



What Has Been Learned About Labor Supply in the Past Twenty Years?

James J. Heckman

The American Economic Review, Volume 83, Issue 2, Papers and Proceedings of the Hundred and Fifth Annual Meeting of the American Economic Association (May, 1993), 116-121.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at <http://www.jstor.org/about/terms.html>, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The American Economic Review is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aea.html>.

The American Economic Review
©1993 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2001 JSTOR

What Has Been Learned About Labor Supply in the Past Twenty Years?

By JAMES J. HECKMAN*

The main advance in the study of labor supply in the past 20 years has come in recognizing and interpreting the variety of different labor supply functions that coexist in the empirical literature. A crucial theoretical distinction with important empirical payoff is that between labor supply choices at the extensive margin (i.e., labor-force participation and employment choices) and choices at the intensive margin (i.e., choices about hours of work or weeks of work for workers). Another important distinction is the one between a descriptive labor supply function and a structural supply function that can be the basis for out-of-sample policy investigations concerning responses to tax and welfare programs and evaluations of welfare losses. A full understanding of the participation-hours dichotomy leads to an appreciation of the importance of the problem of selection bias. What is the wage to impute to a nonworking person? These distinctions are a legacy of the research conducted on labor supply over the past 20 years.

Many important empirical questions have been addressed. (i) How strong are the effects of taxes, welfare, and transfer programs on labor supply? (ii) How important is intertemporal substitution? Can life-cycle labor supply models explain business-cycle labor supply behavior? (iii) Has the consensus view of the 1960's of high labor supply elasticities for married women and low labor supply elasticities for married men held up? All of the empirical evidence assembled on labor supply must be called into question in the light of important research on the basic quality of the data conducted by ana-

lysts at the University of Michigan. I discuss this evidence in a concluding section. Except for my discussion of measurement issues I focus only on the major discoveries about which a general consensus has emerged.

I. A Framework for Thinking about Labor Supply and Missing Wages

Let H be a measure of labor supply and let W be the appropriate wage. Denote by Y the amount of resources transferred to economic agents independent of their work effort and let ϵ denote all other variables that affect labor supply. Some components of ϵ are likely to be unobserved by the analyst but may still affect labor supply choices. For expositional simplicity, I assume that ϵ is scalar and unobserved. In a life-cycle setting, W may be a vector of expected current and future wages. Tax and transfer programs may create complicated functional relationships involving W , Y , and current and past labor supply behavior. In this case it is useful to distinguish marginal wages from average wages. The former are relevant to marginal substitution choices; the latter are relevant to income effects.

An important advance in thinking about labor supply came in distinguishing the following equations. Let E denote expectation. Four different labor supply functions confused in the 1960's literature were:

- (1) $E(H|W, Y, \epsilon)$
- (2) $E(H|W, Y, H > 0)$
- (3) $E(H|W, Y) = E(H|W, Y, H > 0) \times \Pr(H > 0|W, Y)$
- (4) $\Pr(H > 0|W, Y).$

For H defined as hours of work, (1) is the "structural" labor supply equation, holding

*Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637. This research was supported in part by NSF-91-11455 and a grant from the Lynde and Harry Bradley Foundation. I thank V. Joseph Hotz for comments.

tastes ε constant. Derivatives of (1) are the standard neoclassical income and substitution effects of labor supply, providing that there is sufficient freedom to select weeks or hours by marginal calculations. These Slutsky effects are required for counterfactual policy analysis and estimation of the welfare costs of taxes and transfers. Equation (2) characterizes the labor supply of workers. It describes an empirical relationship. It does not control for tastes. Its derivatives combine the derivatives of equation (1) with the parameters of the distribution of tastes. See Heckman (1978) or John Pencavel (1986) for exact expressions. Equation (3) is an aggregate labor supply curve conditional on W and Y . It traces out movements along labor supply curves for workers (the first term) and movements into or out of employment or the labor force (the second term). Equation (4) is the equation characterizing employment (or participation).

Prior to the late 1960's, most analysts of labor supply failed to distinguish between participation and hours of work. For example, Jacob Mincer's (1962) classic study of labor supply defined H as the fraction of available lifetime hours worked. Letting $d = 1$ if a person works in a subperiod, (and zero otherwise) Mincer equated (1) and

$$(5) \quad \Pr(d = 1 | W, Y, \varepsilon)$$

and claimed to estimate income and substitution effects from data on labor-force participation. Strong assumptions are required to justify this procedure (see Heckman, 1978; Pencavel, 1986).

Motivated by the searching unpublished criticism of H. Gregg Lewis (1967), labor economists began to distinguish between choices at the extensive margin and choices at the intensive margin. This distinction is fundamental. Once made, it elevates a problem previously regarded as a nuisance—what to do about missing wages for nonworkers—to a central research problem in economics: the consequences and causes of self-selection bias in estimating wage and labor supply functions on samples of workers.

These problems are relevant to the analysis of all five versions of the labor supply functions listed above. Avoiding the problem of selection bias by fitting labor supply functions only for workers [equation (2)] estimates an equation that is of limited economic interest because of selection bias with respect to the population distribution of labor supply functions. To estimate any of the four remaining types of labor supply equations requires taking a position on the source of missing wages for nonworkers.

These distinctions are empirically important. Participation (or employment) decisions generally manifest greater responsiveness to wage and income variation than do hours-of-work equations for workers. The 1960's characterization of married-female labor supply as much more wage- and income-elastic than male labor supply arose, in part, because participation elasticities for women were being compared with hours-of-work elasticities for men.

George Borjas and Heckman (1978) noted that wage and income estimates for male labor supply based on (3) show more responsiveness to wages and incomes than estimates based on (2). This point is implicit in the careful analysis of Julie Devanzo et al. (1973). At higher wages and for greater hours worked, male labor supply shows little wage and income responsiveness. Virtually all of the wage and income responsiveness found for this group is at or near the zero-hours point. It is the derivative of the participation probability on the right-hand side of (3) that contributes the most to estimated wage and income elasticities. The analysis of Chinhui Juhn et al. (1991) confirms this point in more recent data. It was already evident in labor supply functions fit on data for the 1960's and the 1970's and, hence, is not a new empirical phenomenon. What is new is the greater relevance of participation decisions for describing male labor supply since male labor-force participation rates are now lower, even for prime-age males.

A similar finding appears in the literature on female labor supply. As noted by John Cogan 1981 and Heckman (1980), a substantial portion of the estimated wage and income responsiveness found in estimated

hours-of-work equations for married females comes from tightly constraining hours of work and participation equations using the so-called "Tobit" specification. Incorporating fixed costs of work into the analysis of labor supply breaks the tight constraints across the parameters of labor supply and participation equations and produces selection-corrected female hours-of-work equations that are much more similar to male hours of work equations.

Thomas Mroz's (1987) influential recent study produces estimates of (1) for married women with small wage and income elasticities fairly close to the low elasticities found for males by Thomas MaCurdy et al. (1990). Married female wage and income elasticities for hours-of-work equations are generally still somewhat larger in absolute value than male labor supply elasticities (see the survey by Mark Killingsworth [1983]). Whether labor supply behavior by sex will converge to equality as female labor-force participation continues to increase is an open question.

A revision is in order for George Stigler's dictum that all elasticities are 1 in absolute value. A dictum closer to the truth would be that elasticities are closer to 0 than 1 for hours-of-work equations (or weeks-of-work equations) estimated *for those who are working*. A major lesson of the past 20 years is that the strongest empirical effects of wages and nonlabor income on labor supply are to be found at the extensive margin—at the margin of entry and exit—where the elasticities are definitely not zero.

II. Taxes and Labor Supply

Marvin Kosters's (1967) pioneering study of the effects of taxes on labor supply found very weak tax effects for male hours-of-work equations for those who are working and somewhat stronger, but still small, effects on participation. His estimates are confirmed by MaCurdy et al. (1990). Mroz (1987) finds similar weak tax effects on female hours of work for working women.

In the period between Kosters (1967) and MaCurdy et al. (1990), economists were entertained by the spectacle of anomalously

large estimates of income effects and compensated substitution effects for male hours of work produced from functional-form-dependent estimation schemes designed to estimate the effect of taxes on labor supply (see the survey by Jerry Hausman [1985]). Some of the estimates produced from these schemes were so large that they implied that males consumed fewer goods as their incomes increased! (See Pencavel, 1986.) These schemes have now fallen into disrepute. The econometric procedures used to produce the estimates were econometrically and economically inconsistent in part because they did not properly correct for missing wage data for nonworkers (see Heckman, 1983). Competent analysts have been unable to replicate the earlier findings even using the same data (see MaCurdy et al., 1990). When these models are reestimated using more robust schemes, weak wage and income effects of taxes are found for males in numerous countries. However, most of this literature focuses on choice at the intensive margin. There are few estimates available on the responsiveness of male labor-force participation to taxes.

The economic model underlying the now-suspect tax and labor supply literature ignores the effect of taxes on the price of deductibles (see Heckman, 1983; Robert Triest, 1992). When deductibles are introduced into an empirical analysis of male labor supply, Triest estimates that virtually all of the effect of taxes on male labor supply operates through this channel.

III. The Importance of the Selection Problem

Wages are missing for nonworkers. There is now general agreement about the potential importance of the missing-wage problem. It is especially important in light of the growing evidence that much of the elasticity in estimated labor supply functions comes in entry and exit decisions. There is less agreement about particular solutions to the wage-imputation problem that have been offered in the literature. Each solution invokes different assumptions about unobserved counterfactuals: what wages would have been had nonworking persons worked.

Different assumptions are likely to be appropriate for different problems and data sets.

Selection bias is present in estimating structural labor supply equations [like (1)], wage equations, or both. For it to be absent in both, economic decisions about labor-force participation and employment decisions have to be made on criteria not motivated by economic returns. To show this in the simplest setting, let the reservation wage be W_R . It is the value of W that just sets $H = 0$ in a standard model of labor supply without fixed costs. Persons work if $W > W_R$. There is no selection bias in observed market wages only if wages are independent of $W - W_R$. No selection on wages implies in the general case that there is selection on reservation wages (i.e., that there is selection bias in fitting empirical labor supply functions), since the reservation wage is obtained from the inverse of the labor supply curve evaluated in the neighborhood of $H = 0$. There can be no selection bias in estimated hours of work and wage equations only if the joint distribution of W and W_R is degenerate. For W to be independent of the selection rule ($W - W_R$), $\text{Var}(W) = \text{Cov}(W, W_R)$. For W_R to be independent of the selection rule $\text{Var}(W_R) = \text{Cov}(W, W_R)$. Both conditions cannot hold if the variance-covariance matrix of (W, W_R) is nondegenerate. Except for the degenerate case, one cannot have an economically meaningful model of labor supply in which there is no selection on either wages or reservation wages. The magnitude of the bias is, of course, a separate matter.

IV. Life-Cycle Labor Supply and the Intertemporal-Substitution Hypothesis

An enormous literature ably surveyed by David Card (1991) considers the allocation of working time over the life cycle. Wages are broadened to include current and future values. Empirical research in this area has largely focused attention on estimating annual-hours-of-work equations for those who work sometime each year. The micro studies are primarily based on panel data. The weight of the available empirical evidence

suggests a weak or even nonexistent wage responsiveness of labor supply over the life cycle. There is also evidence that demand-side variables affect labor supply holding measured wages constant. Both facts have been interpreted to mean that there is little scope for intertemporal substitution to explain life-cycle labor supply or labor supply over the business cycle.

Before this conclusion is embraced too strongly, it is important to notice that virtually all empirical studies in this literature ignore entry and exit decisions. Yet as shown by Thomas Coleman (1984), most of the variation in aggregate manhours in U.S. data comes from employment variation, not variation in hours per head. Empirical research by Gregory Mankiw et al. (1985) tests and rejects the intertemporal-substitution hypothesis using the least-important component of aggregate manhours: hours per week. When entry and exit decisions are incorporated as in George Alogoskoufos (1987) (a macro study of variation in employment), the empirical evidence shows much greater support for the intertemporal-substitution hypothesis. Adding nonseparability in preferences over time and across persons also raises estimates of intertemporal-substitution elasticities for time (see V. Joseph Hotz et al., 1988).

Evidence that, controlling for measured wages, demand-side variables help to determine individual labor supply, is consistent with many competing theories. Although it is sometimes used to cast doubt on the validity of the intertemporal-substitution hypothesis, it can only do so for naive versions of that hypothesis cast strictly in terms of measured wages. ("The wage should be a sufficient statistic for the labor market.") Demand-side variables may simply proxy expectations about future wages or real interest rates or may proxy amenities that determine labor supply and wage packages. Assuming that wages are correctly measured—a strong assumption in view of the absence of measured wages for nonworkers—the evidence suggests that naive spot-wage models of labor supply do not explain much of the data on aggregate fluctuations. Much variation in employment comes at

fixed *measured* wages. Whether employment responds elastically to unmeasured wage offers is not known.

V. Better Understanding of the Limits of the Data

Two studies conducted at the Survey Research Center of the University of Michigan provide important evidence on the quality of the labor supply and wage data used in many micro and macro labor supply studies. Two major facts based on the Current Population Survey (CPS) which have attracted wide-scale attention—the decline in real wages for certain skill groups and the small decline in male hours of work per week over the past 25 years—arise, in part, from systematic respondent reporting errors. F. Thomas Juster and Frank Stafford (1991) document that widely used CPS questionnaire formats produce bunching at standard weekly hours of work. More precise time-diary studies show fewer hours worked per week spread out more evenly across persons, even ignoring information on work breaks and consumption of leisure on the job. The persistent bunching of reported time at standard intervals affects estimates of secular trends in wages and labor supply. CPS data show a 2.7-percent decline in work time over the period 1965–1981 for men age 20–65 while the hours from time-diary data show a decline of 13.5 percent over the same period. Biases in the CPS may have obscured a secular trend toward reduced working hours by males. Juster and Stafford's evidence also calls into question the validity of recent CPS-based estimates of declining real wages. The size if not the sign of the recent measured real wage decline may be an artifact of CPS reporting methods. A 10–12-percent upward adjustment in real wages which is suggested by the upward statement of usual weekly hours over actual hours worked that is found in time-diary studies goes a long way toward dampening concern about declining real wages, especially for workers in the middle deciles of the wage distribution.

Further evidence that labor supply and hourly wages are badly mismeasured is sup-

plied in an important paper by John Bound et al. (1989). Interviewing workers and comparing interview responses with firm records, they find systematic biases in labor supply and wage data. Measurement error in hours and hourly wages neither has a mean of zero nor is it uncorrelated with true values of own and other variables as is assumed in textbook discussions. Errors in hourly wages are positively correlated with true values of wages, education, age, and job tenure. Errors in hours of work are negatively correlated with true values. CPS-type wage measures based on usual hours measures have a very low signal-to-noise ratio. The signal is positively correlated with the noise.

The measurement error reported in Bound et al.'s study of a single firm suggests substantial bias toward zero in estimated wage and income effects on labor supply. The low estimated value of the intertemporal-substitution elasticity found in panel-data studies appears to be a consequence of substantial nonstandard measurement-error problems. This evidence casts serious doubt on the validity of the panel-data estimates of labor supply parameters that play a central role in the debate on the importance of the intertemporal-substitution hypothesis. It also casts doubt on estimated cross-section labor supply equations, but the bias for these is less severe. Standard instrumental-variable methods used to correct for measurement errors in previous studies of labor supply may be invalid because of the correlation between the measurement error and the true variable and between measurement error and common instruments like age and education. Expansion of this study to a larger universe of firms would contribute greatly to understanding labor supply and wage functions fit on widely used (but not yet authenticated) CPS measures of wages and hours.

REFERENCES

- Alogoskoufos, George, "On Intertemporal Substitution and Aggregate Labor Supply," *Journal of Political Economy*, October 1987, 95, 938–60.

- Borjas, George and Heckman, James**, "Labor Supply Estimates for Public Policy Evaluation," in Industrial Relations Institute, *Proceedings of the Thirty-First Annual Meeting*, Madison, WI: Industrial Relations Association, 1978, pp. 320-31.
- Bound, John, Brown, Charles, Duncan, Greg and Rogers, Willard**, "Measurement Error in Cross Sectional and Longitudinal Labor Market Surveys: Results from Two Validation Studies," National Bureau of Economic Research (Cambridge, MA) Working Paper No. 2884, January 1989.
- Card, David**, "Intertemporal Labor Supply: An Assessment," National Bureau of Economic Research (Cambridge, MA) Working Paper No. 3602, January, 1991.
- Cogan, John**, "Fixed Costs and Labor Supply," *Econometrica*, December 1981, 49, 945-64.
- Coleman, Thomas**, "Essays on Aggregate Labor Market Business Cycle Fluctuations," Ph.D. dissertation, University of Chicago, 1984.
- Devanzo, Julie, Detray, Dennis and Greenberg, David**, *Estimating Labor Supply Response: A Sensitivity Analysis*, Rand report R-1372-OEO, Santa Monica: The Rand Corporation, December 1973.
- Hausman, Jerry**, "Taxes and Labor Supply," in A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Vol. I, Amsterdam: North-Holland, 1985, pp. 213-63.
- Heckman, James**, "A Partial Survey of Recent Research on the Labor Supply of Women," *American Economic Review*, May 1978 (*Papers and Proceedings*), 68, 200-7.
- _____, "Sample Selection Bias as a Specification Error: An Application to the Estimation of Female Labor Supply Functions," in James Smith, ed., *Female Labor Supply*, Princeton, NJ: Princeton University Press, 1980, pp. 206-48.
- _____, "Comment on the Hausman Essay," in Martin Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Research*, Chicago: University of Chicago Press, 1983, pp. 70-82.
- Hotz, V. Joseph, Kydland, Finn and Sedlacek, Guilherme**, "Intertemporal Preferences and Labor Supply," *Econometrica*, March 1988, 56, 335-60.
- Juhn, Chinhui, Murphy, Kevin M. and Topel, Robert H.**, "Why Has the Natural Rate of Unemployment Increased Over Time?" *Brookings Papers on Economic Activity*, 1991, (2), 75-142.
- Juster, F. Thomas and Stafford, Frank**, "The Allocation of Time: Empirical Findings, Behavioral Models and Problems of Measurement," *Journal of Economic Literature*, June 1991, 39, 471-522.
- Killingsworth, Mark**, *Labor Supply*, Cambridge: Cambridge University Press, 1983.
- Kosters, Marvin**, "Effects of an Income Tax on Labor Supply," in Arnold Harberger and Martin Baily, eds., *The Taxation of Income From Capital*, Washington, DC: Brookings Institution, 1967, pp. 301-21.
- Lewis, H. Gregg**, "On Income and Substitution Effects in Labor Force Participation," mimeo, University of Chicago, 1967.
- MaCurdy, Thomas, Green, David and Paarsch, Harry**, "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply," *Journal of Human Resources*, Summer 1990, 25, 415-90.
- Mankiw, N. Gregory, Rotemberg, Julio and Summers, Lawrence**, "Intertemporal Substitution in Macroeconomics," *Quarterly Journal of Economics*, February 1985, 100, 225-51.
- Mincer, Jacob**, "Labor Force Participation of Married Women: A Study of Labor Supply," in H. G. Lewis, ed., *Aspects of Labor Economics*, Princeton, NJ: Princeton University Press, 1962, pp. 63-97.
- Mroz, Thomas**, "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, July 1987, 55, 765-800.
- Pencavel, John**, "Labor Supply of Men: A Survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, Vol. I, Amsterdam: North-Holland, 1986, pp. 3-102.
- Triest, Robert**, "The Effect of Income Tax Deductions on Labor Supply When Deductions Are Endogenous," *Review of Economics and Statistics*, January 1992, 74, 91-9.