Who Pays the Social Security Tax?*

Haizheng Li  
School of Economics  
Georgia Institute of Technology  
Atlanta, GA 30332-0615  
Phone: (404) 894-3542  
Fax: (404) 894-1890  
Email: haizheng.li@econ.gatech.edu

*I would like to thank Janusz Mrozek and Lynn Yang for helpful discussions.
Abstract

This study applies a difference-in-difference approach to estimate the incidence of the social security tax using panel data from the Panel Study of Income Dynamics (PSID). The existence of the maximum income subject to the payroll tax and substantial exogenous increases of this threshold in some years creates an opportunity to identify the tax incidence. Specifically, an individual with an income above the threshold faces a marginal tax rate of zero, whereas one below the maximum faces a marginal tax rate equivalent to the average tax rate. Therefore, the effect of the payroll tax on wages is different for those above or below the tax threshold. An exogenous and large increase in the threshold, like a natural experiment, moves many individuals from above the threshold to below the threshold. Those individuals form a treatment group. By applying a difference-in-difference procedure based on alternative control groups, the tax effect on wages can be isolated and the tax incidence can be identified. The results show a full shifting back of the tax burden; i.e., based on the sample used, workers bear the full burden of the social security tax. This finding is robust under different assumptions, and the hypothesis of a complete shifting of the tax burden back to employees cannot be rejected.

Key Words: Incidence; Social security tax; Difference-in-difference

J. E. L. code: H22, H55
I. Introduction

This study applies a new approach to identify the incidence of the social security (OASDHI) tax in the United States. Currently a 7.65% social security payroll tax rate applies to both employees and employers. The incidence of the tax--i.e., how the burden of the tax is actually shared between employers and employees--is unobservable, however, and, based on the Liability Side Equivalence principle, independent of the way in which the tax is collected.\(^1\) Theoretically, tax incidence is determined by the relative elasticities of labor demand and supply.

Yet for the social security tax, as a benefit tax, the incidence is also affected by the value employees place on future social security benefits (Summers, 1989).

The tax incidence has important implications concerning the reform of social security. At the heart of the debate is whether the current pay-as-you-go defined benefit system should be switched to a defined contribution private (or partially private) system (see Gramlich 1996 and Diamond 1996 for further discussions). The incidence of the social security tax will indicate the potential gains in efficiency from such a change. More specifically, if the tax burden cannot be fully shifted back to employees, labor costs will increase, total employment will decrease and market distortions will occur. In this case, the establishment of a private system with defined contributions, such as individual accounts, will at least strengthen the linkage between the payroll tax and future benefits, and will thus make it easier for employers to pass on the tax burden to employees; consequently, the efficiency of the social security will be improved. If employees already bear the full tax incidence under the current pay-as-you-go system, then there will be no efficiency loss, and the efficiency issue will be less of a concern in future reforms.\(^2\)

---

\(^1\) The Liability Side Equivalence principle has been recently tested in the laboratory by Kerschbamer and Kirchsteiger (2000). They find that subjects who actually have to pay the tax carry a higher tax burden.

\(^2\) A defined contribution private system is at least as efficient as the defined benefit pay-as-you-go system; see discussion in Summers (1989).
Furthermore, the payroll taxes have become an increasingly important component of the tax burden for many low- and middle-income families. Based on Mitrusi and Poterba (2000), nearly two thirds of families in 1999 paid more in payroll taxes than they did in federal income taxes, while this share was 42 percent in 1979. Therefore, the extent to which the payroll taxes are ultimately borne by the families, i.e., the incidence, will also have important implications concerning income distribution.

Identifying the incidence of the social security tax is by no means an easy task, however. There have been no major studies of social security tax incidence since the early 1980s. Furthermore, these studies have been sharply criticized for failing to identify the true incidence parameter. Approaches based on aggregate data, such as Perry (1970) and Vroman (1974), have been criticized for a lack of theoretical grounding and for insufficient variation in payroll tax rates (Hamermesh, 1979). A different strand of approaches based on estimating production functions, for example, Brittain (1971), has been criticized for misinterpreting the estimated parameter as measuring tax shifting (Feldstein, 1972).

Hamermesh (1979) used individual-level data to estimate the tax incidence and found that at most only a small percentage of the payroll tax is ever shifted back to employees; this finding contradicted the assumption held by many economists that the payroll tax is shifted back to primary workers. Asher (1984) sharply criticized the findings and demonstrated that the incidence estimated by Hamermesh derives in fact from a spurious relationship and is not the true tax-shifting parameter. In addition, Hamermesh only included the portion of the tax paid by employers, and thus the incidence of the employee’s portion is unclear.

In general, previous studies attempted to estimate the tax-shifting parameter directly. A direct estimation of the tax incidence, however, suffers from a number of problems. More
specifically, because of the existence of the maximum taxable income, the social security tax on wages affects those with an income above the threshold differently than it affects those below. Therefore, if individuals at different wage levels are pooled together to estimate the tax incidence, as in Hamermesh (1979), it will result in model mis-specification. More importantly, in general, the maximum taxable income is adjusted following an increase in average wage levels; thus it will be correlated with unobserved annual effects that affect wages. Such an omitted variable bias is generally difficult to fix. As discussed in Asher (1984), cross-section data on individuals cannot provide usable results on tax incidence. In fact, even with panel data, a direct estimation of tax incidence is doomed, as the next section will make clear.

This study develops an indirect procedure, a difference-in-difference estimation (DD), to identify the incidence of the social security tax based on a "natural experiment" approach. “Natural experiment” approach has been widely used in studying policy effects. For example, Gruber (1994) applies the approach to estimate the incidence of mandated maternity benefits in the U.S., based on the fact that different states had different laws regarding such benefits. Gruber (1997) estimates the incidence of the Chilean social security tax using the exogenous changes in Chile-the privatization of its social security system in the early 80s. Anderson and Meyer (2000) investigate the incidence of the unemployment insurance payroll tax, and Lee (2000) estimates the impact of taxing unemployment insurance benefits, both using natural experiments.

The DD estimation procedure in this study is based on the existence of the maximum income subject to the social security payroll tax and on exogenous increases of this threshold. Specifically, individuals with incomes above the threshold face a marginal tax rate of zero, whereas those below the maximum face a marginal tax rate equivalent to the average tax rate.
Therefore, the effect of the payroll tax on wages is different for those above or below the tax threshold. An exogenous and large increase in the threshold moves many individuals from above the threshold to below the threshold. Like a natural experiment, such a large exogenous change creates an opportunity to identify the tax incidence.

More specifically, this study uses panel data from the Panel Study of Income Dynamics (PSID) for the years 1978-80. From 1978 to 1979, the tax limit rose about 30%, the largest increase in the history of the OASDHI tax limit; many individuals who were above the threshold in 1978 ended up below the threshold in 1979. As a result, those who had faced a zero marginal tax rate in 1978 faced a new marginal tax rate of 6.13% (the payroll tax rate) in 1979. Since this increase was set by statute and was not an automatic adjustment following an increase in the average wage level, it should be an exogenous shock to the labor market. Therefore, by applying an iterative difference-in-difference procedure, the tax incidence can be identified.

This procedure uses those individuals with an income above the tax threshold in 1978 but below the threshold in 1979 as a treatment group, and those below the tax threshold in both years as a control group. Therefore, the effect of the payroll tax on wages was different for the control group and the treatment group in 1978, but identical in 1979. By comparing the difference between these two groups in these two years, the tax effect and the incidence can be identified. Furthermore, to check the robustness of the estimates, alternative control groups are used to form a difference-in-difference-in-difference estimation. These controls include the years following the large change in the tax threshold in 1978--i.e., the years 1979 and 1980--as well as government employees who were not covered by the social security system at that time. In all estimations, the results point to a complete shifting of the tax: the burden of the social security tax is shifted back to employees; and the hypothesis of a full shifting cannot be rejected.
The rest of the paper is organized as follows: section II provides a theoretical framework for the payroll tax incidence and for the DD estimation; section III discusses the DD estimation procedure and the results; section IV uses alternative assumptions and a difference-in-difference-in-difference procedure to estimate the tax incidence; section V presents further discussion and conclusions.

II. Theoretical Models

As a benefit tax, the social security tax is different from other general taxes. In particular, besides paying the tax, workers also expect to receive direct future benefits. There are five major categories of benefits paid for through social security taxes: retirement, disability, family, survivor and Medicare. To receive these benefits, workers earn credits that count toward eligibility by paying social security taxes. The tax is equally divided between employees and employers. The linkage between the payroll tax liability and expected social security benefits, however, is very complicated. Feldstein and Samwick (1992) examine this “tax benefit linkage” and present estimates of the distribution of marginal benefit paybacks associated with the social security payroll tax.³ To simplify the analysis on the benefits side, this study follows Summers (1989) and incorporates workers valuations of future social security benefits into their labor supply decisions.

³ For the redistribution under the current social security benefit formula, see a recent study by Gustman and Steinmeier (2001).
labor supply curve will also shift and the labor supply will increase. Therefore, in addition to the relative elasticity of labor demand and supply, the incidence of the social security tax also depends on the value employees assign to the expected benefits.\footnote{It is also possible that employers can shift the tax burden forward to consumers by price hiking. The effect will be reflected by the shift of the labor demand curve through the profit maximization mechanism.} If workers fully value the benefits, the cost of providing social security shifts completely to wages.

In deriving an empirical model for the payroll tax incidence, this study includes the payroll taxes paid by both employees and employers, and allows different specifications for those above and below the tax threshold. Figure 1 shows a general case for payroll tax incidence. Because of the social security tax, labor demand decreases and the demand curve shifts to $D'$, a standard effect for any tax. Furthermore, as discussed above, the social security tax is a benefit tax and workers are expected to receive direct benefits in the future. If employees value such benefits, they will be willing to accept a lower wage for the same amount of work. Thus, the labor supply increases and the supply curve shifts to $S'$. As a result, the market equilibrium moves from $(H, W_O)$ to $(H', W_K)$.

Suppose that $t_f$ and $t_e$ represent, respectively, the employer and worker social security tax rates on earnings below the maximum taxable annual salary, the “threshold,” S. $H$ is the observed annual hours worked. $W$ is the observed taxable hourly wage, and $W_O$ is the hourly wage that would prevail in the absence of the payroll tax.\footnote{In order to focus on the payroll tax, the federal income tax is not included.} $W_G$ is the employer’s hourly cost for an employee when the payroll tax is present, and $W_K$ is the after-payroll-tax “take home” wage.

Assume $\tau$ is the incidence of the tax on the employee: the proportion of the tax burden that falls on the employee. When $\tau = 0$, i.e., when the after-tax wage rate $W_K$ equals the no-tax
wage rate $W_o$, the take-home wage does not change when the tax is implemented. In this case, employers bear the entire payroll tax. Similarly, when $\tau = 1$, employees actually pay all the tax.\footnote{If the shift of labor supply is greater than the shift of demand, $\tau$ can be greater than 1. This will happen if workers value the social security benefits more than the costs they pay. This is possible if private markets fail to provide these types of insurance. For example, in the absence of a private real annuity market, individuals will be willing to pay more than one dollar for one dollar of social security benefits (Gruber, 1997). Clearly, such a situation is unlikely in the United States.}

**Figure 1. The Incidence of Social Security Tax**

The relationships between the unobserved no-tax hourly wage, the observed wage, and the tax incidence depend on whether the annual wage is above or below the salary threshold $S$. Thus, the two cases are treated separately.

**Case 1: Annual Wage below the Tax Threshold, $W < S / H$**

The (combined employer and employee) amount of social security taxes paid per hour of employment is $(W_G - W_K)$, which equals $(t_r + t_e)W$, where $W$ is the observed "declared" wage for
workers (since taxes are calculated based on the observed “declared” wage, and not the take-home wage). The incidence of the tax \( \tau \) is the fraction of the tax borne by the employee:

\[
\tau = \frac{(W_O - W_K)}{((t_f + t_e) \cdot W)}.
\]

For this case \( W_K = (1-t_e) \cdot W \) and \( W_G = (1+t_e) \cdot W \). Rearranging and substituting for \( W_K \):

\[
W = \frac{W_O}{(1 + \tau \cdot t_f - (1-\tau) \cdot t_e)}
\]

For small values of \((\tau \cdot t_f - (1-\tau) \cdot t_e)\), the following mathematical approximation holds:

\[
\ln(1+\tau \cdot t_f - (1-\tau) \cdot t_e) = \tau \cdot t_f - (1-\tau) \cdot t_e.
\]

Taking the log of \( W \) and substituting in the approximation:

\[
\ln W = \ln W_O - \tau \cdot t_f + (1-\tau) \cdot t_e.
\]

Clearly, the observed wage \( W \) is affected by the tax. Without the tax, the actual wage would be \( W_o \). While \( W_o \) is not directly observable, it is determined by the market equilibrium and the individual’s human capital. Assuming the usual specification for wage without tax, \( \ln W_o = X'\beta + \varepsilon \), where \( X \) is determined by the human capital theory and \( \varepsilon \) is the error term, we get the empirical equation:

\[
\ln W = X'\beta - \tau \cdot t_f + (1-\tau) \cdot t_e + \varepsilon,
\]

If \( t_f = t_e = t \), as holds for the current payroll tax in the U.S., the above equation becomes:

\[
\ln W = X'\beta + (1-2\tau) \cdot t + \varepsilon.
\]

Thus, the observed wage consists of two deterministic elements in addition to the random component. The first element is represented as a vector of individual and job characteristics \( X \). The second element is determined by the tax incidence.

---

7 This approximation simplifies notation. In calculating the tax incidence in next sections, the exact formula is used.
In this specification, \( \tau \) cannot be identified with cross-section data if the tax rates \( t_f \) and \( t_e \) are constant. Thus, in cross-sectional data consisting solely of observations for people below the income threshold, incidence cannot be estimated. In panel data, \( t_f \) and \( t_e \) may vary over time, as the tax rate in the U.S. has increased a number of times. However, since the payroll tax rate is identical for all individuals in any given year, the tax rate will be perfectly collinear with an annual fixed effects. Therefore, the estimated tax effect will incorporate other yearly fixed effects that influence wages; the true tax incidence will thus not be identified. In other words, neither cross-section data nor longitudinal data can be used to estimate the tax incidence if the sample consists solely of individuals with an annual income below the maximum taxable income.

**Case 2: Above the Tax Threshold, \( W > S / H \)**

When annual wages are greater than the taxable income threshold \( S \), the amount of social security taxes paid per hour of employment is: \( (t_f + t_e) \cdot S/H \). Then, the incidence \( \tau \) becomes:

\[
\tau = \frac{(W_O - W_K)}{((t_f + t_e) \cdot S/H)}.
\]

For this case \( W_K = W - t_e \cdot S/H \) and \( W_G = W + t_f \cdot S/H \), and then

\[
W = W_O + [(1-\tau) \cdot t_e - \tau \cdot t_f] \cdot (S/H).
\]

When \( t_f = t_e = t \),

\[
W = W_O + (1-2\tau) \cdot t \cdot (S/H).
\]

Assuming \( W_O = X'\beta + \epsilon \), the wage equation in non-logarithmic form becomes

\[
W_O = X'\beta + (1-2\tau) \cdot t \cdot (S/H) + \epsilon,^8
\]

---

^8 This model is different from the model for those below the tax threshold because of the logarithm.
Although it appears that this equation can be used to estimate the tax incidence directly if individuals’ work hours are different, there are in fact potential problems. In particular, the tax limit $S$ generally increases in proportion to increases in average wage level based on an automatic adjustment under Act 1972a, and thus is likely to be correlated with unobserved factors that affect the wage level. Furthermore, it is well known in labor supply literature that working hours may be endogenous, i.e., some unobserved variables may be correlated with both wages and hours worked; a direct estimation of tax incidence will therefore be inconsistent. Although an Instrumental Variable (IV) estimation may be used, it is generally difficult to find instruments that are correlated with both the tax limit $S$ and working hours, but uncorrelated with other unobserved factors.

Therefore, in either case, above or below the tax threshold, there are problems for a direct estimation of the payroll tax incidence using individual level data. Moreover, the models for these two cases are clearly different, and estimations that pool two groups together, such as Hamermesh (1979), will result in model mis-specification. In the following sections, a new indirect method based on difference-in-difference estimation will be discussed.

### III. An Iterative Difference-in-Difference Estimation

The difference-in-difference (DD) estimation in this study relies on panel data. It utilizes the fact that if an individual moves from above the threshold to below the threshold, the marginal tax rate changes from zero to the average tax rate, and then the corresponding wage equation will
change. If this change is exogenous--i.e., the movement across the threshold is exogenous to the individual’s wage--then the change can be viewed as a “natural experiment.”

The maximum taxable earnings of the OASDHI tax have been under constant annual adjustment. Under Act 1972a (as modified by Acts 1973a and 1973b), the change was based on an automatic adjustment in proportion to increases in average wage levels. For a few years, however, the change was set by statute and was not related to average wage increases. In particular, from 1978 to 1979, the threshold increased from $17,700 to $22,900, an increase of 29.4% (Social Security Bulletin, 1999), which was much higher than the average wage increase. This was the largest percentage increase in the tax threshold in the history of the social security tax after the inclusion of Hospital Insurance in the tax. Because this increase was not related to average wage increases, it was exogenous to the labor market.

Therefore, the years 1978 and 1979 are selected for the DD estimation using data from the PSID. The sample is restricted to white male heads of household. This selection is to avoid possible differences in tax incidence between different racial groups, and between men and women. In addition, the PSID data provide the most comprehensive information on household heads. The sample is limited to those with earnings below the social security tax threshold in 1979. Specifically, the “control group” includes individuals with an income below the tax limit in both 1978 and 1979; and the “treatment group” includes individuals with an income above the tax limit in 1978 but below the tax limit in 1979. In the sample, there are 151 individuals in the

---

9 The marginal tax rate here refers to the nominal marginal tax rate, not the effective marginal tax rate that accounts for future social security benefits.

10 From 1965 to 1966, the tax threshold increased 37.5%. This is because the Hospital Insurance (HI) payroll tax was added in 1966. The PSID data were not available prior to 1968. Another large increase in the tax threshold was the 73-74 increase, one of about 22.2%.
“treatment group” and 479 in the “control group”. The descriptive statistics for wages and work hours are provided in Table 1.

For the individuals in the treatment group, the marginal tax rate was zero in 1978 but became 6.13% in 1979. Applying the theory discussed in the previous section, the observed wage for individuals i at year j is $W_{ij}$, the wage in the absence of the payroll tax would be $W_{oij}$, and the payroll tax rate is $t_j$ for both employees and employers. For individuals below the tax threshold, the wage equation is

$$\ln W_{ij} - t_j(1-2\tau) = \ln W_{oij}. \quad (2)$$

For individuals above the threshold, the wage equation becomes

$$W_{ij} = W_{oij} + (1-2\tau) t_j (S_j/H_{ij}).$$

In order to use the same specification, the wage equation for those above the tax limit can be rearranged as follows:

$$\ln [W_{ij} - (1-2\tau) t_j (S_j/H_{ij})] = \ln W_{oij}. \quad (3)$$

This framework provides a uniform specification of wages under the influence of the payroll tax. We can follow human capital theory (such as Mincer, 1974), and assume that an individual’s wage in the absence of the payroll tax would be $\ln W_{oij} = X_{ij}'\beta + \varepsilon_{ij}$, where $X$ includes variables such as education and job experience, and an empirical model can be specified.

The DD procedure proceeds in the following iterative way. First, for 1978, the wage equation for the treatment group is given by equation (3), while for 1979, it is given by equation (2). The wage equation for the control group is determined by equation (2) in both years. Since in general the incidence parameter is in the interval $[0,1]$, thus we can guess a value for $\tau$ for equation (3). With a known tax incidence (by guessing), in 1978 the difference in wages
(logarithm) between the treatment group and the control group consists of the tax effect $t_j(1-2\tau)$ and “other unobserved differences” unrelated to the payroll tax. In contrast, in 1979, the difference in wages between these two groups is not affected by the payroll tax (it is cancelled out). If we assume that the unobserved differences unrelated to the payroll tax between these two groups does not change during these two years, then such differences can be identified in 1979. After differencing it out, we can identify the tax effect $t_j(1-2\tau)$ using data from 1978, and hence get a new $\tau$.

Second, the estimated $\tau$ is clearly based on the guessed $\tau$ for the equation (3). Only if the guessed $\tau$ equals the true $\tau$, asymptotically, will it be identical to the estimated $\tau$. Therefore, the DD procedure will continue until the guessed $\tau$ and the estimated $\tau$ converge (in the sense that the difference is the smallest). Since $\tau$ is within the range of $[0,1]$, a grid search for $\tau$ is feasible. Moreover, by guessing a $\tau$, the possible endogeneity problem caused by the tax threshold and working hours $H_{ij}$ in equation (3) also disappears.

The DD procedure depends on the assumption that the difference between the treatment group and the control group remains constant during these two years in the absence of the payroll tax. This sort of assumption is common for any DD estimation or natural experiment approach, and the assumption seems to be quite reasonable in this case. Moreover, since wages are specified in log form (i.e., the difference in the logarithm is approximately a percentage change), the effect caused by the difference in absolute changes in wages associated with wage levels is eliminated. For example, wages may increase more in absolute amounts for those at higher wage levels, but such a difference does not exist in the percentage sense (if the percentage change does not depend on the levels).
The empirical model for the DD estimation is as follows:\textsuperscript{11}
\[ \ln W_{ij} = X_{ij}'\beta + \gamma_1 \cdot \text{group} + \gamma_2 \cdot \text{time} + \delta \cdot \text{group} \cdot \text{time} + \varepsilon_{ij}, \]
where the variable group is a dummy variable taking value 1 for the control group and 0 for the treatment group. The variable time is a year dummy, equaling 1 for 1978 and 0 for 1979. \( X_{ij} \) includes other explanatory variables affecting wages such as schooling, job experience, and the square of job experience.\textsuperscript{12} To control for possible differences in returns to schooling and job experience between the two groups and in different years, terms interacting education with group and time, and experience with group and time are also included.

The interaction term between group and time captures the tax effect \( t_j(1-2\tau) \).\textsuperscript{13} More specifically, after controlling for the time invariant group difference and the group invariant annual effect unrelated to the tax change, the group and time specific effect is caused only by the tax change. The regression proceeds first with the pooled ordinary least squares (OLS). The grid search on \( \tau \) starts at 0 with an incremental of 0.1.\textsuperscript{14}

The DD estimation using pooled OLS converges at \( \tau = 1 \) (when the difference between the guessed \( \tau \) and estimated \( \tau \) is the smallest). At convergence, the estimated tax effect is \(-0.149\) and is very significant. The calculated \( \tau \) is 1.83. Obviously, the distance between the guessed \( \tau \)

\textsuperscript{11} It is possible to compare the difference in the logarithms of wage without using regression. In a regression framework, however, other variables that affect the outcome of the parameter of interest can be controlled, and thus can reduce sample errors.
\textsuperscript{12} The specification follows a very basic Mincerian model. Other control variables are not included because of the concern of degrees of freedom given the limited sample size. Since a fixed effects model will be used, time invariant variables will be compressed into fixed effects. In addition, the DD estimation results below are not sensitive to model specification.
\textsuperscript{13} As discussed in the previous section, this tax effect is an approximation of \(-\ln(1+(2\tau-1)t_j)\). This approximation is reasonably close when \((2\tau-1)t_j\) is small. When calculating \( \tau \), in order to reduce error, the exact formula instead of the approximation is used.
\textsuperscript{14} The increase in \( \tau \) can be made even smaller in each iteration. It does not change the results, though. The software used in the estimation is SAS. I thank technical support from SAS, especially Michelle Schlude, on the programming code.
and the estimated $\tau$ is still quite large at the convergence.\footnote{15} If the estimated $\tau$ is 1, the estimated tax effect should be $-0.059$. The null hypothesis of $\tau=1$ is rejected in favor of $\tau>1$. The result of the OLS regression at the convergence is given in Table 2. Most other regressors have the expected sign.

Furthermore, the fixed effects model for panel data is also estimated. The fixed effects model will mitigate the omitted variable bias caused by the time invariant heterogeneity in individuals’ wages. Because the data are for only two years, the fixed effects model is estimated using first differencing method.\footnote{16} The iteration again converges at $\tau=1$, and the resulting tax effect is $-0.151$ and is again very significant. The calculated tax incidence $\tau=1.85$. As in the pooled OLS estimation, we reject the null hypothesis of $\tau=1$ in favor of $\tau>1$. Other results are presented in Table 1.\footnote{17}

What we have found from the DD estimation is that $\tau$ converges at 1, and we cannot reject that $\tau$ is greater than 1 in both the OLS estimation and the fixed effects model. Therefore, it shows a full shifting of the payroll tax toward workers. The relatively large estimated $\tau$, however, may indicate that the DD procedure isolates an effect larger than the true tax effect (in absolute value). If this is true, such an effect should be removed before calculating the tax incidence. The approach will be discussed in the next section.

\section*{IV. Difference-in-Difference-in-Difference Estimations}

\footnote{15} The convergence should be interpreted in an asymptotic sense, i.e., as the sample size goes to infinity, the estimated $\tau$ will converge to the true $\tau$, and thus will converge to the guessed $\tau$ if the guess is correct. In a finite sample, however, the difference can be large because of sample errors.

\footnote{16} With two periods, using first differencing v.s. regular fixed effects models using time-demeaning will produce identical results. The former is easier to implement in SAS because of iterative estimation.

\footnote{17} As in any other fixed effects models, time invariant regressors cannot be identified.
In the DD estimation discussed in the previous section, a critical assumption is that the wage difference between the control group and the treatment group remains constant in the absence of a payroll tax for these two consecutive years. While this assumption sounds quite reasonable, it is still possible that wages for these two groups would have grown at a different rate if there had been no payroll tax. As a result, the wage difference (in logarithm) in the absence of tax may change over the year. If this is the case, an additional control should be used in the DD estimation. This is the so-called difference-indifference-in-difference (DDD) estimation (Gruber, 1994).

To control for the possible difference in wage growth at different wage levels, two potential controls are used in the DDD estimation. The first one is to use year 1979 and year 1980 to control for the change in the wage difference between the control group and the treatment group for the year 1978-79. The second one is to use government employees as an additional control group because they were not covered by the social security system before 1984.

More specifically, for 1979 and 1980, the control group and the treatment group became the same group in terms of the payroll tax, i.e., they were both below the tax threshold and thus their wages are determined by equation (2). Moreover, from 1979 to 1980, the payroll tax rate was the same, 6.13%. Thus, the difference of the wage difference between these two groups between 1979 and 1980 was not related to the payroll tax (since the tax effect is differenced out), and thus reflects an annual change of the wage gap.

If the wage difference between the control group and the treatment group changes at the same rate across these three years, then the change in wage difference between 1978-79 and between 1979-80 should be identical if there had been no payroll tax. For example, when the
unobserved labor market forces that affect wage growth for those two groups do not change in
two years in the absence of the payroll tax, the wage difference will have changed at the same
rate in the consecutive two-year periods. Therefore, the change in the wage gap between 1979-80 can be used to control for such a change between 1978-79 to isolate the true tax effect, and thus to remove the growth effect in wages from the tax effect identified by the DD estimation.

To implement the DDD procedure, we treat the same control group and treatment group
as an experiment group in 1978-79, but as a non-experiment group in 1979-80 because there was
no tax change for them in 1979-80. The regression model is as follows:

\[
\ln W_{ij} = X_{ij} \beta + \gamma_1 \cdot \text{group} + \gamma_2 \cdot \text{time} + \gamma_3 \cdot \text{expm} + \xi_1 \cdot \text{group} \cdot \text{time} + \xi_2 \cdot \text{expm} \cdot \text{time} + \xi_3 \cdot \text{group} \cdot \text{expm} + \delta \cdot \text{group} \cdot \text{time} \cdot \text{expm} + \epsilon_{ij}. \tag{18}
\]

In the model, the dummy variable group takes value 1 for the control group and 0 for the
treatment group; the dummy variable year equals 1 for year 78 and 0 for year 79 for the
experiment group, while 1 for 1979 and 0 for 1980 for non-experiment group; and the dummy
variable expm takes value 1 for the experiment group (1978-79) and 0 for the non-experiment
group (1979-80). As in the DD estimation, other variables in \(X_{ij}\) include schooling, job
experience, and the square of job experience, and interaction terms of education with group and
time, and of experience with group and time.\(^{19}\)

\(^{18}\) The regression model is easier to understand if we think of the group for 1979-80 as a separate group (non-experiment group) of individuals as though they had gone through the same tax change between 1978-79 as those in the experiment group but had not been affected by the change.

\(^{19}\) Because both the experiment group and the non-experiment group consist of the same individuals, there is no need to interact education and experience with the experiment dummy.
The result based the pooled the OLS regression is given in Table 2. The convergence again occurs at \( \tau=1 \), and the estimated tax effect is \(-0.1215\) and significant. The tax effect from the DDD procedure is smaller than that derived from the DD estimation; it is possible that the DD procedure overestimates the tax effect by incorporating some other differences into the tax effect. The calculated tax incidence based on the DDD estimation is that \( \tau=1.57 \), which is considerably smaller than the figure from the DD estimation. Moreover, in this case, the test cannot reject the null hypothesis that \( \tau=1 \), with the t-statistic equals to 1.27.

In addition, the DDD procedure can also be implemented as separate regressions for the years 1978-79 and 1979-80. As in the DD estimation, the tax effect based on 1978-79 is \(-0.149\), and the change in the wage difference unrelated to the payroll tax based on 1979-80 is \(-0.0425\). By differencing it out, the net tax effect becomes \(-0.107\), and the corresponding calculated \( \tau \) is 1.43 (converge again at \( \tau=1 \)). Thus, the separate estimation and the pooled estimation in the DDD approach give almost identical results.

The fixed effects model is also estimated in the DDD procedure. The estimated tax effect is \(-0.1223\), very close to that derived from the pooled OLS estimation, and the estimated tax incidence is 1.58. If the fixed effects model is estimated separately for 1978-79 and 1979-80, the estimated \( \tau \) is 1.45. In both estimations, \( \tau \) converges at 1. Therefore, the DDD procedure using 1979-80 as an additional control gives a tax incidence 1 in both the pooled OLS and fixed effects model, and we cannot reject that the true incidence is 1. The DDD results again indicate a full shift of the payroll tax to workers.

This DDD approach depends on the assumption that the change in the wage difference between the control group and the treatment group remained constant in those years. It would be
interesting to see how sensitive the result is using an alternative control for the DDD estimation. Before 1984, federal employees were not covered under the social security system. The 1983 Amendments made numerous changes to the social security programs, including the first coverage of federal employees. Therefore, federal employees can be treated as an additional control group that was unaffected by the change in the payroll tax threshold in 1978-79.

The group of federal employees is also called the non-experiment group, and the group of non-government workers covered by the social security tax is called experiment group. For government employees, those with wages above the threshold in 1978 but below the threshold in 1979 are comparable to the treatment group in the experiment group; and those with wages below the threshold in both years are comparable to the control group. Thus the additional change in wage differences unrelated to the payroll tax between 1978 and 1979 between the treatment group and the control group can be controlled for by using the corresponding difference among government employees. Given the different wage structures of government versus non-government sectors, this assumption seems to be restrictive. However, in the DDD model, the difference in wage growth is only caused by the difference in wage levels. If the labor market has perfect information and mobility, then wage growth at different wage levels should be closely correlated across these two sectors in the equilibrium. Therefore, it still makes sense to use government employees as an additional control.

The DDD estimation using government employees as an additional control is very similar to the DDD procedure using the year 1979-80, except that the wage equation for the non-experiment group becomes \( \ln W_{ij} = \ln W_{oij} \), where the payroll tax does not have an effect. The results are presented in Table 2. The parameter for the third-level interaction of dummy variables is the tax effect, and captures all variation in wages specific to the treatments—that is,
the change in tax limit (relative to controls) in the experimental group (relative to the non-experimental group) in the year before the tax limit change (relative to the year after).

Interestingly, the estimation converges at \( \tau = 0.60 \), and the estimated \( \tau \) equals 0.63. In this case, the estimated \( \tau \) and the guessed \( \tau \) are much closer; moreover, we cannot reject the hypothesis that the true \( \tau \) is 1.

V. Further Discussion and Conclusion

The various difference-in-difference estimations all point to a complete shift of the payroll tax toward workers, i.e., the tax incidence equals one. From an efficiency point of view, this is a desirable outcome. If workers bear the entire tax burden, then there is no increase in labor costs for employers and no reduction in employment. Studies on other incidences also find a full shifting back of the taxes/mandates in the United States; for example, Gruber (1994) on mandated maternity benefits, and Anderson and Meyer (2000) on unemployment insurance.

It is helpful to take a closer look at the mechanism of shifting the tax burden. Generally, as discussed in section II, the incidence of the payroll tax depends on the relative elasticity of labor supply and demand, and the workers’ valuation of future benefits from the taxes. More specifically, suppose that the labor demand and supply are as follows:

\[
D = D(W \cdot (1+t_f))
\]

\[
S = S(W \cdot (1-\alpha \cdot t_e) + \varphi \cdot w \cdot t_e).
\]

where \( W \) is the observed wage and \( t_f \) and \( t_e \) are payroll tax rates on firms and workers, respectively. The variable \( \alpha \) is the extent to which workers discount their tax payments relative to cash income. If \( \alpha = 0 \), employees believe that the entire value of the tax payments will return as social security benefits; and if \( \alpha = 1 \), employees think that the tax payment is the same as
other tax payments considered lost. Similarly, ϕ is the extent to which workers value employer payments for social security, relative to cash income. Thus, when ϕ=1, workers expect the full value of employer contributions to return to them in the form of social security benefits, and when ϕ=0, workers expect to receive nothing back from the employer's portion.

At equilibrium, \(D(w \cdot (1+t_f)) = S(w \cdot (1-\alpha \cdot t_e) + \varphi \cdot w \cdot t_f)\). It can be shown that

\[
\frac{d w / w}{d t_f} = \frac{\eta_d - \varphi \cdot \eta_s}{\eta_s (1 - \alpha \cdot t_e + \varphi \cdot t_f) - \eta_d (1 + t_f)},
\]

where \(\eta_d\) is the elasticity of labor demand and \(\eta_s\) is the elasticity of labor supply.

Clearly, a complete shifting of the payroll tax to employees can happen in three cases. First, when workers value social security benefits fully, i.e., \(\alpha = 0\) and \(\varphi = 1\); second, when the labor supply is perfectly inelastic, \(\eta_s = 0\); and third, when the labor demand is perfectly elastic, \(\eta_d = -\infty\). In all these cases, \(\frac{d w / w}{d t_f} = -\frac{1}{1 + t_f}\), and at \(t_f = 0\), it becomes \(\frac{d w / w}{d t_f} = -1\). In other words, if \(t_f\) increases by one percentage point, the wage will fall by one percent, which indicates a full shifting to workers of the cost of the payroll tax for employers.

Generally a perfectly elastic labor demand is unlikely.\(^20\) Therefore, a complete shifting of the payroll tax will be caused either by a perfectly inelastic labor supply or by the full valuation of social security benefits, or both. It is generally believed that the labor supply is very inelastic for males (Killingsworth, 1983). For the sample used in this study, white male household heads, the labor supply elasticity is likely to be very inelastic. In addition, for the year of 1978-79, workers might value social security benefits that they will receive. Either an inelastic labor supply or a valuation of the social security benefits will contribute to a complete shift of the payroll tax toward workers. As a result, it is not surprising to see that the incidence

---

\(^{20}\) It certainly depends on the degree of substitutability among different types of labor (homogeneous or heterogeneous) in production (see Bhattarai and Whalley, 1999).
of the payroll tax equals one using the current sample. This finding is consistent with the belief held by many economists on the incidence of the social security tax.

If this is true in general, the efficiency issue becomes less of a concern for the current pay-as-you-go social security system, although other problems, such as a financial imbalance remain. It is unclear, however, whether other groups with a more elastic labor supply (such as women) have the same incidence. For such a group, if the valuation of future social security benefits (hence an increase in labor supply) is not large enough to offset the reduction in labor demand caused by the social security tax, the tax burden will not be fully shifted back, and thus inefficiency will result. Therefore, as discussed in Summers (1989), a defined contribution scheme (private or partially private) with a stronger linkage between payments and benefits should be at least as efficient as the current defined benefit system. Unfortunately, the DD approach cannot be directly applied to other groups using PSID data because there were very few individuals who were above the tax threshold in one year and below the threshold in the following years.

This paper attempts to use natural experiment and a difference-in-difference approach to estimate the incidence of the social security tax. It is generally difficult to estimate the tax incidence directly using individual level data. This study attempts to offer a solution using an indirect estimation. By applying DD and DDD procedures and using different controls in the estimation, the results all indicate a full shift-back of the tax burden; i.e., workers bear the full tax burden in the sample. The estimations clearly depend on various assumptions concerning the treatment group and the control groups. Like other studies using the natural experiment and the difference-in-difference method, these sorts of assumptions should be kept in mind when interpreting the results.
Reference List


Appendix  Sample Selection

The sample is created using the PSID’s World Wide Web-based custom merge and subsetting system. Family and individual observations from 1978 to 1981 are selected. The 1981 family and individual data file is needed because employment information for 1980, such as work hours and pay, was surveyed in 1981.

The sample is selected based on the following criteria: household head, white male, either working in the private sector or for the government every year from 1978 to 1980. The sample only includes those with earnings below the social security tax threshold in 1979 ($22,900) and 1980 ($25,900). Since the minimum-wage law caused censoring in the wage structure, the sample is restricted to those whose wages are above the minimum wage every year. In addition, all individuals work at least 500 hours every year (and less than 3000 hours per year). A few observations have been dropped because of invalid information concerning education and experience.

The sample consists of 774 observations, with 630 individuals working in non-government sectors and 144 individuals working for the government. Among workers in non-government sectors, 479 persons had annual earnings below the tax limit during all three years in question, and 151 had earnings above the tax limit in 1978 but were moved to below the tax limit in 1979 and 1980 as a result of the increase in the tax limit. Among government employees who were not covered by the social security system in those three years, 103 persons were under the tax limit for all three years, while 41 were above the tax limit in 1978 and below the limit in 1979-80. Among government employees, however, we do not have information to distinguish among federal government, state-government and local government; additionally there is a relatively small number of government employees comparable to those in the treatment group.
Annual earnings are based on the survey question regarding the household head’s income from wages and bonus, overtime etc. Annual working hours are constructed using questions on average hours worked per week and number of weeks worked per year on the head's primary job. Job experience is measured by the number of years worked since 18 years old.
Table 1  Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978 hourly wage</td>
<td>7.23</td>
<td>2.23</td>
</tr>
<tr>
<td>1979 hourly wage</td>
<td>8.05</td>
<td>2.26</td>
</tr>
<tr>
<td>1980 hourly wage</td>
<td>8.91</td>
<td>2.68</td>
</tr>
<tr>
<td>1978 work hours</td>
<td>2045.79</td>
<td>295.61</td>
</tr>
<tr>
<td>1979 work hours</td>
<td>2007.79</td>
<td>305.88</td>
</tr>
<tr>
<td>1980 work hours</td>
<td>1958.69</td>
<td>341.16</td>
</tr>
</tbody>
</table>

Note: The total number of observations is 774 per year.
Table 2  Estimated Incidence of Social Security Tax

<table>
<thead>
<tr>
<th></th>
<th>DD estimation (pooled OLS)</th>
<th>DD estimation (fixed effects)</th>
<th>DDD estimation (pooled OLS, 79-80 as control)</th>
<th>DDD estimation (fixed effects, 79-80 as a control)</th>
<th>DDD estimation (pooled OLS, government as a control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax effect</td>
<td>-0.149 (-4.336)</td>
<td>-0.151 (-7.674)</td>
<td>-0.122 (-2.463)</td>
<td>-0.122 (-4.399)</td>
<td>-0.0159</td>
</tr>
<tr>
<td>τ at convergence (grid search)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>τ at convergence (calculated)</td>
<td>1.832</td>
<td>1.849</td>
<td>1.569</td>
<td>1.576</td>
<td>0.63</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.117 (18.05)</td>
<td>0.0313 (0.591)</td>
<td>2.149 (24.586)</td>
<td>0.0629 (1.548)</td>
<td>1.919 (13.773)</td>
</tr>
<tr>
<td>Group</td>
<td>-0.533 (-4.544)</td>
<td>-0.472 (-5.366)</td>
<td>-0.584 (-5.071)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-0.0473 (-0.504)</td>
<td>(intercept above)</td>
<td>-0.0707 (-0.976)</td>
<td>(intercept above)</td>
<td>-0.0527 (-0.527)</td>
</tr>
<tr>
<td>Time-group</td>
<td>(Tax effect above)</td>
<td>(Tax effect above)</td>
<td>-0.0351 (-0.995)</td>
<td>-0.0355 (-1.786)</td>
<td>-0.095 (-1.467)</td>
</tr>
<tr>
<td>Education</td>
<td>0.00857 (1.026)</td>
<td>0.0103 (1.685)</td>
<td>0.0205 (2.333)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>0.00735 (2.669)</td>
<td>0.00714 (3.535)</td>
<td>0.0136 (4.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00017 (-2.77)</td>
<td>-0.000125 (-2.840)</td>
<td>-0.000188 (-3.553)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education-group</td>
<td>0.0192 (2.261)</td>
<td>0.0192 (3.083)</td>
<td>0.0209 (2.899)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience-group</td>
<td>-0.00077 (-0.488)</td>
<td>-0.00225 (-1.936)</td>
<td>0.000399 (0.287)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education-time</td>
<td>0.00167 (0.262)</td>
<td>0.00049 (0.137)</td>
<td>-0.00107 (-0.228)</td>
<td>-0.00165 (-0.627)</td>
<td>-0.000107 (-0.019)</td>
</tr>
<tr>
<td>Experience-time</td>
<td>0.00264 (1.93)</td>
<td>0.00256 (3.294)</td>
<td>0.00108 (1.073)</td>
<td>0.00104 (1.836)</td>
<td>0.00187 (1.56)</td>
</tr>
<tr>
<td>Experiment</td>
<td></td>
<td>-0.0667 (-2.191)</td>
<td>0.205 (1.782)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment-time</td>
<td></td>
<td>0.0833 (1.935)</td>
<td>0.0833 (3.435)</td>
<td>-0.00216 (-0.035)</td>
<td></td>
</tr>
<tr>
<td>Experiment-group</td>
<td></td>
<td>-0.0379 (-1.087)</td>
<td>0.0148 (0.284)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>1260</td>
<td>630</td>
<td>2520</td>
<td>1548</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.376</td>
<td>0.114</td>
<td>0.328</td>
<td>0.389</td>
<td></td>
</tr>
<tr>
<td>F-value</td>
<td>75.286</td>
<td>26.953</td>
<td>87.472</td>
<td>61.014</td>
<td></td>
</tr>
</tbody>
</table>

Note:  1. t-values are in parenthesis.
2. The fixed effects model is estimated by the first differencing method, and time invariant variables cannot be identified.