NBER WORKING PAPERS SERIES

ARE RISING WAGE PROFILES A FORCED-SAVING MECHANISM?

David Neumark

Working Paper No. 4213

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1992

Assistant Professor, Department of Economics, University of Pennsylvania, and NBER. I am grateful to Richard Johnson for outstanding research assistance, to Josh Angrist, McKinley Blackburn, Frank Diebold, Richard Johnson, Daniel Radner, David Ribar, Paul Taubman, and seminar participants at Hebrew University and Ben Gurion University for helpful discussions, to Paul Taubman for assistance in tracking down the data used in this paper, and to the NIA for research support. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4213
November 1992

ARE RISING WAGE PROFILES A FORCED-SAVING MECHANISM?

ABSTRACT

This paper tests the hypothesis that rising earnings profiles are a mechanism by which individuals engage in forced saving. It does this by examining the cross-sectional relationship between overwithholding on income tax payments--behavior that is consistent with a preference for forced saving--and the slopes of age-earnings profiles. The force-saving hypothesis receives some support from earnings regression estimates. Individuals who receive tax refunds are on earnings profiles that are steeper and have lower intercepts, although the evidence is statistically significant in only a subset of the specifications estimated. On average, individuals who receive refunds have about one percentage point faster earnings growth per year.

David Neumark
Department of Economics
University of Pennsylvania
Philadelphia, PA 19104
and NBER

I. Introduction

A positive return to experience in standard log earnings equations throughout much of the life-cycle is one of the most robust findings in labor economics. But economists disagree on the source of this relationship. The dominant explanation is based on the human capital model, in which the stock of human capital rises with experience over much of the life-cycle, resulting in increasing earnings (Ben-Porath, 1967; Mincer, 1974). In the general human capital model, earnings rise lock-step with productivity, while in early specific human capital models, earnings rise more slowly than productivity (Becker, 1975). In contrast, in models of long-term incentive-compatible contracts, firms pay workers less than their marginal product when they are young or less-tenured, and more when they are older or more-tenured; thus, earnings will rise with age or tenure even if productivity is constant over a worker's life (Lazear, 1979). Some versions of specific human capital models incorporating seniority rules also imply that earnings will rise faster than marginal product (Carmichael, 1983).[1] Although it is difficult to distinguish among specific alternative models, some researchers have concluded that earnings do in fact rise faster than marginal product (Medoff and Abraham, 1980; Kotlikoff and Gokhale, forthcoming).

A third, more recent explanation of rising age-earnings profiles, independent of age-productivity profiles, is the forced-saving hypothesis, first articulated by Frank and Hutchens (forthcoming). This hypothesis has two components. First, workers prefer rising consumption over the life cycle. Second, individuals are not disciplined savers, and therefore need rising wage profiles to enjoy rising consumption. Frank and Hutchens discuss two occupations--commercial airline pilots and intercity bus drivers--in which age-

[1] Alternative models with this implication are developed in Salop and Salop (1976) and Harris and Holmstrom (1982).

productivity profiles are arguably flat, while age-earnings profiles rise. They argue further that for both of these occupations neither the specific human capital nor the Lazear contract models can explain rising wage profiles. Frank and Hutchens instead offer the forced-saving hypothesis as an explanation. First, they argue that utility depends not only on the level of consumption, but also depends positively on its rate of change. They present some indirect empirical evidence, and findings from other social sciences, to support this assumption about tastes. Second, they assume that individuals find it difficult to save in order to defer present consumption. They report results from a survey of economics students in which 78 percent of respondents preferred a rising wage profile to a flat profile, even when the rising profile had lower present value. Evidence of preferences for rising wage profiles (as well as consumption profiles) is required to support the forced-saving hypothesis, since individual utility maximization with a preference for rising consumption would entail maximization of the present value of income, if individuals were disciplined savers. However, Frank and Hutchens present no direct evidence that preferences for rising wage profiles reflect an inability to save.

Loewenstein and Sicherman (1991) present findings from a more thorough survey in which individuals were asked to rank alternative wage profiles. As in Frank and Hutchens' work, survey respondents demonstrated a widespread preference for increasing wages (with 87 percent of respondents preferring rising profiles), even though these wage profiles had lower present values than the flat or downward-sloping profiles included in the set of choices. Loewenstein and Sicherman consider a number of possible explanations for their findings. One explanation is that rising wages indicate increasing mastery over one's work, which increases utility. This explanation does not require that individuals experience difficulty in saving in order to generate rising age-earnings profiles.[2] However, the survey respondents also expressed preferences for rising non-wage payments (labelled rental income in the survey), which undermines this explanation. The other explanations--anticipated needs for future expenditures, and a utility function that depends on rising consumption--both require a "self-control" or saving problem that "prevents easy transformation of decreasing payment into increasing consumption sequences" (p. 71). Survey respondents' reasons for preferring rising wages were consistent with these explanations, although only a small percentage of respondents directly identified "self-control" as a reason.

In contrast to these existing papers, the present paper examines direct evidence on the link between forced saving and rising earnings profiles. Specifically, it asks whether there is evidence that rising age-earnings profiles are associated with individuals' propensities to engage in forced saving. It does this by looking at the empirical relationship between the steepness of age-earnings profiles and individuals' overwithholding on income tax payments. As Loewenstein and Sicherman point out, overwithholding is one of several varieties of behavior that are anomalous from the perspective of present-value maximization, yet consistent with forced saving. Other examples that they cite include Christmas clubs and simultaneous saving and borrowing at a higher rate of interest.[3,4]

---

[2]It also does not necessarily imply that wages rise <u>faster</u> than productivity.

[3]Christmas clubs, which pay lower interest rates than savings accounts, and prohibit withdrawals before December 1, were considered by Stigler (1966). The simultaneous saving and borrowing phenomenon was considered by Thaler and Shefrin (1981). The notion is that individuals impose self-control by prohibiting dissaving, while still allowing borrowing for certain purchases. One way to interpret this behavior is that it raises the price of large purchases that would otherwise drain savings.

Another possible manifestation of forced saving is the growing substitution of lump-sum payments for base wage increases in union contracts (Bell and Neumark, 1991), even

Evidence of this nature can add significantly to empirical research on the forced-saving explanation of rising earnings profiles, for three reasons. First, the existing research has focused solely on evidence regarding preferences; it has perhaps lent credence to the assumption that individuals prefer rising consumption profiles, but it has not provided direct evidence on the second key assumption needed to generate rising wage profiles, *i.e.*, individuals' inability to defer present consumption. Second, even if it is accepted that individuals prefer rising consumption, and that individuals have trouble saving, this does not necessarily imply that labor markets will generate rising earnings profiles. By exploring directly the link between earnings profiles and overwithholding on taxes, this paper not only asks whether the assumptions underlying the forced-saving hypothesis are true, but also if the hypothesis helps to explain rising wage profiles. Finally, if the hypothesis does receive some support, it is of interest to gauge its quantitative significance, *i.e.*, how much of the upward slope in earnings profiles can be explained by the forced-saving hypothesis.

II. Direct Evidence on Rising Age-Earnings Profiles as Forced-Saving Mechanisms

II.A. The Empirical Approach

This paper exploits a data set that includes variables used to estimate standard earnings regressions, as well as tax return information indicating whether or not a person received a tax refund (and its amount), or equivalently whether or not they overwithheld on their taxes. The data are described in detail in the following section, but the empirical approach can be explained and discussed at this point. In the simplest case, a standard log

---

though the expected present value of compensation under lump-sum plans may be lower.

[4]Other anomalies with respect to present-value maximization are discussed in Loewenstein and Thaler (1989).

earnings regression is augmented to include a variable indicating whether or not an individual received a tax refund, and an interaction between this dummy variable and age or experience:

$$(1) \qquad \ln(w) = \alpha + X\beta + \gamma E + \delta E \cdot R + \lambda R + \epsilon ,$$

where w is annual earnings, X is a set of control variables, E is age or experience, and R is a dummy variable equal to 1 if the individual received a refund.

The spirit of this equation is that R is a proxy for a preference for forced saving; the forced-saving hypothesis then offers predictions as to how this preference should be related to earnings profiles. Under two assumptions discussed below, a finding that the estimate of δ is positive would be consistent with the hypothesis that rising age-earnings profiles are partly a forced-saving mechanism.[5] (Only if the estimate of γ is zero would the evidence suggest that forced saving is the entire explanation of rising earnings profiles.) Since the forced-saving hypothesis implies that individuals <u>sacrifice</u> present value in choosing rising earnings profiles (a result that is supported by the findings of Loewenstein and Sicherman, 1991), the estimated coefficient of R, the refund dummy variable, should be negative.

Age or experience is written here as a linear variable, but more standard non-linear specifications (such as a quadratic) are used in the empirical work that follows. Equation (1) is also estimated defining R as the ratio of the refund (or the negative of the tax payment) to wage and salary income, instead of a simple dummy variable.[6] This definition

[5] An alternative interpretation of this finding would be that older workers prefer to have (larger) refunds for any given level of income. In the absence of longitudinal data, this "age effect" hypothesis cannot be distinguished from the forced-saving hypothesis.

[6] The level of the refund must be scaled by income to eliminate the positive association between levels of income and refunds or levels of income and taxes owed that arises simply because only high income earners can have high levels of refunds or taxes owed. The dummy variable specification provides an alternative "scale-free" measure of refunds.

has the advantage of providing a more informative "index" of individuals' propensity to overwithhold, or to engage in forced saving. In addition to providing information on the relative size of the refund, it also includes information on underwithholding. According to the forced-saving hypothesis, underwithholding should be negatively related to the slope of the earnings profile; for this reason, the continuous refund variable provides considerably more information. But because w appears in the denominator of R, any measurement error in w would likely induce a spurious negative correlation between ln(w) and R, therefore biasing the estimate of λ downward; the direction of bias in the estimate of δ, however, is not clear *a priori*.

Two assumptions are essential to the test of the forced-saving hypothesis using equation (1). The first is necessitated by the availability of data on tax refunds for only one year. Given this, it must be assumed that individuals' propensities to overwithhold are persistent over time, so that individuals who overwithheld for the year for which data are available can be assumed to have overwithheld in previous years (and to do so in future years), and *vice versa*. This assumption permits the relationship between earnings growth and overwithholding to be identified from comparisons across individuals of differences in earnings levels at different ages or levels of experience, for those who do and do not receive tax refunds.[7]

The second assumption is that forced saving through tax refunds and forced saving through rising earnings are not substitutes. Consider first the case in which they are substitutes. In the extreme, all individuals could have "self-control" problems and engage in forced saving. Some could do this by overwithholding, while others do this by choosing

[7]In contrast, if data were available on individuals' earnings growth and refunds, then the earnings growth-overwithholding relationship could be observed directly.

rising earnings profiles. In this case, even though the forced-saving explanation of rising age-earnings profiles is valid for the latter group, no relationship between refunds and the steepness of earnings profiles would be expected in equation (1). The assumption, instead, is that these two forced-saving mechanisms, if they exist, are complementary. Forced saving through rising earnings seems most likely to be saving for older ages and retirement, while forced saving through overwithholding seems more likely to be within-year saving (for example, for the purchase of a durable good). Thus, the assumption is that there is heterogeneity across individuals in the ability to engage in both types of saving, so that, if the forced-saving explanation of rising earnings profiles is valid, individuals who overwithhold are also more likely to have steeper earnings profiles.

II.B. Potential Biases from the Failure to Choose the Amount of Taxes Withheld

The notion that overpayment of taxes reflects forced saving, which is the basis of the empirical test in this paper, assumes that individuals choose the amount of taxes to withhold from their income. From the perspective of the rational, calculating economic agent, this assumption appears to be valid, at least for predictable sources of income. It is true that for each number of withholding allowances entered on an individual's W-4 form, IRS tables or formulas are used to calculate the amount withheld from each paycheck. But individuals can presumably manipulate their number of allowances, estimated non-wage income, and additional amounts to withhold, in order to balance tax payments with estimated tax liabilities.[8] Nonetheless, it is conceivable that individuals do not fine-tune their

[8]Loewenstein and Thaler (1989), in their discussion of the anomaly posed by tax refunds, claim that overwithholding is easily avoided by adjusting the withholding rate. Even if individuals cannot choose the withholding rate exactly, they can choose to underwithhold (by an amount sufficiently small to avoid penalties), and to owe taxes at the

withholding, perhaps out of laziness, an inability to perform the required calculations (or the high cost of doing so), or because there are unpredictable sources of income. If individuals do not choose their withholding to exactly match tax liabilities, then estimates of δ and λ in equation (1) may be biased, either because the level or steepness of the earnings profile determine whether or not one receives a refund, or because factors associated with age affect the relationship between earnings and refunds. The first possibility, that characteristics of the earnings profile affect refunds, obviously calls into question the direction of causality in equation (1).[9] In fact, equation (1) should not be given a causal interpretation. The forced-saving hypothesis does not imply that refunds cause higher wage growth, but just that the two are associated. In this subsection five such sources of bias in the estimation of this association are discussed, and empirical procedures for studying some of these biases are explained.

First, unanticipated shocks to labor income may be negatively correlated with refunds (*e.g.*, individuals who experience an unanticipated income decline are likely to have overwithheld, and therefore to receive a refund). This creates a downward bias in the

---

end of the year. Since this amounts to an interest-free loan from the government, it is the behavior that would be predicted by present-value maximization of income.

[9]Given this potential problem, it would be ideal to instrument for the refund variables in equation (1) with a variable that is correlated with the refund, but not witn the contemporaneous wage equation error. Individuals are allowed to credit a refund to the following year's taxes. If some individuals owing taxes also chose to overpay (crediting taxes to next year), then a "crediting" or "pre-payment" (dummy) variable would be a valid instrument. It would presumably be correlated with a preference for forced saving, but would be unaffected by current wages. However, only those who are due refunds are given the option to "pre-pay" part of next year's taxes. Furthermore, in the samples used in this paper for most of the results, only a handful of individuals due refunds chose to credit the refund to the following year. This rules out using information on this behavior as an additional measure of a preference for forced saving.

estimate of λ, the coefficient of the refund variable.[10] This problem may be most acute for the self-employed, whose income is probably less predictable; of course, along the lines of the above discussion, these individuals could adjust their fourth quarter tax payments to reflect the income shock.

Second, individuals with higher labor income likely have more assets producing non-labor income, on which they tend to owe taxes at the end of the year; *i.e.*, such individuals may find it easier to underwithhold. This would also bias the estimate of λ downward. This non-labor income may also lead to bias in the estimated coefficient of the age- or experience-refund interaction variable (δ), since older workers generally have accumulated more assets and therefore have higher non-labor income. In this case, the estimate of δ is likely to be biased downward, as the correlation between labor income and refunds decreases (becoming more negative) with age. This direction of bias works against the forced-saving hypothesis. In the empirical work that follows, the potential impact of biases from non-labor income and from unanticipated changes in labor income are considered by comparing estimates for subsamples with different sources of taxable income. In particular, the greatest emphasis is placed on earnings regressions estimated for non-self-employed individuals with primarily labor income.

Third, biases could arise if standard withholding rates systematically overwithhold or underwithhold in a manner that differs by level of income. Based on overwithholding tables used in 1972, overwithholding was more prevalent (in either absolute terms, or relative to income) at lower levels of income. This creates a negative association between earnings and the refund variable, or a downward bias in the estimate of λ. Of course, if the

[10]One source of bias working in the other direction is that those with lower income may be liquidity constrained, in which case they would be less likely to overwithhold.

lower overwithholding at higher income levels exists to account for more deductions for higher-income taxpayers, then this effect may not be present. It may also seem that if individuals receive salary increases during the tax year, and do not adjust their withholding, then because of progressive tax rates, they end up overwithholding and receiving a refund. Since individuals on steeper earnings profiles receive more or larger raises, this could lead to a finding of steeper earnings profiles associated with tax refunds. In fact, however, progressive taxes *per se* do not lead to this effect. Individuals ultimately end up paying taxes based on an average rate that is a weighted average (by income) of the tax rates on the lower and higher levels of income.[11]

Fourth, estimates of the relationship between the steepness of earnings profiles and refunds may be biased because of factors correlated with age or experience, which also affect tax refunds. Over much of the life-cycle age (or experience) is positively correlated with marriage and dependents. Because the tax tables used to calculate withholding rates change with marital status and number of dependents, or because of other features of the tax code, the relationship between income and withholding could differ by marital status of number of dependents. Because of the relationship of age with marriage and dependents, the estimate of δ could reflect a difference in the correlation between earnings and refunds

---

[11]To see this, let Y_1 be income earned in, say, the first half of the year, and Y_2 ($> Y_1$) income earned in the second half. Assume that taxes are withheld based on the annual income implied by the current salary, that the average tax rate applied to Y_1 is A_1, and the marginal rate applied to $Y_2 - Y_1$ is M_{12} ($> A_1$). Then A_2, the average rate on Y_2, is given by $\{A_1Y_1 + (Y_2 - Y_1)\cdot M_{12}\}/Y_2$. A', the average rate paid on income for the year, is $\{A_1Y_1 + A_1Y_1 + (Y_2 - Y_1)\cdot M_{12}\}/(Y_1 + Y_2)$, which is just a weighted average of A_1 and A_2, with weights $Y_1/(Y_1 + Y_2)$ and $Y_2/(Y_1 + Y_2)$. This implies that taxes paid during the year exactly equal taxes owed for the year. (Taxes paid during the year amount to $A_1Y_1 + \{[A_1Y_1 + (Y_2 - Y_1)\cdot M_{12}]/Y_2\}\cdot Y_2$, while taxes owed for the year amount to $A_1Y_1 + A_1Y_1 + (Y_2 - Y_1)\cdot M_{12}$.)

by marital status, rather than evidence pertaining to the forced-saving hypothesis.[12] To control for this possibility, specifications are also estimated in which interactions of the refund variable with marital status and number of dependents are added to equation (1).

Finally, another factor that may be correlated with age or experience, and also affect refunds, is the dollar value of itemized deductions that an individual takes on his tax return. *A priori*, it seems more difficult to take account of itemization of deductions in adjusting withholding rates than, for example, to take account of marital status or number of dependents, partly because some itemized deductions may be unanticipated, so this source of bias may be particularly worrisome. The largest itemized deductions are for expenses associated with home ownership (mortgage interest payments and real estate taxes) and health expenses.[13] Because age is probably positively associated with both home ownership and health care expenses,[14] and because these expenses are likely to positively associated with earnings, a positive estimate of δ may reflect a rising correlation between earnings and refunds as itemized deductions increase, rather than an age effect *per se*. In the empirical work that follows, this source of bias is considered by adding to equation (1) a dummy variable for whether or not an individual itemized deductions, and an interaction

[12]On the other hand, because the specifications estimated in the paper include controls for marital status and number of dependents, the estimated coefficient of the non-interactive refund variable is not likely to be biased from this source.

[13]For example, for the sample for which earnings profiles are estimated using taxable Social Security earnings later in the paper (Table 5), home ownership expenses average 33 percent of total deductions ($1114), and medical expenses average six percent of deductions ($193).

[14]One caveat is that mortgage interest payments, which are a major part of the deductible expenses of home ownership, decline with age during the life of a mortgage.

between this dummy variable and the refund variable.[15,16]

III. The Data

The data used in this study come from the Social Security Administration and Census Bureau's 1973 CPS-IRS-SSA Exact Match Study.[17] Part of this project includes an exact match between data from three sources. The starting point for the data sets constructed in this study was the March 1973 Current Population Survey (CPS). An exact match was made between the CPS sampled households and Social Security benefit and earnings records. Then, limited information on 1972 tax returns made available by the IRS was matched to the CPS households. The resulting file is called the 1973 Current Population Survey-Administrative Record Exact Match File. This matched sample has been used to study income distribution and life-cycle earnings (Rosen and Taubman, 1982), but the limited tax return information does not cover overwithholding and refunds.

A second data file constructed as part of the Exact Match Study consists of a sample of 95,000 federal individual income tax returns chosen by subsampling the IRS 1972

---

[15]On the other hand, because the receipt of refunds and the itemization of deductions are closely intertwined, in that it is typically large deductions that lead to large refunds, it is difficult to distinguish between the effects of refunds and the effects of itemizing deductions. For this reason, attention is not focused on trying to sort out separate roles of refunds and deductions, but rather only on the potential for rising itemized deductions with age to lead to a spurious positive estimate of δ in equation (1).

[16]The same warning against a causal interpretation of the estimates applies here. Just as wage levels or growth could potentially affect refunds, they could also potentially affect itemization; one reason is that whether or not one can itemize depends on the size of potential deductions relative to taxable income.

[17]Details provided here are taken in large part from Kilss, *et al.* (1978) and Radner (1978).

Statistics of Income (SOI) sample. This sample contains full information from individual tax returns, including overwithholding. Limited demographic information from SSA records, as well as earnings for the year subject to Social Security taxes, were also matched to the records in this sample. This latter file is used to carry out the empirical tests in this paper. Some research (Radner, 1978) constructs "statistical matches" between the SOI sample and the Exact Match File. This is potentially useful in studying income distribution. In the present context, the potential advantage of such a match would be the ability to estimate equation (1) using CPS data to estimate the earnings equation, augmented by information on overwithholding and refunds from the SOI sample. However, a statistical match would involve imputing which individuals in the CPS sample received refunds, based on whether individuals in the SOI sample with similar characteristics (for the subset of characteristics available in both data sets) received refunds. However, it is not obvious that overwithholding should be predictable based on this set of characteristics. In fact, the safest working assumption was that it was unpredictable, ruling out a statistical match.[18]

Given this decision, only the SOI sample (augmented by SSA information) is available to estimate equation (1). This sample contains limited information with which to estimate standard earnings regressions. Two earnings measures are available: IRS wage and salary income, from the tax returns; and earnings subject to Social Security taxes from the matched SSA data. Age (in five-year intervals), sex, and race are available from the matched SSA data. Number of dependents is available from the tax returns. The sample also contains information on filing status (which subsumes marital status), other sources of

[18]Because there are no overlapping observations between the SOI sample and the Exact Match File, containing all of the required data, standard missing data techniques (see, *e.g.*, Rubin, 1976) are not applicable.

income, and self-employment status. The information on other sources of income is particularly important, since, as discussed in Section II, events such as capital gains or losses could affect whether or not an individual receives a refund, independently of forced-saving motivations. Also, since rising earnings profiles for self-employed individuals cannot represent forced savings, the theory should receive less support (if it is true) when the self-employed are included in the sample.

The SOI sample has three important limitations. First, the IRS wage and salary income available from the tax returns may, in the case of joint filers, include a spouse's income. For joint filers, this can be partially accounted for by controlling for marital status. More importantly, all of the empirical specifications reported in the paper are also estimated for the sample of individuals who file separately (either single individuals, or married individuals filing separately), in which case only the filer's income is recorded. Finally, the specifications are also estimated using earnings subject to Social Security taxes, which can be attributed exclusively to the individual, as an alternative dependent variable.

Second, education is unavailable in the SOI sample. Since education plays an important independent role in earnings determination, and is also frequently used in constructing a measure of potential experience, this is a potentially significant limitation. To lessen the impact of this limitation, information from the Exact Match File was used to impute schooling levels for individuals based on characteristics available in the SOI file (characteristics also available from the exact match of SSA and limited tax return information to the CPS sample). Specifically, cross-tabulations were calculated for the Exact Match File of schooling by five-year age intervals, filing status (which subsumes marital status), race (white, black, or other), and number of dependents. Individual cell means were then used to estimate education in the SOI sample for individuals with the same characteristics as those in each cell.[19] To obtain results both with and without relying on this imputation procedure, estimates of equation (1) are reported using age, and omitting constructed education, as well as using potential experience (using constructed education), along with constructed years of education.

The third limitation is that the Social Security earnings variable is severely top-coded, at $9,000 (the taxable maximum). Since this censors well over half of the sample, Tobit estimates of the earnings equation are reported when this variable is used.[20]

The sample used in the paper is restricted to males aged 25-54, to minimize problems from selection into employment. Sample descriptive statistics are reported in Table 1. Sampling weights have not been used, since the principal interest in this paper is on the estimated regression coefficients.[21] Furthermore, as the next section shows, many of the key results are estimated using subsamples representing a small fraction of the original sample, for which the providing sampling weights are irrelevant.

IV. Empirical Results

IV.A. Preliminaries

As a preliminary to estimating equation (1), Table 2 reports estimates of standard

---

[19]In cases of empty cells, the mean across all filing statuses for the age-race-dependents cell was used. This broader cell was never empty in the data set.

[20]In contrast, IRS wage and salary income is not top-coded. For this measure, OLS estimates of the wage equation are reported.

[21]Average income is higher in the SOI file because high-income individuals were oversampled. In both the Exact Match and SOI files, IRS wage and salary income is not top-coded, while CPS income is top-coded at $50,000. This, plus the fact that the IRS measure may include a spouse's income, probably explains why in the Exact Match file average IRS income exceeds average CPS income.

earnings equations for the Exact Match file. This file offers valuable information of two sorts. First, the Exact Match file contains the IRS and Social Security earnings measures, as well as the more familiar CPS annual earnings measure, permitting a comparison of standard earnings equation estimates with these three earnings measures. Second, this file can be used to provide evidence on whether the use of constructed schooling (along with potential experience), or the omission of schooling (and the substitution of age for potential experience), biases estimates of the parameters describing the shape of the earnings profile.

Table 2 reports estimates of these preliminary regressions. Column (1) of Panel A reports estimates of a standard specification using potential experience (and its square), and actual schooling, in which the dependent variable is CPS annual earnings. The parameter estimates are standard. For comparison, column (2) substitutes constructed schooling for actual schooling, both as an independent variable, and in constructing potential experience. The estimated schooling coefficient is closer to zero, but the coefficients of potential experience and its square are little changed; the point estimates indicate an earnings profile that is slightly steeper initially, with somewhat more curvature. The same columns in Panel B report specifications for the same dependent variable, first including age and its square along with schooling, and then dropping schooling. In this case, the shape of the earnings profile implied by the coefficient estimates of age and its square changes by even less. None of these estimates necessarily implies that the use of constructed schooling, or the omission of schooling, does not bias the estimated coefficient of the interaction between the refund dummy variable and age or potential experience (δ in equation (1)). But the fact that the usual parameters governing the shape of the earnings profile are little changed because of the imputation of the schooling variable makes it plausible that the estimates of δ that follow are unbiased.

Next, columns (3) and (4) report estimates of the same specifications, substituting IRS wage and salary income for CPS annual earnings. The estimates in Panel A indicate that for this earnings measure the schooling coefficient, as well as the coefficients of potential experience and its square, are little changed by substituting constructed for actual schooling. Similarly, earnings profile estimates in Panel B are unaffected by the omission of schooling. In addition, the point estimates are close to those using CPS earnings. The only notable difference occurred for the estimated marriage coefficient (not reported), which was considerably higher for IRS income, presumably because this measure sometimes includes a spouse's income.

Finally, columns (5) and (6) report estimates of similar specifications for earnings subject to Social Security taxes. Recall that in the SOI file this variable is top-coded at $9,000, necessitating the use of Tobit estimation techniques. In the Exact Match file this same variable is not top-coded. However, because interest ultimately lies in the estimates from the SOI file, for comparability Social Security income is treated as if it were top-coded at $9,000 in Table 2. Despite the severe censoring, these estimates convey the same message as do the other estimates in this table. Parameter estimates with this income measure are similar to those for the other measures, and the use of constructed schooling, or the omission of schooling, do not create notable changes in the estimates of parameters governing the shape of the earnings profile.

IV.B. Evidence on the Forced-Saving Explanation of Rising Wage Profiles

Recall that the forced-saving hypothesis predicts a positive estimate of δ, the coefficient of the age- or experience-refund interaction, and a negative estimate of λ, the coefficient of the refund variable, in equation (1). In interpreting the results, the various biases discussed in Section II should be kept in mind. Some of these biases receive explicit attention in the tables that follow. But based on the considerations discussed in Section II, those that do not receive explicit attention appear most likely to bias the estimates of both δ and λ downwards. Thus, a positive estimate of δ should perhaps be regarded as providing stronger evidence in favor of the forced-saving hypothesis than is indicated by standard significance tests. On the other hand, a negative estimate of λ may not in fact provide such evidence. (Because of this, the estimates of λ receive less attention in the ensuing discussion of the results.)

Table 3 turns to the first set of main results, estimates of equation (1) using the SOI file, with IRS wage and salary income as the dependent variable. Panel A reports estimates using constructed schooling as an independent variable, and in the construction of potential experience, while Panel B omits schooling, and substitutes age for potential experience. The various columns restrict the sample along two dimensions, to address two problems in testing the forced-saving hypothesis with these data. The first dimension concerns the source of income. As discussed in Section II, significant capital gains or losses, or other non-labor income, may influence the estimates of δ and λ. Thus, more reliable estimates are likely to result from a subsample restricted to individuals with primarily labor income. In addition, evidence in favor of the forced-saving hypothesis should be strongest, if the hypothesis is true, when individuals with self-employment income are excluded from the sample. Thus, one source of variation in the samples used in Table 3 is the restriction of the source of income, first excluding those with self-employment income, and then restricting the sample further to those for whom labor income is between 90% and 100% of adjusted gross income.

The second dimension is filing status. Because in the case of joint filers (which covers most married couples) IRS income may include a spouse's income, a less contaminated income measure can be obtained by restricting the sample to those who filed individually. (In addition, in the next table, estimates of the same specifications will be reported for Social Security earnings, which can be attributed to the individual tax filer.)

The first four columns report estimates for all filers. Column (1) includes all individuals with IRS wage and salary income. In this sample, the forced-saving hypothesis receives little support. The estimated coefficients of the interactions between potential experience or age and the refund dummy variable are small and statistically insignificant, and actually negative. Recall, though, the arguments in Section II that the presence of significant amounts of non-labor income may bias the estimate of δ downward. In addition, the estimated coefficient of the refund dummy variable, which measures the relationship between the earnings profile intercept and refund status, is positive rather than negative. In column (2), which reports estimates for the subsample excluding those with self-employment income,[22] the statistical evidence is not much more favorable to the forced-saving hypothesis. The estimated coefficients of the interactions between potential experience or age and the refund dummy variable are again statistically insignificant, although they are now positive, and the estimated coefficients of the refund dummy variable are now negative and statistically significant. The change in this latter coefficient is somewhat surprising since it seems likely that self-employment income is less predictable than other sources of income, which would tend to create a downward bias in the estimated coefficient of the refund dummy variable when those with self-employed income are included in the sample.

Results for what is likely a cleaner experiment for the joint filers are reported in

[22]The self-employed are identified as those who paid the self-employment FICA tax, or who reported business or farm gains or losses.

column (3), in which the subsample is further restricted to those whose wage and salary income is between 90% and 100% of adjusted gross income, and who have no capital gains or losses, hence focusing primarily on labor income. For this subsample, the evidence in favor of the forced-saving explanation of rising wage profiles is stronger. In both the potential experience specification (Panel A) and the age specification (Panel B) the estimate of δ is positive and statistically significant at the five percent level; this rise in the estimate of δ is consistent with the discussion in Section II of biases stemming from significant amounts of non-labor income.

The evidence is even more supportive in column (4), in which the refund (or negative tax payment) as a proportion of IRS wage and salary income is substituted for the refund dummy variable (in the interaction term as well). In this case, the estimated coefficient of the interaction between the refund variable and potential experience or age is positive and highly significant. The mean value for the refund variable for this sample is 0.035 (for those who received refunds), which implies that, for example, an individual with this mean value of the refund variable has a return to experience that is 0.0068 (0.035 × 0.1954) higher than an individual receiving no refund (*i.e.*, those who withheld the exact amount of taxes due), and even higher compared to an individual who underwithheld. This point estimate is somewhat larger than the estimate that results from the dummy variable specification in column (3), perhaps because of reduced measurement error in the refund variable.

Columns (5)-(8) report estimates for the same three samples with respect to sources of income, but restricting attention to separate filers, for whom IRS wage and salary income can be attributed solely to the individual. In these estimates, once the self-employed are excluded, the forced-saving hypothesis also receives some statistical support. For the four dummy variable specifications (Panels A and B, columns (6) and (7)), the point estimates of the slope-interaction coefficients are larger than when the joint filers were included. However, because of the smaller sample size the statistical evidence is somewhat weaker; the t-statistics for the estimated coefficients of the interactions between potential experience or age and the refund dummy variable are range from 1.5 to 1.85. In addition, the estimated coefficients of the refund dummy variables are always negative and statistically significant. As for the joint filers, the specification using the continuous refund variable provides much stronger statistical evidence in support of the forced-saving hypothesis; for this specification, individuals receiving the mean refund-to-income ratio have a return to experience that is higher by 0.0088 (the mean of the refund variable is 0.046, multiplied by 0.1922), compared with individuals receiving no refund. In this case, this point estimate is slightly lower than the estimated coefficients of the refund dummy variable.

It seems plausible that small refunds may reflect mistakes or inadvertent behavior, rather than deliberate forced saving. To study this possibility, the models in columns (3) and (7) of Table 3 were reestimated defining the refund dummy variable to be one only if the refund was greater or equal than one percent of wage and salary income. (This eliminated about ten percent of the refunds.) In all four cases, this raised the point estimates of the refund-age or refund-experience interactions, and reduced the standard errors. For the estimates using all filers (paralleling column (3)), this naturally increased the strength of the statistical evidence in favor of the forced-saving hypothesis. For the separate filers (paralleling column (7)), the evidence in favor of the forced-saving hypothesis became significant at the five-percent level for the experience specification, and at the ten-percent level for the age specification. The corresponding estimates (standard errors) of the refund-experience and refund-age interactions were 0.0122 (0.0062) and

0.0127 (0.0067).

Table 4 reports estimates for the SOI file, using earnings subject to Social Security taxes instead of IRS wage and salary income.[23] On one hand, evidence using this variable should parallel that using the individual filers in Table 3, since, again, there is a direct attribution of income to the individual. On the other hand, the sample used in Table 4 includes predominantly married men, which may lead to estimates more similar to those for the joint filers in Table 3. OLS and Tobit estimates are reported to reveal the effects of the correction for censoring. These effects are substantial; the OLS estimates of the earnings equation parameters are nonsensical, while the Tobit estimates parallel the earlier specifications. As such, attention is focused only on the Tobit estimates. Qualitatively, the results parallel those in Table 3. In particular, when the self-employed are excluded, the signs of the coefficient estimates are as predicted by the forced-saving hypothesis. When the sample is further restricted to those with IRS wage and salary income between 90% and 100% of adjusted gross income, the point estimate implies that those receiving refunds have earnings growth that is higher by 0.8 percentage point. However, in these Tobit estimates the interactive slope coefficient estimates are not statistically significant; much of the problem may be the considerably larger standard errors than in the first four columns of Table 3 owing to the severe censoring of the earnings measure. The specification in column (6) was also reestimated defining the refund dummy variable to be one only if the refund was greater than or equal to one percent of wage and salary income, to reduce the role of

[23]Because of the top-coding of this income variable, only specifications using the refund dummy variable are estimated. It may appear that Tobit estimates of a standard earnings equation could be used to impute earnings for the top-coded cases, and these imputed earnings could then be used to construct the ratio of the refund to earnings. But if equation (1) is the correct model, then the earnings measure is also required to construct the independent variables involving this ratio.

inadvertent refunds. As for the OLS estimates discussed above, this leads to statistically significant evidence in favor of the forced-saving hypothesis. The estimated coefficient (standard error) of the refund-experience interaction was 0.0123 (0.0050), and that of the refund-age interaction was 0.0119 (0.0053).

IV.C. Potential Biases

The empirical analysis to here has considered some of the sources of bias in the estimates of equation (1) discussed in Section II, by presenting evidence based on different sources of income, and focusing attention on the non-self-employed with primarily labor income. Section II also discussed potential biases stemming from the correlation of age with marital status, number of dependents, and itemized deductions, variables that may influence the relationship between earnings and refunds, and hence bias the estimate of δ in equation (1). Table 5 reports results that address these sources of bias. In all cases, models are estimated for the subsamples excluding the self-employed, excluding joint filers (for specifications using IRS wage and salary income), and excluding those with IRS wage and salary income less than 90% or more than 100% of adjusted gross income. OLS estimates using IRS wage and salary income, with the dummy variable and ratio versions of the refund variable are reported, as are Tobit estimates using the Social Security income measure and the refund dummy variable. Panel A reports specifications using potential experience and constructed schooling, while Panel B reports results using age.

Columns (1)-(3) report results adding interactions between marital status and number of dependents and the refund variable. The estimated coefficients of the marital status interactions are never statistically significant, and those of the dependents interactions are significant only in the Tobit estimates using the Social Security earnings measure. In all cases, though, adding these interaction variables does weaken the evidence in favor of the forced-saving hypothesis, compared with the estimates for the corresponding samples in Tables 3 and 4. The point estimates of the coefficients of the interaction between the refund dummy variable and age or experience decline slightly in column (1), and more so in column (3) (the same specification in which the dependents-refund interaction has a significant coefficient estimate). But in column (2), using the ratio of the refund to wage and salary income, the estimated coefficient of the refund-age (or experience) interaction remains positive and statistically significant.

Roughly similar results are obtained in column (4)-(6), in which instead the itemized deduction variables are added. The evidence in favor of the forced-savings hypothesis is again weakened in all specifications, although it remains statistically significant in column (5), using the ratio of the refund to wage and salary income, and marginally significant in the Tobit estimates using the Social Security earnings measure. Thus, overall, the evidence in favor of the forced-saving hypothesis is weakened, but not eliminated, by augmenting the specification of equation (1) to account for other factors associated with age, earnings, and tax refunds.

To this point, the statistical evidence in favor of the forced-saving hypothesis is stronger in the specifications using the continuous refund variable. As discussed in Section II, however, estimates using this variable are potentially biased because wage and salary income appears in the denominator of the refund variable (the "division bias" problem). One way to ask whether these biases affect the results is to compare the partial derivatives of ln(w), in equation (1), with respect to R, using the estimates obtained using the dummy and continuous versions of the refund variable. In column (7) of Panel A of Table 3, $\partial \ln(w)/\partial R$ (at zero years of potential experience) is -0.624. In contrast, in column (8) the implied derivative is -12.930, multiplied by the mean of the refund variable for those receiving refunds (0.046), or -0.595. Similarly, it was reported earlier that $\partial^2 \ln(w)/\partial R \cdot \partial \text{Pot.Exp.}$ in column (8) is 0.0088, compared with 0.0112 in column (7). Results for the same columns in Panel B (using age) are similar. If the differences between columns (7) and (8) are interpreted as reflecting this "division bias," then it appears that estimates of λ are biased upward, and estimates of δ are biased downward, both of which biases work against the forced-saving hypothesis. This suggests that division bias does not generate spurious evidence in favor of the hypothesis.

Finally, Table 6 explores the robustness of the key results from Tables 3 and 4 to alternative functional forms. The models are estimated for the same subsamples and earnings variables as the models in Table 5. Two variations of the functional form are considered. First, in columns (1)-(3) of both panels, dummy variables corresponding to five-year intervals of potential experience or age are substituted for the linear and quadratic terms. In all cases this variation in the functional form results in a slight strengthening of the statistical evidence in favor of the forced-saving hypothesis, compared with the corresponding estimates in Tables 3 and 4. Also, in the OLS estimates, in three of the four cases the adjusted-R^2 increases, suggesting that this is an appropriate enrichment of the functional form. Second, in columns (4)-(6) the refund dummy variable is interacted with the linear potential experience or age variable, as well as the quadratic. For the dummy variable specifications, this renders estimated coefficients of the interaction variables statistically insignificant; however, in neither set of OLS estimates does the adjusted-R^2 improve. For the specifications using the continuous refund variable the linear and quadratic interactions are statistically significant with, if anything, stronger effects at the means.

V. Conclusions

The hypothesis that rising wage or earnings profiles are a forced-saving mechanism receives some support from the earnings regression estimates reported in this paper. The main finding is that individuals who receive tax refunds in the one year for which data are available are on steeper earnings profiles, as predicted by the forced-saving explanation of rising wage profiles, although the evidence is statistically significant in only a subset of the specifications estimated. On average, individuals who receive refunds have about one percentage point faster earnings growth per year. In addition, the earnings profiles of the same individuals have lower intercepts, which is also consistent with the forced-saving hypothesis, although the estimates of this intercept may be biased towards this conclusion. The result that individuals who receive refunds are on steeper earnings profiles is robust to variations in the functional form of the earnings regression, but is somewhat less robust to the addition to the regression of variables potentially associated with age, which may themselves influence the relationship between earnings and refunds.

These results lead to two further questions regarding the forced-saving explanation of rising wage profiles. First, an implication of the theory is that individuals who engage in forced-saving by "choosing" steeper wage profiles sacrifice present value of earnings. Present value calculations could, in principle, be used to compare present values of the earnings streams of those who did and did not receive tax refunds. However, it was argued that the estimated coefficient of the tax refund dummy variable has a negative bias, which in turn biases the present value calculations towards supporting the forced-saving hypothesis.

Second, the question naturally arises as to how much of the upward slope in earnings profiles can be accounted for by the forced-saving explanation. Faster earnings growth of one percentage point per year is substantial. To interpret this magnitude, consider that the estimated coefficients of the linear potential experience or age variables are on the order of 0.03 to 0.13, indicating that at the outset, before the negative effect of the quadratic term kicks in, earnings grow by three to thirteen percent per year for those not receiving refunds. In addition, at the mean levels of age and potential experience, rates of earnings growth for those not receiving refunds are around one percent per year. These estimates suggest that forced saving could be a substantial part of the explanation of rising earnings profiles.

Furthermore, these estimates may provide a lower bound in the sense that receipt of a tax refund (and the value of the refund) is only one indicator, and undoubtedly a noisy one, of a propensity to engage in forced saving. It remains to be seen whether data can be uncovered that can more accurately identify this propensity, and its association with earnings growth. This is particularly important because in nearly all of the specifications reported in the paper, the standard quadratic effects of age or experience persist for those who do not receive refunds; without other data on forced saving, the conclusion at this point must therefore be that the forced-saving hypothesis is at best a partial explanation of rising earnings profiles. Nonetheless, the evidence in this paper, as well as the survey evidence of other researchers indicating that workers prefer rising wage profiles, implies that the forced-saving hypothesis should receive further attention as a potentially important source of rising earnings over the life-cycle.

References

Becker, Gary S. 1975. Human Capital (Chicago: University of Chicago Press).

Bell, Linda A., and David Neumark. 1991. "Lump Sums, Profit Sharing, and Labor Costs in the Union Sector." NBER Working Paper No. 3630.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." Journal of Political Economy 75(2): 352-65.

Carmichael, H. Lorne. 1983. "Firm-Specific Human Capital and Promotion Ladders." Bell Journal of Economics 14(1): 251-8.

Frank, Robert H., and Robert M. Hutchens. "Wages, Seniority, and the Demand for Rising Consumption Profiles." Forthcoming in Quarterly Journal of Economics.

Harris, Milton, and Bengt Holmstrom. 1982. "A Theory of Wage Dynamics." Review of Economic Studies LXIX: 315-33.

Kilss, Beth, Fritz Scheuren, Faye Aziz and Linda DelBene. 1978. "The 1973 CPS-IRS-SSA Exact Match Study: Past, Present, and Future." In Policy Analysis with Social Security Research Files (U.S. Department of Health, Education, and Welfare), pp. 163-194.

Kotlikoff, Laurence J., and Jagadeesh Gokhale. "Estimating a Firm's Age-Productivity Profile Using the Present Value of Workers' Earnings." Forthcoming in Quarterly Journal of Economics.

Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" Journal of Political Economy 87(6): 1261-84.

Loewenstein, George, and Nachum Sicherman. 1991. "Do Workers Prefer Increasing Wage Profiles?" Journal of Labor Economics 9(1): 67-83.

Loewenstein, George, and Richard H. Thaler. 1989. "Anomalies: Intertemporal Choice." Journal of Economic Perspectives 3(4): 181-94.

Medoff, James L., and Katharine G. Abraham. 1980. "Experience, Performance, and Earnings." Quarterly Journal of Economics 95: 703-36.

Mincer, Jacob. 1974. Schooling, Experience, and Earnings (New York: Columbia University Press).

Radner, Daniel. 1978. "Age and Family Income." In Policy Analysis with Social Security Research Files (U.S. Department of Health, Education, and Welfare), pp. 195-224.

Rubin, Donald B. 1976. "Inference and Missing Data." Biometrika 63(3): 581-92.

Rosen, Sherwin, and Paul Taubman. 1982. "Changes in Life-Cycle Earnings: What Do Social Security Data Show?" Journal of Human Resources 17(3): 321-38.

Salop, Steven C., and June Salop. 1976. "Self-Selection and Turnover in the Labor Market." Quarterly Journal of Economics XC: 619-28.

Stigler, George J. 1966. The Theory of Price (New York: Macmillan).

Thaler, Richard H., and H.M. Shefrin. 1981. "An Economic Theory of Self-Control." Journal of Political Economy 89(2): 392-406.

Table 1

Sample Descriptive Statistics, Exact Match and SOI Samples[a]

	Exact Match File Full Sample (1)	SOI File Full sample (2)	SOI File Refund (3)	SOI File No Refund (4)
CPS variables:	N = 17,659	N = 32,645	N = 20,124	N = 12,521
CPS wage and salary income	11035 (47.69)	...	...	...
Schooling	12.20 (0.02)	...	...	...
Potential experience (with actual schooling)	20.40 (0.07)	...	...	...
Age[b]	38.60 (0.07)	41.23 (0.05)	40.01 (0.06)	43.19 (0.07)
Black	0.07	0.04	0.05	0.02
Non-black/non-white	0.01	0.01	0.02	0.01
Percent married	0.89	0.93	0.92	0.95
Number of dependents (IRS data)[c]	1.86 (0.01)	2.07 (0.01)	2.07 (0.01)	2.07 (0.01)
IRS variables:	N = 18,391			
IRS wage and salary income	12670 (52.88)	20385 (139.25)	20749 (170.39)	19799 (238.26)
Percent filed jointly	0.86	0.92	0.91	0.95
Proportion receiving refund	...	0.62	1.0	0.0
Refund (or negative tax payment)/wage and salary income[d]	...	...	0.035 (0.0003)	...
Constructed schooling[e]	12.18 (0.007)	12.16 (0.005)	12.19 (0.006)	12.11 (0.007)
Potential experience (with constructed schooling)	20.57 (0.07)	22.66 (0.05)	21.44 (0.06)	24.62 (0.07)
SSA variables:	N = 17,612	N = 29,913	N = 18,058	N = 11,855
Social Security earnings (< \$9,000)	7450 (18.15)	8135 (11.76)	8072 (15.78)	8230 (17.34)
Proportion top-coded[f]	0.56	0.77	0.75	0.79

Table 1 (continued)

a. Standard errors of means are reported in parentheses.
b. The samples are restricted to males, aged 25-54. Age is coded in five-year intervals as it is in the SSA data on the SOI file.
c. This includes children, parents, and other dependents, both inside and outside home.
d. For sample with no capital gains or losses, not self-employed, and wage and salary income between 90% and 100% of adjusted gross income (N=7,988).
e. Imputed from CPS data.
f. Social Security earnings is top-coded at $9,000 in the SOI file, but not in the Exact Match file. Here, the figures are presented as if the $9,000 top-coding exists in both data sets.

Table 2

Comparison of OLS Earnings Equations in the Exact Match CPS-SSA-IRS File, with CPS, IRS, and SSA Income

	A. Potential Experience/Constructed Schooling Specifications					
	(1)[a]	(2)[a]	(3)[b]	(4)[b]	(5)[c]	(6)[c]
Potential experience[d]	0.037 (0.002)	0.045 (0.003)	0.033 (0.003)	0.040 (0.004)	0.050 (0.005)	0.064 (0.007)
Potential experience^2/100	-0.058 (0.006)	-0.083 (0.008)	-0.050 (0.006)	-0.065 (0.008)	-0.073 (0.012)	-0.112 (0.016)
Constructed schooling[e]	...	0.046 (0.008)	...	0.088 (0.008)	...	0.084 (0.016)
Schooling	0.077 (0.002)	...	0.076 (0.002)	...	0.106 (0.004)	...
$\bar{R}^2$	0.172	0.090	0.142	0.082	...	...
Log-likelihood	...	...	...	...	-17699.7	-18007.5
Estimation method	OLS	OLS	OLS	OLS	Tobit	Tobit
	B. Age Specifications					
Age	0.078 (0.006)	0.083 (0.007)	0.072 (0.007)	0.075 (0.007)	0.120 (0.014)	0.120 (0.014)
Age^2/100	-0.085 (0.008)	-0.095 (0.008)	-0.077 (0.009)	-0.086 (0.009)	-0.130 (0.018)	-0.137 (0.018)
Schooling	0.066 (0.002)	...	0.066 (0.002)	...	0.090 (0.004)	...
$\bar{R}^2$	0.172	0.089	0.143	0.078	...	...
Log-likelihood	...	...	...	...	-17691.4	-18017.4
Estimation method	OLS	OLS	OLS	OLS	Tobit	Tobit

a. Dependent variable is logarithm of CPS wage and salary income. All independent variables are from CPS. Sample is restricted to males aged 25-54. There are 17,659 observations. All specifications include an intercept, and dummy variables for married, black, and other (non-black/non-white), and dummy variables for one, two, three, four, and five or more dependents. Age is defined as midpoint of five-year intervals (beginning with 25-29), to correspond with coding of age in the SOI file.

b. Dependent variable is logarithm of IRS wage and salary income. There are 18,391 observations.

c. Dependent variable is logarithm of taxable Social Security income. There are 17,612 observations, of which 9,878 are top-coded.

d. In specifications using constructed schooling, potential experience uses this schooling measure.

e. Imputed from CPS data.

Table 3

OLS Earnings Equations for Survey of Income File, Alternative Sources of Income and Filing Statuses
(Dependent Variable: Log of IRS Wage and Salary Income)[a]

A. Potential Experience/Constructed Schooling Specifications

	(1)[b] (N=32,645)	(2)[c] (N=16,960)	(3)[d] (N=9,156)	(4)[d] (N=9,156)	(5)[e] (N=2,465)	(6)[f] (N=1,739)	(7)[g] (N=1,179)	(8)[g] (N=1,179)
Potential experience	0.049 (0.006)	0.053 (0.004)	0.033 (0.004)	0.024 (0.004)	0.065 (0.014)	0.055 (0.014)	0.025 (0.016)	0.025 (0.013)
Potential experience^2/100	-0.062 (0.012)	-0.072 (0.008)	-0.048 (0.009)	-0.029 (0.009)	-0.128 (0.031)	-0.104 (0.029)	-0.042 (0.034)	-0.030 (0.030)
Potential experience × refund dummy	*-0.0003 (0.0017)*	*0.0002 (0.0015)*	*0.0038 (0.0019)*	...	*0.0073 (0.0054)*	*0.0102 (0.0057)*	*0.0112 (0.0076)*	...
Potential experience × refund/wage and salary income[h]	...	...	...	*0.1954 (0.0168)*	...	...	...	*0.1922 (0.0554)*
Refund dummy variable	0.676 (0.043)	-0.392 (0.037)	-0.377 (0.046)	...	0.001 (0.125)	-0.728 (0.132)	-0.624 (0.168)	...
Refund/wage and salary income	...	...	...	-10.090 (0.406)	...	...	...	-12.930 (1.104)
Constructed schooling	0.101 (0.012)	0.078 (0.008)	0.060 (0.009)	0.033 (0.008)	-0.005 (0.027)	0.024 (0.025)	0.032 (0.028)	0.018 (0.024)
$\bar{R}^2$	0.074	0.215	0.221	0.303	0.067	0.175	0.114	0.308

B. Age Specifications

	(1)[b]	(2)[c]	(3)[d]	(4)[d]	(5)[e]	(6)[f]	(7)[g]	(8)[g]
Age	0.070 (0.010)	0.079 (0.008)	0.056 (0.008)	0.036 (0.007)	0.147 (0.031)	0.120 (0.030)	0.051 (0.034)	0.037 (0.029)
Age^2/100	-0.067 (0.012)	-0.079 (0.009)	-0.061 (0.010)	-0.034 (0.009)	-0.174 (0.038)	-0.143 (0.037)	-0.060 (0.043)	-0.033 (0.038)
Age × refund dummy	*-0.0002 (0.0018)*	*0.0008 (0.0015)*	*0.0048 (0.0020)*	...	*0.0079 (0.0059)*	*0.0115 (0.0062)*	*0.0127 (0.0083)*	...
Age × refund/wage and salary income[h]	...	...	...	*0.2082 (0.0174)*	...	...	...	*0.2252 (0.0601)*
Refund dummy variable[i]	0.667 (0.034)	-0.405 (0.030)	-0.372 (0.037)	...	0.045 (0.099)	-0.677 (0.105)	-0.567 (0.134)	...
Refund/wage and salary income[i]	...	...	...	-9.050 (0.316)	...	...	...	-12.036 (0.841)
$\bar{R}^2$	0.072	0.212	0.219	0.302	0.069	0.177	0.115	0.310

Table 3 (continued)

a. All variables are from SOI file. All specifications include an intercept, dummy variables for married, black, and other (non-black/non-white), and dummy variables for one, two, three, four, and five or more dependents. See Tables 1 and 2 for variable definitions and further details.
b. All individuals with earnings.
c. Excludes self-employed individuals.
d. Excludes individuals with capital gains or losses, self-employed, and those with wage and salary income less than 90% or more than 100% of adjusted gross income.
e. Same as column (1), excluding joint filers.
f. Same as column (2), excluding joint filers.
g. Same as column (3), excluding joint filers.
h. The interactions between potential experience or age and the refund variable are defined as age or experience multiplied by the refund variable minus its mean. Thus, the experience or age coefficients measure the returns to experience or age evaluated at the sample mean of the refund variable.
i. The age × refund variable interactions are defined as the refund variable multiplied by (age - 25). Thus the refund dummy measures the log wage differential associated with refunds at age 25, the minimum age in the sample.

Table 4

Tobit Earnings Equations for Survey of Income File, Married Men, Alternative Sources of Income
(Dependent Variable: Log of Taxable Social Security Earnings)[a]

A. Potential Experience/Constructed Schooling Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[c]	(5)[d]	(6)[d]
Potential experience	-0.001 (0.008)	0.094 (0.009)	-0.002 (0.013)	0.098 (0.013)	-0.004 (0.015)	0.070 (0.015)
Potential experience^2/100	-0.010 (0.018)	-0.151 (0.020)	-0.001 (0.027)	-0.160 (0.027)	0.005 (0.031)	-0.112 (0.031)
Potential experience × refund dummy	*-0.0010* *(0.0026)*	*-0.0028* *(0.0029)*	*0.0015* *(0.0051)*	*0.0048* *(0.0049)*	*0.0021* *(0.0066)*	*0.0079* *(0.0064)*
Refund dummy	-0.089 (0.063)	-0.018 (0.071)	-0.019 (0.120)	-0.484 (0.119)	-0.104 (0.151)	-0.384 (0.152)
Constructed schooling	-0.045 (0.018)	0.049 (0.020)	-0.039 (0.024)	0.044 (0.025)	-0.025 (0.026)	0.040 (0.028)
$\bar{R}^2$	0.008	...	0.012	...	0.016	...
Log-likelihood	...	-21569.0	...	-9396.5	...	-6221.4
N	6972	29913	3049	15363	2240	8175
Estimation method	OLS	Tobit	OLS	Tobit	OLS	Tobit

B. Age Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[c]	(5)[d]	(6)[d]
Age	0.006 (0.016)	0.161 (0.017)	-0.003 (0.025)	0.172 (0.024)	-0.011 (0.028)	0.124 (0.028)
Age^2/100	-0.010 (0.019)	-0.170 (0.021)	0.004 (0.030)	-0.184 (0.029)	0.016 (0.035)	-0.133 (0.034)
Age × refund dummy	*-0.0016* *(0.0028)*	*-0.0034* *(0.0030)*	*0.0012* *(0.0055)*	*0.0045* *(0.0051)*	*0.0010* *(0.0071)*	*0.0081* *(0.0068)*
Refund dummy[e]	-0.084 (0.049)	-0.026 (0.055)	0.000 (0.096)	-0.451 (0.094)	-0.072 (0.119)	-0.336 (0.121)
$\bar{R}^2$	0.007	...	0.012	...	.016	...
Log-likelihood	...	-21566.5	...	-9394.5	...	-6220.4
N	6972	29913	3049	15363	2240	8175
Estimation method	OLS	Tobit	OLS	Tobit	OLS	Tobit

Table 4 (continued)

a. All variables are from SOI file. All specifications include an intercept, dummy variables for married, black, and other (non-black/non-white), and dummy variables for one, two, three, four, and five or more dependents. See Tables 1 and 2 for variable definitions and further details. OLS estimates exclude individuals at or above the top-coded value of $9,000 for the dependent variable.
b. All individuals with earnings.
c. Excludes self-employed individuals.
d. Excludes individuals with capital gains or losses, self-employed, and those with IRS wage and salary income less than 90% or more than 100% of adjusted gross income.
e. The age × refund dummy variable interaction is defined as the refund dummy variable multiplied by (age - 25). Thus the refund dummy variable measures the log wage differential associated with refunds at age 25, the minimum age in the sample.

Table 5

OLS and Tobit Earnings Equations for Survey of Income File, with Interactions of Potential Experience or Age with Marital Status, Dependents, and a Dummy Variable for Itemized Deductions[a]

A. Potential Experience/Constructed Schooling Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[b]	(5)[b]	(6)[c]
Potential experience	0.026 (0.016)	0.026 (0.013)	0.074 (0.015)	0.026 (0.015)	0.029 (0.010)	0.070 (0.015)
Potential experience^2/100	-0.042 (0.034)	-0.033 (0.030)	-0.111 (0.031)	-0.041 (0.032)	-0.041 (0.024)	-0.112 (0.031)
Potential experience × refund dummy	*0.0099* *(0.0079)*	...	*0.0043* *(0.0066)*	*0.0076* *(0.0072)*	...	*0.0096* *(0.0065)*
Potential experience × refund/wage and salary income	...	*0.1857* *(0.0561)*	...	...	*0.1125* *(0.0449)*	...
Refund dummy variable	-0.621 (0.170)	...	-0.488 (0.211)	-0.255 (0.172)	...	-0.092 (0.165)
Refund/wage and salary income	...	-12.742 (1.117)	...	...	-9.291 (0.936)	...
Constructed schooling	0.035 (0.028)	0.020 (0.025)	0.048 (0.028)	0.034 (0.026)	0.018 (0.019)	0.045 (0.028)
Married × refund variable[d]	-0.020 (0.194)	-2.069 (1.664)	0.038 (0.193)	...	...	...
Number of dependents × refund variable	0.051 (0.064)	0.406 (0.509)	0.112 (0.042)	...	...	...
Itemized deduction dummy variable[e]	...	...	...	-0.621 (0.050)	...	-0.283 (0.046)
Itemized deductions/wage and salary income	...	...	...	...	-0.340 (0.014)	...
Itemized deduction dummy variable × refund dummy variable	...	...	... (0.148)	-0.372	...	-0.444 (0.129)
Itemized deductions/wage and salary income × refund/wage and salary income	...	...	...	...	-0.432 (0.223)	...
$\bar{R}^2$	0.113	0.308	...	0.221	0.561	...
Log-likelihood	...	...	-6217.6	...	...	-6197.3

Table 5 (continued)

B. Age Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[b]	(5)[b]	(6)[c]
Age	0.052 (0.034)	0.040 (0.029)	0.128 (0.028)	0.052 (0.032)	0.051 (0.023)	0.123 (0.028)
$Age^2/100$	-0.061 (0.043)	-0.037 (0.038)	-0.132 (0.034)	-0.059 (0.040)	-0.053 (0.030)	-0.133 (0.034)
Age × refund dummy	*0.0114 (0.0086)*	...	*0.0044 (0.0070)*	*0.0087 (0.0078)*	...	*0.0099 (0.0068)*
Age × refund/wage and salary income	...	*0.2186 (0.0605)*	...	...	*0.1363 (0.0488)*	...
Refund dummy variable	-0.570 (0.136)	...	-0.454 (0.192)	-0.222 (0.144)	...	-0.041 (0.139)
Refund/wage and salary income	...	-11.923 (0.864)	...	...	-8.820 (0.723)	...
Married × refund variable[d]	-0.030 (0.193)	-1.966 (1.659)	0.033 (0.194)	...	...	...
Number of dependents × refund variable	0.045 (0.064)	0.462 (0.500)	0.108 (0.042)	...	...	...
Itemized deduction dummy variable[e]	...	...	...	-0.621 (0.050)	...	-0.283 (0.046)
Itemized deductions/ wage and salary income	...	...	...	...	-0.340 (0.014)	...
Itemized deduction dummy variable × refund dummy variable	...	...	...	-0.367 (0.148)	...	-0.439 (0.129)
Itemized deductions/ wage and salary income × refund/ wage and salary income	...	...	...	...	-0.425 (0.222)	...
$\bar{R}^2$	0.114	0.310	...	0.221	0.562	...
Log-likelihood	...	...	-6216.9	...	...	-6196.4

a. All variables are from SOI file. All specifications include an intercept, dummy variables for married, black, and other (non-black/non-white), and dummy variables for one, two, three, four, and five or more dependents. See Tables 1 and 2 for variable definitions and further details.
b. Dependent variable is logarithm of IRS wage and salary income. Estimation method is OLS. Excludes individuals with capital gains or losses, self-employed, and those with wage and salary income less than 90% or more than 100% of adjusted gross

Table 5 (continued)

income. Also excludes joint filers (N = 1,179).
c. Dependent variable is logarithm of taxable Social Security earnings. Estimation method is Tobit. Exclusion restrictions are the same as in footnote b, except that joint filers are not excluded (N = 8,175).
d. This is a dummy variable in columns (1), (3), (4), and (6), and the refund/wage and salary income in columns (2) and (5). This variable is entered as the deviation from its mean, so that the effects of the non-interactive refund variables, at the sample means, are comparable to those in Tables 3 and 4.
e. An individual was assumed to have itemized deductions if total deductions amounted to more than 15% of adjusted gross income (since the standard deduction was 15% of adjusted gross income), or if they filed jointly (separately) with deductions exceeding $2000 ($1000), since this was the maximum standard deduction for joint (separate) filers.

Table 6

OLS and Tobit Earnings Equations for Survey of Income File, Alternative Functional Forms[a]

A. Potential Experience/Constructed Schooling Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[b]	(5)[b]	(6)[c]
Potential experience 10-14	0.214 (0.082)	0.208 (0.067)	0.399 (0.070)	...	...	...
Potential experience 15-19	0.142 (0.116)	0.145 (0.083)	0.454 (0.090)	...	...	...
Potential experience 20-24	0.130 (0.132)	0.194 (0.079)	0.444 (0.112)	...	...	...
Potential experience 25-29	0.233 (0.166)	0.329 (0.085)	0.520 (0.135)	...	...	...
Potential experience $\geq$ 30	0.249 (0.193)	0.353 (0.096)	0.611 (0.162)	...	...	...
Potential experience	...	...	...	0.046 (0.038)	0.026 (0.013)	0.055 (0.034)
Potential experience^2/100	...	...	...	-0.092 (0.088)	-0.034 (0.030)	-0.077 (0.077)
Potential experience × refund dummy	*0.0113* *(0.0073)*	...	*0.0102* *(0.0063)*	*-0.0135* *(0.0407)*	...	*0.0251* *(0.0357)*
Potential experience^2/100 × refund dummy	...	...	...	*0.0582* *(0.0942)*	...	*-0.0398* *(0.0812)*
Potential experience × refund/wage and salary income	...	*0.1943* *(0.0554)*	...	...	*0.8028* *(0.2597)*	...
Potential experience^2 /100 × refund/wage and salary income	...	...	...	...	*-1.580* *(0.657)*	...
Refund dummy variable	-0.632 (0.164)	...	-0.438 (0.150)	-0.433 (0.352)	...	-0.524 (0.325)
Refund/wage and salary income	...	-12.976 (1.104)	...	...	-17.485 (2.190)	...
Constructed schooling	0.041 (0.028)	0.023 (0.024)	0.040 (0.027)	0.032 (0.028)	0.019 (0.024)	0.039 (0.028)
$\bar{R}^2$	0.116	0.309	...	0.114	0.311	...
Log-likelihood	...	...	-6214.9	...	...	-6221.3
$\partial ln(w)/\partial Exp\partial Refund$, at sample mean experience	...	...	...	*0.0076*	*0.2299*	*0.0088*

Table 6 (continued)

B. Age Specifications

	(1)[b]	(2)[b]	(3)[c]	(4)[b]	(5)[b]	(6)[c]
Age 30-34	0.198 (0.082)	0.197 (0.065)	0.384 (0.070)	...	...	...
Age 35-39	0.064 (0.114)	0.077 (0.076)	0.446 (0.092)	...	...	...
Age 40-44	0.098 (0.139)	0.200 (0.073)	0.414 (0.115)	...	...	...
Age 45-49	0.171 (0.170)	0.304 (0.070)	0.553 (0.162)	...	...	...
Age 50-54	0.081 (0.196)	0.269 (0.079)	0.540 (0.162)	...	...	...
Age	...	...	...	0.126 (0.086)	0.040 (0.029)	0.111 (0.072)
$Age^2/100$	...	...	...	-0.157 (0.110)	-0.039 (0.038)	-0.116 (0.091)
Age × refund dummy	*0.0132* *(0.0083)*	...	*0.0090* *(0.0068)*	*-0.0749* *(0.0928)*	...	*0.0233* *(0.0761)*
$Age^2/100$ × refund dummy	...	...	...	*0.1128* *(0.1191)*	...	*-0.0193* *(0.0963)*
Age × refund/wage and salary income	...	*0.2237* *(0.0600)*	...	...	*1.228* *(0.636)*	...
$Age^2/100$ × refund/wage and salary income	...	...	...	...	*-1.334* *(0.842)*	...
Refund dummy variable	-0.590 (0.134)	...	-0.355 (0.121)	-0.450 (0.182)	...	-0.360 (0.169)
Refund/wage and salary income	...	-12.065 (0.840)	...	...	-13.183 (1.109)	...
$\bar{R}^2$	0.117	0.312	...	0.115	0.311	...
Log-likelihood	...	...	-6214.3	...	...	-6220.4
$\partial^2 ln(w)/\partial Age \partial Refund$, at sample mean age	...	...	...	*0.0074*	*0.2542*	*0.0082*

a. All variables are from SOI file. All specifications include an intercept, dummy variables for married, black, and other (non-black/non-white), and dummy variables for one, two, three, four, and five or more dependents. See Tables 1 and 2 for variable definitions and further details.

b. Dependent variable is logarithm of IRS wage and salary income. Estimation method is OLS. Excludes individuals with capital gains or losses, self-employed, and those with wage and salary income less than 90% or more than 100% of adjusted gross income. Also excludes joint filers (N=1,179).

c. Dependent variable is logarithm of taxable Social Security earnings. Estimation method is Tobit. Exclusion restrictions are the same as in footnote b, except that joint filers are not excluded (N=8,175).