wage rate and costlessly select the offer yielding the highest utility. The types of models considered here ignore many factors such as search costs, imperfectly elastic labor demand schedules, labor force participation and dynamic behavior, and nonpecuniary benefits. Understanding the influence of these factors may be crucial for a more complete understanding of married women's labor supply.

Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637

Manuscript received April, 1985; final revision received August, 1986.

APPENDIX 1

Construction of the Sample

The data used in this analysis come from the twelve year Michigan Panel Study of Income Dynamics tape (Wave XII tape). The following sequential criteria were used to select this sample.

Selection Criteria	Number of Observations Deleted
Total number of records: 6373	
1. Nonrandom low income sample	2876
2. Not living in the U.S. in both 1975 and 1976	13
3. Change in head or wife in household from 1975 to 1976	331
4. Duplicate records due to later splitoffs	379
5. Female head of household in 1975	590
6. Not classified as married in both 1975 and 1976	212
7. Head retired, permanently disabled or did not work in 1975	362
8. No wife interview in 1976	15
9. Incomplete information on birthdates of children	5
10. Nonwhite head of household	146
11. Wife less than 30 years old in 1975	613
12. Wife older than 60 in 1975	31
13. Husband less than 30 years old in 1975	9
14. Husband older than 60 in 1975	28
15. Husband works more than 5200 hrs. in 1975	3
16. Husband's education not reported	2
17. Family income in upper bracket in 1975 (Fam. income > $99,999, not known exactly)	3
18. Unknown taxable income in 1975	1
19. Husband reported no labor earnings in 1975	1

Remaining sample: 753 observations.

APPENDIX 2

Covariance Matrices for Correlated Estimators

In this appendix we report the formulae used to construct the covariance matrices for sets of correlated estimators. Formal derivations of these estimators of the covariance matrices can be found in MaCurdy and Mroz (1984). First, we present the formulae for two instrumental variable estimators. Second, we report the formulae for the estimators when the instrumental variable estimators depend upon pre-estimated quantities, such as conditional expected values derived from first stage estimates of the labor force participation function. Third, we report the formulae when one of the estimators is a multi-stage estimator (as in case 2) and the other estimator is a maximum likelihood or quasi-maximum likelihood estimator.

Case 1: *Two Instrumental Variables Estimators*: Under the assumption that the observations are independent, an exact first order Taylor series expansion of the derivatives of the instrumental variables' objective functions,

$$F_1 = .5(Y - Xb_1)'Z_1(Z_1'Z_1)^{-1}Z_1'(Y - Xb_1),$$
$$F_2 = .5(Y - Xb_2)'Z_2(Z_2'Z_2)^{-1}Z_2'(Y - Xb_2),$$

yields the following asymptotic formula for the variance-covariance matrix of the estimates $(\hat{b}_1, \hat{b}_2)$:

$$\operatorname{Var}[(\hat{b}_1', \hat{b}_2')'] = A^{-1}BCB'A^{-1}$$

where

$$A = \begin{pmatrix} X'Z_1(Z_1'Z_1)^{-1}Z_1'X & 0 \\ 0 & X'Z_2(Z_2'Z_2)^{-1}Z_2'X \end{pmatrix},$$

$$B = \begin{pmatrix} X'Z_1(Z_1'Z_1)^{-1} & 0 \\ 0 & X'Z_2(Z_2'Z_2)^{-1} \end{pmatrix},$$

$$C = \sum_{i=1}^{N^*} \begin{pmatrix} Z_{1i}Z_{1i}'\hat{u}_{1i}^2 & Z_{1i}Z_{2i}'\hat{u}_{1i}\hat{u}_{2i} \\ Z_{2i}Z_{1i}'\hat{u}_{1i}\hat{u}_{2i} & Z_{2i}Z_{2i}'\hat{u}_{2i}^2 \end{pmatrix},$$

and X is a N^* by k matrix of the explanatory variables, Z_1 is a N^* by k_1 matrix of the instrumental variables (set Z_1), Z_2 is a N^* by k_2 matrix of the instrumental variables (set Z_2), Z_{ji} is a vector of the ith observation's instrumental variables in the jth set, and $\hat{u}_{ji}$ is the ith disturbance in the jth equation. N^* indexes the observations with observed values of the dependent variable Y_t.

CASE 2: *Instrumental Variables Estimators With Pre-estimated Parameters*: Consider the case where pre-estimated quantities (e.g. conditional expected values) are included as regressors and in the sets of instrumental variables. Suppose that the pre-estimated quantities are constructed using estimates $\hat{g}_1$ and $\hat{g}_2$ and that these estimates are obtained by maximizing the functions $G_1(g_1)$ and $G_2(g_2)$. Assume that the first derivatives of these functions are asymptotically independent across observations. Thus,

$$\frac{\partial G_j}{\partial g_j} = \sum_{i=1}^{N} \frac{\partial G_{ji}}{\partial g_j}, \quad j = 1, 2,$$

where the terms inside the summation are independent. The joint covariance matrix of the estimates used to construct the pre-estimated quantities can be written as

$$V = \operatorname{Var}[(\hat{g}_1', \hat{g}_2')'] = R^{-1}\Lambda R^{-1}$$

where

$$R = \begin{pmatrix} \dfrac{\partial^2 G_1}{\partial g_1\,\partial g_1'} & 0 \\ 0 & \dfrac{\partial^2 G_2}{\partial g_2\,\partial g_2'} \end{pmatrix},$$

$$\Lambda = \sum_{i=1}^{N} \begin{pmatrix} \dfrac{\partial G_{1i}}{\partial g_1}\dfrac{\partial G_{1i}}{\partial g_1'} & \dfrac{\partial G_{1i}}{\partial g_1}\dfrac{\partial G_{2i}}{\partial g_2'} \\ \dfrac{\partial G_{2i}}{\partial g_2}\dfrac{\partial G_{1i}}{\partial g_1'} & \dfrac{\partial G_{2i}}{\partial g_2}\dfrac{\partial G_{2i}}{\partial g_2'} \end{pmatrix},$$

and $\partial G_{ji}/\partial g_j$ is the derivative of the function G_j with respect to the parameters g_j for the ith observation. The covariance matrix for the instrumental variables estimates of b_1 and b_2 is given by

$$\operatorname{Var}[(\hat{b}_1', \hat{b}_2')'] = A^{-1}BCB'A^{-1} + A^{-1}KVK'A^{-1},$$

where

$$K = \begin{pmatrix} \dfrac{\partial^2 F_1}{\partial b_1 \partial g_1'} & 0 \\ 0 & \dfrac{\partial^2 F_2}{\partial b_2 \partial g_2'} \end{pmatrix}$$

and the other matrices are as defined above. All matrices are evaluated at the consistent estimates, and the cross derivative matrix, K, was estimated by numerically differentiating the first derivatives of F_1 and F_2 with respect to the pre-estimates $\hat{g}_1$ and $\hat{g}_2$. Note that the construction of these covariance matrices does not depend upon the assumption of an efficient estimator nor does it depend upon homoscedasticity of the disturbances u.

CASE 3: *Instrumental Variables and Quasi-Maximum Likelihood Estimates*: We now consider the case where the first set of estimates is as in case 2 and the second set of estimates is obtained by maximizing a likelihood or quasi-likelihood function. This second set of estimates does not depend upon pre-estimated parameters, but the extension to this case is straightforward. Suppose that this second set of estimates is obtained by choosing $\hat{b}_2$ to maximize the function $\Phi(b_2)$ and that the derivatives of Φ are independent across observations. As above,

$$\frac{\partial\Phi_j}{\partial b_2}=\sum_{i=1}^{N}\frac{\partial\Phi_{ji}}{\partial b_2}.$$

The covariance matrix of $(\hat{b}_1, \hat{b}_1)$ is given by

$$\operatorname{Var}[(\hat{b}_1', \hat{b}_2')'] = A^{-1}BCB'A^{-1}+A^{-1}KVK'A^{-1},$$

where

$$A=\begin{pmatrix} X'Z_1(Z_1'Z_1)^{-1}Z_1'X & 0 \\ 0 & \dfrac{\partial^2\Phi}{\partial b_2\partial b_2} \end{pmatrix},$$

$$B=\begin{pmatrix} X'Z_1(Z_1'Z_1)^{-1} & 0 \\ 0 & I \end{pmatrix},$$

$$C=\sum_{i=1}^{N^*}\begin{pmatrix} Z_{1i}Z_{1i}'\hat{u}_{1i}^2 & Z_{1i}\dfrac{\partial\Phi_i}{\partial b_2'}\hat{u}_{1i} \\ \dfrac{\partial\Phi_i}{\partial b_2}Z_{1i}'\hat{u}_{1i} & \dfrac{\partial\Phi_i}{\partial b_2}\dfrac{\partial\Phi_i}{\partial b_2'} \end{pmatrix},$$

$$V=\begin{pmatrix} \dfrac{\partial^2 G_1}{\partial g_1\partial g_1'} & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$K=\begin{pmatrix} \dfrac{\partial^2 F_1}{\partial b_1\partial g_1'} & 0 \\ 0 & 0 \end{pmatrix}.$$

REFERENCES

AMEMIYA, T. (1974): "The Non-linear Two-stage Least-Squares Estimator," *Journal of Econometrics*, 2, 105–110.

BASMANN, R. (1960): "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, 650–659.

BORJAS, G. (1980): "The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias," *Journal of Human Resources*, 15, 424–434.

BORJAS, G. AND J. HECKMAN (1978): "Labor Supply Estimates for Public Policy Evaluation," *Proceedings of the Thirty-first Annual Meeting of the Industrial Relations Research Center*, pp. 320-331.

BOSKIN, M. (1973): "The Economics of Labor Supply," in *Income Maintenance and Labor Supply*, ed. by G. Cain and H. Watts. Chicago: Rand McNally, pp. 163–181.

CAIN, G., AND H. WATTS (1973): "Towards a Summary and a Synthesis of the Evidence," in *Income Maintenance and Labor Supply*, ed. by G. Cain and H. Watts. Chicago: Rand McNally, pp. 328-367.

COGAN, J. (1980a): "Married Women's Labor Supply: A Comparison of Alternative Estimation Procedures," in *Female Labor Supply*, ed. by J. Smith. Princeton: Princeton University Press, pp. 90–118.

——— (1980b): "Labor Supply with Costs of Market Entry," in *Female Labor Supply*, ed. by J. Smith. Princeton: Princeton University Press, pp. 327-359.

——— (1981): "Fixed Costs and Labor Supply," *Econometrica*, 49, 945–964.

DAVANZO, J., D. DETRAY, AND D. GREENBERG (1973): *Estimating Labor Supply Response: A Sensitivity Analysis*, The Rand Corporation, R-1372-OEO.

DURBIN, J. (1954): "Errors in Variables," *Review of the International Statistical Institute*, 22, 22–32.

GARFINKEL, I. (1973): "On Estimating the Labor-Supply Effects of a Negative Income Tax," in *Income Maintenance and Labor Supply*, ed. by G. Cain and H. Watts. Chicago: Rand McNally, pp. 205-264.

GREENBERG, D., AND M. KOSTERS (1973): "Income Guarantees and the Working Poor: The Effect of Income-Maintenance Programs on the Hours of Work of Male Family Heads," in *Income Maintenance and Labor Supply*, ed. by G. Cain and H. Watts. Chicago: Rand McNally, pp. 14-101.

GREENHALGH, C. (1980): "Participation and Hours of Work for Married Women in Great Britain," *Oxford Economic Papers*, 32, 296-318.

HALL, R. (1973): "Wages, Income, and Hours of Work in the U.S.," in *Income Maintenance and Labor Supply*, ed. by G. Cain and H. Watts. Chicago: Rand McNally, pp. 102-162.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1272.

——— (1981): "Labor Supply," in *How Taxes Affect Economic Behavior*, ed. by H. Aaron and J. Pechman. Washington: Brookings Institution.

HECKMAN, J. (1974): "Shadow Prices, Market Wages and Labor Supply," *Econometrica*, 42, 679-694.

——— (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection Bias and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475-492.

——— (1980): "Sample Selection Bias as a Specification Error," in *Female Labor Supply*, ed. by J. Smith. Princeton: Princeton University Press, pp. 206-248.

HOTZ, V. J. (1980): "A Life Cycle Model of Fertility and Married Women's Labor Supply," Working Paper No. 17-80-81, Graduate School of Industrial Administration, Carnegie-Mellon University.

KILLINGSWORTH, M. (1983): *Labor Supply*. Cambridge: Cambridge University Press.

LAYARD, R., M. BARTON, AND A. ZABALZA (1980): "Married Women's Participation and Hours," *Economica*, 47, 51-72.

LEE, L.-F. (1982): "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 59, 355-372.

LEUTHOLD, J. (1978): "The Effect of Taxation on the Hours Worked by Married Women," *Industrial and Labor Relations Review*, 31, 520-526.

MACURDY, T., AND T. MROZ (1984): "An Intertemporal Model of Household Hours of Work Behavior and Labor Force Participation," in progress.

MASTERS, S., AND I. GARFINKEL (1977): *Estimating the Labor Supply Effects of Income-Maintenance Alternatives*. New York: Academic Press.

MOFFITT, R., AND K. KEHRER (1981): "The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments," in *Research in Labor Economics*, Vol. 4, ed. by R. Ehrenberg. Greenwich: JAI Press, Inc.

MROZ, T. (1984): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," unpublished Ph.D. Dissertation, Stanford University.

——— (1986): "A Note On Exogeneity Tests In Censored Regression Models," University of Chicago, in progress, 1986.

NAKAMURA, A., AND M. NAKAMURA (1981): "A Comparison of the Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes," *Econometrica*, 49, 451-489.

PECHMAN, J. A., AND B. A. OKNER (1974): *Who Bears the Tax Burden*? Washington, D.C.: The Brookings Institution.

ROSEN, H. (1976): "Taxes in a Labor Supply Model with Joint Wage-Hours Determination," *Econometrica*, 44, 485-507.

SARGAN, J. (1959): "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society*, Series B, 21, 91-105.

SCHULTZ, T. P. (1980): "Estimating Labor Supply Functions for Married Women," in *Female Labor Supply*, ed. by J. Smith. Princeton: Princeton University Press, pp. 25-89.

SMITH, J. (1980): "Assets and Labor Supply," in *Female Labor Supply*, ed. by J. Smith. Princeton: Princeton University Press, pp. 166-205.

STERN, N. (1986): "On the Specification of Labor Supply Functions," in *Unemployment, Search and Labour Supply*, ed. by R. Blundell and I. Walker. Cambridge: Cambridge University Press, pp. 143-189.

WHITE, H. (1982a): "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483-499.

——— (1982b): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

WU, D. (1973): "Alternative Tests of Independence Between Stochastic Regressors and Disturbance," *Econometrica*, 41, 733-750.