# Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?

***Abstract -*** *The Social Security earnings test reduces payments to beneficiaries whose labor income exceeds a given threshold. We investigate the impact of this rule by studying the significant changes in its structure over the past 25 years. We find that the earnings test exerts no robust influence on the labor supply decisions of men, although there is some suggestive evidence for a labor supply response among women. We also find that loosening the earnings test accelerates benefits receipt among the eligible population, lowering benefits levels, and heightening concerns about the standard of living of these elderly at very advanced ages.*

**Jonathan Gruber**
*Massachusetts Institute of Technology, Department of Economics, Cambridge MA 02142*

**Peter Orszag**
*The Brookings Institution, Washington, DC 20034-2103*

**National Tax Journal**
Vol. LVI, No. 4
December 2003

## INTRODUCTION

On April 7, 2000, President Clinton signed into law the "Senior Citizens Freedom to Work Act of 2000," which eliminated the unpopular earnings test that applied to those over the Social Security normal age of retirement (currently age 65 and 4 months). The earnings test, a version of which still applies to those ages 62–64, reduces immediate payments to beneficiaries whose labor income exceeds a given threshold. Although benefits are subsequently increased to compensate for any such reduction, the earnings test is typically viewed as a tax on working. As a result, it is commonly viewed as an important disincentive to paid work for older Americans. For example, when President Clinton signed the legislation that removed the earnings test for beneficiaries at or above the normal retirement age, he noted, "because of the Social Security retirement earnings test, the system withholds benefits from over 800,000 older working Americans and discourages countless more—no one knows how many—from actually seeking work."[1] Similarly, Alan Greenspan stated that with the elimination of the earning test, "the presumption, of course, is that you'll get an increase in the number of retired people coming back into the work force."[2]

Despite this rhetoric, a careful reading of the literature on the earnings test produces very mixed conclusions as to its labor supply effects. Two principal types of past studies ex-

[1] President Clinton, "Remarks at Bill Signing for Senior Citizens Freedom to Work Act of 2000," April 7, 2000, U.S. Newswire.

[2] Alan Greenspan, Humphrey–Hawkins Testimony, as quoted in transcript # 00022306FN–l05 by CNNfN, February 23, 2000.

amine the labor supply impact from the earnings test. One part of the literature involves studying the earnings "bunching" at the earnings test limits. For example, echoing earlier findings, Friedberg (2000) documents that significantly more workers earned amounts just below the earnings test limit than either significantly below that limit or above it. The other approach to examining the earnings test involves more sophisticated econometric analyses of the implications for the labor supply decisions of older workers of the kinked budget constraint that results from the earnings test. Traditionally, these studies suggested negligible effects of the earnings test on labor supply. Friedberg (2000) found more sizeable impacts, which suggest that removal of the earnings test for workers over age 65 would raise the hours of work of affected workers by 5 percent.

Each of these approaches has important weaknesses, however. The "bunching" literature is cleanly identified, particularly when it examines how the bunching changes as the earnings test limits change. But this literature is uninformative about the aggregate labor supply impacts of the earnings test, or indeed even about the sign of the impact. The "kinked budget constraint" literature addresses the aggregate labor supply impacts of the earnings test among those working, but does so using an econometric framework that imposes a variety of structural assumptions which have been strongly criticized in labor economics in recent years. Moreover, even this literature speaks only to the impact of the earnings test on conditional hours worked, and not on the decision to supply labor in the market at all. If there are labor market rigidities which impede the free adjustment of hours to their desired level near retirement age, then the earnings test can affect not only hours but also the retirement decision itself. As the quotations above indicate, much of the policy interest in the earnings test arises from the prospect that it is deterring older workers from seeking or keeping jobs in the first place. And the previous literature has focused almost exclusively on the impact of the earnings test on men, despite the fact that women increasingly make up an important part of the labor force at older ages.

Finally, the past literature on the U.S. has not considered the impact of the earnings test on benefits receipt. The major concern of opponents of removing the earnings test, particularly for workers between ages 62 and 64, is that it will lead to increased early claiming of benefits (Gruber and Orszag, 1999). The argument is that workers may myopically claim benefits early if they can do so while still working, and as a result end up with a lower standard of living for themselves (and their widows / widowers) later in life due to actuarial benefits reductions. But there is no evidence to date regarding the impact of the earnings test *per se* on benefits receipt.

The purpose of our paper is to update and complement the previous literature on the earnings test by examining the impact of changes in the earnings test on the decision to work, aggregate hours supplied, and claiming behavior for both men and women. Over the past three decades, the structure of the earnings test has changed significantly. To examine the impact of these changes on labor supply and benefits receipt, we use data from twenty–five years of the March Supplement to the Current Population Survey (CPS), which provide large samples of observations on the elderly. We first present a simple graphical analysis, in order to illustrate the relationship between program parameter changes and labor supply decisions. We then examine regression models that combine the information across years in a simple reduced form framework to estimate earnings test impacts. In contrast to much of the previous literature, we pursue simple reduced–form models that use the variation in earnings test parameters to identify the impact of the earnings test on labor supply. This has the advantage that our identification strategy is straightforward and, we argue, clearly convincing. But it has the disadvantage that we are not able to answer richer structural questions about the impact of the earnings test, such as disentangling the income and substitution effects of any change. Thus, we view our findings here as complementary to the existing literature in terms of male hours of work, and more novel with respect to female labor supply, the work decision, and benefits claiming.

Our analysis suggests two major conclusions. First, the earnings test exerts no robust influence on the labor supply decisions of men. Neither graphical analyses of breaks in labor supply trends, nor regression estimates that control for underlying trends in labor supply by age group, reveal any significant impact of changes in earnings test parameters on aggregate employment, hours of work, or earnings for men. For women, there is some suggestive evidence that the earnings test may be affecting labor supply decisions. Second, loosening the earnings test appears to accelerate benefits receipt among the eligible population, lowering benefits levels, and heightening concerns about the standard of living of these elderly at very advanced ages Our findings are focused on changes in earnings test parameters, and not on the removal of the test altogether, so it is possible that there are nonlinear effects of removing the test that are not captured by our framework. But the results nevertheless suggest some cause for caution before rushing to relax or remove the earnings test at younger ages.

Our paper proceeds as follows. The first section briefly describes the earnings test and its history. The second section summarizes the literature on the labor supply effects from the test. The third section presents our methodology for examining behavioral impacts, and the fourth section presents simple graphical evidence on the effects of the earnings test as well as more formal econometric results. The fifth section concludes by summarizing our findings and their implications for policy.

## BACKGROUND ON THE EARNINGS TEST

When the Social Security system was created in 1935, the Act stipulated that *no* benefits would be paid to a beneficiary who had received "wages with respect to regular employment." Subsequent statutes have relaxed the rules significantly, allowing higher earnings and reducing benefits only partially for earnings above the limit. The earnings test as it currently operates has three components: the earnings thresholds above which benefits are reduced, the percentages by which benefits are reduced for earnings exceeding those thresholds, and the increase in future benefits that is designed to compensate (on a lifetime basis) for the benefit reduction while working.

Some of the more important changes to the earnings test are presented in Table 1. Figure 1 also graphs against time the earnings test thresholds for workers age 62–64, 65–69, and 70–71, in constant (2000) dollars, for the 1973–1998 period used for our study. The thresholds for all three groups were identical until 1978, when there was a significant jump in the thresholds for those age 65 and over. Then, in 1983, the earnings test was removed for those aged 70 and 71. Finally, starting in 1996, the threshold was rapidly increased in real terms (and then removed altogether in 2000) for those aged 65–69. Relative to those aged 62–64, the earnings threshold for those aged 65 and over rose by 34 percent from 1978 to 1981, by another 4 percent by 1995, and then by more than 20 percent over the next three years.

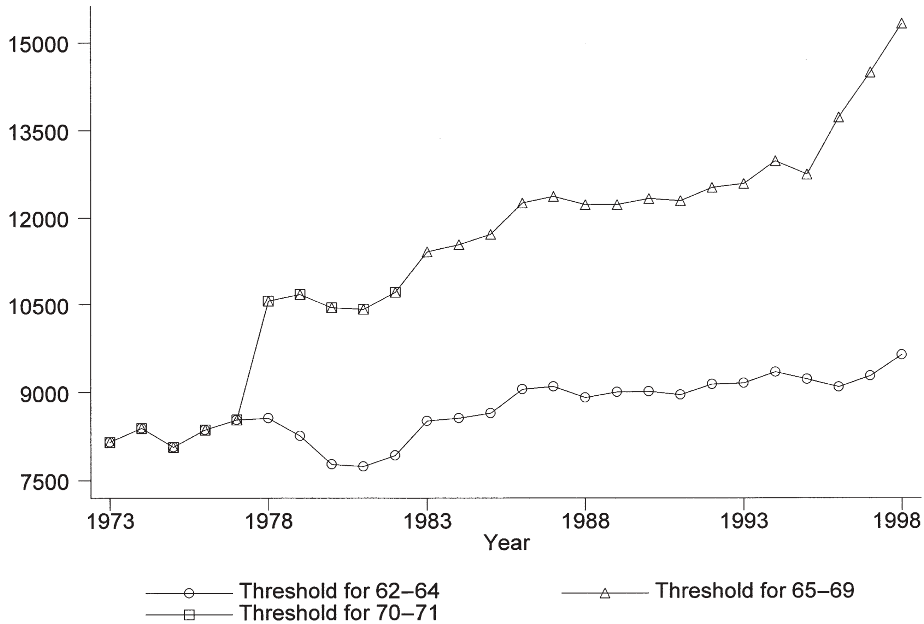
In 2000, the earnings test was removed for those at or above the normal retirement age (currently 65 years and 4 months, and

**TABLE 1**
MAJOR CHANGES IN THE EARNINGS TEST

| Year effective | Change |
|---|---|
| 1940 | Beneficiaries allowed to earn up to $14.99 per month without reduction in benefits |
| 1951 | Beneficiaries aged 75 or older exempted from the earnings test, other beneficiaries allowed to earn up to $50 per month without reduction in benefits |
| 1955 | Beneficiaries aged 72 or older exempted from the earnings test |
| 1961 | Initial benefit reduction rate $1 for every $2 of earnings above $1,200 a year but below $1,500 per year (and $1 for every $1 of earnings above that) |
| 1973 | Initial benefit reduction rate $1 for every $2 of earnings for all earnings above threshold of $2,100 (i.e., removal of 100 percent benefit reduction rate) |
| 1978 | Earnings test threshold for those at or above the normal retirement age increased above the threshold for those age 62–64 |
| 1983 | Beneficiaries age 70 and above exempted from the earnings test |
| 1990 | Initial benefit reduction rate $1 for every $3 of earnings above the threshold for those at or above the normal retirement age |
| 1996 | Earnings threshold for those at or above normal retirement age increased in real terms by statute |
| 2000 | Beneficiaries at or above normal retirement age exempted from the earnings test |

Note: Source—Social Security Administration, *Annual Statistical Supplement*, Table 2.A.29.

*Figure 1.* Real Thresholds Over Time



scheduled to increase to 67 by 2022). The earnings test thus now applies only to beneficiaries below the normal retirement age.[3] Under current rules, Social Security beneficiaries below the normal retirement age can earn up to a threshold amount without any reduction in retirement benefits. The threshold in 2000 was $10,080.[4] If beneficiaries earn more than the threshold amount, their current–year Social Security benefits are reduced. Under the earnings test that still applies to those under the normal retirement age, benefits are reduced by $1 for every $2 of earnings over the earnings threshold. For example, a beneficiary aged 63 who earns $12,080 would be earning $2,000 more than the limit of $10,080 and his benefit would be reduced by $1 for every $2 of earnings above the limit. His benefit would therefore be reduced by $1,000. As Table 1 shows, the initial benefit reduction rate for those at or above 65 was reduced in 1990, with the rate applying to those aged 65 and over falling from 50 to 33 percent.

Any such benefit reduction, however, would be refunded to the beneficiary through a benefit increase later in life. Specifically, any benefits reduced for the three years before the normal retirement age (currently 62–64; ultimately, under current legislation, 64–66) are subsequently increased by 6.67 percent of one's Primary Insurance Amount (PIA) for each year of full benefits lost due to the earnings test;[5] benefits are increased by 5.5 percent of the PIA for each dollar lost to the earnings test after the normal retirement age (rising to 8 percent by 2008). These adjustments, paralleling the adjustments for delayed benefits claiming, are roughly actuarially fair.

Despite this subsequent benefit adjustment, the earnings test is widely viewed as a pure tax, likely due to misinformation about how it functions. Most popular tax guides do not note the subsequent adjustment in benefits under the earnings test. *J.K. Lasser's Your Income Tax 1998* guide, for example, warned readers that if "you are under age 70, Social Security benefits are reduced by earned income," but did not note the subsequent benefit adjustment.

## EXISTING LITERATURE ON THE EARNINGS TEST

### *Theory*

Most analyses assume that the earnings test represents a pure tax, which is consistent with the popular misunderstanding of how the earnings test functions. Given that assumption, (which we also adopt in this paper), theoretical analysis of the effects of the earnings test on labor supply is straightforward (see, e.g., Vroman (1971), pp. 1–9). The natural presumption that the earnings test reduces labor supply reflects the substitution effect of the high implicit (perceived) tax rate on earnings in the earnings test range. This perceived tax compounds the substitution effect from existing payroll and income taxes. For example, for a 62 year old worker in the 15 percent federal bracket in the state of Massachusetts, the marginal tax rate on earnings above the threshold would be almost 80 percent.[6]

[3] A slightly different test applies in the year during which beneficiaries reach the normal retirement age. In that year, earnings before the month in which the beneficiary reaches the normal retirement age are subjected to a different earnings test than the one that applies at younger ages.

[4] This limit is raised each year by the percentage that (lagged) average wages have risen, as computed by the Social Security actuaries.

[5] Benefits reduced in the fourth or fifth year before the normal retirement age would subsequently be increased by 5 percent of the Primary Insurance Amount.

[6] This statement presumes that the worker also treats the Social Security payroll tax as a pure tax, which may not be true due to tax–benefit linkages later in life; see Feldstein and Samwick (1992) for a further discussion of these linkages and their implications for net payroll tax rates.

But this substitution effect is offset by several considerations. First, these high tax rates also have income effects, which naturally offset the reduction in labor supply. When the earnings test is loosened, marginal tax rates fall, but effective income rises as well. Moreover, as emphasized by Disney and Tanner (2000) in their study of the removal of the earnings test in the U.K., lifting the earnings test can lead to more claiming of benefits by those already working, which has further negative (cash) income effects on work. On net, then, the impact of the earnings test on labor supply is theoretically ambiguous. Only empirical evidence on its effects can inform policy–makers as to the importance of this test for the labor supply decisions of older workers.

One particularly important question is the impact on the earnings test on the intensive vs. extensive margins of labor supply. With no impediments to the adjustment of hours worked to their ideal level, the earnings test should impact only hours of work, but not the work decision. But, if there are rigidities such as fixed schedules which make hours impossible to adjust perfectly, then the earnings test can push workers from work to retirement. This latter response is of particular concern to policy–makers, as highlighted in the introduction.

An additional behavioral response of some concern is benefits receipt. As highlighted by Gruber and Orszag (1999), loosening the earnings test may lead more individuals to claim benefits early. As Coile, Diamond, Gruber and Jousten (1999) show, it is optimal to delay claiming beyond the date of initial benefits entitlement for a wide class of preference parameters, since benefits are actuarially adjusted for delay.[7] "Excess" early claiming could arise for two reasons. The first is a lack of understanding of the value of delayed claiming in terms of benefit increases, which would be consistent with the parallel misunderstanding of the functioning of the earnings test. The second is time–inconsistent behavior, whereby individuals claim early due to high short–run discount rates, but regret this decision at the end of their life due to lower long–run discount rates. This argument is developed by Diamond and Koszegi (1998).

Moreover, Coile, Diamond, Gruber and Jousten (1999) show that, among those who are eligible for full benefits under the existing earnings test, claiming is very high. In particular, among those who have earnings at their 62$^{nd}$ birthday below the earnings test threshold, roughly 90 percent claim within a year of turning 62. Among those whose earnings are below this level after their 62$^{nd}$ birthday, claiming is even higher, with essentially no one who retires from age 65 onwards delaying claiming.

The concern raised by these findings is that loosening the earnings test would lead to increases in early claiming and, as a result, lower living standards later in life. This concern is heightened by the fact that the average annual benefit among older widows whose spouses had claimed early benefits was slightly below the poverty line in 1998, whereas the average annual benefit among older widows whose spouses had not claimed early benefits was more than $1,800 above the poverty line.[8] Similarly, Social Security Administration (SSA)(1999) finds that if elimination of the earnings test at age 62 induced all beneficiaries to claim at that age, elderly poverty would increase by roughly 700,000 people, or from 12.0 percent of the relevant population to 13.9 percent. In es-

[7] The reward to delay (in terms of the value of higher benefits) is increased if one incorporates the fact that Social Security provides a real annuity that cannot be purchased in the private market.

[8] Some, but not all, of this mean differential reflects differences in covered earnings and therefore benefits across the two groups. Roughly one–third of the mean difference between the two groups remains after controlling for primary insurance amounts.

sence, then, a looser earnings test potentially leads to higher living standards among the young old (who can work and receive benefits at the same time) but lower living standards among the oldest old, among whom elderly poverty is concentrated.[9]

### *Past Evidence*

A sizeable literature exists regarding the labor supply effects of the earnings test, and this literature fairly uniformly concludes that there is a small impact of the earnings test on labor supply. Leonesio (1990) summarizes this established view; he writes that "numerous scholarly studies have examined the effect of the Social Security retirement test on the labor supply of older workers. Virtually all of this research indicates that the effect is probably small and that eliminating the test would have a minor impact on the work activity of older Americans."

There is clear evidence from past work that workers tend to bunch immediately at or below the earnings test threshold, and the bunching moves as the threshold does, suggesting some effect from the earnings test on labor supply; recent convincing evidence on this point is found in Friedberg (2000). But this approach does not speak to a potentially important question of interest: the aggregate labor supply impact of the earnings test. As Gruber and Orszag (1999) point out, the number of bunched workers is quite small relative to the elderly population. For example, the clustering for 66–69 year olds represents only 0.4 to 0.7 percent of the elderly population, and only 1.6 to 2.9 percent of elderly workers. Moreover, the net impact on earnings could be very small if this small subset of bunched workers would raise their earnings by only one or two thousand dollars in the absence of the earnings test.

Another class of studies in this literature relies on a structural approach that uses the kinked budget constraint created by the initial benefit reduction schedule to identify the labor supply effects of the earnings test. Vroman (1971) and Pellechio (1978) adopt this approach. In an important recent study that significantly advances this type of analysis, Friedberg (2000) revisits non–linear budget constraint estimation using the 1983 change in the earnings test that eliminated it for those age 70 and over. Her results suggest statistically significant impacts of the earnings test on labor supply, but they are modest in magnitude. For example, removing the earnings test would raise hours of work by those at or above the earnings test level by 5.3 percent. Those 65–69 year olds at or above the earnings test level in 1998, the last year of our data, worked 64 percent of the hours worked by all 65–69 year olds. So this implies an aggregate hours effect of 3.4 percent of removing this test.

Even this small estimated effect, moreover, is subject to some limitations. First, despite the improvements over previous non–linear budget constraint estimates, Friedberg must still assume that the gross wage which drives the shape of the budget constraint is exogenous. While some of her specification tests suggest that this is the case, others (such as the fact that her model yields very similar answers if simply estimated on the period after the "natural experiment" in question) do not. Second, by necessity, given the modeling framework employed, she does not examine the effect of the earnings test on the decision to work. If there is an effect on the decision to work, it could bias her findings through compositional effects on the

[9] This argument presumes that increases in benefits receipt would be consumed, not saved. This presumption is supported by recent research which suggests a strong connection between the consumption of retirees and their post–retirement income replacement rate (Bernheim, Skinner and Weinberg, 2001); the evidence is suggestive that higher income replacement is not saved, but rather consumed.

sample that is working. For example, if the persons joining the labor force when the earnings test is removed have higher than average hours of work, average hours of work among those in the labor force will rise mechanically and not through any behavioral response.

The evidence from elimination of earnings tests in other countries is much stronger than in the U.S. Baker and Benjamin (1999) examine the impact of removal of the earnings test under Canada's public pension plans in the mid–1970s, exploiting the fact that the earnings test was changed sequentially in Quebec and in the rest of Canada (which sets up natural treatment and control groups). They find no impact on the retirement decision, but an increase in weeks worked conditional on working. They also find a very large rise in benefits claiming. Disney and Tanner (2000) study the abolition of the U.K. earnings test in October 1989. They also find no impact of this abolition on the decision to work, but a large impact on hours worked, with hours rising roughly 20 percent among the affected population when the test was removed. Unlike the Canadian case, however, they find no increase in claiming when the test was removed. This may be because the UK had a fairly generous actuarial adjustment for delayed claiming (7.5 percent per year of delay), while in Canada there was no actuarial adjustment.

To summarize, the older literature on the earnings test for the U.S. implies fairly small effects on labor supply, while the literature for other nations suggests larger impacts. The best recent U.S. evidence from Friedberg (2000) implies significant effects, albeit modest ones, of complete removal of the earnings test. But this past evidence suffers from five limitations. First, the "bunching" literature does not provide insights into the aggregate labor supply impacts of the earnings test on the elderly. Second, the structural estimation of Friedberg (2000) and others requires assumptions on endogeneity of the wage rate (that may or may not be met) in order to identify the estimates. Third, there has been little work in the U.S. on the effects of the earnings test on the decision to work. Fourth, previous work has generally not examined the impact of the earnings test on women as well as on men. Finally, none of the work on the U.S. has examined the impact of the earnings test on the decision to receive benefits, which, as we argue above, is an important parameter for welfare analysis of earnings test reforms. We propose an approach below that tries to address these issues in a straightforward manner.

## EMPIRICAL STRATEGY

Our empirical strategy is to use the significant variation in earnings test parameters over the past 25 years to identify the aggregate impact of the earnings test on employment, weeks of work, earnings, and benefits receipt. As illustrated by Table 1 and Figure 1, there have been several notable relative changes in the earnings test across groups subject to the test: the 1978 relative upward shift in the test threshold for those over 65 relative to those under 65; the 1983 elimination of the test for those ages 70 and 71; the 1990 reduction in the tax rate for those ages 65–69; and the 1996–1998 dramatic increase in the threshold for those ages 65–69. In addition, there have been more minor movements in the real value of the thresholds over time. Thus, our strategy is to model aggregate behavioral responses for different age groups of elderly persons to the changes in the relative level of the earnings test that they face.

The key advantage of this approach is that we do not impose any structural assumptions on the underlying estimates in order to obtain our effects, but rather simply examine how aggregate labor supply (and benefit receipt) responds to these significant changes. The disadvantage of the approach is that we are not able to identify income and substitution effects separately, nor are we able to examine effects at an individual worker level. Since our focus is the aggregate labor supply and benefit receipt responsiveness to the earnings test, the advantages appear to us to outweigh the disadvantages.

### *Data*

Our data for this exercise come from the 1974–1999 March Supplements to the CPS. These supplements gather data on labor force attachment in the previous calendar year, so our sample period runs from 1973 through 1998. From each supplement, we computed year by sex by age (or by age group) cell means on several key indicators of labor supply: employment in the previous year; total hours worked last year; total earnings last year; and hours and earnings, conditional on being positive.[10] The total hours and earnings measures do not condition on work, thereby allowing employment to respond to the earnings test changes without inducing sample selection bias in our estimates. (The cost of this approach is that if employment does not respond, we potentially mask responses within the working population by including a large mass of non–workers who are not responding to the policy change.) We also measure receipt of Social Security income in the previous year to capture benefit receipt behavior.

One limitation of the CPS data is that they ask about age in March, while our labor supply and income measures of interest refer to the previous year. For our regression analysis, we therefore take the conservative approach of discarding those ages for which ambiguity exists about the earnings test regime that workers faced in the previous year (ages 62, 65, 70, and 72). We also explored an alternative where we assume that the typical person is born in October, so that they are on average between birthdays when interviewed in March, and take an appropriately weighted average of the earnings test by the share of the previous year that the person is of each age. The results from doing so are almost identical to those reported below, with only slightly more precisely estimated standard errors. We therefore focus on our "cleaner" results.

The means of our data for the entire 1973–1998 period, by age group and sex, are shown in Table 2. A clear trend exists toward reduced labor supply and more Social Security recipiency across ages. There is non–trivial Social Security recipiency even among those under age 62, despite the fact that the CPS question is designed to elicit recipiency of retirement benefits only (and not disability benefits). This may reflect either misrepresentation of disability insurance benefit (or other program benefit) receipt, or (particularly for women) survivor's benefits. In any case, it is for reasons such as this potential age–specific misreporting that our model will include a full set of age dummies, and our central specifications will also include age–specific trends to capture any changes through time in this reporting behavior.

Among those who work, hours and earnings decline more modestly with age. Women work significantly less than men at every age, but even over this long period and even at older ages, they are a non–trivial share of labor force partici-

[10] Our hours measure is somewhat imprecise because the CPS only began asking about hours worked per week in the previous year in the March 1981 survey, and before that we only have data on hours worked in the survey week. Hours in the survey week are highly correlated with usual hours last year in the years for which both are available (correlation over time and across age for men = 0.995), so this should not be an important issue. Moreover, we have replicated all of our findings here using weeks worked last year, and they are similar.

**TABLE 2**
MEANS BY SEX AND AGE GROUP

| | Males | | | | | |
|---|---|---|---|---|---|---|
| | All Ages | 59–61 | 63–64 | 66–69 | 71 | 73–75 |
| Work Last Year | 0.431 | 0.803 | 0.627 | 0.391 | 0.276 | 0.195 |
| | +(0.216) | (0.064) | (0.108) | (0.094) | (0.068) | (0.063) |
| Hours Worked | 647 | 1498 | 943 | 423 | 266 | 179 |
| | (512) | (99) | (94.1) | (86.3) | (47.4) | (54.5) |
| Hours Worked if > 0 | 1514 | 2036 | 1807 | 1372 | 1246 | 1185 |
| | (353) | (65.3) | (115) | (108) | (103) | (129) |
| Earnings | 7102 | 16918 | 11485 | 4535 | 2533 | 1714 |
| | (5630) | (1452) | (1992) | (1240) | (737) | (627) |
| Earnings if > 0 | 14595 | 21395 | 18626 | 12045 | 9578 | 9091 |
| | (5195) | (1730) | (1654) | (2461) | (2278) | (2295) |
| SS Recipiency | 0.669 | 0.102 | 0.50 | 0.822 | 0.886 | 0.906 |
| | (0.314) | (0.026) | (0.10) | (0.127) | (0.130) | (0.137) |
| | Females | | | | | |
| | All Ages | 59–61 | 63–64 | 66–69 | 71 | 73–75 |
| Work Last Year | 0.255 | 0.501 | 0.361 | 0.204 | 0.131 | 0.080 |
| | (0.156) | (0.045) | (0.036) | (0.038) | (0.021) | (0.027) |
| Hours Worked | 318 | 783 | 470 | 198 | 109 | 65.4 |
| | (284) | (123) | (78.6) | (56.4) | (23.5) | (26.3) |
| Hours Worked if > 0 | 1235 | 1632 | 1454 | 1123 | 990 | 1001 |
| | (276) | (60) | (80.5) | (100) | (122) | (148) |
| Earnings | 2140 | 5050 | 3176 | 1242 | 605 | 377 |
| | (1896) | (998) | (594) | (418) | (178) | (181) |
| Earnings if > 0 | 7171 | 10269 | 8986 | 6133 | 4744 | 4666 |
| | (2597) | (1245) | (1155) | (1319) | (1087) | (1414) |
| SS Recipiency | 0.689 | 0.116 | 0.60 | 0.829 | 0.859 | 0.859 |
| | (0.298) | (0.036) | (0.11) | (0.150) | (0.161) | (0.155) |

Note: Author's tabulations of CPS data described in text; standard deviations in parentheses.

pants. The role of females in the labor force highlights the value of examining the response of women as well as men to the earnings test.

### *Regression Framework*

To examine the aggregate responses to major changes in the earnings test, we use data on mean behavior by age to estimate models of the form:

$$[1]\quad Y_{at} = \alpha + \beta_1 THRESH_{at} + \beta_2 TDUM_{at} + \beta_3 X_{at} + \beta_4 \delta_a + \beta_5 \tau_t + \varepsilon_{at},$$

where $a$ indexes ages and $t$ indexes years; $Y$ is one of our measures of labor supply or claiming behavior; $THRESH$ is the real earnings test in 1982 dollars threshold for that age group in that year (set to zero for groups not subject to the test); $TDUM$ is a dummy for whether the age group is subject to an earnings test (1 = yes, 0 = no); $X$ is a set of control variables ( percent white, black; percent high school dropout, high school graduate, some college; percent married, never married, divorced, separated; percent veteran; and dummies for each of the nine U.S. divisions); $\delta_a$ is a full set of age dummies; and $\tau_t$ is a full set of year dummies. We also estimate regression models where we add the age–specific tax (initial benefit reduction) rate, $TAX_{at}$.

This regression controls for secular differences across ages and over time to parse out the impact of changes in the earnings test parameters on behavior. But one limitation of this approach is that age–group specific time trends in behavior could confound the analysis: there are strong differential age trends in labor supply and benefits receipt. We cannot allow for fully non–parameteric age group–specific time effects, since relative changes over time in the earnings test are the source of identification for our model. But we do explore the sensitivity of our model to the inclusion of age group–specific linear and quadratic time trends. Of course, one limitation of adding extra controls is that it reduces the identifying variation for earnings test effects, reducing the precision of our estimates. We therefore compromise on the model with age–specific linear trends as our base specification. Since the underlying means and trends in behavior are so different for men and women, we estimate our models separately by sex.

Our regression framework implicitly interacts the *THRESH* variable with the *TDUM* variable, since we set *THRESH* to zero when there is no earnings test. This implies that the correct interpretation of the effect of removing the earnings test depends on the level of the threshold before it was removed. That is, the full effect of removing the earnings test is $-\beta_1 THRESH_{at} - \beta_2$. The threshold for the earnings test was at \$5917 in real terms for 70–71 year olds before it was removed in 1983. So, when evaluating the impact of this change, we will interpret the effect relative to that preexisting level.

## EVIDENCE

### Graphical Evidence

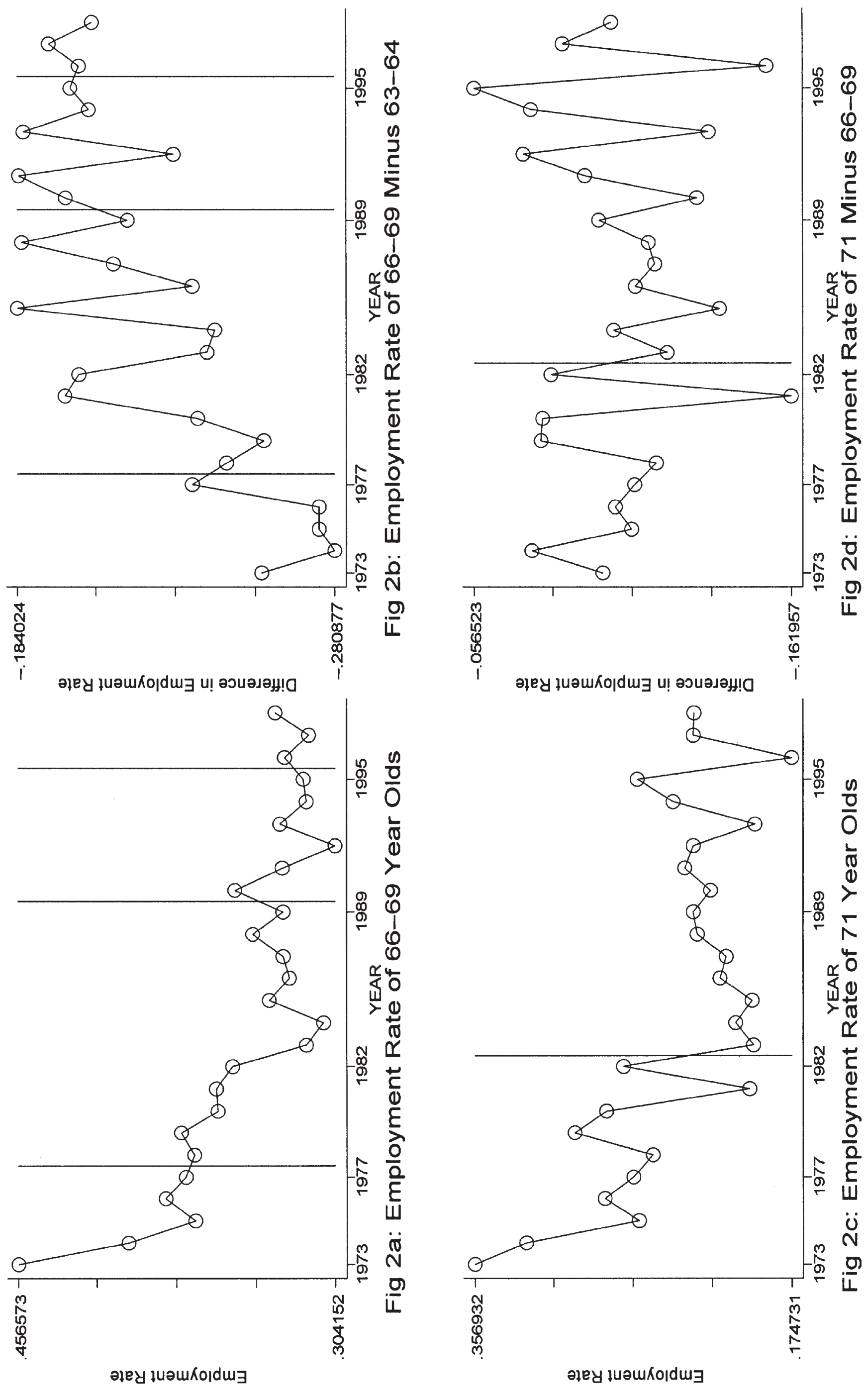
We begin with some simple graphical evidence to illustrate the patterns in the data that underscore our main conclusion below: there is little immediate labor supply response to changes in earnings test parameters. Figures 2a–2b show the evolution of employment for 66–69 year olds over time. Figure 2a shows the absolute employment trends for this group, which are steadily declining until the mid–1980s, then flatten. From this figure, there is little evidence of a major break in employment trends when the earnings test parameters are changed in 1978, 1990, or 1996. Figure 2b shows the trend for 66–69 year olds relative to 63–64 year olds. In relative terms, there is once again no glaring evidence of a relative rise in employment when the earnings test is loosened over time. There is a large jump in 1990, when the tax rate is lowered, but it is followed within one or two years by a very steep decline that eliminates the gains. Overall, little evidence exists of a break from the general upward trend in employment of 66–69 year olds relative to this other age group over time.

Figures 2c–2d pursue a similar analysis for 71 year old men, focusing on the 1983 removal of the earnings test for this group. In absolute terms, there is some potential indication of a rise in employment after the 1983 change; there is a fall in 1983, but a subsequent slow rise relative to a pre–existing declining trend. But, relative to 66–69 year olds, there is no apparent break in 1983. The lack of any aggregate response for those 70 and above to the 1983 removal of the earnings test is consistent with the conclusions reached by Packard (1990).

### Regression Results for Men

The graphs presented above highlight the importance of long run behavioral trends, even in relative measures of labor supply and benefits receipt across age groups. This suggests that it is important to control for age–specific time trends which may otherwise confound the analysis; we do so in our base specification below.

*Figure 2.* Employment Rates Over Time—Men



Our regression results for men are presented in Table 3. We show in the first two columns our estimates from estimating equation [1], for each of our dependent variables of interest. The first column shows the coefficient on the threshold level, and the second on a dummy for whether the earnings test is present. As noted above, we show the implied total effect of threshold removal, at its 1982 level for 70–71 year olds, along with the associated standard error.

The second set of columns then adds to the regression specification a separate linear trend by age. This allows for a slowly evolving age–specific trend in labor supply or benefits claiming decisions. Finally, the third set of columns includes a quadratic trend for each age, which allows for a more flexible modeling of the baseline trends in work and claiming behavior over time. The coefficients on the threshold level are multiplied by 1,000, so that they show the impact of a $1,000 increase in the threshold.

We focus first on employment decisions. Consistent with the evidence from the U.K. and Canada, we find no impact of earnings test parameters on decisions to work for men. In the first set of columns, we find that higher thresholds lead to more work, as expected, but the estimate is not significant. In the next two sets of columns, the coefficient on the threshold switches signs, with the potentially perverse implication that the earnings test encourages work, although the standard errors rise so that the coefficient remains statistically insignificant. In all cases, the coefficient on the threshold dummy is perversely signed and insignificant.

Our confidence intervals for removing the threshold are particularly large, which continues to be true throughout the re-

**TABLE 3**
BASIC RESULTS FOR MEN

| | Basic Model | | Age–Specific Linear Trends | | Age–Specific Quadratic Trends | |
|---|---|---|---|---|---|---|
| Dependent Variable | Threshold Level/1000 | Threshold Dummy | Threshold Level/1000 | Threshold Dummy | Threshold Level/1000 | Threshold Dummy |
| Work Last Year | 0.0019 (0.0025) | 0.003 (0.023) | –0.0046 (0.0068) | 0.030 (0.058) | –0.0087 (0.0078) | 0.030 (0.070) |
| Hours Worked | 12.9 (5.2) | 3.97 (48.7) | –22.7 (14.8) | 30.6 (127) | –4.2 (17.0) | 273 (151) |
| Hours Worked if >0 | 5.5 (9.5) | –54.7 (93.4) | –12.3 (24.9) | –18.2 (220) | –25.5 (26.5) | 17.0 (243) |
| Earnings Last Year | 238 (72.2) | –47.0 (677) | 11.9 (209) | 365 (1790) | –410 (233) | 922 (2075) |
| Earnings if > 0 | 272 (145) | –269 (1424) | –21.5 (414) | –266 (3677) | –439 (456) | 809 (4173) |
| Receive SS Benefits | 0.0069 (0.0026) | –0.012 (0.024) | 0.0119 (0.0065) | –0.028 (0.056) | 0.0159 (0.0068) | –0.041 (0.057) |
| Number of Obs | 338 | 338 | 338 | 338 | 338 | 338 |

Notes: Standard errors in parentheses. Estimates are from model such as equation [1] in text. Each set of estimates in columns (1) and (2), (3) and (4), and (5) and (6), come from separate models, and each row is from a separate model. First set of columns include all regressors from equation [1], but no age–specific trend terms. Second set of columns include age–specific linear time trend terms. Third set of columns include age–specific quadratic time trend terms. For threshold dummy, results show implied effect of removing threshold at 1982 level for 70–71 year olds, and standard error on this implied effect.

mainder of the results. This is not surprising given the potentially tenuous identification of the threshold dummy, which relies solely on one change for one age group (age 71). Indeed, these estimates are sufficiently imprecise that we do not focus on them at all in the remainder of the discussion. Thus, the discussion below is focused solely on the estimates for the level of the threshold, conditional on their being an earnings test. This does have the important implication that, to the extent that there are nonlinear responses to removing the threshold, we cannot identify them very effectively in this framework. Rather, our models are more precisely focused on the implications of changes within the existing earnings test regime.

In addition, these results show the sensitivity of our findings to the inclusion of age–specific time trends. While we continue to show the results below both with and without time trends, it seems clear to us from the graphical analysis that the results with trends are the most reliable. On the other hand, the results with quadratic time trends are particularly imprecise, and the conclusions generally agree with what we find in the linear time trend case. As a compromise, we focus our attention on the middle set of columns of Table 3, which include the linear age–specific time trends.

We next turn to hours of work, both unconditionally, and conditional on some labor force attachment. For both of these variables, the effects of the earnings test are wrong–signed and insignificant. Similar results obtain for total earnings and earnings conditional on positive earnings.

Thus, our conclusion is that there is no robust evidence for a large immediate labor supply response to changes in the level of the earnings test, once one allows for either linear or quadratic trends in labor supply by age. Given the strong underlying trends through time in relative labor supply by age group shown in the earlier figures, models that control for such trends would appear to be more reliable. On the other hand, such models have the disadvantage that the resultant estimates are not very precise, so that we cannot rule out with our estimates modest responsiveness to changes in the earnings test threshold. In addition, our models examine only the immediate impact of the earnings test changes. It is possible that the changes have more substantial labor supply effects after a lag.

We next turn to recipiency of Social Security benefits. Here, in contrast to the labor supply measures, we find evidence that loosening the earnings test leads to increased benefits recipiency, and the results are actually strengthened by the inclusion of controls for age–specific trends. Our estimates from the age–specific linear trend specification in the middle columns indicate that a $1,000 rise in the earnings test threshold would increase the share of the elderly receiving Social Security benefits by 1.19 percent. Thus, we do find robust evidence that the level of the threshold matters for benefits receipt. This finding is consistent with the notion that individuals do not understand the actuarial adjustment implicit in the benefits reduction of the earnings test.

Our models thus far have excluded any control for the earnings test tax rate, which we incorporate in Table 4 for the model with age–specific linear time trends. The conclusion from these models is very similar to that from the middle columns of Table 3; the coefficient on the threshold level is either wrong–signed or insignificant (or both). The tax rate itself is also insignificant with respect to these labor supply and earnings measures. For Social Security recipiency, we continue to find a positive and (marginally) significant coefficient on the threshold level, and we obtain an insignificant coefficient on the tax rate. Thus, it appears that benefit receipt decisions are sensitive to the threshold, but not to the tax rate beyond that

**TABLE 4**
INCLUDE TAX RATE—MODEL WITH AGE–SPECIFIC TRENDS

| | Threshold Level/1000 | Threshold Dummy | Tax Rate |
|---|---|---|---|
| Work Last Year | –0.0093 | 0.025 | 0.094 |
| | (0.0075) | (0.073) | (0.062) |
| Hours Worked | –27.3 | 25.6 | 92.4 |
| | (16.2) | (149) | (134) |
| Hours Worked if > 0 | 23.2 | 118 | –286 |
| | (27.5) | (263) | (228) |
| Earnings Last Year | –40.0 | 315 | 1042 |
| | (229) | (2153) | (1887) |
| Earnings if > 0 | 170 | –92.1 | –3943 |
| | (454) | (4410) | (3863) |
| Receive SS Benefits | 0.0132 | –0.026 | –0.025 |
| | (0.0071) | (0.067) | |
| (0.059) | | | |
| Number of Obs | 338 | 338 | 338 |

Notes: Standard errors in parentheses. Estimates are from model such as equation [1] in text, with inclusion of earnings test tax rate and age–specific linear time trend terms. For threshold dummy, results show implied effect of removing threshold at 1982 level for 70–71 year olds, and standard error on this implied effect.

threshold level (at least to the extent that it varies in our sample, between 33 and 50 percent).

Another possible confounder of our findings could be the business cycle. For example, in 1983, the economy was emerging from the early 1980s recession. If the oldest workers were the least sensitive to the business cycle in their work decisions, this could explain the lack of labor supply response of 71 year olds to the lifting of the earnings test (which identifies the dummy for the presence of an earnings test). This age–specific effect of business cycle conditions would not necessarily be captured by age–specific trends. To control for this alternative hypothesis, we have reestimated all of our models including an interaction of the sex–specific adult unemployment rate with the full set of age dummies; this model controls for any age–specific sensitivity to cyclical conditions. Our estimates are very similar when these controls are included.

### *Regression Results for Females*

As noted earlier, previous work on the earnings test has focused largely on males. But females represent a non–trivial share of older workers, even in our sample period. In 1970, for example, women represented 42 percent of all fully insured workers between the ages of 65 and 69; by 1998, this ratio had risen to 47 percent (SSA, 1999). Therefore, it is important to assess the impact of the earnings test on female, as well as male, behavior. We do so by estimating the same models as above for women (with the exclusion of the control for veteran status).

Table 5 replicates for women our findings from Table 3, for the model with age–specific linear time trends. For employment, in contrast to men, we find an economically significant impact of variations in the earnings test threshold, but the coefficient is not statistically significant. For hours, the coefficient is not of the expected sign. For earnings, the coefficient on the

**TABLE 5**
BASIC RESULTS FOR WOMEN

| | Age–Specific Linear Trends | |
|---|---|---|
| Dependent Variable | Threshold Level/1000 | Threshold Dummy |
| Work Last Year | 0.0065<br>(0.0051) | –0.007<br>(0.044) |
| Hours Worked | –7.8<br>(8.7) | –14.3<br>(74.5) |
| Hours Worked if > 0 | –21.4<br>(26.6) | –9.69<br>(232) |
| Earnings Last Year | 167<br>(78.4) | –83.4<br>(670) |
| Earnings if > 0 | 26.0<br>(224) | 120<br>(1957) |
| Receive SS Benefits | 0.0334<br>(0.0063) | –0.002<br>(0.054) |
| Number of Obs | 338 | 338 |

Note: Standard errors in parentheses. Estimates are from model such as equation [1] in text. Each set of estimates in columns (1) and (2) comes from separate models, and each row is from a separate model. All results include age–specific linear time trend terms. For threshold dummy, results show implied effect of removing threshold at 1982 level for 70–71 year olds, and standard error on this implied effect.

threshold level is significant and of the expected sign, suggesting that each $1,000 rise in the earnings test threshold increases mean earnings for women by $167. Much of this large effect appears to arise through the (statistically insignificant) employment effects that we document in the first row; for earnings conditional on work, the effects are much weaker.

For Social Security recipiency, we once again see significant and sizeable impacts of the threshold level. Indeed, the coefficient is quite large, suggesting that each $1,000 rise in the earnings test would raise recipiency by 3.34 percentage points.

Thus, for women, the story is more mixed. There is a significant impact of the earnings test on earnings, but not on employment or hours. We also once again confirm the finding that Social Security benefit recipiency appears to be sensitive to earnings test parameters.

## CONCLUSIONS AND IMPLICATIONS

While fundamental reform of the Social Security system continues to be the subject of heated debate with no legislative action, the Federal government has recently enacted—with little fanfare—the largest reform of the Social Security system in almost 20 years, removing the earnings test for those at or above the normal retirement age. This removal is unlikely to end the debate over the value of the earnings test, which is still in force at the early retirement age of 62, for two reasons. First, as the normal retirement age rises to 67, more retirees will continue to face the earnings test. Second, the baby boom will shortly enter the ages at which the earnings test at 62 begins to bind, and the political pressure that succeeded in removing the test at the normal retirement age is likely to be replicated for removal at age 62.

While our analysis is not specifically focused on removal of the earnings test, our findings suggest some cause for caution before rushing to remove the earnings test at younger ages, for two reasons. First, using more recent data, we confirm the findings of the previous literature that suggests no large labor supply response of men to earnings test parameter changes. While our confidence intervals are large, our straightforward empirical framework provides a number of advantages for assessing the behavioral implications of the earnings test. On the other hand, we do find some suggestive evidence of a labor supply response among women, with earnings rising significantly as the threshold is increased.

Second, we provide the first confirmation for the U.S. that changing the earnings test may increase early recipiency of Social Security benefits. Throughout our analysis, we find highly significant evidence of a response of benefit recipiency among both men and women to loosening the earnings test. As highlighted earlier, increased benefit recipiency at younger ages could translate to higher poverty rates at older ages. According to SSA (2000), in 1998, 48 percent of insured workers claimed benefits within one month of turning 62. They find that if the fraction of those delaying past age 62 were reduced by 20 percent, so that 58 percent claimed benefits at age 62, elderly poverty rates would rise from 12.0 to 12.4 percent. Our findings suggest that each $1000 rise in the earnings threshold raises claiming by 2.3 percentage points (averaging the male and female benefits receipts results). Thus, interpolating the SSA figures, our results imply that each $1000 rise in the earnings test threshold would lead to an increase in the number of elderly living in poverty of 32,000, including an increase in the number of poor widows (and widowers) of 18,000. The SSA results, combined with our results, thus suggest that loosening the earnings test at age 62 could produce a non–trivial increase in elderly poverty.

Our work has four potential limitations. First, we do not attempt to uncover the underlying structural mechanisms behind any labor supply response (or lack thereof) to changes in the earnings test. In that respect, we view this effort as largely complementary to the existing literature. Given the weight of the evidence from earlier studies, and our new findings from this very different framework, it now appears fairly clear that there is little short run response to earnings test changes among males. There is more of a suggestion of an effect on the labor supply of females, although it is present only for earnings and not for hours or employment.

Second, and related, this is by definition a short– to medium–run analysis. It is quite possible that the real impacts of the earnings test are in the long run. As new cohorts of workers reach retirement age, retirement norms may shift along with the earnings test, leading to larger long run effects. We have attempted in our models to assess the importance of this effect by using lagged values of the earnings test parameters; even with these lags we find no evidence of a significant labor supply response. But merely including lags is likely not sufficient to measure the true long–run effect of earnings test changes.

Third, none of the historical changes that we examine replicates the recent complete removal of the earnings test for those over age 65, nor provides any direct insight into the impacts of possible changes in the test for those ages 62 to 64. In particular, we were unable to estimate even reasonably precisely the impact of removing the earnings test for 71 year olds. Thus, our results apply strictly to changes in the threshold for the earnings test, and it is not clear that they can be extrapolated to complete removal.

Finally, we have not attempted to uncover how the effects of the earnings test vary across the population. Even if there are small effects in aggregate, there may be large impacts on particular subgroups of interest to policymakers. Given our fairly large confidence intervals, our approach does not seem a fruitful one for disentangling individual–specific responses, but this is a useful goal for future work in this area. Relatedly, our analysis does not consider how the response to the earnings test might differ over time, as the level of the test changes. The response to a $2000 increase in the threshold may differ depending on the base level from which that change occurs, but we do not have enough power in our approach to distinguish such differential effects. This could be true, for example, due to the point at which fixed costs of work cause workers to retire rather than to adjust their hours of work.

Welfare interpretation of these benefit receipt results is difficult in the absence of other evidence regarding who is claiming earlier and on how they use the extra income from earlier claiming. Further research is needed on the implications of earlier benefits receipt for standards of living. Does earlier recipiency lead to more consumption at younger ages and less at older ages, as is argued to be likely by Gruber and Orszag (1999), or do recipients smooth their consumption over their remaining lives? If the former, what does this imply for the standard of living of the oldest elderly? Nevertheless, the finding of no robust evidence of an immediate labor supply response, but clear evidence for earlier benefit receipt, appears to weaken the case for relaxing the remaining earnings test at younger ages.

#### Acknowledgments

We are grateful to Michael Frakes, Cindy Perry, and Diane Whitmore for excellent research assistance, and participants at conferences convened by the National Bureau of Economic Research and the Center for Retirement Research, along with Therese McGuire and two anonymous referees, for excellent comments. We gratefully acknowledge funding from the Social Security Administration through the Center for Retirement Research at Boston College. Gruber also acknowledges financial support from the National Institute on Aging and the National Science Foundation.

## REFERENCES

Baker, Michael, and Dwayne Benjamin.
"How Do Retirement Tests Affect the Labor Supply of Older Men?" *Journal of Public Economics* 71 No. 1 (January, 1999): 27–51.

Bernheim, B. Douglas, Jonathan Skinner, and Steven Weinberg.
"What Accounts for the Variation in Retirement Wealth Among U.S. Households?" *American Economic Review* 91 No. 4 (September, 2001): 832–57.

Coile, Courtney, Peter Diamond, Jonathan Gruber, and Alain Jousten.
"Delays in Claiming Social Security Benefits." NBER Working Paper No. W7318. Cambridge, MA: National Bureau of Economic Research, 1999.

Diamond, Peter, and Botond Koszegi.
"Hyperbolic Discounting and Retirement." Massachusetts Institute of Technology. Mimeo, 1998.

Disney, Richard, and Sarah Tanner.
"The Abolition of the Earnings Rule for UK Pensions." Institute of Fiscal Studies Unpublished Draft, London, 2000.

Feldstein, Martin, and Andrew Samwick.
"Social Security Rules and Marginal Tax Rates." *National Tax Journal* 45 No. 1 (March, 1992): 1–22.

Friedberg, Leora.
"The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics* 82 No. 1 (February, 2000): 48–63.

Gruber, Jonathan, and Peter Orszag.
"What to Do About the Social Security Earnings Test?" Issue in Brief Number 1, Center for Retirement Research, July 1999.

Lasser Institute, J.K.
*J.K. Lasser's Your Income Tax.* New York: MacMillan, 1997.

Leonesio, Michael.
"The Effects of the Social Security Earnings Test on Labor–Market Activity of Older Americans: A Review of the Evidence." *Social Security Bulletin* 53 No. 5 (May, 1990): 2–21.

Packard, Michael.
"The Effects of Removing the 70– and 71–Year–Olds from Coverage Under the Social Security Earnings Test." ORS Working Paper No. 44. Washington, DC: Office of Research and Statistics, Office of Policy, Social Security Administration, 1990.

Pellechio, Anthony.
"The Social Security Earnings Test, Labor Supply Distortions, and Foregone Payroll Tax Revenue." NBER Working Paper 272. Cambridge, MA: National Bureau of Economic Research, August, 1978.

Social Security Administration.
*Annual Statistical Supplement.* Washington, DC, 1999.

Social Security Administration. Office of Policy.
"The Impact of Repealing the Retirement Earnings Test on Rates of Poverty." http://www.ssa.gov/policy/pubs/policypapers/RET_repeal.html., 2000.

Vroman, Wayne.
"Older Worker Earnings and the 1965 Social Security Amendments." Research Report 38. Office of Research and Statistics, Social Security Administration, Washington, DC, 1971.