Do the Elderly Respond to Taxes on Earnings?

Evidence from the Social Security Retirement Earnings Test

Steven J. Haider* and David S. Loughran

Draft: July 2005

Abstract

The effective tax on earnings embodied in the Social Security retirement earnings test has been as high as 50 percent. Surprisingly, among the numerous empirical studies that have examined the earnings test, there is little agreement about whether the earnings test affects elderly labor supply at all. In this paper, we examine new and reconsider existing empirical evidence on the earnings test and conclude that, at least for men, the earnings test has a substantial effect on labor supply and claiming behavior.

* Haider (corresponding author): Dept. of Economics; Michigan State University; East Lansing, MI 48103; haider@msu.edu. Loughran: RAND; 1700 Main Street; Santa Monica, CA 90401; loughran@rand.org. We gratefully acknowledge the financial support of the Social Security Administration through the Michigan Retirement Research Center and from the National Institute on Aging under grant 1 P01 AG022481-01A1, as well as useful advice on a previous draft from Jeff Biddle, Tom DeLeire, Alan Gustman, Kathleen McGarry, Paul Menchik, Gary Solon, Mel Stephens, and Steve Woodbury. The findings and conclusions expressed are solely those of the authors and do not represent the views of the Social Security Administration, any agency of the Federal government, or the Michigan Retirement Research Center.

1. Introduction

The retirement earnings test is a provision of the Social Security system that reduces the benefits of current beneficiaries who earn above a specified threshold. In 2004, the provision reduced current Social Security benefits by $1 for every $2 earned above $11,640 for individuals who claim benefits before their normal retirement age (age 65 and four months for those turning age 65 in 2004). Popular opinion has long viewed the earnings test as an unfair tax on the earnings of older workers that dramatically reduces their incentive to work.[1] Echoing these concerns, Congress voted to eliminate the earnings test for workers ages 70-71 in 1983 and for workers between the normal retirement age and age 69 in 2000.

Understanding how the earnings test affects labor supply is important for several reasons. First, the earnings test provides an opportunity to examine how the elderly respond to taxes on earnings and, more generally, to changes in wages. This is of interest because the elderly may face health and institutional constraints that limit their ability to respond to changes in the pecuniary rewards to work. In addition, because the earnings test still applies to those ages 62 through the normal retirement age, the currently scheduled increase in the normal retirement age implies that that the earnings test will apply to an increasing age range of elderly individuals in the years to come. Evaluating the budgetary and welfare implications of the earnings test requires estimates of how sensitive the elderly are to both incremental and large-scale changes to the earnings test.

Employing a variety of methods and study periods, a lengthy empirical literature on the earnings test has come to mixed conclusions about whether the elderly respond at all. Leonesio

[1]Although the forgone benefits are reimbursed at later ages, it is widely believed that the taxed benefits have been reimbursed in an actuarially unfair manner (e.g., Leonesio 1990, Tran 2003). Recent changes in the reimbursement rate, however, aims to make the reimbursement actuarially fair.

(1990) provides a comprehensive literature review of the early scholarly research and concludes that the earnings test is empirically unimportant in the United States.[2] This conclusion was echoed by Gruber and Orszag (2003) who, in examining labor supply behavior under the earnings test from 1973 to 1998, concluded, "We find that the earnings test exerts no robust influence on the labor supply decisions of men" (p. 755). Examining a similar time period but employing a more structural model of labor supply, Friedberg (2000) comes to a much different conclusion about the earnings test: "The estimation yields significant elasticities that suggest considerable deadweight loss suffered by working beneficiaries" (p. 48). Her conclusion is consistent with two studies that find that the elimination of the earnings test in 2000 affected employment outcomes among men ages 65-69 (Tran 2003; Song 2004) and two studies that examine the elimination of similar earnings tests in Canada and the United Kingdom (Baker and Benjamin 1999; Disney and Tanner 2000). These four studies each rely on reduced-form methods similar to those employed by Gruber and Orszag (2003).

In this paper, we seek to reconcile conflicting conclusions about the earnings test by revisiting economic theory, replicating previous empirical studies, and presenting new empirical evidence. We revisit economic theory in order to clarify theoretical predictions regarding the labor supply and claiming response to the earnings test and how those predictions relate to the commonly employed empirical approaches. By replicating previous empirical studies and extending those analyses to more recent changes in the earnings test, we highlight how those analyses complement one another. In addition, we add to the empirical evidence on the earnings test by showing (1) measurement error in earnings leads to a substantial underestimate of the number of individuals constrained by the earnings test, (2) earnings substantially increase as

[2] Among the early published studies reviewed by Leonesio (1990) are Burkhauser and Turner (1978), Viscusi (1979), Burtless and Moffitt (1984), Vroman (1985), and Gustman and Steinmeier (1985).

individuals age past the earnings test, and (3) the response of women to the earnings test differs from that of men. In the end, we argue that the empirical evidence presented here and in previous studies tells a highly consistent story regarding the extent to which men reduce labor supply in response to the earnings test. For women, the empirical evidence is less consistent and so our conclusions are more circumspect.

2. A Theoretical Analysis of the Retirement Earnings Test

The retirement earnings test is a provision of the Social Security system that reduces the benefits of current Social Security recipients for each dollar earned above a given threshold. The ages covered by the earnings test, the level of the threshold, and the rate at which benefits are reduced have varied considerably over the last three decades. Table 1 shows how these various provisions of the earnings test varied between 1975 and 2002. All beneficiaries ages 62-71 faced the same earnings threshold between 1975 and 1977, and then in 1978, the threshold was increased more for individuals ages 65-71 compared to individuals ages 62-64. The earnings test was eliminated for beneficiaries ages 70-71 in 1983 and for beneficiaries older than the normal retirement age in 2000. In nominal terms, the threshold increased steadily over the entire period, with especially large percentage increases in 1978 and the late 1990s. The earnings test reduced Social Security benefits $1 for every $2 earned above the threshold during most of our sample period. In 1990, the rate of benefit reduction was lowered to $1 for every $3 earned above the threshold for beneficiaries ages 65-69.

The Earnings Test in a One-Period Model of Labor Supply

Following previous studies, we begin our theoretical analysis by considering a standard one-period model of labor supply in which individuals are offered a wage and then choose the

quantity of hours they work.[3] We assume throughout that leisure and consumption are normal goods. With an earnings test in place, individuals can either claim Social Security benefits and be subject to the benefit reduction of the earnings test or not claim benefits and avoid these reductions. An individual that does not claim benefits faces a standard budget constraint that is determined by other income, Y_1, and the wage rate, w, denoted as line segment AC in Figure 1.

For an individual who claims Social Security benefits when the earnings test is in place, the budget constraint is altered by the additional income the benefits represent and the rate of benefit reduction for earnings above the threshold. The benefit reduction affects the budget constraint as a tax does, but this tax is applied only to those hours worked where the earnings test binds. More specifically, we denote the sum of other income and Social Security benefits as Y_2 and the threshold as T. Define h_1 to be hours worked such that an individual earns exactly the amount of the threshold ($h_1=T/w$). The effective wage rate for hours worked less than h_1 is still w because total earnings are less than the threshold. For hours worked above h_1, the earnings test reduces benefits by the rate τ, resulting in an effective wage rate of $w(1-\tau)$. The worker faces this lower effective wage until he or she works hours such that current benefits have been fully taxed away. We denote this point as h_2 in Figure 1. The effective wage rate then returns to its previous level w for hours worked above h_2. Thus, the full budget constraint for an individual who claims Social Security benefits and is covered by an earnings test is denoted by the line segment DEBC in Figure 1. If the earnings test were eliminated, then the effective wage would be w everywhere, resulting in the budget constraint denoted as line segment DF.

---

[3] Similar presentations can be found in Blinder, Gordon, and Wise (1980), Ehrenberg and Smith (1988) and Friedberg (2000). Throughout this discussion, we ignore all other taxes and transfers to focus our attention on the impact of the earnings test. We will extend the discussion to consider multiple periods and labor market rigidities below.

With the earnings test in place, the one-period labor supply model has two predictions. Individuals who choose to work fewer than h_2 hours should claim Social Security benefits and individuals who choose to work more than h_2 hours should be indifferent to claiming benefits. This prediction follows because, compared to the budget constraint for not claiming, the budget constraint for claiming is higher for those working fewer than h_2 hours and is coincident for those working greater than h_2 hours. The model also predicts that workers will disproportionately locate, or "bunch," at the convex kink in the budget constraint (point E) because that point is consistent with a range of indifference curves. For small tax changes, the amount of bunching at the convex kink is proportional to the compensated labor supply elasticity (e.g., Saez 2002).

If the earnings test is eliminated, all eligible workers will now claim benefits because the budget constraint when claiming (DF) strictly dominates the budget constraint when not claiming (AC). Importantly, the labor supply response will vary according to individual preferences for work and leisure. For those who would have worked fewer than h_1 hours under the earnings test, hours worked should not change because the relevant part of the budget constraint is unaffected. For those who would have located at the kink (point E), hours worked should increase because these individuals can now locate on the segment of the budget constraint labeled EF. Individuals who were working more than h_2 hours should unambiguously decrease hours worked because the elimination of the earnings test shifts the budget constraint out in a parallel fashion, inducing a pure income effect. For individuals who would have chosen hours between h_1 and h_2, the prediction of the simple model is ambiguous because these workers experience both income and substitution effects and we do not know *a priori* which effect will dominate. Thus, the aggregate response to the elimination of the earnings test is an average over individuals in these four

regions of the budget constraint, implying a strict prediction regarding the sign of the aggregate effect is not available.

Extensions to the One-Period Model

There are a number of reasons why actual labor supply and claiming behavior might deviate from the predictions of the one-period model. One reason is that two provisions of the Social Security system link the budget constraint across periods.[4] For workers above the normal retirement age, benefits lost to the earnings test are eventually refunded though the delayed retirement credit (DRC). It is generally believed, however, that the DRC is not actuarially fair for most of our sample period and, as many researchers have pointed out, the vast majority of beneficiaries, their financial advisors, and the media perceive the earnings test to be a pure tax.[5] In addition, Social Security also recomputes benefits to take into account earnings after retirement. Specifically, if a beneficiary's current earnings are higher than any of the annual earnings used to compute current benefits, then the beneficiary's future benefits are automatically recomputed using these higher annual earnings. To the extent workers are forward-looking and consider these two provisions, the actual tax inherent in the earnings test is less than the full benefit reduction rate.

A multi-period model of labor supply also predicts that individuals intertemporally substitute labor supply away from periods of high-taxes to periods of low-taxes. For example, individuals may increase their hours in periods before the earnings test applies and reduce their

---

[4] There is a provision related to claiming that we ignore here. The Actuarial Reduction Factor (ARF) reduces benefits for individuals who claim before the normal retirement age. We ignore the ARF in our discussion because it is widely believed to be actuarial fair and individuals who claim early are still subject to the retirement earnings test and the DRC.

[5] Regarding actuarial fairness, the Social Security Administration has computed that the actuarially fair DRC is eight percent (Leonesio 1990). The actual DRC was far below this rate for most of the older workers in our sample. See Pingle (2003) for details regarding the DRC and recent changes.

hours until after the earnings test expires.[6] In either case, we would expect the measured change in earnings between high- and low-tax periods to be greater than in a one-period model (e.g., Blundell and MaCurdy 1999).

The second reason why behavior may not conform to the predictions of the one-period model is that we have assumed workers can freely choose hours conditional on wages. However, a number of studies argue that labor market rigidities prevent workers from freely choosing hours (e.g., Gustman and Steinmeier 1983; Lundberg 1985; Hurd 1996; Rust and Phelan 1997). Part time work, for example, may pay lower wages than full time work. If older workers cannot choose hours freely, they may not be able to choose hours precisely at the convex kink, and it may be optimal for some workers to withdraw from the labor force entirely. Such labor market rigidities might make bunching at the threshold less distinct, even if individuals perceive the earnings test as a tax. Moreover, when the earnings test is eliminated, we can no longer make clear predictions about how labor supply will respond. For example, individuals who were induced to leave the labor force might return following the elimination of the earnings test. Similarly, constrained individuals who were induced to choose a different wage/hours package may increase or decrease their hours worked depending on the new wage/hours packages that become available when the earnings test is eliminated.

Labor market rigidities might also affect whether older individuals can freely exit and enter the labor market. Such rigidities could arise if entry and exit costs are high, if human capital depreciates quickly, private pensions create strong incentives to exit the labor market, or if older workers demand certain non-pecuniary job characteristics that are not widely available. For example, to the extent that individuals cannot freely exit and enter the labor market, targeted

[6] See Burkhauser and Turner (1978).

workers might not be able to respond immediately to the elimination of the earnings test. Instead, the earnings test elimination might create incentives for younger individuals to remain in the labor force, even if faced with a severe earnings test, so that they can eventually enjoy working without the earnings test at later ages. Such concerns would suggest that the short-run labor supply response among the targeted age group might be smaller than the long-run response. As another example, if the cost of re-entering the labor market increases with time out of the labor force, then the effect of eliminating the earnings test might be smaller for targeted individuals who are older.

Implications for Empirical Analysis

Previous studies have primarily used one of two methods to analyze the effects of the earnings test on labor supply, neither of which is completely satisfactory. The first method is based on examining whether there is bunching in the vicinity of the earnings test threshold (i.e., the kink in the budget constraint). Bunching is used as an indication of the number of individuals who respond to the retirement earnings test (e.g., Leonesio 1990) and to estimate behavioral parameters of labor supply (e.g., Burtless and Moffitt 1984; Friedberg 2000). However, focusing on bunching near the threshold overlooks the fact that theory predicts all individuals above the threshold are affected by the earnings test, and thus, bunching represents a lower bound on the number who respond. Moreover, labor market rigidities might prevent some individuals from locating precisely at the threshold and measurement error in earnings can obscure the true number located at the threshold.

The other method is to examine aggregate labor supply responses to changes in the earnings test as in Gruber and Orszag (2003). However, this approach estimates the average labor supply response over individuals who, according to theory, should respond differently to

changes in the earnings test. A small aggregate labor supply effect, for example, could represent consistent labor supply responses among a few individuals or offsetting responses among many individuals.

In the following sections, we conduct both bunching and aggregate analyses in an effort to provide a comprehensive assessment of how the earnings test affects labor supply. We show results for time periods that have been previously examined and for the more recent elimination. We complement these methods with several new empirical analyses based upon longitudinal administrative earnings records. These include an assessment of how measurement error in earnings obfuscates the degree of bunching observed in survey data and how individual-level labor supply responds as an individual ages past the earnings test. We also present new evidence on how women, a group rarely studied in this literature, respond to changes in the earnings test. In Section 7, we discuss how these various pieces of empirical evidence fit together.

3. Data

In this section, we briefly describe the three data sets that we use for our empirical analysis. We provide a table of summary statistics for each data set in the Appendix.

New Beneficiary Data System (NBDS)

The NBDS is a nationally representative panel survey that was collected by the Social Security Administration (SSA). The first interview was conducted in 1982 and a follow-up interview was conducted in 1991. The data contain extensive information on respondents' demographic characteristics, labor supply, health, household income, and wealth.

A major advantage of the NBDS for our purpose is the availability of matched administrative records of covered earnings from 1951 to 1992. These data allow us to examine the earnings of individuals who are subject to the Social Security earnings test over time and

relatively error free. In particular, it is exactly the earnings that are reported to SSA that matter for the earnings test, and because the administrative earnings data are reported to four digits of significance, we observe Social Security earnings to the dollar in the neighborhood of the earnings test.

Our analysis uses the Retired Workers sample. These respondents represent a random sample of individuals who first received retired worker benefits during the sample selection period (mid-1980 through mid-1981), qualified for benefits based on their own earnings history, and did not receive Social Security Disability Insurance payments before they retired.[7] We follow these men and women between the ages of 63 to 76 over the years 1983 to 1992 for each full calendar year they are alive.[8] Our analysis sample includes 4,769 men followed for 38,498 person-years and 3,841 women followed for 33,044 person-years. Because the NBDS is a stratified random sample, we use the NBDS-provided weights for all of our analyses.

Current Population Survey (CPS) March Demographic Files

We use data from the March CPS for the years 1976 through 2004. Each CPS survey provides earnings information for the previous year, implying we have earnings data for the years 1975 through 2003. Throughout the rest of the paper, we will refer to the data by the year for which earnings are reported and define age to be one year less than the reported age in the survey year. We restrict our CPS sample to individuals who were ages 63 to 76 during the year

---

[7] The Retired Workers sample also contains those with "Dual Entitlement"—entitlement based on their own earnings and their spouses' earnings. The NBDS also contains three other samples: the Disabled Workers sample comprised of individuals who participated in the Social Security Disability Insurance program, the Wife and Widow Beneficiary sample comprised of individuals who qualified for benefits based on their spouse's earnings records, and the Medicare sample comprised of individuals who qualified for retirement benefits but had not yet claimed those benefits. The sampling scheme of the NBDS does not permit one to combine these samples.

[8] We make two additional refinements to our sample. First, we restrict our sample to individuals who were born after 1912, so that we observe each sample member at an age less than 70 years old at least once during our sample period. Second, we include 63 and 64 year olds in 1982, so that we have a larger sample of younger individuals.

of reported earnings. We measure total labor earnings in the CPS as the sum of wage and salary, self-employment, and farm earnings. Earnings in the CPS are self-reported by a household respondent. Because the CPS is a stratified random sample, we use the CPS-provided weights for all of our analyses.

We use the entire CPS sample to examine aggregate labor supply and claiming effects. One limitation of using the entire CPS sample is that some individuals are not eligible to receive Social Security. These individuals should not respond to the earnings test and their inclusion in our sample will therefore cause us to understate the impact of the earnings test on the eligible population. However, SSA estimates that the fraction of men and women who do not receive benefits as a retired worker, the spouse of a retired worker, or the survivor of a retired worker is under eight percent and, therefore, any understatement of program effects should be minor (SSA, 2004).

1978 CPS-Social Security Summary Earnings Records (CPS-SER) Exact Match File

The 1978 CPS-SER Exact Match File was created through the joint effort of the Census Bureau and SSA and contains the responses from the 1978 CPS March Demographic File linked to SSA administrative earnings records. We make similar restrictions to our CPS-SER sample as we made for the CPS samples, selecting all individuals who were ages 63 to 76 during 1977. We use the CPS-provided weights for all of our analyses.

The benefit of the CPS-SER is that it contains self-reported and administrative earnings for the same person. Such information allows us to directly examine the difference between self-reported and administrative earnings. The drawback is that our available sample size is about one-tenth of that which is available from the NBDS administrative data.

4. The Retirement Earnings Test and Bunching at the Threshold

In this section, we first use the NBDS and the CPS-SER to demonstrate that the use of self-reported earnings in the CPS substantially underestimates the proportion of workers who choose earnings near the threshold. We then show, similar to Friedberg (2000), that the bunching behavior systematically responds to changes in the earnings test between 1975 and 2003.

We present our results using earnings histograms. Each histogram plots the fraction of workers reporting earnings within a certain percentage of a given threshold. For all histograms, we use the number of individuals who fall into any of the specified bins as the denominator when we compute percentages.[9] Unless otherwise noted, we divide each histogram into bins representing increments of ten percentage points relative to a given threshold. For example, a bin labeled "–100" contains individuals who have earnings between 100 percent below and 90 percent below a given threshold, and a bin labeled "–90" contains individuals who have earnings between 90 and 80 percent below the threshold.[10]

Comparing Bunching in the NBDS and CPS

To assess the extent of measurement error in the self-reported CPS earnings, we begin by comparing the administrative earnings data in the NBDS to the self-reported earnings data in the CPS.[11] We make several additional restrictions to obtain comparable samples from the CPS and

---

[9] Other potential denominators exist, including the total number of individuals in the population and the total number of workers. Because of this ambiguity about the correct denominator, we make specific comparisons across histograms by comparing the ratio of bin percentages. Such comparisons make the specific choice of a denominator irrelevant.

[10] More precisely, let E_i^* be the earnings of individual *i* divided by the specified threshold and let $\{\gamma_b\}_{b=1}^{B}$ be a sequence of bin starting values. We consider individual *i* to be in bin γ_b if $\gamma_b < E_i^* \times 100 \leq \gamma_{b+1}$.

[11] Unfortunately, self-reported earnings in the NBDS covers only the previous three months, preventing us from making internal comparisons between the NBDS self-reported and administrative earnings data.

the NBDS. We restrict the samples to male respondents because the NBDS Retired Workers sample is representative only of individuals who receive benefits based on their own earnings, whereas the CPS sample is representative of all beneficiaries. Restricting to males should make the samples more comparable because males almost always qualify for benefits based on their own earnings. We restrict both analysis samples to current Social Security beneficiaries because individuals must have claimed benefits at least once to enter the NBDS sample. We only examine the years 1983-1989, during which time there were no major changes to the earnings test. Finally, we drop individuals ages 65 and 70 because their earnings for the calendar year are covered by two different earnings test regimes. We refer to these restricted samples as the "NBDS-r" and the "CPS-r" and present basic descriptive statistics for each in Table 2.

The top panel of Table 2 summarizes basic demographic characteristics for both samples. By these characteristics, the samples appear to be quite similar. The average birth year is one year older in the CPS (1918 vs. 1917), consistent with the NBDS being a panel of younger individuals who age over the sample period. The average education (10.8 years) and the percent white (88 percent) are very similar across the two samples. In the bottom panel, we report labor force participation rates by age. Labor force participation is broadly comparable across the two samples, although there is variation year-to-year.

We compare bunching in the NBDS-r and the CPS-r in Figure 2. Panels A and C of Figure 2 show that a considerable fraction of NBDS workers choose earnings at or just below the threshold. In Panel A, 15.8 percent of male workers ages 63-64 are located in the –10 bin compared to 5.8 and 2.8 percent of the male workers located in the 0 and 10 bins, respectively. Panel C shows a somewhat smaller, but nonetheless distinct, spike at the –10 bin for workers ages 66-69. In both panels, the proportion of workers in the age groups affected by the earnings

test is close to zero in bins above the threshold, which is consistent with the theoretical prediction that bunched workers relocate from bins above the threshold.[12] For purposes of comparison, we also graph earnings histograms for 71-74 year olds, individuals who are not covered by the earnings test during this period. There is no spike at the threshold for these workers and the proportion of workers above the threshold does not fall to zero.

We observe less bunching at the threshold in the CPS-r sample in Panel B; 11.0 percent of male workers ages 63-64 are located in the –10 bin compared to 4.0 and 3.3 percent of the workers located in the 0 and 10 bins, respectively. The ratio of the percent of male workers ages 63-64 located in the –10 bin to the 10 bin is 3.3 (11.0 vs. 3.3) in the CPS-r sample and 5.6 (15.8 vs. 2.8) in the NBDS-r sample. Thus, the spike in the –10 bin relative to the 10 bin is 70 percent larger in the NBDS-r than it is in the CPS-r. We see a 67 percent difference in bunching between the NBDS and CPS-r when we focus on workers ages 66-69 in Panels C and D, with a ratio of 3.0 (8.8 v. 2.9) in the CPS-r and 5.0 (10.5 v. 2.1) in the NBDS-r. The difference between these histograms suggests that the self-reported CPS data understate the degree of bunching near the threshold by about 70 percent.

Measurement error in the CPS is even more apparent when we examine smaller bins near the threshold. In Panels A and B of Figure 3 we show one percent bins (labeled –10, –9, etc.) around the threshold for 66-69 year olds. In the NBDS-r sample, we see that the spike in the fraction of workers with earnings just below the threshold in Figure 2 (based on 10 percentage point bins) is driven almost entirely by workers locating in the –1 bin (i.e., workers with earnings between one and zero percent below the threshold). This result demonstrates a remarkable

[12] Another possibility is that the threshold could simply serve as a focal point for earnings of affected individuals. The observation that the excess mass of workers at the threshold appears to be due to workers reducing hours is contrary to the focal-point hypothesis.

degree of programmatic knowledge and employment flexibility among workers because one percent of the earnings test during the 1980s is about $80. There is no distinct spike at the –1 bin in the CPS-r sample (Panel B).

Most every study that examines bunching with survey data acknowledges the problem of measurement error. Friedberg (2000) notes that much of the problem arises because individuals report earnings to just one or two digits of significance, and thus, she argues that using $1,000 bins around the earnings test will minimize measurement error problems. In Panels C and D of Figure 3, we present histograms based on $1,000 bins for the NBDS-r and the CPS-r. Even using this method, the NBDS sample shows more bunching near the threshold than does the CPS-r sample. The ratio of the percentage of male CPS-r workers ages 66-69 located in the –$1,000 versus the $1000 bins is 2.9 (21.7 vs. 7.4) in the CPS-r sample and 5.3 (24.5 vs. 4.6) in the NBDS sample. These ratios suggest that self-reported earnings understate bunching by 83 percent, an understatement that is even larger than when we use percentage bins.

Comparing Bunching within the CPS-SER

There are two potential problems with the above analyses that could undermine our conclusion that CPS self-reported earnings data underestimate the degree of bunching. One potential problem is that the differences we observe between the CPS and NBDS are attributable to the different sampling frames used for each data set (see Section 3). Another potential problem is that, in order to avoid benefit reductions, individuals choose to work in uncovered employment or they or their employers illegally misreport earnings to SSA.[13] If so, administrative earnings data could systematically underestimate actual labor supply, especially in the neighborhood of the earnings test threshold. Such behavior would imply that the

[13] SSA estimates that about 89 percent of total earnings are in covered employment (SSA 2004).

measurement error with respect to labor supply is in the administrative earnings, not the self-reported earnings. We address both of these potential problems by analyzing the CPS-SER data, which contain administrative and self-reported earnings for the same sample.

Panels A and B of Figure 4 present earnings histograms for 1977 using administrative and self-reported earnings, respectively. At first glance, it might appear that the degree of bunching is remarkably similar between the two panels, casting doubt on the NBDS-CPS comparisons in the previous subsection. However, the similarity between Panels A and B is driven by each panel using just one year of data (1977) and the particular value of the threshold for that year. The self-reported earnings exhibit a saw-toothed pattern because individuals tend to report earnings to round numbers. For example, there are a disproportionate number of workers locating in the −70, −40, and −10 bins, and these bins contain the round earnings reports of $1000, $2000, and $3000, respectively. The saw-toothed pattern is more pronounced in Figure 4 than in Figures 2 and 3 because Figure 4 does not use multiple years in which the earnings threshold (and thus the bins that contain round earnings reports) changes. In addition, one of the key bins we have been examining, the −10 bin, contains a round earnings value.

Because of the difficulties regarding round numbers, we use two methods to assess the extent to which measurement error in self-reports obfuscate the peak. The first method is to compare the gap between the −10 and −40 bins, both of which are affected by rounding. In Panel A, the peak in the −10 bin is 1.95 times larger than the peak in the −40 bin (11.3 percent vs. 5.8 percent), whereas the similar difference in Panel B is 1.24 (12.3 percent vs. 9.9 percent). This method implies that the peak is 57 percent greater in the administrative data than it is in the self-reported data. Alternatively, we make use of the fact that the threshold itself is a round

number and switch the inequalities that we use to define the bins.[14] This switch moves the round reports of the threshold itself into the 0 bin, allowing us to focus on those who report earnings just below the threshold as before. Panels C and D of Figure 4 present these histograms. Now, in the administrative data (Panel C), the number of workers locating in the −10 bin is 3.75 times greater than the number locating in the 10 bin (11.20 vs. 2.99 percent). In the self-reported data, the same difference is 2.42. This method implies that the peak is 55 percent greater in the administrative data than it is in the self-reported data, a difference similar to what we find using the first method (57 percent). Thus, the large discrepancy between administrative and self-reported earnings data remains even when we use data drawn from the same sample.

The second potential problem is that the CPS-NBDS differences could arise if individuals or their employers do not report all earnings to SSA (whether legally or illegally), but do report all earnings when surveyed by the CPS. However, if this were true, we would expect the correlation between administrative and self-reported earnings data to be smaller for individuals subject to the earnings test than for individuals not subject to the earnings test. Individuals who are not subject to the earnings test do not have an incentive to choose uncovered employment or deliberately underreport earnings to SSA. Because we have self-reported and administrative earnings data for the same individuals in the CPS-SER, we can examine this correlation directly. Contrary to this hypothesis, the correlation between administrative and self-reported earnings is

[14] In a previous footnote, we defined an individual to be in bin γ_b if $\gamma_b < E_i^* \times 100 \leq \gamma_{b+1}$. For this comparison, we instead define an individual to be in bin γ_b if $\gamma_b \leq E_i^* \times 100 < \gamma_{b+1}$; we continue to restrict our sample to positive earners.

larger for individuals covered by the earnings test than it is for individuals who have aged past the earnings test, casting doubt on this alternative hypothesis.[15]

Bunching over Time in the CPS

Although the previous results suggest that the self-reported CPS data obfuscates between 55 and 70 percent of the bunching that occurs around the threshold, only the self-reported data allow us to consistently examine how bunching changes over a long time period. Figure 5 presents histograms of relative male earnings for four earnings test regimes (1975-77, 1978-1982, 1983-1989, and 1990-1999) for the full CPS sample.

Workers bunch exactly as the simple theory predicts under each earnings test regime. In the 1975-77 regime presented in Panel A, the youngest three age groups all bunch at the same place; there is no evidence of bunching for the oldest age group. In the second regime (1978-1982, Panel B), the middle two age groups continue to bunch together but the youngest age group bunches at a lower level of earnings. The bunching behavior changes again in the third earnings test regime (1983-1989, Panel C) with the 70-71 year olds behaving like the 72-76 year olds, consistent with the elimination of the earnings test for 70-71 year olds in 1983. The younger two age groups continue to bunch at their respective thresholds. The 1990-99 regime in Panel D is similar to the 1983-89 regime, although the bunching for the 65-69 year olds is less pronounced. The less pronounced bunching for 65-69 year olds in panel D when compared to Panel C is consistent with three policy differences between the two regimes: a decrease in the benefit reduction for excess earnings ($1 of reduced benefits for each $3 of excess earnings

[15] For our basic analysis, we restrict our sample to those workers who have positive earnings (because zeroes are presumably easier to report accurately and there are more zero earners in the older sample) below the taxable limit ($16,500 in 1977). The correlation is 0.74 for workers ages 63-71 and 0.45 for workers ages 73-76. Including the zeroes makes the differences less pronounced, 0.73 for ages 63-71 and 0.54 for ages 73-76. Going to tighter ages also makes it less pronounced (0.68 for ages 69-71 and 0.55 for ages 73-75), but in no case is the correlation higher for the older workers.

rather than the previous rate of $1 for each $2 of excess earnings), an increase in the real value of the threshold, and an increase in the DRC for some cohorts.[16]

Figure 6 presents a similar set of earnings histograms for women. In Panel A (1975-77), we observe women bunching in a manner quite similar to men. Bunching for women is much less apparent in Panel B (1978-82), and non-existent in Panels C and D (1983-99).

Finally, in Figure 7, we compare earnings histograms for 66-69 year olds and 71-75 year olds between the 1990-99 and 2000-03 periods. Panels A and B present results for men and panels C and D present results for women. For men, the bunching we observe at ages 66-69 during 1990-99 disappears during 2000-03, at which point the earnings test has been eliminated for that age group. For women, there is no evidence of bunching under either earnings test regime.

5. The Effect of Aging Past the Earnings Test

The administrative earnings data in the NBDS allow us to examine how earnings change longitudinally when the earnings test is eliminated at age 70. Over the period of our NBDS sample (1983-1992), the earnings test applied to individuals ages 62-69 but not to individuals ages 70 and above. Longitudinal data allow us to test the hypothesis presented in Section 2 that the change in labor supply between age 69 and 70 should depend on earnings relative to the threshold at age 69. Specifically, there should be little or no relative change in labor supply for those with earnings less than the threshold at age 69, an increase in relative labor supply for those with earnings equal to the threshold, and an indeterminate response for those with earnings

[16] See Gustman and Steinmeier (1985) and Gustman and Steinmeier (1991) for a structural model of retirement that estimates the impact of these and similar proposed changes to the retirement earnings test.

just above the threshold.[17] In this section, we provide evidence on the extent to which NBDS workers increase or decrease labor supply in each of these regions of the budget constraint.

To examine the effect of aging past the earnings test, we present results using two-year changes. The motivation for such a comparison is that, during the calendar year in which an individual turns age 70, earnings in months before the month of birth are subject to the earnings test but earnings in subsequent months are not. Thus, to capture the full effect of facing the earnings test to not facing the earnings test, we compare 69 year olds to 71 year olds. We account for the secular decline in labor supply with age by comparing two-year changes in earnings for those aging past the earnings test (workers age 69 in the initial period) to two-year changes in earnings for workers ages 65-67 and workers ages 71-75 in the initial period. We present results for men in Panel A and for women in Panel B of Figure 8.

The figures show that two-year earnings growth is negative across most of the earnings distribution for all three age groups, presumably because workers in this age range are generally reducing hours worked as they age. Focusing on the results for men in Panel A, male workers with age-69 earnings near the threshold experience relatively high earnings growth between ages 69 and 71. For workers with age-69 earnings in the –20 and –10 bins (earnings between 20 and zero percent below the threshold), earnings increase by almost 20 percent between ages 69 and 71. In comparison, earnings decline by almost 20 percent at the same points in the earnings distribution for workers ages 65-67 and workers ages 71-75. Statistical tests indicate that these

[17] Because the administrative earnings in the NBDS are censored at the Social Security taxable limit, we are not able to use these data to examine the change in earnings for those at the highest hours. Moreover, this censoring causes all of our results on earnings growth to be downward biased.

differences in earnings growth rates are statistically significant.[18] For women (Panel B), there is no evidence that two-year changes in earnings differ systematically between age groups.

It is worth noting that one explanation for the relatively large earnings increase in the bins adjacent to the –10 bin is that some workers were unable to choose earnings exactly at the threshold, perhaps because of labor market rigidities that constrain their choice of hours. Such an explanation is consistent with the elevated concentrations of workers in bins adjacent to the –10 bin in the bunching analysis. In Figure 2, for example, we observe elevated concentrations of age 63-64 workers in the –20 bin and elevated concentrations of age 66-69 workers in the –40 and –30 bins.

6. The Impact of the Earnings Test on Aggregate Labor Supply and Claiming

The results of Sections 4 and 5 indicate that workers covered by the earnings test disproportionately bunch just below the threshold and that those who bunch experience considerably higher rates of earnings growth as they age past the earnings test. Thus, the earnings test clearly affects labor supply, at least locally and at least for some individuals. However, these results do not tell us whether relaxing the earnings test affects aggregate labor supply. In this section, we test whether the 1983 elimination of the earnings test for 70-71 year olds and the 2000 elimination of the earnings test for 65-69 year olds affected aggregate labor supply and claiming behavior. The difference-in-differences approach we employ here is similar to that used by Gruber and Orszag (2003), Tran (2003), and Song (2004).

---

[18] To examine whether the differences in earnings growth rates in Panel A were statistically significant across age groups, we computed a series of *t*-tests bin by bin. Given the similarity in earnings changes among the 65-67 year olds and the 71-75 year olds, we grouped these ages together and tested whether mean two-year earnings changes were different between the 69 year olds and all others. The *t*-tests indicated that the differences were statistically significant for the –40, –20, –10, and 0 bins at the 0.05 confidence level.

As an initial indication of how aggregate labor supply changed following the 1983 and 2000 eliminations, we graph in Figure 9 average annual earnings in the years surrounding 1983 and 2000 for men and women. For both policy changes, we compare the affected age group (workers ages 70-71 in 1983 and workers ages 65-69 in 2000) to an older age group. We use an older age group as the comparison group for two reasons. First, in both cases, the older age group is not covered by the earnings test in the period we analyze. While younger age groups were not affected by the policy change in those years either, there may have been some confusion that would lead these younger individuals to respond, nonetheless. Second, a life-cycle model of labor supply predicts that individuals who are younger than the ages for which the earnings test is eliminated should also respond to the change.[19] Figure 9 suggests that earnings of the affected age group increased relative to the earnings of the older age group for both men and women following the 2000 elimination. No discernible pattern in earnings is evident following the 1983 elimination.

To formally test for aggregate effects, we estimate reduced-form regressions for each repeal. Our basic regression model is

$$Y_{it} = \beta_0 + \beta_1 ELIM_{it} + \beta_2 Age_{it} + \beta_3 Year_t + \beta_4 X_{it} + \varepsilon_{it}, \quad (1)$$

where $ELIM_{it}$ is an indicator variable for whether individual i in year t is in an age group for whom the earnings test was eliminated, Age_{it} is a vector of age dummies, $Year_t$ is a vector of year dummies, and X_{it} is a vector of demographic controls including race/ethnicity (white, black, Hispanic, and other), completed education (less than high school, high school, some college, and 4+ years of college), and marital status (married, divorced, widowed, and never married). We examine six outcome variables (Y_{it}): annual earnings, worked at all, log hourly wages,

[19] See Burkhauser and Turner (1978) and Section 2 above.

weeks/year, hours/week, and receipt of Social Security benefits. In the case of log hourly wages, weeks/year, and hours/week, we limit our sample to those who reported positive earnings. We use five years of data on either side of 1983 and 2000 (1978 to 1987 and 1995 to 2003). Our analysis of the 1983 reform includes individuals ages 70-76 and our analysis of the 2000 reform includes individuals ages 66-76 (excluding individuals age 70). The coefficient of interest is β_1, which measures the impact of eliminating the earnings test on the affected age group.[20]

The regression results reported in Table 3 suggest that the 1983 reform had little effect on the aggregate labor supply of individuals ages 70-71, whereas the 2000 reform had a considerable effect on the labor supply of individuals ages 66-69. For the 1983 elimination, the effect on annual earnings is small and statistically insignificant. Although the results indicate that some of the components of earnings changed with the elimination of the earnings test (e.g., fewer males worked, males worked fewer weeks, females earned lower wages), no consistent patterns emerge. The 1983 elimination increased claiming among men by 1.5 percentage points, a magnitude statistically significant at conventional levels.

The pattern of results is much different for the 2000 elimination. Concentrating on results that are generally statistically significant, we estimate that the 2000 reform increased the fraction of men ages 66-69 who work by 1.5 percentage points and the fraction of women who work by 1.8 percentage points. The addition of these new workers to the labor force led to mean increases in annual earnings of $1,644 for men and $461 for women (corresponding to 23 and 22 percent of mean earnings in these populations). The fraction of men and women claiming benefits increased a statistically significant 2.5 and 1.6 percentage points, respectively.

---

[20] The regression model we use is a simplified version of the one used in previous studies. For example, Gruber and Orszag (1983) use variation in the earnings test that stems from the 1982 elimination and the incremental increases in the threshold. Nonetheless, we obtain results that are similar to those reported in the other reduced-form studies.

7. Reconciling Empirical Findings on the Earnings Test

As mentioned in the Introduction, numerous studies have come to varying conclusions regarding the effect of the earnings test on labor supply. In this section, we attempt to reconcile our results and those from previous studies to answer the question: Does the earnings test affect labor supply? In short, we believe the available empirical evidence consistently indicates that the answer is "yes", at least for men.

As Friedberg (2000), Burtless and Moffitt (1984), Vroman (1985) and others have shown and our results confirm, male workers have disproportionately chosen earnings at or just below the threshold (Figures 5 and 7). This fact is consistent with how theory suggests the earnings test affects the budget constraint, creating an incentive for those who would prefer to locate just above the threshold (in the absence of the earnings test) to locate at the threshold. Also consistent with this simple prediction, we present new evidence that the workers who were locating at the threshold substantially increase their earnings as they age past the earnings test. Importantly, the extent to which workers bunch at the threshold has declined considerably over time, which is consistent with the increase in the real value of the threshold, the decline in the benefit reduction rate, and the elimination of the earnings test for some age groups.

Leonesio (1990) and Gruber and Orszag (2003) dismiss the bunching evidence as being empirically unimportant because the number of bunched workers is small relative to the elderly population. We disagree with this assessment of the bunching evidence. Our results suggest that measurement error in earnings leads to a substantial underestimate of the fraction of bunched workers in self-reported data and, indirectly, a substantial underestimate of the compensated labor supply elasticity of older workers. Moreover, labor market rigidities may keep some workers from locating precisely at the convex kink in the budget constraint. Our analysis of the NBDS supports this later hypothesis. More workers locate in the –30, –20, and 0 bins than

would otherwise be expected (Figure 2), and the workers in the –20 and 0 bins similarly increase their earnings as they age past the earnings test (Figure 7). Finally, theory suggests that it is not just the individuals who bunch at the threshold who are affected by the earnings test. The earnings test affects all individuals who would have located above the threshold in absence of the earnings test, some of whom may still locate above the threshold and some of whom may have exited the labor market. Given these concerns, any bunching found in self-reported data should be interpreted as just a small indication of the number of individuals who are actually responding to the earnings test.

Our aggregate analysis indicates that the 2000 elimination of the earnings test increased employment and earnings among men ages 66-69.[21] We emphasize, however, that these aggregate estimates represent the average response across the population, some of which could be positive and some of which could be negative. These estimates are consistent with those estimated by Tran (2003) using CPS data. Song (2004) estimates a specification similar to ours using SSA data and finds an earnings effect but no employment effect for the 2000 repeal. Song, however, uses only one year of post-repeal data.

These results for the 2000 elimination are also consistent with the estimated effects of eliminating the earnings test in Canada and the United Kingdom. Baker and Benjamin (1999) report that the 1970s elimination of the Canadian earnings test for 65 to 69 year olds led to an increase in weeks/year worked, primarily through an increase in employment. Disney and Tanner (2000) report that the 1989 elimination of the United Kingdom earnings test for 65 to 69

[21] The 2000 elimination had a small negative effect on weeks/year and no statistically significant effect on hours/week. One explanation for this result could be that individuals induced to enter the labor market upon elimination of the earnings test on average supply fewer hours/week or weeks/year than individuals who were already in the labor market. Changes in the composition of the working population might also explain why hourly wages increase among male 65-69 year-olds following the 2000 elimination.

year old men led to a labor supply increase of three to four hours/week; their results do not allow one to determine whether the increase is on the intensive or extensive margin of working.

Consistent with the results of Gruber and Orszag (2003) but contrary to the results for the 2000 elimination, we find little evidence that the aggregate earnings for males responded to the 1983 elimination.[22] A potential explanation for the different results between 1983 and 2000 is that the relatively young workers affected by the 2000 elimination (66-69 year olds) could more easily reenter the workforce following the repeal of the earnings test than could the 70-71 year olds affected by the 1983 elimination.[23] To examine this hypothesis, we re-estimate the earnings and employment regressions for the 2000 elimination allowing the estimated effect to vary by age using two specifications. The first specification allows separate policy effects for each age 66-69 and the second specification constrains the policy effects to be linear in age. The results in Table 3 show that the effects of the 2000 elimination are concentrated among younger individuals. Although the age-specific coefficients are not monotonically declining in age, the coefficients are generally larger and statistically significant at younger ages.[24] The specification with the linear age trend verifies this pattern.

In this large empirical literature, only Gruber and Orszag (2003) analyze how women respond to the earnings test, despite the fact that women comprise nearly half of fully insured workers (SSA 2004). We find far less bunching for women than for men in the CPS, especially after 1977 when evidence of bunching for women largely disappears (Figures 6 and 7). In

---

[22] Leonesio (1990) reports that Packard (1988) also found no effect of the 1983 elimination on the employment of 70-71 year old men.

[23] Another potential explanation is that the 1983 elimination affected a population that was relatively weighted towards individuals who would have had an incentive to reduce their labor supply following elimination. However, the data suggest that the population of workers who would have had an unambiguous incentive to reduce their labor supply was large for the 2000 elimination when compared to the 1983 elimination. There are more high earners among the 65-69 year olds in 1999 than there are among the 70-71 year olds in 1982.

[24] Tran (2003) also reports finding larger employment effects for younger workers in his CPS sample.

addition, we find little evidence that women increase their earnings as they age past the earnings test (Figure 8). However, Gruber and Orszag (2003) find some evidence of an aggregate response among women, and our aggregate analysis of the 2000 elimination finds earnings, employment, and claiming effects that are large and comparable in size to that found for men. Thus, women appear not to make labor supply changes on the intensive margin in response to the earnings test, but they do appear to make discrete changes in employment when the earnings test is eliminated. We do not have a good explanation for these seemingly contradictory findings.[25]

As in Gruber and Orszag (2003), we find strong evidence that the elimination of the earnings test induces earlier Social Security claiming for both men and women. We estimate that the elimination of the earnings test in 1983 increased male claiming by 1.5 percentage points and female claiming by a statistically insignificant 0.3 percentage points. In 2000, the corresponding percentages were 2.5 and 1.6. Given that few individuals were not claiming benefits in the relevant ages prior to the eliminations—nine percent of 70-71 year old males in 1975-1977 and 14 percent of 65-69 year olds males in 1990-1999 had not claimed—these claiming effects are very large.

Gruber and Orszag (1999, 2003) worry that earlier claiming induced by the earnings test will lead to greater levels of poverty, especially among female widows, because earlier claiming results in lower annual benefits. We do not share this concern. The empirical results imply that the elimination of the earnings test, if anything, increases employment, which would lead to higher lifetime income, and theory suggests that the increase in claiming should be concentrated

[25] Differences in the prevalence of spousal benefits and retirement benefits between men and women does not offer an explanation because the earnings test applies to both. It is possible that the effective tax inherent in the earnings test is less for women than men because of benefit recomputation rules and the DRC. The benefit recomputation would be more valuable to women if they have more years of low earnings, and the DRC might be closer to being actuarially fair because women have a longer life expectancy. However, if the earnings test is less of a tax for women, then it remains a puzzle why they nonetheless respond in aggregate to its elimination.

among individuals with relatively high earnings (i.e., earnings greater than the amount at which all benefits have been taxed away by the earnings test). Individuals with such strong attachment to the labor force are likely to be healthier and more educated than average and have above average private savings. In fact, such distributional concerns were the focus of much of the early debate about eliminating the earnings test (e.g., Honig and Reimers 1989).

To examine whether there is empirical support for this last theoretical prediction, we report results in Table 5 for regressions in which we interact the policy dummy variable (*ELIM*) with indicators for four educational groups in models of earnings and claiming. The results are generally supportive of the prediction that both earnings and claiming effects are concentrated among more educated workers.

8. Conclusion

Does the retirement earnings test affect labor supply? For a significant number of individuals, it absolutely does. We conclude that the earnings test both depresses hours worked for those who work and induces others to leave the labor force. This conclusion is supported by the empirical results we report here and by a large body of previous empirical research both in the United States and in other countries. Our conclusion should come as no surprise because, at least for most of our sample period, the earnings test represents a substantial tax on earnings.

The earnings test still covers individuals ages 62 to the normal retirement age, and the normal retirement age is currently scheduled to increase to age 67 by 2022. Consequently, the age range subject to the earnings test will continue to rise and pressure to reform the earnings test yet again will likely mount. Because our findings indicate that it is the youngest covered workers who are most responsive to the earnings test, future reforms that target individuals under the normal retirement age could have even larger effects than those for the 2000 elimination.

Given our results, we suggest that future research on the earnings test focus on two issues. One issue is that the welfare implications of the earnings test depend critically on whether individuals respond to its incentives in a static or dynamic setting. For example, if the estimated labor supply response to the earnings test is mostly attributable to intertemporal substitution, then the net impact of the earnings test on lifetime wealth would be less than if the response were attributable to a permanent reduction in labor supply. We suggest that future research focus on distinguishing between permanent and intertemporal labor supply responses.

The second issue rests with the following question: if the combination of the earnings test and the delayed retirement credit is approximately revenue-neutral in the long-run, then what is the rationale for having the earnings test?[26] One possibility is that the earnings test addresses paternalistic concerns regarding appropriate levels of consumption and saving in old age. However, if the earnings test largely affects the labor supply and claiming behavior of the most educated workers for whom retirement is generally secure, these paternalistic concerns may be misplaced.

[26] See Tran (2004) for illustrative calculations regarding whether the DRC represents a tax. See Leonesio (199) for a discussion of the actuarially fair DRC. See Gustman and Steinmeier (2004) for an analysis of the short-run and long-run budgetary impacts of eliminating the earnings test.

Appendix Tables

Table A1. Descriptive Characteristics of the Full NBDS Retired Workers Sample

	Men		*Women*	
Cross-sectional sample	N=4,769		N=3,841	
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
Birth year	1917.0	1.59	1917.4	1.45
Education	10.80	3.47	11.22	2.83
White	0.88	--	0.88	--
Black	0.08	--	0.08	--
Hispanic	0.03	--	0.03	--
Pooled sample	N=38,498		N=33,044	
Age	*LFPR*		*LFPR*	
63	0.254		0.199	
64	0.240		0.209	
65	0.209		0.187	
66	0.216		0.173	
67	0.209		0.190	
68	0.223		0.190	
69	0.219		0.177	
70	0.209		0.158	
71	0.181		0.138	
72	0.170		0.122	
73	0.166		0.114	
74	0.172		0.140	
75	0.159		0.138	
76	0.164		0.123	

Data source: NBDS.

Table A2. Descriptive Characteristics of the 1978 CPS-SER Sample for Men

	Full 1978 CPS Sample				Matched 1978 CPS Sample			
	N	LFPR	Earn-ings	SS Receipt	N	LFPR	Earn-Ings	SS Receipt
63-71	4,624	0.374	3,156	0.819	3,428	0.368	3,131	0.827
72-76	1,554	0.198	1,068	0.929	1,194	0.199	1,052	0.932

Data source: 1978 CPS-SER.

Table A3. Descriptive Characteristics of the Full CPS Sample

	Men				Women			
Year/Age	N	LFPR	Earn-ings	SS Receipt	N	LFPR	Earn-ings	SS Receipt
1975-1977								
63-64	3,344	0.566	5,714	0.611	3,889	0.292	1,362	0.688
65-69	6,892	0.347	2,362	0.850	8,756	0.183	595	0.818
70-71	2,051	0.261	1,342	0.907	3,103	0.111	320	0.832
72-76	4,166	0.218	1,183	0.922	5,977	0.078	193	0.817
1978-1982								
63-64	6,296	0.512	7,469	0.627	7,437	0.283	2,029	0.731
65-69	13,196	0.332	3,324	0.861	17,054	0.169	816	0.892
70-71	4,305	0.224	1,760	0.916	5,937	0.104	415	0.926
72-76	8,335	0.184	1,296	0.932	11,611	0.073	228	0.920
1983-1989								
63-64	8,600	0.448	9,606	0.665	10,303	0.273	2,822	0.752
65-69	18,473	0.300	4,458	0.872	23,170	0.168	1,266	0.904
70-71	6,071	0.197	2,318	0.938	8,145	0.105	599	0.940
72-76	11,665	0.157	1,774	0.942	17,221	0.069	410	0.934
1990-1999								
63-64	10,019	0.457	13,998	0.666	11,930	0.312	4,906	0.732
65-69	23,264	0.302	7,344	0.862	28,928	0.193	2,331	0.889
70-71	8,297	0.206	4,290	0.916	10,824	0.124	1,244	0.923
72-76	16,921	0.155	3,196	0.928	23,297	0.083	758	0.932
2000-2003								
63-64	5,021	0.481	19,739	0.657	5,653	0.352	7,246	0.698
65-69	10,564	0.333	12,449	0.871	12,605	0.232	4,200	0.876
70-71	3,756	0.238	7,408	0.908	4,700	0.139	2,200	0.896
72-76	8,202	0.163	4,818	0.912	10,798	0.090	1,420	0.909

Data source: 1976-2003 CPS.

References

Baker, Michael and Dwayne Benjamin. 1999. How Do Retirement Tests Affect the Labour Supply of Older Men? Journal of Public Economics 71(1): 27-52.

Blinder, Alan S., Gordon, Roger H., and Donald E. Wise. Reconsidering the Work Disincentive Effects of Social Security. National Tax Journal 33(4): 431-42.

Blundell, Richard and Thomas MaCurdy. Labor Supply: A Review of Alternative Approaches in Ashenfelter, Orley and David Card (eds.) *Handbook of Labor Economics v. 3A*. Amsterdam: Elsevier Science.

Bound, John and Alan B. Krueger. 1991. The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? Journal of Labor Economics 9(11): 1-24.

Burkhauser, Richard V. and John A. Turner. 1978. A Time Series Analysis on Social Security and Its Effects on the Market Work of Men at Younger Ages. Journal of Political Economy 86(4): 701-16.

Burtless, Gary and Robert A. Moffitt. 1984. The Effect of Social Security Benefits on the Labor Supply of the Aged in Aaron, Henry J and Gary Burtless (Eds.) Retirement and Economic Behavior. Washington: The Brookings Institution.

Burtless, Gary and Robert A. Moffitt. 1985. The Joint Choice of Retirement Age and Post-Retirement Hours of Work. Journal of Labor Economics 3: 209-36.

Disney, Richard and Sarah Tanner. 2000. The Abolition of the Earnings Rule for UK Pensioners. IFS WP 00/13. London: Institute for Fiscal Studies.

Ehrenberg, Ronald G. and Robert S. Smith. 1988. Modern Labor Economics: Theory and Public Policy, 3rd Edition. Glenview: Scott, Foresman and Company.

Friedberg, Leora. 2000. The Labor Supply Effects of the Social Security Earnings Test. The Review of Economics and Statistics 82(1): 46-63.

Gruber, Jonathan and Peter Orszag. 1999. What To Do About the Social Security Earnings Test? Issue Brief No. 1. Boston: Center for Retirement Research, Boston College.

Gruber, Jonathan and Peter Orszag. 2003. Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt? National Tax Journal (December): 755-73.

Gustman, Alan and Thomas Steinmeier. 1983. Minimum Hours Constraints and Retirement Behavior. Contemporary Policy Issues (Supplement to Economic Inquiry) 3: 77-91.

Gustman, Alan and Thomas Steinmeier. 1985. The 1983 Social Security Reforms and Labor Supply Adjustments of Older Individuals in the Long Run. Journal of Labor Economics 3(2): 237-53.

Gustman, Alan and Thomas Steinmeier. 1991. Changing the Social Security Rules for Work after 65. Industrial and Labor Relations Review 44(4):733-745.

Gustman, Alan and Thomas Steinmeier. 2004. The Social Security Retirement Earnings Test, Retirement and Benefit Claiming. NBER Working Paper 10905. Cambridge, MA: National Bureau of Economic Research.

Honig, Marjorie and Cordelia Reimers. 1989. Is it Worth Eliminating the Retirement Earnings Test? American Economic Review Papers and Proceedings 79(2):103-107.

Hurd, Michael. 1996. The Effect of Labor Market Rigidities on the Labor Force Behavior of Older Workers in Wise, David (Ed.) Advances in the Economics of Aging. Chicago: University of Chicago Press, 11-58.

Leonesio, Michael V. 1990. The Effects of the Social Security Earnings Test on the Labor-Market Activity of Older Americans: A Review of the Empirical Evidence. Social Security Bulletin 53(5): 2-21.

Lundberg, Shelly. 1985. Tied Wage-Hours Offers and the Endogeneity of Wages. Review of Economics and Statistics 405-410.

Packard, Michael D. 1988. The Effects of Removing 70- and 71-Year-Olds from Coverage Under the Social Security Earnings Test. Unpublished manuscript. Washington: Office of Policy, Social Security Administration.

Pingle, Jonathan. 2003. Social Security's Delayed Retirement Credit and the Labor Supply of Older Workers. Manuscript, University of North Carolina-Chapel Hill, Department of Economics.

Rust, John and C. Phelan. 1997. How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. Econometrica 65: 781-831.

Saez, Emmanuel. 2002. Do Taxpayers Bunch at Kink Points? Unpublished manuscript. Berkeley: Department of Economics, University of California Berkeley.

Social Security Administration. 2004. Annual Statistical Supplement. Washington: Office of Policy, Social Security Administration.

Song, Jae G. 2004. Evaluating the Initial Impact of Eliminating the Retirement Earnings Test. Social Security Bulletin 65(1): 1-15.

Tran, Bac V. 2003. The Effect of the Repeal of the Retirement Earnings Test on the Labor Supply of Older Workers. Unpublished manuscript. College Park: Department of Economics, University of Maryland.

Viscusi, Kip W. 1979. Welfare of the Elderly: An Economic Analysis and Policy Prescription. New York: John Wiley & Sons.

Vroman, Wayne. 1985. Some Economic Effects of the Retirement Test in Ehrenberg, Ronald (Ed.) Research in Labor Economics, v. 7. Greenwich: JAI Press Inc.

Table 1: The Social Security Retirement Earnings Test Threshold by Year and Age

	Age		
Year	62-64	65-69	70-71
1975	2,520	2,520	2,520
1976	2,760	2,760	2,760
1977	3,000	3,000	3,000
1978	3,240	4,000	4,000
1979	3,480	4,500	4,500
1980	3,720	5,000	5,000
1981	4,080	5,500	5,500
1982	4,440	6,000	6,000
1983	4,920	6,600	--
1984	5,160	6,960	--
1985	5,400	7,320	--
1986	5,760	7,800	--
1987	6,000	8,160	--
1988	6,120	8,400	--
1989	6,480	8,880	--
1990	6,840	9,360*	--
1991	7,080	9,720*	--
1992	7,440	10,200*	--
1993	7,680	10,560*	--
1994	8,040	11,160*	--
1995	8,160	11,280*	--
1996	8,280	12,500*	--
1997	8,640	13,500*	--
1998	9,120	14,500*	--
1999	9,600	15,000*	--
2000	10,080	--**	--
2001	10,680	--**	--
2002	11,280	--**	--
2003	11,520	--**	--
2004	11,640	--**	--

Notes: All figures are in current dollars. The dashed-lines denote major changes to the retirement earnings test. The loss in benefits is $1 for every $2 earned above the threshold for every group except those groups that are denoted with a single asterisk (*); for these groups, the loss is $1 for every $3 earned. Beginning in 2000 (years marked with two asterisks), individuals are subject to a different threshold (e.g., $17,000 in 2000, $25,000 in 2001, $30,000 in 2002) and lower benefits tax ($1 for every $3 of labor earnings) in the year in which they reach the normal retirement age. In the double asterisk years, this age was 65.

Table 2. Comparability of the NBDS-r and CPS-r Samples

	NBDS-r		*CPS-r*	
Cross-sectional sample	N=4,767		N=34,780	
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
Birth year	1917.0	1.59	1917.9	3.85
Education	10.80	3.47	10.81	3.65
White	0.879	--	0.875	--
Black	0.084	--	0.078	--
Hispanic	0.025	--	0.032	--
Pooled sample	N=31,077		N=34,780	
Age	*LFPR*		*LFPR*	
63	0.252		0.263	
64	0.239		0.300	
65	--		--	
66	0.214		0.269	
67	0.208		0.237	
68	0.221		0.248	
69	0.217		0.234	
70	--		--	
71	0.192		0.192	
72	0.210		0.194	
73	0.200		0.175	
74	0.206		0.156	

Notes: See text for definition of NBDS-r and CPS-r samples. *Data source*: NBDS and 1984-1990 CPS.

Table 3: The Effect of Eliminating the Earnings Test on Aggregate Labor Supply and Claiming

	Earnings	Working	Lwages/hr | working	Weeks/yr | working	Hours/wk | working	Claiming SS
1983 elimination for 70-71 year old men						
Dep. Mean	2784	0.182	2.09	38.2	27.8	0.935
ELIM	21	-0.016*	-0.060	2.46**	0.19	0.015**
	(285)	(0.010)	(0.098)	(0.99)	(0.94)	(0.006)
R-squared	0.040	0.023	0.045	0.017	0.014	0.021
N	25,268	25,268	4,730	4,730	4,730	25,268
2000 elimination for 66-69 year old men						
Dep. Mean	7086	0.230	2.55	41.8	31.1	0.897
ELIM	1644**	0.015*	0.117*	-1.00	1.01	0.025**
	(552)	(0.008)	(0.063)	(0.668)	(0.640)	(0.006)
R-squared	0.043	0.048	0.054	0.012	0.024	0.040
N	40,043	40,043	9,456	9,456	9,456	40,043
1983 elimination for 70-71 year old women						
Dep. Mean	626	0.082	1.77	37.2	23.4	0.928
ELIM	-18	0.003	-0.187*	0.197	-2.08*	0.003
	(86)	(0.006)	(0.097)	(1.27)	(1.11)	(0.005)
R-squared	0.010	0.018	0.31	0.020	0.027	0.042
N	35,810	35,810	2,950	2,950	2,950	35,810
2000 elimination for 66-69 year old women						
Dep. Mean	2117	0.141	2.20	40.2	26.6	0.904
ELIM	461**	0.018**	-0.072	-0.752	-1.08	0.016**
	(192)	(0.006)	(0.061)	(0.803)	(0.688)	(0.005)
R-squared	0.027	0.049	0.032	0.018	0.029	0.043
N	51,162	51,162	7,376	7,376	7,376	51,162

Notes: * and ** indicate statistical significance at the 0.10 and 0.05 level, respectively. *Data source*: 1979-88 and 1996-04 CPS.

Table 4: The Effect of Eliminating the 2000 Earnings Test on Aggregate Earnings and Employment by Age

	Earnings		Work	
Results for men				
Dep. mean	7086	7086	0.230	0.230
ELIM × age 66	2125**		0.005	
	(900)		(0.014)	
ELIM × age 67	1503*		0.051**	
	(919)		(0.014)	
ELIM × age 68	2123**		0.019	
	(933)		(0.014)	
ELIM × age 69	776		-0.016	
	(938)		(0.014)	
ELIM		1918**		0.025**
		(602)		(0.009)
ELIM × age trend		-228		-0.008**
		(200)		(0.003)
R-squared	0.043	0.043	0.049	0.048
N	40,043	40,043	40,043	40,043
Results for women				
Dep. mean	2117	2117	0.141	0.141
ELIM × age 66	468		0.022**	
	(324)		(0.011)	
ELIM × age 67	760**		0.028**	
	(324)		(0.011)	
ELIM × age 68	402		0.001	
	(326)		(0.011)	
ELIM × age 69	207		0.023**	
	(324)		(0.011)	
ELIM		546**		0.017**
		(211)		(0.007)
ELIM × age trend		-68		0.001
		(70)		(0.002)
R-squared	0.028	0.027	0.056	0.056
N	51,162	51,162	51,162	51,162

Notes: * and ** indicate statistical significance at the 0.10 and 0.05 level, respectively. *Data source*: 1996-04 CPS.

Table 5: The Effect of Eliminating the Earnings Test on Aggregate Earnings and Claiming by Educational Attainment

	Earnings		Claiming	
	1983	2000	1983	2000
Results for men				
Dep. mean	1621	6861	0.935	0.897
ELIM × less than high school	77	-972	0.020**	0.013
	(193)	(790)	(0.007)	(0.009)
ELIM × high school	74	1562**	-0.008	0.014*
	(269)	(741)	(0.009)	(0.008)
ELIM × some college	1181**	965	0.024	0.036**
	(353)	(895)	(0.013)	(0.010)
ELIM × college	1452**	6168**	0.029**	0.040**
	(362)	(801)	(0.013)	(0.009)
R-squared	0.039	0.045	0.022	0.040
N	25,268	40,043	25,268	40,043
Results for women				
Dep. mean	369	2056	0.928	0.904
ELIM × less than high school	46	-460	0.003	0.009
	(61)	(282)	(0.006)	(0.008)
ELIM × high school	20	116	0.004	0.023**
	(77)	(235)	(0.008)	(0.007)
ELIM × some college	175	2043**	0.002	0.011
	(111)	(310)	(0.012)	(0.009)
ELIM × college	520**	1792**	0.006	0.015*
	(136)	(345)	(0.014)	(0.008)
R-squared	0.010	0.030	0.042	0.043
N	35,810	51,162	35,810	51,162

Notes: * and ** indicate statistical significance at the 0.10 and 0.05 level, respectively. *Data source*: 1979-88 and 1996-04 CPS.

Figure 1: The Retirement Earnings Test and the Budget Constraint

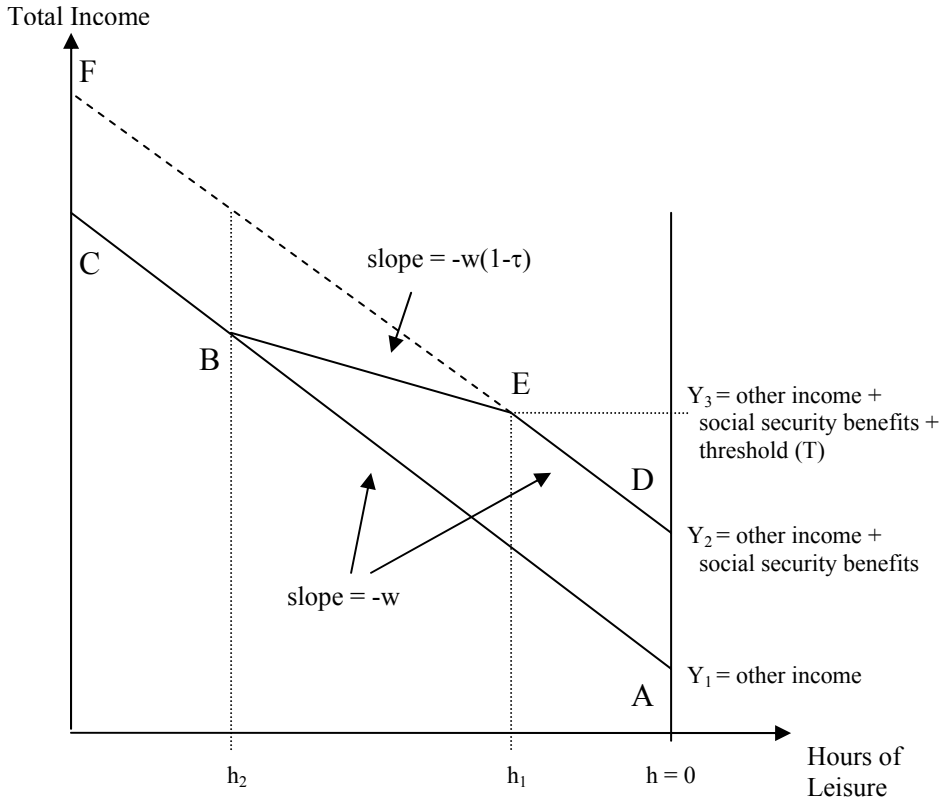
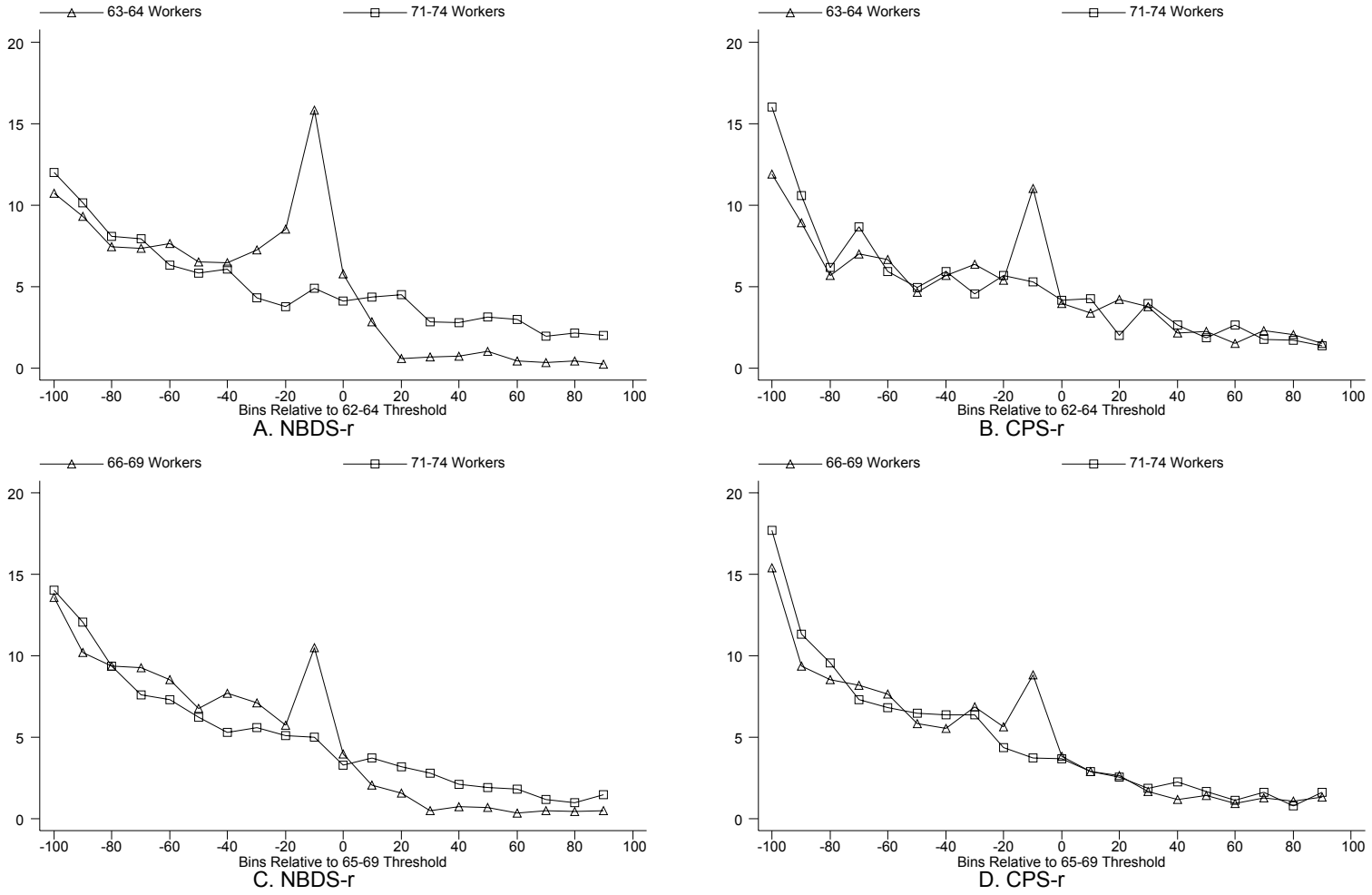
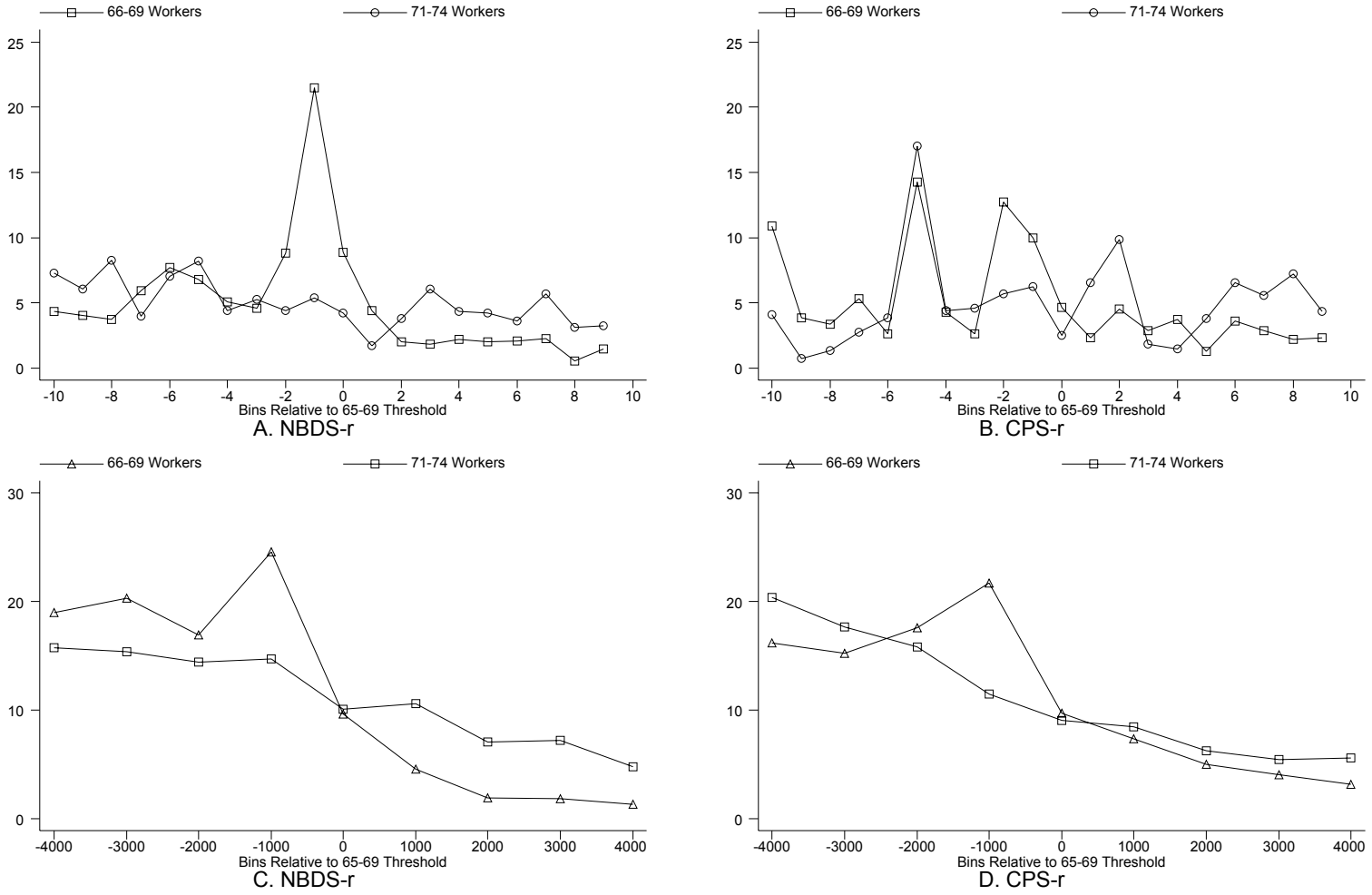


Figure 2: Histograms of Relative Earnings for Men, 1983-1989



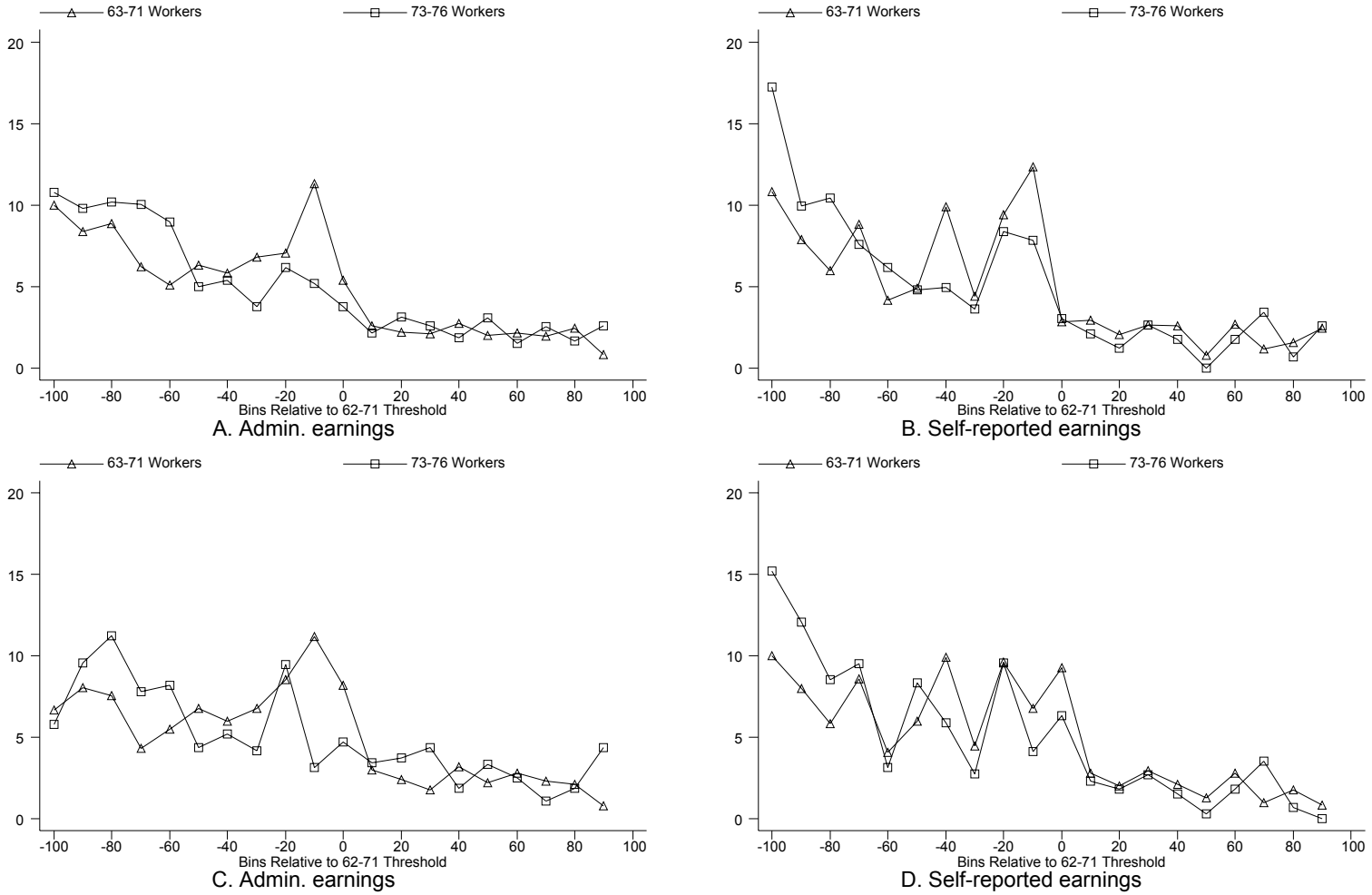
Notes: Each bin represents 10 percentage point intervals relative to an age-specific threshold.

Figure 3: Alternative Histograms of Relative Earnings for Men, 1983-1989



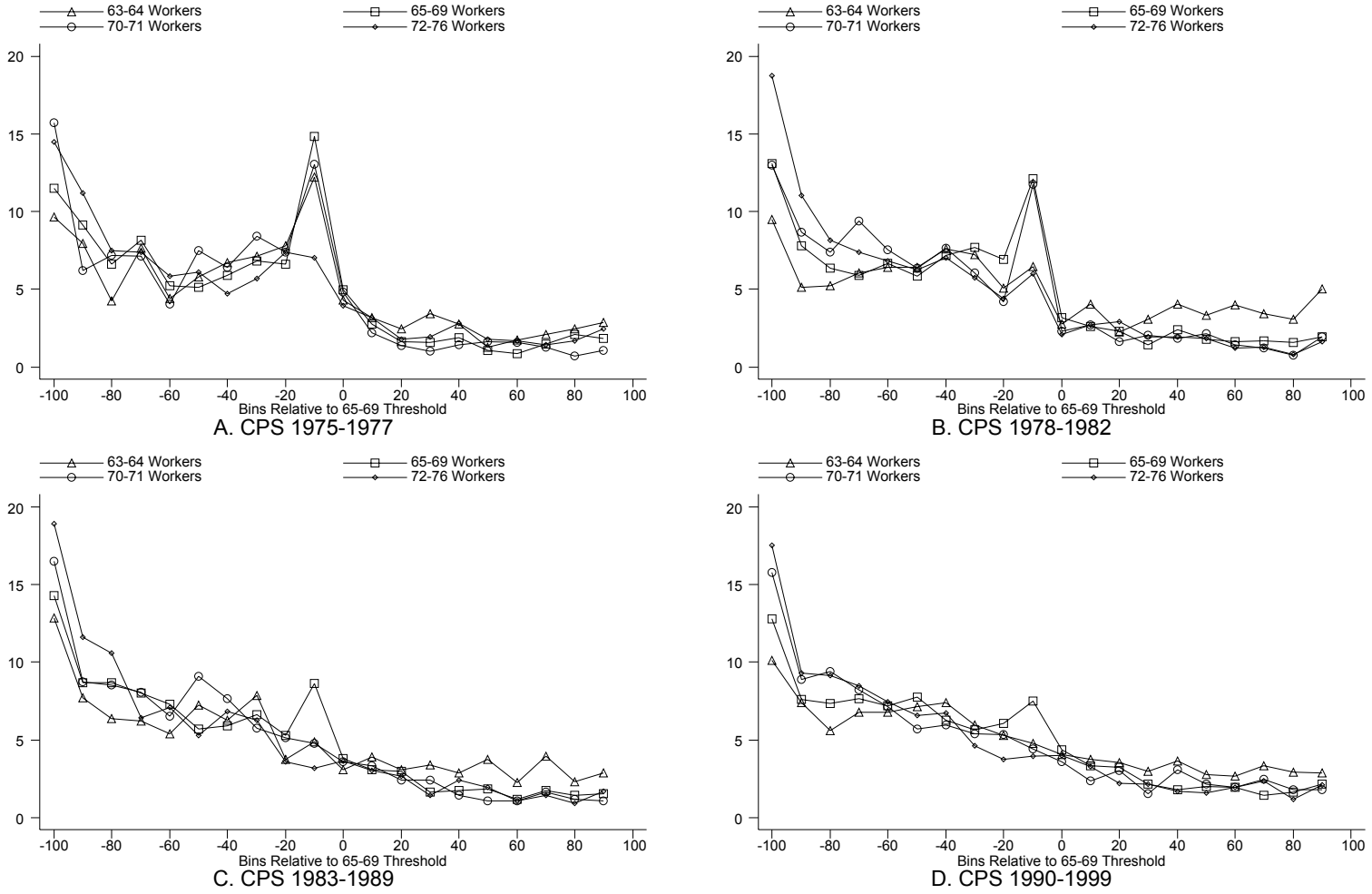
Notes: Panels A and B use 1 percentage point intervals and Panels C and D use $1000 intervals relative to the age 65-69 threshold.

Figure 4: Histograms of Relative Earnings for Men, 1977



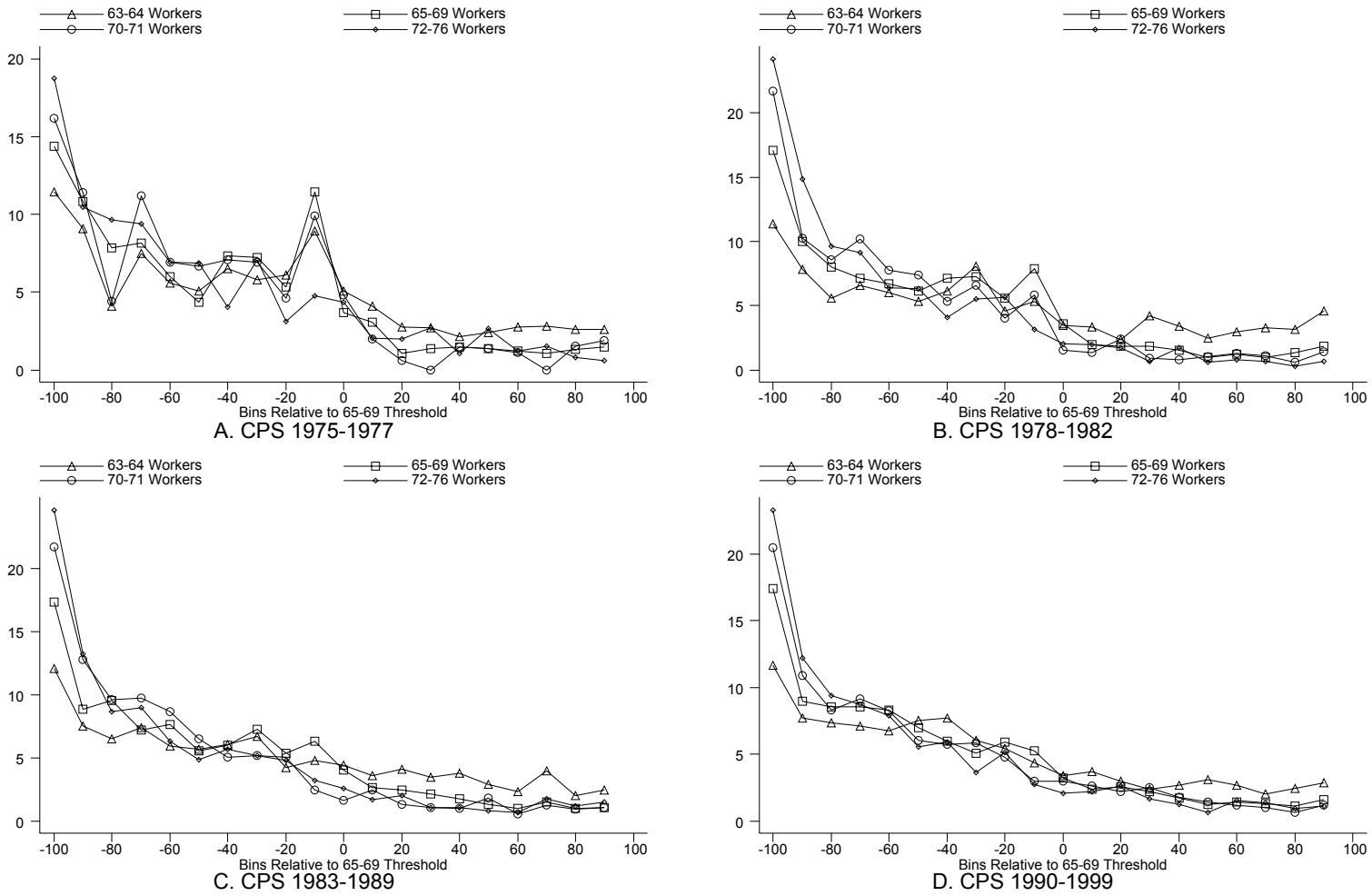
Notes: Each bin represents 10 percentage point intervals relative to the 62-71 earnings threshold. Data Source: 1978 CPS-SER.

Figure 5: Histograms of Relative Earnings for Men, 1975-1999



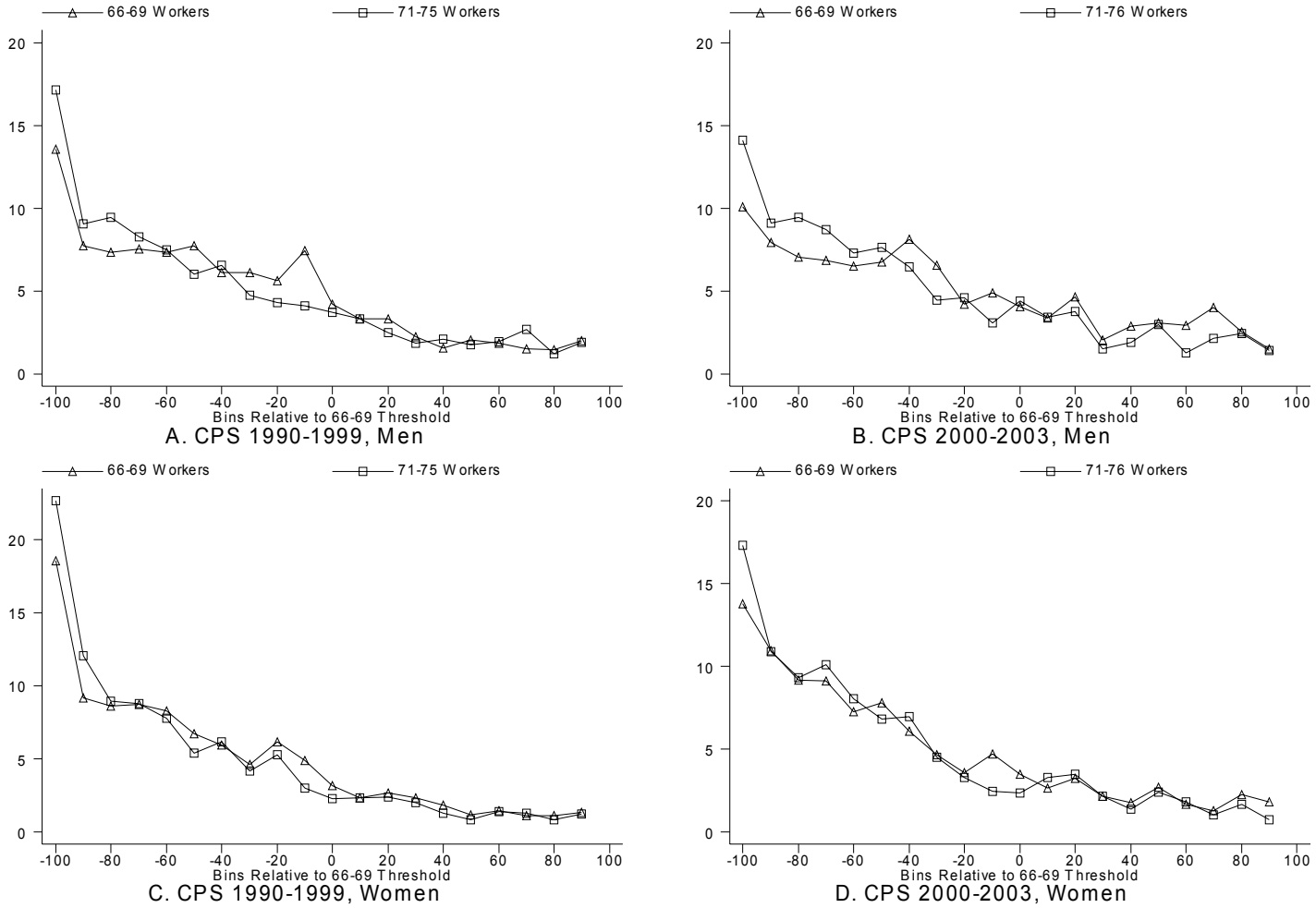
Notes: Notes: Each bin represents 10 percentage point intervals relative to an age-specific threshold. Data source: 1976-2000 CPS.

Figure 6: Histograms of Relative Earnings for Women, 1975-1999



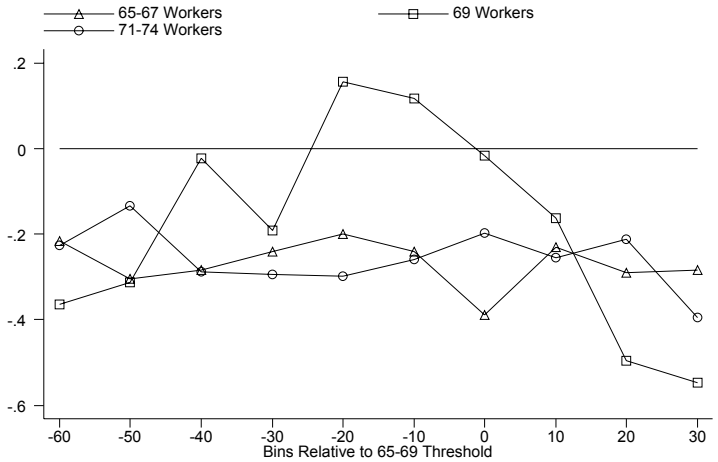
Notes: Each bin represents 10 percentage point intervals relative to an age-specific threshold. Data source: 1976-2000 CPS.

Figure 7: Histograms of Relative Earnings for Men and Women, 1990-2002

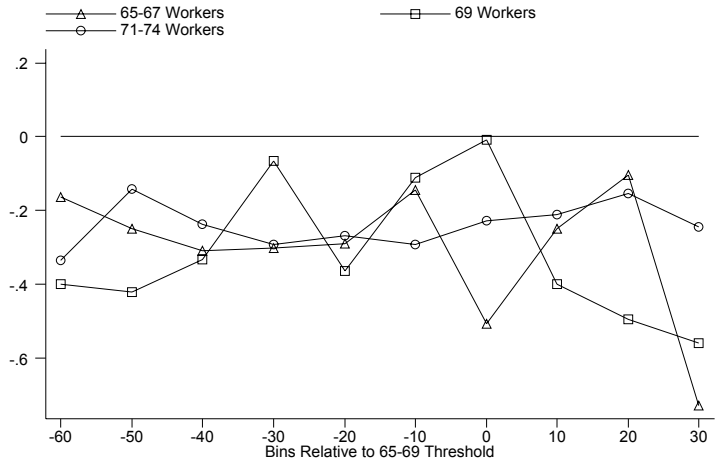


Notes: Each bin represents 10 percentage point intervals relative to an age-specific threshold. Data source: 1991-2003 CPS.

Figure 8: Change in Earnings by Age-69 Bin and Age, 1983-1989



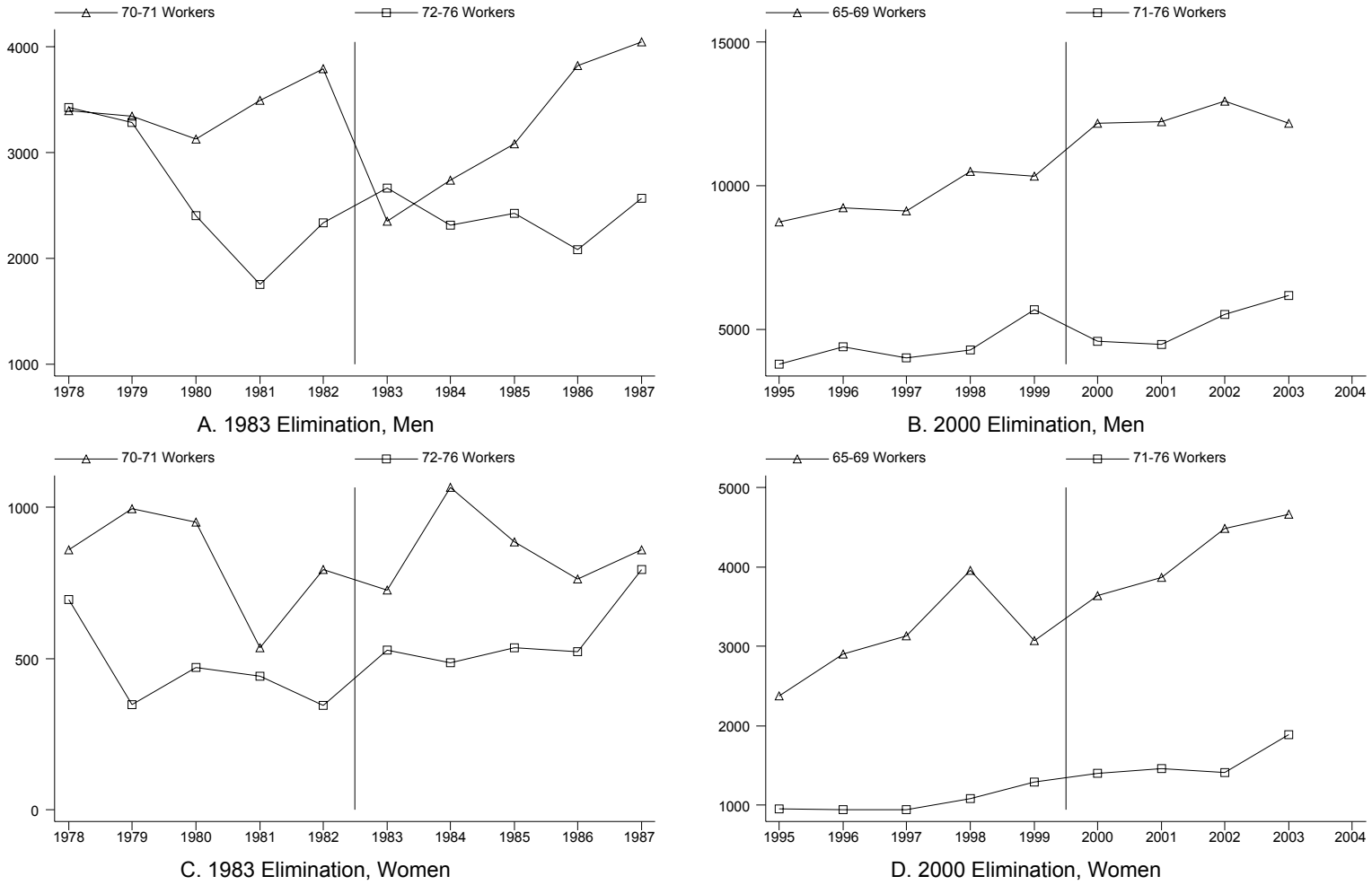
A. Growth between t and t+2, Men



B. Growth between t and t+2, Women

Notes: Each bin represents 10 percentage point intervals relative to the threshold at age 69. Data source: NBDS.

Figure 9: Annual Earnings by Age and Year



Data source: 1979-88 and 1996-2003 CPS.