Volume Title: Empirical Foundations of Household Taxation

Volume Author/Editor: Martin Feldstein and James Poterba, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-24097-5

Volume URL: http://www.nber.org/books/feld96-1

Conference Date: January 20-21, 1996

Publication Date: January 1996

Chapter Title: Labor Supply and the Economic Recovery Tax Act of 1981

Chapter Author: Nada Eissa

Chapter URL: http://www.nber.org/chapters/c6235

Chapter pages in book: (p. 5 - 38)

1 Labor Supply and the Economic Recovery Tax Act of 1981

Nada Eissa

1.1 Introduction

U.S. personal income tax rates changed dramatically during the 1980s, especially at the top of the income distribution. In 1980, the top marginal tax rate (at the federal level) was 70 percent. The Economic Recovery Tax Act of 1981 (ERTA) reduced that rate to 50 percent, and the Tax Reform Act of 1986 (TRA86) reduced it further to 28 percent. A dominant motivation for the initial law was to alleviate the disincentives for individuals to supply labor and to save that were generated by the high marginal tax rates. For labor supply, ERTA pursued this goal by introducing a deduction for the secondary earner in the household and, more generally, by reducing marginal tax rates by 23 percent within each tax bracket.

By providing large and potentially exogenous variation in marginal tax rates, these tax laws provide fertile ground for analyzing the responsiveness of individual behavior to taxes. Evidence suggests that individual behavior did respond to the incentives in these tax laws. Lindsey (1987) and Navratil (1994) use tax return data and find that the marginal tax rate reductions in ERTA had a significant effect on taxable income. Feldstein (1993) and Auten and Carroll (1994) find similar results for TRA86. Burtless (1991) and Bosworth and Burtless (1992) study the labor supply responses to the tax reforms of the 1980s. These studies analyze the trend in labor supply for different demographic groups using Current Population Survey (CPS) data for 1968–88 and 1968–90, respectively. They find significant responses in hours of work (relative to trend)

Nada Eissa is assistant professor of economics at the University of California, Berkeley, and a faculty research fellow of the National Bureau of Economic Research.

The author is grateful to Dan Feenberg for assistance with the TAXSIM calculations, to conference participants for useful suggestions, and to David Cutler, Douglas Elmendorf, and Martin Feldstein for helpful comments.

by married women at the top and bottom of the income distribution. Eissa (1995) also finds a strong response by upper-income women to TRA86.

This paper examines whether married women responded to the incentives to increase labor supply in ERTA, using individual-level data rather than aggregated data (as in the case of Bosworth and Burtless 1992). I focus on married women for two reasons. First, ERTA implicitly targeted this group in the introduction of the secondary earner deduction. Second, this group is believed to be more responsive to changes in the tax rate than any other group (men and female heads of households). While the empirical literature is in less agreement on the overall responsiveness of married women than of other groups, it is generally accepted that it is the participation decision rather than the hours-of-work decision for working women that is responsive. I therefore analyze these two margins separately.

I use the time variation in marginal tax rates to estimate both difference-in-difference regression models (where I compare the change in labor supply for upper-income women with the change in labor supply for lower-income women) and standard labor supply models (where labor supply is a function of the after-tax wage). This approach is appealing in that it allows me to address a standard criticism of the empirical analysis of taxation and labor supply. Because cross-sectional variation in marginal tax rates derives primarily from differences in income and family structure, the existing literature faces an identification problem. Separating the tax effect from a nonlinear income (or family structure) effect is difficult in the cross section. ERTA, however, generated potentially exogenous time variation in marginal rates that can be used to evaluate the responsiveness of labor supply to taxes.

Using data from the 1981 and 1985 CPS, I find weak evidence that labor force participation of upper-income married women is responsive to taxes. The point estimates suggest that following ERTA, upper-income married women increased their labor force participation by up to 2.6 percentage points (from a predicted base of 47 percent). That estimate suggests an elasticity of 0.79. For working women, the most likely values show a response of between 20 and 49 hours per year, but these are estimated with such imprecision that it is not possible to rule out no response at all. Finally, standard labor supply estimates predict participation and hours-of-work responses for upper-income women that are at the lower end of the observed responses.

This paper is organized as follows: Section 1.2 presents a quick review of the tax and labor supply literature. To motivate the regressions estimated in the paper, I first review the basic model of labor supply and outline some basic assumptions maintained for the analysis of labor supply within the household. To place the current paper in context, I present a brief review of the empirical labor supply literature. Section 1.3 reviews the provisions in ERTA relevant for the treatment of earned income. Section 1.4 presents the identification strategy, with particular attention to the difference-in-difference approach. Section 1.5

discusses the data and presents basic labor supply results. The difference-in-difference specification and regression results are presented in section 1.6, and the standard labor supply estimates are presented in section 1.7. Finally, section 1.8 concludes.

1.2 Tax and Labor Supply Literature

1.2.1 Model of Labor Supply

To motivate the labor supply equations that I estimate, I sketch below the simple model of taxes and labor supply. I use the basic static, two-good labor supply model. In this model, the worker has a utility function defined over consumption and leisure and chooses her hours of work at a fixed wage. The optimization problem is typically characterized as follows:

$$\max u(c,l) \text{ such that } [wh - G(wh,y)] + y = c, \tag{1}$$

where $u(c, l)$ is the utility function, c is consumption of a composite commodity (the numeraire), l is leisure, wh is the woman's labor income (the product of the wage and hours of work), y is unearned income, and $G(wh, y)$ is the tax liability. The partial derivative of utility with respect to leisure, u_l, is positive, and the second partial, u_{ll}, is negative.

The function $G(\cdot)$ can incorporate nonlinearities and nonconvexities of the budget set due to several features of the tax system, such as the social security payroll tax, the earned income tax credit, and transfer programs. The marginal tax rate, θ, is the derivative of the function $G(\cdot)$ with respect to labor income. At interior solutions, the equality of the marginal rate of substitution between consumption and leisure (MRS) and the net wage determines labor supply: $u_l/u_c = w(1 - \theta)$.

With a nonlinear budget set, the choice of hours and consumption determines the segment of the budget constraint on which the individual locates. For local movements, behavior is equivalent to that arising from utility maximization subject to a linear budget constraint with the net wage, $w(1 - \theta)$, and "virtual" income given by the intercept of that budget segment with the consumption axis. Theory predicts that the income effect is negative because greater income leads one to purchase more leisure (assuming that leisure is a normal good). Theory provides little guidance on the effect of the net wage, however, because the substitution effect from a tax cut leads to an increase in labor supply while the income effect leads to greater consumption of leisure.

This paper analyzes separately two measures of labor supply: participation and hours of work conditional on working. Because no attempt is made to explicitly address the nonlinearity of the budget set, one can characterize the participation decision as a function of the MRS at zero hours and the after-tax market wage.[1] If the individual has unearned income, an uncompensated change in the tax rate has an ambiguous effect on the participation margin.

Unearned income for a married woman includes her spouse's earned income and is therefore affected by the spouse's labor supply decisions. I assume that the interaction between the husband's and wife's labor supply is governed by the chauvinist model. In that model, the wife conditions her labor supply on her spouse's labor supply decision, making her the secondary earner in the household. The husband's earned income affects the wife's labor supply only through an income effect. This model is useful for two reasons: the household's capital income and the husband's total labor income generate an exogenous measure of unearned income, and the first-hour marginal tax rate faced by the wife is her spouse's last-hour marginal rate. The chauvinist model is clearly not valid for some households. One could use more general models that allow, for example, an interaction between one partner's wage and the other partner's labor supply. Evidence suggests, however, that such models do not produce very different estimates of the wage elasticity of labor supply (Hausman and Ruud 1984).

1.2.2 Empirical Labor Supply Literature

The empirical literature on taxation and labor supply is extensive and employs several approaches to estimating labor supply equations. The early literature posited a linear budget constraint (i.e., a proportional tax) and estimated a structural labor supply equation of the form

$$h_i = \delta_0 X_i + \delta_1 w_i^n + \delta_2 y_i + \varepsilon_i , \tag{2}$$

where h_i is annual (or weekly) hours of work, X_i is a set of individual characteristics, w_i^n is the after-tax hourly wage, and y_i is unearned income.

Researchers used ordinary least squares (OLS) and two-stage least squares (2SLS) to estimate the hours-of-work equations (Boskin 1973; Hall 1973). Estimates from this early work generally found large wage and income elasticities for married women. Killingsworth and Heckman's (1986) review of the literature cites a range of estimated wage elasticities of -0.3 to 14, with a tendency toward 1. More specific to taxation, Hausman (1985) cites a range of estimates of -0.3 to 2.3. Mroz (1987) showed that this elasticity is closer to zero for *working* married women, but his estimates also suggest that the participation decision may be quite sensitive to the wage. This result is also found in recent work using the nonlinear budget set approach identified with Hausman (1981).

Hausman (1981) assumes a functional form for taxpayers' preferences and then estimates preference parameters by solving an optimizing model in which the nonlinear and nonconvex budget constraints facing taxpayers are carefully

1. Nonconvexities in the budget constraint invalidate this decision process because the entire budget constraint and not just the marginal tax rate on the participation margin determine whether the individual enters the labor force (Hausman 1980).

modeled. Using maximum likelihood methods and cross-sectional data from the 1975 Panel Study of Income Dynamics (PSID), Hausman estimates a net wage elasticity of approximately 1 for married women. Using a similar methodology and 1984 PSID data, Triest (1990) estimates a total labor supply elasticity of 1.1 for married women. However, he estimates an elasticity of only 0.2 for *working* married women. Triest's results provide further support for the view that the participation decision is more responsive to changes in the net wage than are hours conditional to working.

While the careful modeling of the nonlinearity of the budget set is appealing, this approach faces two critical problems. First, the results are quite sensitive to the specification of preferences chosen (Blundell and Meghir 1986), and even under similar preference specifications, results do not seem to be replicable across different data sets and time periods (MaCurdy, Green, and Paarsch 1990). Second, constraints that make the models tractable appear to be binding and heavily influence the results (Heckman 1982; MaCurdy et al. 1990). MacCurdy et al. show that the nonlinear budget set approach imposes the Slutsky condition, which amounts to restricting the income effect to be negative. Even when the nonlinear budget constraint approach is not used, structural labor supply models are extremely sensitive to the specification chosen (Mroz 1987).

An alternative approach to identifying labor supply responsiveness is to examine the response of taxpayers to changes in tax laws (Eissa 1995; Blundell, Duncan, and Meghir 1995). Eissa analyzes the responsiveness of upper-income married women to the large tax reductions in TRA86. That paper uses a difference-in-difference approach and compares the labor supply response of married women at the 99th percentile of the CPS income distribution to women at the 75th and 90th percentiles of the same distribution. Blundell et al. use the several tax reforms in England during the 1980s to estimate a structural model of labor supply that allows them to distinguish between income and substitution effects. The advantage of this approach is that it relies on minimal and transparent assumptions for identification.

This paper follows in the line of the natural experiment approach. Before I discuss the methodology, I review the relevant features of ERTA.

1.3 Economic Recovery Tax Act of 1981

Before 1981, the U.S. federal tax schedule consisted of 16 brackets, ranging from 11 to 70 percent. The highest rate applied to individuals with taxable income over $215,400. Table 1.1 presents the statutory federal income tax schedule for 1980. At higher income tax brackets, earned income was taxed at a lower rate than unearned income. The lower tax rate on earned income was due to the "maximum tax," passed as part of the Tax Reform Act of 1969. The maximum tax provided tax relief to taxpayers with substantial earned income so that the marginal rate on earned income did not exceed 50 percent. In fact,

Table 1.1 Statutory Federal Individual Income Tax Schedule for Married, Joint Tax Filers, 1980 and 1984

Taxable Income (nominal dollars)	Tax Year 1980	Tax Year 1984
0–3,400	0	0
3,400–5,500	14	11
3,400–7,600	16	12
7,600–11,900	18	14
11,900–16,000	21	16
16,000–20,200	24	18
20,200–24,600	28	22
24,600–29,900	32	25
29,900–35,200	37	28
35,200–45,800	43	33
45,800–60,000	49	38
60,000–85,600	54 (50)	42
85,600–109,400	59 (50)	45
109,400–162,400	64 (50)	49
162,400–215,400	68 (50)	50
215,400 and over	70 (50)	50
CPI adjustment factor (1980 $)	1.00	1.26

Sources: Tax schedule, Pechman (1987); CPI, *Economic Report of the President 1994* (Washington, D.C.: Government Printing Office, 1994).

complications in the law meant that the maximum tax on *earned* income could have exceeded 50 percent.[2]

In addition to the federal income tax, taxpayers faced the social security payroll tax (equal to 6.13 percent on the employee)[3] and state taxes. Although the statutory state tax rate could have been as high as 16 percent (in Minnesota), the deductibility of state taxes from federal taxable income reduces the effective marginal tax rate. Nonetheless, for many secondary earners in high-income households, the marginal tax rate for the first hour of work could have far exceeded the federal rate of 50 percent.[4]

2. Not all taxpayers with taxable income in the relevant brackets were eligible for the maximum tax: married taxpayers had to file joint tax returns and have taxable earned income more than the amount that faced the 50 percent rate ($60,000 in 1980) to be eligible. In addition, eligible taxpayers who were in the relevant brackets because they had substantial unearned income did not see marginal tax rates on earned income fall to 50 percent (see Lindsey 1981 for a discussion of the workings of the maximum tax). Lindsey estimates that most taxpayers eligible for the maximum tax faced marginal rates of more than 50 percent in 1977.

3. The social security tax is a payroll tax and therefore applies even to secondary earners in the highest tax brackets if their own earnings are below the social security maximum taxable earnings ($25,900 in 1980).

4. For a taxpayer with $55,000 of taxable earned (and $25,000 unearned) income, the marginal tax rate on earned income would be 69 percent ((55 percent federal tax + 12.26 percent FICA tax + 6 percent state tax)/1.0613).

ERTA reduced marginal tax rates by 23 percent within each bracket over a period of three years: by 10 percent in 1982, another 10 percent in 1983 (such that they were 19 percent below their pre-ERTA levels), and by 5 percent in 1984.[5] ERTA abolished the maximum tax on earned income and set the top marginal tax rate at 50 percent as of 1982. ERTA also provided two-earner married couples with a deduction equal to 10 percent of the income of the lower-earning spouse (up to $30,000).[6] The aim of this provision was to reduce the disincentive effects of high marginal rates on secondary earners. In effect, the deduction provided larger tax reductions for secondary earners at higher levels of the income distribution: it reduced the marginal rate by 5 percentage points for a secondary earner in the 50 percent bracket, but only 1 percentage point for a secondary earner in the 11 percent bracket. The federal statutory income tax schedule for 1984 is also presented in table 1.1.

Because the tax code was not indexed for inflation during this period, taxpayers were likely to find themselves in higher brackets with no real increase in income. In practice, bracket creep is important in this analysis for two reasons. First, inflation was high during the period in which ERTA took effect: between 1980 and 1984, prices increased by 26 percent. Second, jumping into higher brackets was easy because income tax brackets were narrow. Table 1.1 includes the CPI adjustment factor for 1984 to allow comparison of marginal tax rates between years. Consider the taxpayer with taxable income of $20,000 in 1980. Her marginal tax rate was 24 percent in 1980. If her income increased by the CPI each year and ERTA was not passed, she would have faced a marginal rate of 32 percent in 1984. ERTA reduced the marginal rate in the 32 percent bracket to 25 percent. Rather than fall by 23 percent, this taxpayer's federal marginal rate *increased* by 4 percent, from 24 to 25 percent.[7] In effect, bracket creep eroded most of the tax gains for lower-income and middle-income taxpayers (Lindsey 1987), thus leaving upper-income taxpayers as the main beneficiaries of the 1981 tax law. How much very high income individuals gained from ERTA depends heavily on whether the maximum tax capped marginal rates at 50 percent in 1980, however.

Unlike TRA86, ERTA represented an uncompensated tax change. ERTA contained few provisions that affected the tax base except for the expanded eligibility for IRAs and the secondary earner deduction. Thus, net-of-tax incomes rose because the tax on the family fell. Lindsey (1987) estimates that ERTA reduced tax liability by 26.8 percent in 1984. This feature of the tax law generates an income effect and affects the interpretation of the results, an issue I return to later.

5. In 1981 a credit of 1.25 percent was given against regular taxes.
6. The secondary earner deduction was only 5 percent in 1982.
7. If eligible for the secondary earner deduction, her marginal tax rate in 1984 would have been 22.5 percent.

1.4 Identification Strategy

I compare the change in labor supply of women "most" affected by ERTA (the treatment group) with women "less" affected by the law (the control group) before and after the tax reform. The treatment here is the change in the marginal tax rate, or more generally the change in the budget set. Marginal tax rates are not available in the survey data, however. I use income as a proxy for the marginal tax rate because ERTA provided greater tax reductions for women in higher-income households than for women in lower-income households.

The choice of groups from different points on the income distribution generates a potential endogeneity problem. If the allocation is made based on family income, those who respond to the tax changes (high earners after 1981) are the treatment group. This selection process biases upward the estimated response and labor supply elasticity of the treatment group. To remove this bias, the choice of the treatment and control groups before and after the tax law is based on "other household income"—the sum of husband's labor income and any nonlabor income received by the family. I choose women with real other household income of at least $50,000 as the treatment group and women with other household income between $30,000 and $50,000 as the control group.[8] The effect of the tax law is then the difference between the change in labor supply of these two groups.

This difference-in-difference approach requires that assignment into the treatment and the control groups be random. Since these groups are at different points along the income distribution, there are likely to be systematic differences in their characteristics. In addition, the composition of working women may change over the period if there is a participation response and new entrants are different from those already in the labor force. With nonrandom assignment, differences in labor market outcomes may reflect the noncomparability of the two groups rather than the effect of the tax law. To guard against this possibility, I estimate regressions in which I control for the relevant demographic characteristics.[9] With this adjustment, we need a weaker assumption: conditional on observable characteristics, allocation into the treatment and the control groups is random.

Identification is based on the assumption that there is no contemporaneous shock to relative labor market outcomes of women with large tax cuts and women with small tax cuts. This assumption is somewhat fragile in this period. Evidence suggests that wage inequality grew significantly between 1979 and 1987 (Katz and Murphy 1992). In addition, women with higher-income husbands tend to be more educated on average than women whose husbands have

8. The choice of the treatment group is guided by data limitations, which I describe in the next section.

9. Controlling demographic characteristics in a regression framework will solve the problem if new participants differ only in observable characteristics from those in the labor force.

less income. A rise in the relative wages of more-educated individuals would generate a response similar to that of a relative reduction in tax rates (Rosen 1976). Only part of the estimated response would be due to tax reductions in that case. I test for shocks to *relative* labor market outcomes by allowing the impact of education on labor market outcomes to vary over the period of the reforms. If upper-income women increase their participation or work more hours because there is greater demand for them, this test will generate a more precise estimate of the tax effect. This test will also produce better estimates of the tax effect if there are any unobservable shocks to labor supply over the period that are correlated with education. An example of such a shock would be that higher-income women prefer to work more hours at any wage following the tax law.

A more basic identification condition is that the difference in the *change* in the after-tax wage for the two groups is not zero. Wages for nonparticipants do not exist, however. To avoid imputing market wages to nonparticipants, I assume instead that the difference in the after-tax share (1-marginal tax rate) between the two groups is not zero. In other words, I implicitly assume that relative wages remain unchanged. This assumption may not be valid given the endogeneity of the wage to tax reforms and the changes in returns to education during this period. I test this assumption later using the sample of working women.

Under the identifying assumptions, I can calculate the uncompensated elasticity of labor supply for the upper-income group as

$$\eta = \frac{\Delta l^s_H - \Delta l^s_L}{\Delta(1-\theta)_H - \Delta(1-\theta)_L}, \tag{3}$$

where H indexes the treatment group (high income), L indexes the control group (lower income), η is the elasticity of labor supply, l^s refers to the labor supply measure, $1-\theta$ is the after-tax share (1-marginal tax rate), and Δx is the percentage change in x.

Because ERTA reduced the tax liability of taxpayers, the tax law has an income effect and we estimate an uncompensated elasticity. This uncompensated elasticity is relevant for estimating tax revenue. Estimates of the elasticity are presented in the next section. I do not attempt to isolate the compensated effect, which is what matters for deadweight loss calculations. Eissa (1995) estimates compensated labor supply elasticities using TRA86, which was a revenue-neutral and a distributionally neutral tax change.

1.5 Data and Basic Difference-in-Difference Results

1.5.1 Sample

The data I use come from the 1981 and 1985 March CPS. The CPS provides annual labor market and income information for the year preceding the survey, so the data are for tax years 1980 and 1984. I use 1984 as the postreform period because ERTA was phased in over three years, with most of the reductions taking place by 1983.

The advantage of the CPS is that it is the largest data set with income and hours information available for the relevant years. The disadvantage is that income fields are top-coded and the top code changes between 1980 and 1984, from $50,000 to $100,000. The low top code in 1980 affects the marginal tax rate calculation for that year and therefore biases estimates of the effect of ERTA on marginal tax rates. The direction of the bias is not clear, however. On one hand, the top code leads one to underestimate the marginal tax rate in 1980 and thus the reduction in the tax rate over the period. On the other hand, top-coding of income leads one to overestimate the marginal tax rate if these taxpayers are eligible for the maximum tax.[10] Approximately 29 percent of the 1980 sample with real other income of at least $50,000 has wage and salary income that is top-coded. To avoid misclassifying individuals because of top-coded income, I define the treatment group to have at least $50,000 in other income, the sum of both husband's earned income and family unearned income.[11]

The CPS has information on households, families, and individuals. However, the relevant unit of analysis for this study is the tax-filing unit. The tax-filing unit is based on CPS families. Therefore, subfamilies (both related and unrelated) are allocated to separate tax-filing units from the primary family. Any member of the tax-filing unit who is under age 19 (or under age 24 and a full-time student) is considered a dependent child for tax purposes. Tax-filing unit income does not include children's earned income or unearned income.

The sample is made up of married women between ages 19 and 64, residing with their employed spouses at the time of the interview. I exclude women who report being self-employed because interpreting hours of work for this group is difficult. I also exclude women who report being out of the labor force because of an illness or disability, or who report working more than 4,160 hours per year (52 weeks at 80 hours per week). Finally, I exclude women with zero or negative other household income since these women are primary earners. The resulting sample size is 54,381 observations.

Table 1.2 presents the characteristics of the sample. Column (1) presents the

10. Recall that to be eligible for the maximum tax, a married taxpayer filing a joint return had to have taxable earned income in excess of $60,000 in 1980. An individual with $75,000 of taxable earned income will be classified as having $50,000 in my data. This individual's calculated marginal tax rate is higher than his true marginal rate.

11. Individuals may also be misclassified if they misreport their income to the CPS. Here, the difference-in-difference estimates of the labor supply response are inconsistent. Scholz (1990) finds that tax units with wage and salary income more than $50,000 reported, on average, less income to the CPS than to the Internal Revenue Service (IRS) in 1984. Tax units with wage and salary income between $25,000 and $50,000 reported very similar incomes in the two data sets. The estimate of the tax effect is biased downward if individuals underreport income to the CPS. Without a match between IRS and CPS data, however, correcting for this bias is difficult.

Table 1.2 Summary Statistics of Data Sample

	All		Employed	
Variable	Before ERTA81 (1)	After ERTA81 (2)	Before ERTA81 (3)	After ERTA81 (4)
Age	38.45	38.61	37.32	37.56
	(11.77)	(11.38)	(11.28)	(10.78)
Education	12.28	12.56	12.58	12.87
	(2.61)	(2.63)	(2.48)	(2.50)
Nonwhite	.083	.090	.090	.098
	(.275)	(.029)	(.029)	(.030)
Preschool children	.419	.416	.344	.357
	(.72)	(.712)	(.636)	(.650)
Family size	3.323	3.233	3.210	3.142
	(1.31)	(1.24)	(1.24)	(1.16)
Other household income	20,819	21,574	19,656	20,430
	(12,393)	(15,341)	(11,313)	(13,737)
Interest and dividend income	741.4	1,219.2	630.0	1,018.8
	(3,176)	(5,153)	(2,825)	(4,287)
Labor force participation	.656	.687	1	1
	(.475)	(.464)	(0)	(0)
Hours	948.6	1,039.5	1,447.1	1,512.5
	(912.2)	(931.0)	(740.3)	(738.7)
Observations	29,269	25,112	19,186	17,259

Notes: All income figures are in 1980 dollars. Numbers in parentheses are standard deviations. Means are unweighted.

characteristics of all married women before ERTA. The average married woman in the pretax change period is 38.45 years old, has a high school degree, and has approximately 1.3 children. Her family has approximately $21,000 of other household income (defined as tax-filing unit income less the wife's wage income) and $740 of interest and dividend income. The probability that she is employed is two-thirds. Column (2) shows that after ERTA, she has similar characteristics.[12]

Because I analyze the hours response for working married women, I present their characteristics as well. Column (3) of table 1.2 shows that working women are not very different from nonworking women: they are younger, are more educated, have a smaller family size, and are less likely to be white but only slightly so. Also, employed women have less other household income than women out of the labor force.

Columns (1) and (2) show that the participation rate of all married women increased following the tax change: after ERTA, the labor force participation of married women increased by 3.1 percentage points, from 65.6 to 68.7 per-

12. The tax reductions, and especially the secondary earner deduction, should reduce the marriage penalty and therefore affect marriage incentives. Therefore, either there was no marriage response, or newly married women are similar to those already married.

cent. The increase in labor force participation suggests that the populations of working women before and after the tax reforms may not be directly comparable. New participants may enter the labor force at different points on the hours-of-work distribution. For the analysis, however, I assume that the distribution of hours of work for new entrants is similar to that of the pre-ERTA participants. While verifying this assumption with repeated cross-sectional data is impossible, we can check that the demographic characteristics of the two groups look similar. Columns (3) and (4) show that the two populations are quite similar, except for the number of preschool children they have. Whereas all married women have fewer preschool children after 1981, working married women have more preschool children. Therefore, new entrants into the labor force are more likely to be women with young children.[13]

1.5.2 Marginal Tax Rate

I define the marginal tax rate variable as the sum of the federal, state, and social security payroll taxes on the individual's marginal revenue product:

$$\text{tmtr} = [\text{fmtr} + (1 - \text{pitem}\cdot\text{fmtr})\cdot\text{smtr} + 2\cdot\text{ssmtr}]/(1 + \text{ssmtr}),$$

where tmtr is the total marginal tax rate, fmtr is the federal tax rate, pitem is the probability that the individual itemizes deductions for the federal income tax, smtr is the state tax rate, and ssmtr is the employer's (also employee's) share of the social security payroll tax.

The federal and state income tax rates are calculated using the NBER TAXSIM model. Several income sources are used in the calculation: wage and salary, interest, dividend, pension, self-employment, farm, and public assistance income. The CPS does not provide information on tax-filing status; therefore I assume that all couples file jointly.[14] TAXSIM computes the marginal tax rate from the tax liability incurred from an additional $100 of wage and salary income. The deductibility of state taxes from the federal tax for taxpayers who itemize their deductions reduces the contribution of the state marginal tax rate. Since the CPS does not have information on itemization and deductions, I impute the probability that the individual itemizes from the *Statistics of Income* as the share of tax returns that itemize deductions within each income class.[15]

I assume full incidence of the FICA payroll tax on the worker.[16] In 1980, the

13. Women with preschool children tend to work fewer hours than women with no young children once they enter the labor force. Thus new participants should be entering at a lower point in the hours distribution, shifting it to the left and reducing the estimated hours response.

14. In 1980, 96.8 percent of married couples filed a joint tax return; in 1984, 98.2 percent did (*Statistics of Income*).

15. The income classes (in thousands of dollars) are 0–5, 5–10, 10–20, 20–30, 30–50, 50–75, 75–100, 100–200, and 200–500.

16. The social security payroll tax is a tax only to the extent that the present value of taxes exceeds that of benefits. Feldstein and Samwick (1992) argue that married women face the full social security tax.

rate was 12.26 percent (6.13 percent on the employer). In 1984, the rate was 14.00 percent. These rates are zero for any woman whose earnings exceed the social security maximum taxable earnings.

Between 1980 and 1984, the average federal marginal rate for women in the sample fell by only 1 percentage point, from 40.2 to 39.2 percent. The average reduction in the sample is much smaller than the statutory reduction contained in ERTA because of bracket creep and increases in the social security payroll tax.

In the estimation, I rely on variation in marginal tax changes across the income distribution. Table 1.3 presents data on marginal tax rates before and after ERTA, disaggregated by the tax unit's other household income. Two observations are noteworthy. First, the largest tax reduction went to individuals at the top of the income distribution: 5.2 percentage points for women in families with at least $50,000 in other household income. Second, bracket creep and the social security tax completely offset the reductions in ERTA for women in the $10,000–$20,000 and $20,000–$30,000 income groups. At the very bottom of the other-income distribution, taxpayers received a 0.6 percentage point reduction. Table 1.4 transforms these figures to percentage changes in the after-tax share (1-marginal tax rate). It shows that women at the top received a 12.33 percent increase in the after-tax share, whereas those at the very bottom received a 1.12 percent increase.

1.5.3 Basic Labor Supply Results

In this section, I present basic results on labor force participation and hours of work for married women at different points along the income distribution.

Table 1.3 **Marginal Tax Rate**

Group[a]	Before ERTA81	After ERTA81	Change
$y \geq 50$	.599	.547	−.052
	(.001)	(.001)	(.001)
$30 \leq y < 50$	.520	.496	−.024
	(.001)	(.001)	(.001)
$20 \leq y < 30$	.437	.437	.00
	(.001)	(.001)	(.001)
$10 \leq y < 20$	.364	.364	.00
	(.001)	(.001)	(.001)
$y < 10$	.274	.268	−.006
	(.002)	(.001)	(.002)

Notes: Federal and state tax rates are calculated by TAXSIM. I assume all couples file jointly and assign each unit the average itemized deductions for the income class. I assume that the full incidence of the payroll tax falls on the worker. See text for details. Numbers in parentheses are standard errors.

[a]Other household income in thousands of dollars.

Table 1.4 After-Tax Share

Group[a]	Change (%)
$y \geq 50$	12.33
$30 \leq y < 50$	5.44
$20 \leq y < 30$	.22
$10 \leq y < 20$	.17
$y < 10$	1.12

Note: I assume that the growth rate of the real market wage is constant across groups. The reported figure is the difference in the group average of the log of the after-tax share between 1980 and 1984.

[a]Other household income in thousands of dollars.

Table 1.5 Labor Supply of Married Women Before and After ERTA81

Group[a]	Before ERTA81	After ERTA81	Change
	A. *Labor Force Participation*		
$y \geq 50$	.419 (.014)	.499 (.015)	.080 (.020)
	[1,221]	[1,143]	{19.0}
$30 \leq y < 50$	.563 (.008)	.618 (.008)	.055 (.011)
	[3,947]	[3,644)	{9.7}
$20 \leq y < 30$	.649 (.005)	.695 (.006)	.046 (.008)
	[8,307]	[6,512]	{6.9}
$10 \leq y < 20$	.704 (.004)	.723 (.005)	.019 (.007)
	[11,313]	[9,072]	{2.7}
$y < 10$	.690 (.007)	.707 (.007)	.017 (.010)
	[4,481]	[4,739]	{2.5}
	B. *Annual Hours Conditional on Working*		
$y \geq 50$	1,265.9 (35.6)	1,395.9 (34.7)	129.0 (49.7)
	[512]	[570]	{10.3}
$30 \leq y < 50$	1,369.7 (16.3)	1,428.6 (16.5)	58.9 (23.2)
	[2,223]	[2,253]	{4.3}
$20 \leq y < 30$	1,432.6 (10.1)	1,527.5 (10.8)	94.9 (14.8)
	[5,396]	[4,525]	{6.6}
$10 \leq y < 20$	1,488.7 (8.0)	1,535.9 (8.9)	47.2 (12.0)
	[7,692]	[6,558]	{3.2}
$y < 10$	1,450.7 (13.6)	1,522.6 (12.7)	71.9 (18.6)
	[3,093]	[3,352]	{5.0}

Notes: Each cell contains the mean for that group, along with standard error in parentheses, and number of observations in brackets or percentage increase in braces. Means are unweighted.

[a]Other household income in thousands of dollars.

Table 1.5 contains those results. The sample is disaggregated in the same way as in tables 1.3 and 1.4 to allow for comparison. Panel A presents participation results, and panel B presents hours of work for working women. Each cell presents the average participation rate (or hours of work) for that group, standard errors, and size of the sample.

The primary observation from panel A is that labor supply of married women was changing dramatically during the period for nontax reasons. Labor force participation increased even for those groups that saw no change in the marginal tax rate.

The second observation is that women whose marginal rates fell showed larger increases in labor force participation than those whose marginal rates were unchanged: those with spouses earning more than $50,000 increased their labor force participation by 8 percentage points, from 41.9 to 49.9 percent.[17] Those with spouses earning between $30,000 and $50,000 increased their participation rate by 5.5 percentage points.

Clearly the 8 percentage point increase in participation at the top of the income distribution reflects factors other than the effect of ERTA. Using women below that income bracket, one can generate several different estimates of the response to ERTA. Each estimate entails different assumptions about the comparability of the groups. I use the $30,000–$50,000 group because it is closest to the treatment group. While having a control group whose marginal tax rate is unaffected by the tax law (such as the $20,000–$30,000 group) is preferable, the raw means suggest that the results would be similar.

A comparison of the treatment and control groups suggests that participation of upper-income married women increased by 2.5 percentage points, or 6 percent. Over the same period, the relative after-tax share increased by approximately 6.9 percent. The implied elasticity of participation is 0.86.[18]

Panel B of table 1.5 presents average annual hours of work for *working* women. It shows that, while the highest income group increased their hours of work significantly over the period (129.4 hours), so did working women married to the poorest men (72 hours). This pattern is similar to that found by Bosworth and Burtless (1992) using data stratified into quintiles by family income. Overall, the pattern of annual work hours is far less continuous than that of participation. Recall, however, that it is a different group of women working after 1981 and that new participants may work fewer hours than existing participants.

A comparison of the treatment and control groups suggests that annual hours of work by upper-income married women increased by 70.1 hours. This figure represents a 5.5 percent increase in annual hours of work. Dividing by the corresponding relative increase in the after-tax share for working women (7.2 percent) produces an elasticity of 0.77. The question to ask is whether this is a response to tax reduction or to a combination of wage growth and tax reductions. If wages grew faster for the treatment group than the control group, estimates of the effect of ERTA are biased upward. Table 1.6 shows that the *relative* gross hourly wage increased by 2.2 percent between 1980 and 1984 for

17. All dollar amounts are in 1980 dollars.

18. An alternative measure of the effect of the tax law is the share of *non*participants drawn into the labor force. Because the two groups have different participation rates before the tax law, this measure need not generate similar conclusions. In this sample, the elasticity of nonparticipation is 0.62.

Table 1.6 **Change in Gross Hourly Wage**

Group[a]	Change (%)
$y \leq 50$	7.25
$30 \leq y < 50$	5.02
$20 \leq y < 30$	4.93
$10 \leq y < 20$	2.29
$y < 10$	−.57

Note: Reported figures are for workers with hourly wages between $1 and $100 (1980 dollars). The change is the difference in the average log gross wage between 1980 and 1984.

[a]Other household income in thousands of dollars.

the relevant groups.[19] The relative increase in the after-tax wage for upper-income women becomes 9.4 percent. Using the difference in the after-tax wage suggests an annual hours-of-work elasticity of 0.58 (5.5/94).[20]

Even after adjusting the estimates for wage growth, the responses suggested by the raw means are much larger than expected. Using very different methodologies, Mroz (1987) and Triest (1990) estimate uncompensated hours-of-work elasticities that are close to zero for working married women. In the next section, I address the concern that differences in observable characteristics or changes in these characteristics over time may bias both the hours-of-work and the participation responses.

1.6 Difference-in-Difference Regressions

Because women in the treatment group differ from women in the control group in characteristics that are relevant for labor supply, the observed relative differences in participation and hours of work may reflect the noncomparability of the groups rather than a response to ERTA. I control for such a possibility in this section. After presenting the specification used, I discuss the regression results.

1.6.1 Specification

Assuming that disutility of labor is normally distributed in the population generates the probit model for participation, specified as follows:

$$(4) \qquad P(\text{lfp}_{it}) = \Phi(\alpha_1 Z_{it} + \alpha_2 T_t + \alpha_3 \text{High}_k + \alpha_4 (T^*\text{High})_{kt}).$$

The hours-of-work equation is specified as follows:

19. This number cannot be added to the results in table 1.4 because it is based on the sample of working women. Table 1.4 presents the figures for the entire sample.

20. If lower taxes lead women to choose better-paying but less-attractive jobs, and the higher wage represents a compensating differential, then hours of work should not adjust to the higher wage. Here the response is both the higher hours and the greater pay per hour. The hours-of-work increase is 30 percent higher (2.2/7.2), and the elasticity remains 0.77.

(5) $$h_{it} = \beta_1 X_{it} + \beta_2 T_t + \beta_3 \text{High}_k + \beta_4 (T^*\text{High})_{kt} + \mu_{it},$$

where i indexes individuals, t indexes time, k indexes the group, Z_{it} and X_{it} are individual characteristics, T_t is a dummy equal to 1 for 1984 and equal to 0 for 1980, High_k equals 1 if real other household income is at least $50,000 (1980 dollars) and equals 0 if real other household income is between $30,000 and $50,000, and μ_{it} is an error.

The set of covariates Z and X are assumed to adequately control for allocation into the treatment group. The variables included are age, age squared, education, education squared, the number of preschool children, family size, a dummy for self-employed spouse, a race dummy (equal to 1 if the woman is nonwhite), 50 state dummies, and a dummy for residence in a central city. Any unobservable differences in labor supply preferences between the various groups will be picked up by the income class dummy High_k. The coefficient on this variable is expected to be negative because higher-income women will purchase more leisure than their counterparts further down the income distribution. To control for common macroeconomic factors affecting the labor supply of married women, I include a year dummy. Because participation and hours are increasing over time, α_2 and β_2 should be positive. The behavioral response to ERTA will be reflected in the coefficients on the interaction T^*High. A test that ERTA increased the labor supply of upper-income women is a test that α_4 and β_4 are greater than zero.

Thus far, the tax unit's other income determines the group assignment. Because the income distribution shifts over time with productivity growth, the number of families with real other household income of at least $50,000 should be greater in 1984 than in 1980. Classifying individuals using income may generate groups that differ in characteristics over the period. Note that if the included covariates capture all differences, estimates of the labor supply response will not be biased. Nonetheless, I generate different estimates by sorting individuals based on their percentile position in the income distribution. I classify women at or above the 95th percentile of the other income distribution as the treatment group and women between the 80th and 95th percentiles of the income distribution as the control group.

1.6.2 Difference-in-Difference Regression Results

Labor Force Participation

Table 1.2 showed that ERTA reduced taxes most for women with real other household income of at least $50,000. The raw data show that this group increased its participation by 6 percent and hours of work (by those employed) by 5.5 percent following the passage of the tax law.

This section presents results for regressions that control for various observable characteristics. Two sets of estimates are presented in table 1.7. In the first set of regressions, women are classified by their tax unit's real other household

Table 1.7 **Difference-in-Difference Probit Results: Labor Force Participation**

	Level Classification			Percentile Classification	
Variable	Coefficient (1)	Marginal Effect (2)	Coefficient (3)	Coefficient (4)	Coefficient (5)
Age	.038 (.012)	−.0116	.038 (.012)	.036 (.011)	.036 (.012)
Age^2	−.001 (.000)		−.001 (.000)	−.001 (.000)	−.001 (.000)
Education	−.010 (.046)	.0284	−.010 (.046)	.001 (.044)	.001 (.044)
Education^2	.003 (.002)		.003 (.002)	.003 (.002)	.003 (.002)
Children under age 6	−.440 (.029)	−.1732	−.440 (.029)	−.443 (.028)	−.443 (.028)
Nonwhite	.183 (.069)	.0709	.184 (.069)	.206 (.065)	.206 (.065)
Education*T			.012 (.012)		.010 (.011)
T	.098 (.030)	.0386	−.058 (.161)	.073 (.029)	−.059 (.153)
High	−.410 (.044)	−.1621	−.404 (.044)	−.392 (.041)	−.388 (.041)
T*High	.080 (.062)		.071 (.062)	.039 (.058)	.030 (.059)
Log-likelihood	−6,242		−6,242	−6,805	−6,804
Observations		9,995		10,871	
Predicted responses	.030 (.023)		.026 (.023)	.014 (.021)	.011 (.022)
Implied elasticity	0.91		0.79	0.42	0.33

Notes: Regressions include family size, 50 state dummies, a central city dummy, and a dummy for self-employed spouse. Data are from March CPS 1981 and 1985. High$_t$ equals 1 if real other household income exceeds $50,000 in cols. (1)–(3) (the control group is women with other income between $30,000 and $50,000) and at or above the 95th percentile in cols. (4) and (5) (the control group is women with other income between the 80th and 95th percentiles). Numbers in parentheses are standard errors.

income. These results are comparable to those generated from tables 1.3, 1.4, and 1.5. The treatment group is women with real other household income of at least $50,000. The control group is women with real other household income between $30,000 and $50,000. In the second set of regressions, women are classified by their tax unit's position in the other-income distribution. Columns (1)–(3) present the former set of results, and columns (4) and (5) present the latter set of results.

Column (1) presents the probit coefficients for equation (4). Because the probit is a nonlinear model, these coefficients are not equivalent to the marginal effects of the variables on participation. Column (2) presents the marginal probabilities.[21] All estimated coefficients in the regression have the expected signs. Older women are less likely to be in the labor force: the marginal probability is −1.16 percentage points. One year of education increases the probability of entering the labor force by 2.84 percentage points. The number

21. To generate the marginal effects, I multiply the normal density function (evaluated at the individual characteristics) by the coefficient on the after-tax share. The estimates presented are sample averages.

of preschool children reduces the likelihood that the mother enters the labor force, as does having a high-income spouse.

The probit coefficient for the interaction variable is 0.080 (with a standard error of 0.062). I generate the treatment effect using the sample of upper-income married women observed after the tax change. For each woman in that sample, I predict participation assuming that β_4 is zero; I then predict participation at the estimated value of β_4. The difference in the sample average of the participation probabilities is the treatment effect. The predicted increase in participation is 3.0 percentage points (from a base of 47 percent), with a standard error of 2.3 percentage points.[22] A simple calculation shows that the implied elasticity of labor force participation with respect to the after-tax share is 0.91 (6.3/6.9).[23]

These calculations assume that the distribution of potential market wages of the treatment and the control group grew at the same rate between 1980 and 1984. I test that labor demand explains part of the response by adding a variable that interacts education with the time dummy (T). The results are in column (3).

The education interaction is small and statistically insignificant and does not alter the overall results. The predicted response at the top of the income distribution falls to 2.6 percentage points (with a standard error of 2.3), and the implied participation elasticity falls to 0.79. Changes in the returns to education, therefore, explain only 13 percent of the estimated response. This finding should not be surprising. The difference in average education between two groups is very small, less than one year. It seems that distinguishing between a common time effect and a differential education response in this sample is not possible. Note that the inclusion of the education interaction makes the time dummy negative and insignificant. In fact, when evaluated at the average education level in the sample (13.58), the sum of the time and the education coefficients (0.105) is very similar to the time coefficient in the basic regression (0.098).

Using income percentiles to define the various groups generates insignificant participation responses that are of the same order of magnitude. Women at or above the 95th percentile of other-income distribution constitute the high-

22. The asymptotic variance of the estimated treatment effect is given by

$$V(G(\theta)) = \left[\frac{\partial G(\hat{\theta})}{\partial \hat{\theta}}\right] V(\hat{\theta}) \left[\frac{\partial G(\hat{\theta})}{\partial \hat{\theta}}\right]',$$

where $G(\hat{\theta})$ is the treatment effect given by

$$G(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}[\Phi(X_i\hat{\theta} \mid D = 1) - \Phi(X_i\hat{\theta} \mid D = 0)]$$

and $\hat{\theta}$ = estimated parameters, Φ = normal cumulative distribution function, D = treatment interaction dummy, X_i = regressors for individual i, and ϕ = normal density function.

23. Note that this estimate does not necessarily imply that a 1 percent rise in the marginal tax rate will reduce participation by 0.91 percent.

income group. Women who fall between the 80th and 95th percentiles of the distribution constitute the control group. The results are presented in columns (4) and (5). The predicted response falls to a statistically insignificant 1.4 percentage points. Accounting for changes in the returns to education reduces those figures further to 1.1 percentage points. These results imply a participation elasticity of 0.33 with respect to the after-tax share.

Controlling for demographic characteristics does not alter the basic difference-indifference estimates of ERTA's effect on married women's labor force participation. The range of participation responses is 1.1 to 2.6 percentage points.

Specification checks. There remains significant variation in income, both within groups and over time. Part of the time variation in income is artificial, however. The change in the top code (from \$50,000 in the 1981 March CPS to \$100,000 in the 1985 March CPS) increases the treatment group's other income relative to that of the control group. This spurious increase in income generates a reduction in participation by upper-income women (by the income effect) and, as a result, increases the estimated effect of ERTA. To remove this bias, I top-code the husband's wage and salary income at \$50,000 (adjusted for inflation) in the 1985 March CPS and generate an adjusted other income, which I then include as a regressor in the participation equation. The inclusion of the adjusted measure of other income does not affect the results. The estimated treatment effect increases to 3.2 percentage points in the level classification and 1.6 percentage points in the percentile classification.

That upper-income women are observationally different from lower-income women leaves open the possibility that the estimated responses are due to a contemporaneous shock correlated with observable characteristics. If the estimated response varies by race, we would be suspicious of the interpretation that ERTA caused the observed shift in labor supply. To check this possibility, I interacted age, race, family size, and children younger than six variables with the time dummy. The predicted response to ERTA remained unaffected. The results are also not sensitive to the use of education and cohort dummies.

An additional concern is that the nonlinearity of the probit model drives the estimated response. The difference-indifference approach relies heavily on the linearity of the model. To gauge the bias from the nonlinearity of the model, I estimate linear probability models of the participation decision. The estimated responses mimic closely the predicted responses using the probit model: 3.2 percentage points using the level classification and 1.4 percentage points using the percentile classification.

Annual Hours of Work

Table 1.8 presents the hours-of-work regression results for *working* married women. Because the evidence suggests a participation response, we should be careful in interpreting the hours results without any correction for the selection

Table 1.8 **Differences-in-Differences OLS Results: Annual Hours Conditional on Employment**

	Level Classification				Percentile Classification			
Variable	(1)		(2)		(3)		(4)	
Age	35.30	(9.23)	35.89	(9.23)	33.17	(8.75)	32.98	(8.75)
Age^2	−.449	(.109)	−.446	(.109)	−.422	(.104)	−.420	(.104)
Education	−147.33	(42.66)	−146.35	(42.42)	−123.49	(39.81)	−122.16	(39.48)
Education^2	6.29	(1.51)	5.91	(1.51)	5.51	(1.41)	5.13	(1.41)
Children under age 6	−125.23	(24.02)	−126.59	(24.01)	−133.80	(22.49)	−135.06	(21.48)
Nonwhite	221.61	(45.20)	222.51	(45.15)	220.92	(42.09)	221.88	(42.06)
Education*T			18.49	(9.07)			18.17	(8.53)
T	24.15	(22.39)	−228.89	(124.42)	24.74	(21.41)	−223.42	(117.85)
High	−124.66	(38.37)	−118.66	(38.56)	−111.04	(35.21)	−105.01	(35.36)
T*High	61.76	(52.15)	49.91	(52.65)	32.89	(48.72)	21.02	(49.12)
R^2	.090		.091		.089		.089	
Observations	5,558				6,146			
Elasticity	0.56		0.45		0.34		0.22	

Notes: Regressions include family size, 50 state dummies, a central city dummy, and a dummy for self-employed spouse. Data are from March CPS 1981 and 1985. High_i equals 1 if real other household income exceeds $50,000 in cols. (1) and (2) (the control group is women with other income between $30,000 and $50,000) and at or above the 95th percentile in cols. (3) and (4) (the control group is women with other income between the 80th and 95th percentiles). Numbers in parentheses are standard errors.

effect. If one could identify a shock to participation but not to hours of work, one could identify a selection model. ERTA, however, does not include any provisions that affect participation separately from hours of work. With that caveat in mind, I present the regression results.

Column (1) of table 1.8 presents results of equation (5) using OLS. The evidence for an hours-of-work response is weak. The table shows that, relative to women in the next income group, upper-income married women worked 61.76 more hours per year (with a standard error of 52.15). Purging the effect of changes in returns to education reduces the response to 48.8 hours. Using the percentile definitions, the estimate falls to 20 hours per year (with a standard error of 49.12). Simple calculations suggest that the uncompensated elasticity of hours of work is between 0.22 and 0.45. Therefore, controlling for observable characteristics reduces the hours-of-work response at the top of the income distribution.

1.7 Standard Labor Supply Estimates

I have argued that the difference-in-difference approach is preferable to the more standard approach of estimating labor supply equations because it does not rely explicitly on any measure of the net wage. That is an advantage because the net wage is measured with error. Heckman (1993) notes that "CPS-type wage measures have a very low signal-to-noise ratio." Of course, the marginal tax rate will also be measured with error since survey data generally does not include deductions or exemptions. Therefore, the net wage coefficient in a standard labor supply equation will be biased downward. If the measurement error averages to zero in the defined income classes, however, then the difference-in-difference results are unbiased. In this section, I compare the previous results with standard labor supply estimates.

1.7.1 Labor Force Participation

Table 1.9 presents the results for the labor force participation equation:

$$P(\text{lfp}_{it} = 1) = \Phi(\beta_1 Q_{it} + \beta_2 \ln(1 - \theta)_{it} + \beta_3 y_{it}). \quad (6)$$

The covariate set Q includes the same covariates as in the difference-in-difference regressions, θ is the marginal tax rate on the participation margin,

Table 1.9 Standard Labor Supply Model Probit Results: Labor Force Participation

Variable	Coefficient (1)	Marginal Effect[a] (2)	Coefficient (3)	Coefficient (4)
Age	.053 (.004)	−.0076	.052 (.004)	.052 (.004)
Age²	−.001 (.000)		−.001 (.000)	−.001 (.000)
Education	.046 (.010)	.0327	.037 (.010)	.037 (.010)
Education²	.002 (.000)		.003 (.000)	.003 (.000)
Children under age 6	−.374 (.011)	−.1318	−.376 (.011)	−.376 (.011)
Nonwhite	.161 (.024)	.0547	.164 (.024)	.164 (.024)
After-tax income (thousand $)	−.001 (.001)	−.0047	.001 (.002)	.001 (.002)
log (After-tax share)	.622 (.078)		.330 (.095)	.293 (.099)
High*log(After-tax share)				.375 (.315)
Time dummy	Yes		Yes	Yes
Income dummies	No		Yes	Yes
Log-likelihood	−30,684		−30,607	−30,606
Observations		54,373		
Marginal effect of tax variable	.219 (.027)		.116 (.033)	.213 (.112)
Elasticity	0.32		0.17	0.32[b]

Notes: Other covariates include family size, a dummy for self-employed spouse, 50 state dummies, and a central city dummy. Data are from March CPS 1981 and 1985. In cols. (3) and (4), I include 10 income class dummies (see text for definition). Numbers in parentheses are standard errors.

[a]Marginal effect is given by $[\phi(x\beta)]^*\beta$, where ϕ is the standard normal density evaluated using the estimated parameters and β is the estimated coefficient.

[b]The elasticity in col. (4) refers to high-income women (other income of at least $50,000).

and y_{it} is the tax unit's after-tax income, excluding the wife's income. I estimate several specifications of this model.

The probit coefficients of equation (6) are in column (1) of table 1.9, and the marginal probabilities are in column (2). The marginal effects are similar to those estimated in the difference-in-difference regression. The estimate suggest that an additional year of education increases the likelihood that a married woman enters the labor force by 3.27 percentage points, as compared to 2.86 percentage points in the difference-in-difference regression. Older, white women with preschool children are less likely to be in the labor force, again in similar magnitudes to what the previous estimates suggest.

The coefficient on the log of the after-tax share is 0.622, with a standard error of 0.078. The marginal effect, presented at the bottom of the table, is 0.219 (with a standard error of 0.027).[24] To calculate the elasticity of participation, I divide the marginal probability by the average participation rate in the sample (0.671). A 1 percent increase in the after-tax share leads to a 0.32 percent increase in labor force participation.[25]

Both cross-sectional and time variations in marginal tax rates identify the tax effect. The cross-sectional variation in taxes derives largely from differences in income and family size, creating a potential identification problem. If the relationship between these variables and hours of work is nonlinear, then the after-tax share variable may reflect these nonlinearities rather than the tax effect. To account for this possibility, I reestimated the participation equation with 10 dummies for other-household income.[26] Column (3) of table 1.9 shows that the tax coefficient declines by almost 50 percent, from 0.622 to 0.330 (with a standard error of 0.095). Including income class dummies reduces the marginal effect to 11.6 percentage points and the elasticity to 0.17.

One explanation for these results might be that income dummies removed time (and cross-sectional) variation in marginal tax rate. Tables 1.3 and 1.4 show that the tax changes were correlated with income. If this were the entire story, however, we should have observed an increase in the standard error and no change in the tax coefficient. The more plausible explanation is that the relationship between other household income and labor force participation

24. The result is not very different if calculated at the characteristics of the average woman.

25. It is convenient to use the after-tax share because wages are not observed for those who are not working. Nonetheless, wages can be imputed for nonworkers. I do so by using Heckman's (1979) technique for correcting for sample selection bias. I predict the hourly wage by estimating a wage equation (using the sample of working women) in which I include the following demographic variables: age and education (in levels and higher-order terms), an age-education interaction, dummies for time, state, race, and central city residence, and a sample selection term. Using the estimated coefficients from the wage equation, I predict an hourly wage for each woman in the sample. The results for the regression that includes the net wage variable suggest that taxes have a much stronger effect on participation: the elasticity on the participation margin is 0.59. The problem with this procedure is that identification is derived from functional form assumptions.

26. The dummies are defined for the following group: $y \geq \$50,000$, and at \$5,000 intervals for incomes below \$50,000. The excluded dummy is $\$40,000 \leq y < \$45,000$.

is nonlinear and that the coefficient on the after-tax share reflects this nonlinearity.

The difference-in-difference results suggest a participation elasticity much larger than the 0.17 estimated here. Might it be the case that upper-income women are more responsive than the "average" woman? The results in column (4) show that upper-income women do, in fact, have higher participation elasticities than the average woman, although the difference is not statistically significant. The coefficient on the after-tax share (for high-income women) is 0.668, and the implied elasticity is 0.32.

Measurement error in the after-tax share resulting from the fact that deductions and exemptions are imputed may still bias this estimate downward. If this measurement error averages to zero in the defined income classes, then the difference-in-difference estimates are preferable to the standard model estimates.

1.7.2 Annual Hours of Work

Table 1.10 presents results for the hours-of-work equation:

$$(h_{it} \mid X_{it}, w^n_{it}, y_{it}, h_{it} > 0) = \alpha_1 X_{it} + \alpha_2 \ln(w^n_{it}) + \alpha_3 y_{it} + \varepsilon_{it}, \tag{7}$$

where X_{it} includes the same covariates as the participation regression. The income variable used in these regressions is "virtual" income. The wage variable is constructed in the usual method: by dividing wage and salary income by annual hours of work. The after-tax wage is the product of the hourly wage and the after-tax share.

Table 1.10 presents the basic regression results.[27] Column (1) presents the OLS results. Again all estimated coefficients in the regression have the expected signs. Women work fewer hours as they get older. Nonwhite women and women with preschool children also work fewer hours, as do women with more virtual income.

I estimate the regressions using the observed net wage without correcting for self-selection. This regression produces an after-tax wage coefficient of 53.89 hours. To translate this into an elasticity, I divide the coefficient by the average annual hours worked in the sample (1,478). The uncompensated elasticity of hours of work with respect to the after-tax wage is 0.03.

To address the endogeneity of the marginal tax rate and virtual income to hours worked, I use the net wage and the unit's after-tax income at the zero hours margin as instrumental variables.[28] The 2SLS estimates are presented in

27. The sample size for these regressions is smaller than that reported in table 1.2 because the log of the wage is undefined for women who have a zero wage (volunteers).

28. The first-hour marginal tax rate is a valid instrument if it is correlated with the actual marginal rate and uncorrelated with the error in the hours equation. In my sample, it is easy to defend the first assumption: the correlation between the marginal rate and its instrument is 0.44. It is not so easy to defend the second assumption, however. The error in the labor supply equation may be correlated with the first-hour rate if we imagine an assortative mating process. Suppose that higher-

Table 1.10 **Standard Labor Supply Model OLS and 2SLS Results: Annual Hours Conditional on Employment**

Variable	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)
Age	54.34 (2.80)	52.27 (2.81)	52.11 (2.83)	52.02 (2.83)
Age^2	−.688 (0.04)	−.662 (0.04)	−.660 (0.04)	−.659 (0.04)
Children under age 6	−134.54 (7.24)	−136.79 (7.25)	−137.56 (7.25)	−137.56 (7.25)
Nonwhite	135.34 (13.86)	133.19 (13.88)	134.58 (13.89)	134.00 (13.89)
Virtual income (thousand $)	−6.03 (0.45)	−6.00 (0.46)	−6.09 (1.69)	−5.75 (1.69)
log (Net wage)	53.89 (4.69)	102.88 (4.72)	102.39 (4.73)	106.09 (4.90)
High*log(Net wage)				−54.18 (18.15)
Time dummy	Yes	Yes	Yes	Yes
Income dummies	No	No	Yes	Yes
Adjusted R^2	.085	.082	.084	.083
Observations		35,851		
Elasticity	0.03	0.07	0.07	0.04[a]

Notes: Regressions also include family size, a dummy for self-employed spouse, 50 state dummies, and a central city dummy. Data are from March CPS 1981 and 1985. Numbers in parentheses are standard errors. The sample here includes only women with hourly wages between $1 and $100 (1980 dollars). The instrument in the 2SLS regression is the first-hour, after-tax wage. In cols. (3) and (4), I include 10 income class dummies (see text for definition).

[a]The elasticity in col. (4) refers to high-income women (other income of at least $50,000).

column (2). The coefficient on the log of the net wage increases to 102.88 (statistically significant at the 99 percent confidence interval). Nonetheless, the elasticity of hours of work, 0.07, remains very small. In column (3), I reestimate the hours-of-work equation by adding 10 income class dummies (defined as in the participation equation). Here, time variation in after-tax wages rather than cross-sectional variation identifies the hours equation. The inclusion of income dummies does not affect the results: the after-tax wage coefficient remains 102.39 (with a standard error of 4.73). The hours-of-work equations do not seem to exhibit the nonlinear relationship with income found in the participation equations.

The uncompensated elasticity estimates in columns (2) and (3) predict hours-of-work responses to ERTA that are much smaller than those estimated in table 1.8. Moreover, allowing a separate coefficient for upper-income women does not reduce this divergence (col. [4]). The coefficient on the interaction variable (High*log(Net wage)) is negative and statistically significant:

income men have stronger tastes for work and they tend to marry women that are like them. Because it implies a positive correlation between tastes for work and first-hour tax rates, assortative mating would bias the coefficient on the net wage variable upward.

the elasticity of labor supply for high-income women is only 0.04, smaller than for the "average" woman in the sample.

Elasticities derived from basic hours-of-work equations are generally lower than those generated using difference-in-difference methods. One explanation for this divergence is measurement error in the net wage. Evidence suggests that there is significant measurement error in the wage and that it is negatively correlated with hours of work (Mroz 1987). Such error would bias the coefficients in the standard model toward zero. If this measurement error averages to zero in the income classes defined, then the difference-in-difference regression results are unbiased (Wald 1940; Angrist 1991). In addition, the standard labor supply models estimated were the most basic models. No attempt was made to address biases due to wage endogeneity or sample selection in the hours-of-work equations. Nonetheless, the results generated are consistent with those of Mroz (1987), who carefully controlled for sample selection and wage endogeneity.

1.8 Conclusion

The Economic Recovery Tax Act of 1981 reduced marginal tax rates by 23 percent within each tax bracket. In addition, ERTA introduced a tax deduction of 10 percent of the secondary earner's income up to $30,000. Together, these changes produced a significant reduction in marginal tax rates for upper-income individuals and a smaller reduction for lower-income individuals. I use the variation in marginal tax rates to estimate both difference-in-difference regression models (where I compare the change in labor supply for upper-income women with change in labor supply for lower-income women) and standard labor supply models (where labor supply is a function of the after-tax wage).

Using data from the 1981 and 1985 Current Population Survey, I find weak evidence that labor force participation of upper-income married women is responsive to taxes. The point estimates suggest that following ERTA, upper-income married women increased their labor force participation by up to 2.6 percentage points (from a predicted base of 47 percent). That estimate suggests an elasticity of 0.79. For working women, the most likely values show a response between 20 and 49 hours per year, but these are estimated with such imprecision that it is not possible to rule out no response at all. Finally, standard labor supply estimates predict participation and hours-of-work responses for upper-income women that are at the lower end of the observed responses. A likely explanation for the divergence between the difference-in-difference results and the standard model results is measurement error in the marginal tax rate and in the gross wage. This measurement error biases the standard estimates downward. The difference-in-difference results, however, would be unbiased if the measurement error averages to zero for the income groups defined.

References

Angrist, Joshua. 1991. Grouped-data estimation and testing in simple labor supply models. *Journal of Econometrics* 47:243–66.

Auten, Gerald, and Robert Carroll. 1994. Taxpayer behavior and the 1986 Tax Reform Act. Washington, D.C.: Department of the Treasury, Office of Tax Analysis. Mimeograph.

Blundell, Richard, Alan Duncan, and Costas Meghir. 1995. Estimating labour supply responses using tax reforms. London: Institute for Fiscal Studies. Mimeograph.

Blundell, Richard, and Costas Meghir. 1986. Selection criteria for a microeconometric model of labour supply. *Journal of Applied Econometrics* 1:55–80.

Boskin, Michael. 1973. The economics of labor supply. In *Income maintenance and labour supply,* ed. Glen Cain and Harold Watts. Chicago: Markam.

Bosworth, Barry, and Gary Burtless. 1992. Effects of tax reform on labor supply, investment, and saving. *Journal of Economic Perspectives* 6(1): 3–26.

Burtless, Gary. 1991. The supply side legacy of the Reagan years: Effects on labor supply. In *The economic legacy of the Reagan years: Euphoria or chaos?* ed. A. Sahu and Ronald Tracy. New York: Praeger.

Eissa, Nada. 1995. Taxation and labor supply of married women: The Tax Reform Act of 1986 as a natural experiment. NBER Working Paper no. 5023. Cambridge, Mass.: National Bureau of Economic Research.

Feenberg, Daniel, and James Poterba. 1993. Income inequality and the incomes of very high income taxpayers. In *Tax policy and the economy,* vol. 7, ed. James Poterba, 145–77. Cambridge: MIT Press.

Feldstein, Martin. 1993. The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act. NBER Working Paper no. 4496. Cambridge, Mass.: National Bureau of Economic Research.

Feldstein, Martin, and Andrew Samwick. 1992. Social security rules and marginal tax rates. *National Tax Journal* 45(1): 1–22.

Hall, Robert. 1973. Wages, income and hours of work in the U.S. labor force. In *Income maintenance and labor supply,* ed. Glen Cain and Harold Watts. Chicago: Markam.

Hausman, Jerry. 1980. The effects of wages, taxes, and fixed costs on women's labor force participation. *Journal of Public Economics* 14:161–94.

———. 1981. Labor supply. In *How taxes affect economic behavior,* ed. Henry Aaron and Joseph Pechman. Washington, D.C.: Brookings Institution.

———. 1985. Taxes and labor supply. In *Handbook of public economics,* vol. 1, ed. Alan Auerbach and Martin Feldstein. Amsterdam: Elsevier.

Hausman, Jerry, and James Poterba. 1987. Household behavior and the Tax Reform Act of 1986. *Journal of Economic Perspectives* 1(1): 101–20.

Hausman, Jerry, and Paul Ruud. 1984. Family labor supply with taxes. *American Economic Review* 74(2): 242–49.

Heckman, James. 1979. Sample selection as a specification error. *Econometrica* 47(6):153–61.

———. 1982. Comment. In *Behavioral simulation methods in tax policy analysis,* ed. Martin Feldstein. Chicago: University of Chicago Press.

———. 1993. What has been learned about labor supply in the past twenty years? *American Economic Review* 83(2): 116–21.

Katz, Lawrence, and Kevin Murphy. 1992. Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics* 107(1): 35–78.

Killingsworth, Mark. 1983. *Labor supply.* Cambridge: Cambridge University Press.

Killingsworth, Mark, and James Heckman. 1986. Female labor supply: A survey. In

Handbook of labor economics, vol. 1, ed. Orley Ashenfelter and Richard Layard. Amsterdam: Elsevier.
Lindsey, Lawrence. 1981. Is the maximum tax on earned income effective? *National Tax Journal* 34(2): 249–55.
———. 1987. Individual taxpayer response to tax cuts, 1982–1984: With implications for the revenue maximizing tax rate. *Journal of Public Economics* 33:173–206.
MaCurdy, Thomas, David Green, and Harry Paarsch. 1990. Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources* 25(3): 415–90.
Mroz, Thomas. 1987. Sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55(4): 765–99.
Murphy, Kevin, and Finis Welch. 1992. The structure of wages. *Quarterly Journal of Economics* 107(1): 285–326.
Navratil, John. 1994. Evidence of individual taxpayer behavior from panel tax return data. Cambridge: Harvard University. Mimeograph.
Pechman, Joseph. 1987. *Federal tax policy.* Washington, D.C.: Brookings Institution.
Rosen, Harvey. 1976. Taxes in a model of joint wage-hours determination. *Econometrica* 44(3): 485–508.
Scholz, John Karl. 1990. The participation rate of the earned income tax credit. Discussion Paper no. 928-90. Madison: University of Wisconsin, Institute for Research on Poverty.
Triest, Ronald. 1990. The effect of income taxation on labor supply in the United States. *Journal of Human Resources* 25(3): 491–516.
Wald, Abraham. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11:284–300.

Comment James J. Heckman

This paper adopts an atheoretical stance toward measuring the effect of taxes on labor supply. It offers a dramatic contrast to the paper by Hausman that I discussed at the 1981 NBER conference held in Florida on measuring the effect of taxes on behavior (see Heckman 1982).

The earlier Hausman paper offered a tightly structured model of taxes and labor supply that exploits all the information in the data and in the theory and adds a lot of econometric structure to produce tax estimates that are not credible. I pointed out that Hausman lacked information about the true budget constraint facing potential workers and his assumptions produced statistically inconsistent estimators even granting the arbitrary distributional and functional form assumptions. My concerns were validated in a paper by MaCurdy, Green, and Paarsch (1990) who found (1) that they could not even reproduce Hausman's estimates using Hausman's methods and Hausman's sample, (2) that ro-

James J. Heckman is professor of economics at the University of Chicago and a research associate of the National Bureau of Economic Research.

The author thanks Jim Poterba for advice on ERTA and for general comments and Ann Ferris for editorial assistance. This work was supported by NSF-93-21-048 and a grant from the Russell Sage Foundation.

bust estimates show essentially zero wage and income effects for male labor supply, and (3) that simpler and more plausible methods of estimation produce estimates of labor supply that agree with the estimates from the complex methods properly applied.

The absurd labor supply estimates produced by the "structural" econometric approach led a whole generation of empirically oriented scholars to reject formal econometric methods and to adopt a series of substitutes for rigorous econometrics. The move toward social experiments, natural experiments, difference-in-difference methods, and Wald estimators represents a yearning for simplicity, familiarity, and robustness in frameworks for conducting empirical work in economics.

In contrast to a more economically explicit style of doing empirical work, the economics in this new empirical methodology is kept implicit, and the discussion of crucial identifying assumptions is also kept implicit. Many people like to keep their econometrics at an intuitive level and to agree, collectively, on what constitutes a "natural experiment" or an "instrument." Given the power of networks in our profession, this agreement to suppress explicit discussion of identifying assumptions and to suppress use of explicit economic measuring frameworks is likely to have a long life.

The new conventions should be recognized as just that: agreements among groups of like-minded persons to keep things simple and intuitively plausible. There is another way to settle these issues, however, and that is to uniformly apply the standards of credibility. Thus while Hausman's framework and empirical evidence is properly dismissed as arbitrary and unconvincing, it should also be noted that the widely used difference-in-difference method is also strongly functional form dependent. It requires additive separability between observed and unobserved variables. It requires that the unobservables have a special time-series structure. It assumes that common trends operate on both treatments and controls, and it rarely identifies parameters of economic interest. Blundell, Duncan, and Meghir (1995), cited by Eissa, demonstrate how very strong functional form assumptions are required to justify the application of difference-in-difference methods to estimate economically interpretable parameters. I amplify this point below. The available experimental evidence speaks strongly against the difference-in-difference method. LaLonde's widely acclaimed study (1986), which contributed to the distrust of econometric methods and the call for experiments, documented that the method gave very poor estimates. Heckman and Smith (1995) report similar evidence. The only thing going for the method and the closely related fixed-effects strategy is computational convenience.

The economic parameter being estimated is never defined in terms of conventional income and substitution effects measured in other studies. Thus it is difficult to compare Eissa's estimates with those from other studies, even other difference-in-difference studies, since the method is so strongly dependent on particular sample paths for conditioning variables. The 1981 tax reform had the effect of raising the after-tax wages of women from families with high income compared to those from families with low income. It also had the effect of raising the after-tax income of their husbands. It reduced the marginal tax rate on capital income from a top rate of 70 percent to 50 percent and changed other rates below the top as well. The net effect of the reforms on labor supply is ambiguous because the after-tax wage of women was higher, encouraging an increase in their labor supply, but the after-tax income of their husbands was higher, encouraging a reduction in female labor supply. The labor supply response estimated by Eissa is neither a compensated nor an uncompensated effect of taxes on female labor supply as conventionally defined because different tax changes apply to wage and capital gains income. Therefore, the author cannot reasonably compare her estimates to those from the previous literature that identified those effects.

The estimation strategy adopted in this paper relies critically on the classification of women into "high" and "low" cells based on pretax household income, excluding the wife's earnings but including other joint capital income, and assumes that taxes do not affect membership in this classification—the conditions required for application of the Wald estimator. Even if male labor supply has a zero wage elasticity, as the author assumes and as MaCurdy et al. effectively demonstrate, capital income is well known to have a high tax elasticity. If some women change categories as a result of the reduced tax on asset income, the effect is to violate the fixed grouping assumption of the Wald estimator. Presumably the net shift is that some "low" women become "high" women as their families adjust capital incomes.

Bias in the Estimator

To be more precise about the nature of the bias resulting from this violation, consider the following table for log hours of work:

	Low Income	High Income
New tax	C	A
Old tax	D	B

Let P_T be the proportion of women who move from D to A in response to the effect of the tax change on capital gains and other income. Assume no shifting across other cells. Let E be log labor supply. $E^0(\mathrm{l}, \mathrm{T})$ is the log of hours worked under the old tax regime by low-income people who will transfer to the high-income group. $E^0(\mathrm{l}, \tilde{\mathrm{T}})$ is the log hours worked in the old tax regime by low-income people who will stay in the low regime. $E^n(\mathrm{l}, \tilde{\mathrm{T}})$ is the labor supply in the new regime for those initially low-income people who do not switch status. Let $E^0(\mathrm{h})$ be the log hours worked by high-income people in the old tax regime. Let $E^n(\mathrm{h})$ be the log hours worked in the new regime.

Eissa defines her parameter of interest to be

$$[(A - B) - (C - D)]/[\Delta\ln(1 - t)(\mathrm{h}) - \Delta\ln(1 - t)(\mathrm{l})]$$

where A, B, C, and D stand for mean of log hours worked in each cell. Define

$$\Delta \ln (1 - t)(\mathrm{h}) - \Delta \ln (1 - t)(\mathrm{l}) = \Delta t$$

to be the change (in logs) of the after-tax share of wages between high income women and low income women.

$$A = E^{\mathrm{n}}(\mathrm{h}), \qquad B = E^{0}(\mathrm{h})$$

$$C = P_{\mathrm{T}}E^{\mathrm{n}}(\mathrm{l}, \mathrm{T}) + (1 - P_{\mathrm{T}})E^{\mathrm{n}}(\mathrm{l}, \tilde{\mathrm{T}})$$

$$D = P_{\mathrm{T}}E^{0}(\mathrm{l}, \mathrm{T}) + (1 - P_{\mathrm{T}})E^{0}(\mathrm{l}, \tilde{\mathrm{T}}).$$

For simplicity I use geometric means. What Eissa actually estimates is

$$[(A^* - B) - (C^* - D)]/\Delta t,$$

where, letting N_{h} be the number of people in the high-income regime and N_{l} the number in the low-income regime, both measured in the base state,

$$A^* = wE^{\mathrm{n}}(\mathrm{h}) + (1 - w)E^{\mathrm{n}}(\mathrm{l}, \mathrm{T}),$$

$$C^* = E^{\mathrm{n}}(\mathrm{l}, \tilde{\mathrm{T}}),$$

and

$$w = \frac{N_{\mathrm{h}}}{N_{\mathrm{h}} + P_{\mathrm{T}}N_{\mathrm{l}}}.$$

Then the bias for her parameter is

$$\{[(A^* - B) - (C^* - D)] - [(A - B) - (C - D)]\}/\Delta t$$
$$= [(A^* - A) - (C^* - C)]/\Delta t$$
$$= [(1 - w)\,(E^{\mathrm{n}}(\mathrm{l}, \mathrm{T}) - E^{\mathrm{n}}(\mathrm{h})) - P_{\mathrm{T}}(E^{\mathrm{n}}(\mathrm{l}, \tilde{\mathrm{T}}) - E^{\mathrm{n}}(\mathrm{l}, \mathrm{T}))]/\Delta t.$$

It is plausible that income effects on labor supply yield the ordering $E^{\mathrm{n}}(1, \mathrm{T}) - E^{\mathrm{n}}(\mathrm{h}) > 0$, but it is also plausible that $E^{\mathrm{n}}(1, \tilde{\mathrm{T}}) - E^{\mathrm{n}}(1, \mathrm{T}) > 0$. Thus the direction of the bias for Eissa's parameter depends on the disparity in the new tax situation between the mean log labor supply of the transferees relative to those who stay in the high and low cells. The women shifting into cell A raise the labor supply there but by leaving cell C they raise the average there as well. The larger the transferees are as a proportion of women in the postreform high cell, and the larger the gap is between transferee labor supply and the labor supply of the initially high-income women, the more likely it is that her estimate is upward biased. The smaller the transferees are as a proportion of the pretax low-income households, and the farther apart the new tax regime labor supply of transferees and nontransferees is, the more likely it is that her estimate is downward biased for her parameter.

Dependence on Functional Form and Assumptions about Time Paths of Regressors

The extreme dependence of the difference-in-difference estimator on the functional form of the labor supply equation and implicit assumptions about movement of the exogenous variables over time between the high and low groups can also be exhibited within this framework. To discuss this issue, I ignore the crossover problem just discussed.

Write $\Delta E(\mathrm{h})$ as the change in log labor supply between the old and new tax regime for persons in the high group. $\Delta E(\mathrm{l})$ is the change in log labor supply for the low group. As before, $\Delta \ln (1 - t)(\mathrm{h})$ is the change in log marginal tax rates for the high group, and $\Delta \ln (1 - t)(\mathrm{l})$ is the change in the log marginal tax rate for the low group. Let $\Delta \ln X(\mathrm{h})$ be the change in other characteristics for the high group, and let $\Delta \ln X(\mathrm{l})$ be the change in the other characteristics for the low group. Let $\Delta U(\mathrm{h})$ be the change in the unobservables for the high group, and $\Delta U(\mathrm{l})$ be the change in the unobservables for the low group.

In finite changes,

$$\Delta E(\mathrm{h}) = \alpha_1(\mathrm{h})\Delta \ln (1 - t)(\mathrm{h}) + \alpha_2(\mathrm{h})\Delta \ln X(\mathrm{h}) + \Delta U(\mathrm{h})$$

$$\Delta E(\mathrm{l}) = \alpha_1(\mathrm{l})\Delta \ln (1 - t)(\mathrm{l}) + \alpha_2(\mathrm{l})\Delta \ln X(\mathrm{l}) + \Delta U(\mathrm{l})\,.$$

Eissa's estimator of wage response is the difference of the average of the changes within each group:

$$\frac{\Delta \overline{E(\mathrm{h})} - \Delta \overline{E(\mathrm{l})}}{\Delta \ln (1 - t)(\mathrm{h}) - \Delta \ln (1 - t)(\mathrm{l})} = \frac{\Delta \overline{E(\mathrm{h})} - \Delta \overline{E(\mathrm{l})}}{\Delta t},$$

where the overbar denotes average. She implicitly assumes that $\alpha_1(\mathrm{h}) - \alpha_1(\mathrm{l})$ (no wealth effect on the response of a change in taxes on labor supply), $\alpha_2(\mathrm{h}) = \alpha_2(\mathrm{l})$ (other variables have the same marginal effect on log labor supply at different wage levels), and $\Delta \overline{\ln X}(\mathrm{h}) = \Delta \overline{\ln X}(\mathrm{l})$ (the other characteristics, such as child bearing, age, wages, etc. change in the same way in logs between the groups) and

$$\operatorname*{plim}_{N \to \infty} [\bar{\Delta} U(\mathrm{h}) - \bar{\Delta} U(\mathrm{l})] = 0,$$

so that sample differences in the changes in unobservables between high and low converge to zero. Thus she implicitly makes strong functional form assumptions as well as assumptions about the time profiles in logs of the explanatory variables $X(\mathrm{h})$ and $X(\mathrm{l})$ in the two groups. She ignores the effects of taxes on the entry and exit of persons into the workforce and the effects of these compositional changes on estimated labor supply parameters—a major theme of the literature on selection bias surveyed in Killingsworth's survey of labor supply.

By adopting an atheoretical approach, the author throws away a potentially important source of information for identifying wage and tax effects on labor supply: the demand-induced change in real wages that highly educated women experienced in her sample period. The only advantages in not using the wage information are that she can assume that her women suffer from tax illusion and she can avoid standard measurement error and simultaneous equation problems in the use of wages in labor supply equations. However, all the evidence in the literature argues that the after-tax wage is relevant to labor supply decisions. Wage and tax effects in logs should have the same effects on labor supply. Averaging as she does should greatly attenuate mean zero measurement error and simultaneous equations problems. Thus it is not clear that she gains anything by not using the wage data.

Instead of using wage growth to help identify tax effects on labor supply, she adopts an ad hoc method for eliminating the effects of demand growth on wages and hence on labor supply. She throws wages into the X(h) and X(l) variables and therefore is forced to take steps to undo the consequences of the false assumption that high-income women have had the same wage growth as low-income women. Positive assortative mating on education coupled with the greater trend in wage growth for more educated women argues strongly against such an assumption. Part of her estimated tax effect is due to differential wage growth, but it would be better to constrain the tax coefficients and the wage coefficients to allow the variation in the wage growth to inform the estimation of tax effects.

The pendulum has swung too far away from using economics as a means of interpreting economic data. Wage variation should be used as a source of identifying information and not as a problem to be eliminated, especially when a substantial component of the growth is demonstrably exogenous to individual decisions, as was the wage growth of 1980s. Trends and demand shocks would provide "natural instruments" for wages.

Finally, the comparison of Eissa's estimate with those obtained from her interpretation of conventional econometric models of labor supply is not convincing. Econometrics has advanced beyond the simple probit model, and discrete-choice models that allow for more general forms of nonlinearity are now widely available.[1] Even the probit model could be made more flexible by incorporating nonlinearities in the arguments of the model. It is not an essential feature of probit analysis to constrain the estimates to be linear functions within the probit argument or to make tax or wage effects uniform across income levels. The contrast between the constrained probit estimates and her estimates is thus somewhat contrived.

Many methods are available for handling the problem of measurement error in wages. The labor supply equation estimated in this paper is much more like the "first generation" studies (as labeled by Killingsworth) than like the second-generation studies which consider simultaneity, selectivity, and measurement error. Her specification does not distinguish between self-selection effects of wages and the effects of wages on labor supply. Models and methods

1. See Todd (1995), who demonstrates that in many cases the models of nonparametric discrete choice make little difference in a wide array of applications. Her results suggest that probit models are not restrictive.

for estimating labor supply functions under more robust conclusions are available. It would be of interest to estimate these models before the strategy of estimating interpretable economic models is rejected out of hand.

Summary

Difference-in-difference methods are functional form dependent and produce interpretable economic parameters only under very special conditions. Difference-in-difference estimates are not even comparable across studies of the same type because policy-invariant structural parameters are not estimated and different studies condition on different variables and different levels of the variables. In this application, the estimates produce a tax effect on labor supply that has no clear economic interpretation and that cannot be compared with estimates from the structural labor supply literature. Moreover, the classification scheme of "high" and "low" family incomes is likely to be affected by taxes, violating a key assumption of the Wald estimation method used in this paper.

The reaction to the implausible labor supply models of the early 1980s has gone too far. It is time to bring economics back to the study of labor supply and taxes. Credible methods exist to estimate economically interpretable labor supply parameters that can be compared across studies and that can be used to address problems of economic interest.

References

Blundell, Richard, Alan Duncan, and Costas Meghir. 1995. Estimating labour supply responses using tax reforms. London: Institute for Fiscal Studies. Mimeograph.

Heckman, James. 1982. Comment. In *Behavioral simulation methods in tax policy analysis,* ed. Martin Feldstein. Chicago: University of Chicago Press.

Heckman, James, and Jeffrey Smith. 1995. Ashenfelter's dip and the determinants of program participation. Chicago: University of Chicago, March. Manuscript.

LaLonde, Robert. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76(4): 604–20.

MaCurdy, Thomas, David Green, and Harry Paarsch. 1990. Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources* 25(3): 415–90.

Todd, Petra. 1995. Estimation of binary choice models under choice-based sampling by semiparametric least squares. Chicago: University of Chicago. Manuscript.