Social Security and elderly labor supply: Evidence from the Health and Retirement Study

James P. Vere *

School of Economics and Finance, University of Hong Kong, Pokfulam Road, Hong Kong

A R T I C L E   I N F O

Article history:
Received 12 February 2010
Received in revised form 24 January 2011
Accepted 12 February 2011
Available online 19 February 2011

JEL classification:
H55
H31
J22

Keywords:
Social Security
Labor supply
Notch cohorts

A B S T R A C T

This study uses panel data from the Health and Retirement Study (HRS) to estimate the effects of Social Security income on elderly labor supply in the 1990s and early 2000s. The identification strategy takes advantage of the 1977 amendments to the Social Security Act, which led to a large, unanticipated reduction in Social Security benefits for those born after January 1, 1917. Despite the advanced age of the notch cohorts, there is a significant, negative and surprisingly elastic relationship between Social Security income and hours of work. This suggests that currently proposed reductions in benefits would induce Social Security recipients to work more hours in retirement, even through their 70s and early 80s.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The Social Security system of old age, survivorship and disability benefits is the largest and best-known social insurance program in the United States. Unfortunately, it is also well-known that the Social Security system is fiscally unsustainable at current levels of benefits and payroll taxes. Because of this, numerous proposals have been made to reform it in some way, most of which involve reducing benefits in one form or another (e.g., Moynihan and Parsons, 2001; Diamond and Orszag, 2005).

An important factor to understand when assessing the welfare effects of these proposals is the elasticity of elderly labor supply. If older workers can offset lost Social Security income with labor earnings, either by delaying retirement or working part-time in retirement, they will be less negatively affected by reductions in benefits than otherwise. Models of the joint retirement and post-retirement hours of work decision predict this response (Burtless and Moffitt, 1985), and there may even be health benefits from doing so (Snyder and Evans, 2006). On the other hand, continuing an attachment to the labor force at later ages may be difficult, particularly for the less-educated or those in more physically intensive jobs.

This paper contributes to an understanding of these welfare effects with estimates of how workers beyond traditional retirement ages respond to legislated changes in Social Security benefits. At present, very little is known about the labor supply behavior of this segment of the population. Nevertheless, the matter is of critical public interest: since the typical retirement extends well past the range of traditional retirement ages, any assessment of the welfare effects of a possible across-the-board benefit cut must necessarily take this information into account. From an academic perspective, the results are also important because elderly labor supply has generally been thought of as fairly inelastic to changes in Social Security benefits (Krueger and Pischke, 1992). However, this does not appear to be the case any longer.

A priori, there are a number of reasons to believe that the relationship between Social Security income and elderly labor supply has changed since the 1970s and 1980s, the years studied by much of the earlier literature (Hurd and Boskin, 1984; Burtless, 1986; Krueger and Pischke, 1992; Snyder and Evans, 2006). On a policy level, the enactment of laws against age discrimination and the abolition of mandatory retirement have also encouraged the employment of older workers (Neumark and Stock, 1999). Moreover, the implicit tax imposed by the Social Security earnings test, which discourages recipients of retirement benefits from earning labor income, was progressively reduced and ultimately eliminated for beneficiaries at or above the normal retirement age (Gruber and Orszag, 2003; Benitez-Silva and Heiland, 2007; Haider and Loughran, 2008; Engelhardt and Kumar, 2009). Finally, a critical
background element of Krueger and Pischke's (1992) analysis – the fact that elderly labor supply had been declining, notwithstanding legislatively imposed reductions in Social Security wealth – is simply no longer the case. This trend reversed itself in the early 1990s, and elderly workers are now a rapidly growing segment of the labor force (Gendell, 2008).

This paper examines the link between Social Security income and elderly labor supply with detailed Health and Retirement Study data and a well-established instrumental variables strategy based on the structure of the Social Security benefits formula. The intuition behind the estimation strategy is to exploit the fact that Social Security benefits are determined by different sets of rules depending on beneficiaries’ years of birth. Because of this, individuals who are otherwise identical – i.e., individuals with the same real earnings profile, but born in different years – will receive different amounts of benefits. Numerous recent studies have taken advantage of this aspect of the Social Security rules to quantify the effects of retirement income on various economic outcomes (e.g., Engelhardt et al., 2005; Engelhardt and Gruber, 2006; Moran and Simon, 2006; Snyder and Evans, 2006; Engelhardt, 2008; Gustman and Steinmeier, 2008). The estimation strategy is conservative on purpose: since the “natural experiment” involved is very well known, the empirical results can be interpreted and accepted by the widest possible audience of economists.

The empirical results show that the relationship between Social Security income and elderly labor supply is strongly negative, and much more elastic than one would expect given the advanced age of workers in the sample (who are, on average, 79).1 For instance, a $1000 increase in Social Security income (in 2009 dollars) reduces beneficiaries’ labor force participation by 0.9 percentage points (to put this number in perspective, the overall labor force participation rate for this group is 12.3%). Among married couples, wives’ labor supply is more responsive to changes in Social Security income than husbands’ labor supply. In addition, the labor supply of less-educated workers is more sensitive to variation in Social Security income than the labor supply of more-educated workers, particularly on the extensive margin (i.e., the decision whether or not to work).

This paper is organized as follows. The next section provides a brief overview of changes in the Social Security benefits formula. The third section describes the data, instruments and first-stage relationships; the fourth section details the empirical results; and the last section concludes.

2. The Social Security notch

The Social Security notch has been widely studied in the empirical literature (for a detailed description, see Krueger and Pischke, 1992). Because of this, the full details of the legislation that created varying benefits formulas for different cohorts of Social Security beneficiaries will not be rehashed here. Nevertheless, as motivation for the estimation strategy, it is useful to summarize the changes that occurred.

Before 1972, Social Security benefits were related to beneficiaries’ average monthly earnings (AME), which were calculated over a worker’s entire earnings history. The mapping of AMEs to actual monthly benefits was set by a fixed formula, which Congress updated from time to time to keep up with inflation. In 1972, Congress replaced this ad hoc adjustment mechanism with a system of automatic adjustments based on changes in the consumer price index.

The flaw in this system was that benefits were linked to workers’ average nominal monthly earnings. As a result, retirees from later cohorts, with higher nominal average wages than their predecessors, received greater real benefits than retirees from earlier cohorts even if their earnings were the same in real terms. Before the 1972 changes, this feature of the benefits formula was relatively benign because Congress based its adjustments on accumulated surpluses. However, unlike the previous system, the automatic adjustments were not contingent on continuing fiscal solvency. Worse, the automatic adjustments were implemented just as the United States entered a period of high inflation. Since inflation outstripped the wage growth that would have otherwise financed the increases, the system accumulated a long-run deficit that would bankrupt it by the early 1980s.

The 1977 amendments to the Social Security Act solved this problem by tying benefits to average indexed monthly earnings (AIME), where the index used to adjust nominal earnings is based on changes in average wages. This eliminated the double-indexation problem, by which new beneficiaries enjoyed the benefits of nominal wage growth (which was reflected in their AMEs) in addition to inflation-based increases. However, the new formula was not applied to everyone—current beneficiaries and those near retirement (i.e., those born before January 1, 1917) were grandfathered under the old rules. This aspect of the Social Security formula, in combination with the fact that different cohorts were exposed to different levels of wage growth over their working lives, generated significant variation in Social Security benefits by recipients’ years of birth.

From an analytical perspective, this source of exogenous variation in Social Security benefits is attractive for several reasons. First, the entire population is subject to variation in the Social Security rules. This means that instrumental variables estimates based on this identification strategy will broadly reflect the behavior of the general population.2 Second, the changes in the Social Security rules were unanticipated, which removes an important potential source of endogeneity bias. Were the changes anticipated, it is conceivable that individuals would have altered their pre-retirement consumption and labor supply, which would have mitigated the effects of these changes. However, the level of public anger over the notch issue strongly suggests that this was not the case. Third, since the Social Security rules are based simply on year of birth, they are not related to other endogenous determinants of income at the individual level, such as unobserved ability or health status. Finally, the changes were large enough in real terms that instruments based on these changes have a strong relationship with beneficiaries’ actual Social Security benefits (i.e., there is no “weak instruments” problem (Bound et al., 1995)). This means that estimates based on these instruments will be more efficiently identified.

3. Data, instruments and first-stage relationships

3.1 Data

The empirical work in this paper uses data3 from the five cohorts that participate in the Health and Retirement Study (HRS), an ongoing panel survey sponsored by the National Institute of Aging. These cohorts have been surveyed over multiple waves from 1992 to 2006. The earliest are the initial HRS cohort, born between 1931 and 1941.

1 It is worth noting that labor force participation among the elderly in general is significantly higher than most would expect. In 2009, the average labor force participation rate for men ages 75 and higher was 10.3%; for women it was 5.3% (Bureau of Labor Statistics).

2 A popular criticism of instrumental variables estimates is that they identify local average treatment effects (Angrist and Imbens, 1995). The canonical example is that of using draft lottery numbers as an instrument for military service, to study the effects of military service on wages. The instrumental variables estimate identifies the causal effect of military service only for the population of “compliers,” or people who joined the military only because they were drafted. Such people may be quite different from draft-dodgers (who are called but do not go) or volunteers (who go whether or not they are drafted). However, unlike the draft, the Social Security formula determining one’s benefits cannot be avoided. Nor, for that matter, can someone volunteer for a formula other than the one to which he or she was assigned based on his or her year of birth.

3 RAND HRS data file, version 1 (June 2010).
and the Assets and Health Dynamics Among the Oldest Old (AHEAD) cohort, born before 1924; these cohorts have participated in the HRS for all seven waves of the study. The next two are the Children of Depression (CODA) and War Baby cohorts, born between 1924 and 1930 and 1942 and 1947, respectively; these have participated in the four survey waves since 1998. Finally, the Early Baby Boomer cohort, born between 1948 and 1953, has only participated since 2004.

The HRS data have several advantages that relate to the analysis at hand. First, the survey is extensive and a considerable number of observations are available. Further, not only do the data include detailed information on income and labor force participation, but also they include respondents’ exact years of birth, which is critical for identifying which Social Security rules respondents fall under. Past studies with CPS data (e.g., Krueger and Pischke, 1992; Engelhardt et al., 2005; Engelhardt and Gruber, 2006; Engelhardt, 2008) have been limited by the fact that, given only data on respondents’ ages and the current time, it is not possible to identify respondents’ years of birth without some degree of error. Second, unlike the CPS data, the HRS data have a more extensive panel structure. This permits the use of more efficient estimation methods than the pseudo-panel strategies the CPS has required of previous studies.

Since exogenous variation in Social Security benefits attributable to the benefits notch largely affects retirees born between 1901 and 1930, prior research using HRS data to study the implications of this variation has mostly focused on the AHEAD cohort (Moran and Simon, 2006). However, it is worth noting that many of the other cohorts also contain respondents born within the years of interest. This is because, while one member of each household usually (but not always) has a date of birth within the “target” interval, spouses are also included in the survey. Therefore, to create as large a data set as possible, it is useful to combine data from all five HRS cohorts. Doing so results in a data set that is 50% larger than that from the AHEAD cohort alone (however, see the Appendix A for robustness checks using the AHEAD and CODA cohorts, and only the AHEAD cohort). While this does mean that the estimates will rely more heavily on functional form assumptions—particularly where selection on age is concerned—these assumptions are quite reasonable, as a graph discussed at the end of this subsection will show.

In order to identify the Social Security formula that applies to each household, it is necessary to designate an individual in each household as the “Social Security beneficiary” (note that all HRS “households” consist of couples or single individuals). In doing so, this paper adopts an algorithm that is now well-established by prior work (e.g., Engelhardt et al., 2005; Engelhardt and Gruber, 2006; Moran and Simon, 2006; Engelhardt, 2008). First, the Social Security beneficiary is designated as the male present who is age 65 or older. If no man matches this description, the designation passes to the never-married woman present, provided that she is at least age 65 herself. Finally, if no one falls into either of these categories, the Social Security beneficiary is taken to be the widowed or divorced woman in the household age 62 or older.

When a widow or divorced woman is the Social Security beneficiary, the ‘birth year of the Social Security beneficiary’ is deemed to fall 3 years prior to her actual birth year. This is done because Social Security survivor’s benefits relate to the birth year and earnings history of the former spouse. This assumption, also adopted by Moran and Simon (2006), is motivated by Engelhardt et al.’s (2005) calculation of the median age difference between widowed and divorced female Social Security recipients and their former spouses. Engelhardt et al. (2005) arrived at this figure by tabulating data from the 1982 Social Security New Beneficiary Survey. In total, 10,414 households in the HRS data contained an individual, born between 1901 and 1930, who could be designated as the Social Security beneficiary under these rules. Table 1 indicates exactly how many households are contributed by each year of birth. On average, each household participated for four waves, leading to 42,089 household-wave observations in total.

Many previous studies of the Social Security notch make use of cohort level data. An early motivation for using cohort level data was simply that no panel data set covering the cohorts of interest was available (Krueger and Pischke, 1992). Another useful aspect of cohort level data is that the analysis of such data allows one to sidestep potentially dicey issues of heterogeneity across individuals. One must still believe that, after controlling for differences in observed characteristics, the cohorts are identical apart from variation in the level of Social Security benefits they receive; but this proposition is easier to accept at the cohort level than it is at the individual level. That said, aggregating the data is not without cost. Although the instrument itself varies only by cohort and year, the micro-level results are nevertheless expected to be more precise because the coefficients on the exogenous covariates are estimated more efficiently when the micro-level data are used.

For these reasons, the estimation results to follow are presented for two versions of the HRS data: the full HRS microsample, and a cohort level HRS sample that consists of 387 birth year-calendar year cells. In the aggregate sample, the relevant variables are expressed as within-cell means. It is worth noting that both types of analysis have been undertaken by prior researchers; Moran and Simon (2006) analyze HRS microdata, while Snyder and Evans (2006) analyze microdata from the National Health Interview Survey. At the same time, a number of others have analyzed cohort level CPS data (Krueger and Pischke, 1992; Engelhardt et al., 2005; Engelhardt and Gruber, 2006).6

Table 1
Distribution of Social Security beneficiaries’ years of birth in the HRS sample.

<table>
<thead>
<tr>
<th>Year of birth</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1901</td>
<td>61</td>
</tr>
<tr>
<td>1902</td>
<td>85</td>
</tr>
<tr>
<td>1903</td>
<td>82</td>
</tr>
<tr>
<td>1904</td>
<td>140</td>
</tr>
<tr>
<td>1905</td>
<td>158</td>
</tr>
<tr>
<td>1906</td>
<td>212</td>
</tr>
<tr>
<td>1907</td>
<td>191</td>
</tr>
<tr>
<td>1908</td>
<td>215</td>
</tr>
<tr>
<td>1909</td>
<td>270</td>
</tr>
<tr>
<td>1910</td>
<td>323</td>
</tr>
<tr>
<td>1911</td>
<td>309</td>
</tr>
<tr>
<td>1912</td>
<td>348</td>
</tr>
<tr>
<td>1913</td>
<td>341</td>
</tr>
<tr>
<td>1914</td>
<td>375</td>
</tr>
<tr>
<td>1915</td>
<td>396</td>
</tr>
</tbody>
</table>

4 The person weights provided by the HRS assume that the spouse of an HRS respondent will be included in an analysis of the spouse’s birth cohort, even if the spouse was not originally sampled as a member of that birth cohort.

5 Since HRS couples are strictly opposite-sex pairs, there is at most one male and one female in any household.

6 Deaton (1985) describes in detail how repeated cross-sections can be used to calculate “pseudo-panel” estimates of parameters of interest. The basic idea is that, when repeated cross-sections are sampled from the same cohort, one can still control for cohort-level fixed effects, even though the individual members of each cohort are not the same from one sample to the next. The Social Security notch literature has generally focused on analyzing single-year birth cohorts close enough together that this fixed effect is plausibly the same for all of them. When this is the case, estimates from a standard 2SLS regression on the aggregate data can be interpreted as average, cohort-level responses to exogenous variation in the Social Security rules.

7 Unfortunately, though a fixed-effects estimation strategy comes to mind as an obvious solution to this problem, this method is incompatible with the use of the Social Security notch as an instrument because the variation in benefits induced by the notch is across cohorts; virtually none of this variation is of the “within-person” variety. Hence, the fixed-effects transformation, when attempted, annihilates nearly all the power of the instrument, which renders the resulting estimates useless. Fully exploiting the panel structure of the data thus necessitates the use of a random effects-based estimator (e.g., Balestra and Varadhharajan-Krishnakumar’s (1987) G2SLS estimator).
2006). Therefore, presenting both sets of results is useful because it facilitates comparison with earlier work. Table 2 contains some basic summary statistics describing the data set, for the full pooled sample (in both its full and cohort level forms) and also subsamples by marital status and education of the Social Security beneficiary. In addition to these summary statistics, Fig. 1 plots the age distribution of Social Security beneficiaries over the sample (counting each household-wave observation as a separate observation). It can be seen that most (about 84%) of the respondents are in their 70s and 80s.

To take a closer look at labor supply behavior among the Social Security beneficiaries, Fig. 2 graphs labor force participation by age for the pooled data and each of the subsamples. This graph is useful for two reasons. First, since the policy variation affected people who were in their late 70s and early 80s when the HRS was conducted, it is important to check whether beneficiaries in this age range still supply a meaningful amount of labor. As it turns out, they do; the labor force participation rate of 80 year olds in the sample is about 7.5%. For those who are married or who have at least 1 year of college education, this rate is just over 10%.

A second reason the graph is useful is that it provides some guidance as to what functional form is reasonable when controlling for age effects. For efficiency reasons, it is necessary to impose some parametric structure on the relationship between age and labor supply (though see the Appendix A for robustness checks with alternative structures). Fig. 3 shows that a quadratic specification is a very good fit to the data (for the pooled sample, R² = 0.991) and so it is reasonable to use age and age squared terms in the subsequent regression analysis.

Finally, before proceeding to the instrumental variables analysis, it is useful to look at the raw data and examine whether or not there are discernible differences in labor supply by cohort. To this end, Fig. 4 plots male labor force participation by age for three cohorts of interest: a “peak” cohort, born between 1915 and 1917, and two “trough” cohorts born four years before and after these dates. The 1919–21 trough cohort was chosen because, as the next section describes, real benefits did indeed reach a trough in these years. The 1911–13 trough cohort was chosen for reasons of symmetry and also because the real benefits enjoyed by both trough cohorts, as measured by the instrument, are approximately similar.

Looking at Fig. 4, one can see that, at similar ages, the labor force participation of the peak cohort is visibly less than the labor force participation of the trough cohorts—and often by more than one percentage point. The sole exception to this is age 81, which is when the peak cohort experienced the Asian financial crisis. Apart from this outlier year, the pattern is quite consistent, and interestingly becomes more pronounced with age.
3.2. Instruments

Prior instrumental variables studies that use the Social Security notch to examine various economic outcomes have adopted two general strategies when constructing the instrumental variable. The first is to use year of birth as an indicator for which set of Social Security rules an individual falls under. For instance, Snyder and Evans (2006), in studying the effects of retirement income on mortality, use a difference-in-difference strategy that compares mortality rates of individuals born immediately before and after January 1, 1917 (the cutoff date for retirees to be fully grandfathered under the pre-notch rules).

In a similar vein, Moran and Simon (2006) specify their instrument as a binary variable equal to one for individuals born between 1915 and 1917. They choose this interval because it generates the strongest first-stage relationship between the instrument and Social Security income. While this strategy is certainly valid, there is nevertheless room for improvement because, in virtually all cases, the actual variation in the Social Security rules is too complex to be fully captured – or even mostly captured – by a binary instrument.

The second strategy is to use the structure of the Social Security rules to directly measure the intensity of the rules’ effects on the endogenous variable. This is the method used by Engelhardt et al. (2005), Engelhardt and Gruber (2006), and Engelhardt (2008). These authors calculate the monthly Social Security benefits earned by individuals with identical real earnings histories but born in different years. Variation in this instrument thus directly reflects variation in Social Security income that can be attributed solely to differences in the Social Security rules applied to different cohorts.

Although the Social Security benefits of individuals with hypothetical earnings histories and dates of birth are complex to calculate, the efficiency gain from using the full range of variation in the Social Security rules is substantial. In exploratory estimates with the binary version of the instrument and the HRS micro-level data, the first-stage $F$ statistic testing the significance of the Social Security instrument was 0.4. The binary instrument also had an unexpected negative sign. By contrast, the ‘complex’ version of the instrument yielded a first-stage $F$ statistic of 401.2. This improvement occurs because year-of-birth indicators are highly correlated with controls for age and calendar year; hence, any instrument based purely on year of birth will have very limited explanatory power when age and calendar year variables also appear in the model. For this reason, the second strategy was adopted.

The values of the instrument were calculated in the following manner. First, the 1918 cohort was chosen as the baseline cohort. This choice is sensible for two reasons. One is that 1918 is near the midpoint of the 1901–1930 birth interval that defines the population of interest; as such, it is close to when reductions in benefits induced by the Social Security notch started to take effect. This cohort is also convenient because individuals born in 1918 fall exactly at the midpoint of the five-year age categories for which median earnings statistics are available from the 2005 Annual Statistical Supplement to the Social Security Bulletin. This publication contains a table of workers’ median annual Social Security earnings, by age and sex, during selected years (4.86).

Earnings figures were assigned to the 1918 cohort directly from the published figures, for male workers, for the years that are reported (generally, years that are evenly divisible by 5). For example, median annual Social Security earnings for male workers aged 30 to 34 in 1950 (who were born between 1916 and 1920) were $2918. This number was then assigned as earnings for the 1918 cohort in 1950. Although some Social Security beneficiaries in the data are never-married women, male earnings histories were used to construct the instrument in all cases because never-married women’s earnings are much more highly correlated with men’s earnings than they are with women’s earnings generally.

Up to age 60, earnings were assumed to progress linearly between points specified by the table; past age 60, earnings were assumed to grow at the rate of the Consumer Price Index (CPI). Earnings in the table for workers past age 60 were not used because a fair proportion of workers in this age range are no longer engaged in their primary careers; because of this, earnings of these workers would not be representative of those who stay in their main career until retirement. Earnings were also assumed to start at age 21. Therefore, the earnings history is that of a median male wage earner, born in 1918, who remains in his primary career from his 21st birthday through retirement. It is worth noting that, since Social Security benefits depend on workers’ earnings during their highest $n$ working years ($n$ is 35 today but lower for workers born near the start of the program; for those born in 1901, $n = 10$), the assumed age for starting work does not actually affect the values of the instrument, as long as the starting age falls within a fairly wide range of “reasonable” values.

With this earnings history in hand, Social Security’s ANYPIA calculator was used to calculate the retirement benefits of a worker with these characteristics, for each single-year retirement age from 62 to 68 and each calendar year from 1991 to 2006 (the years for which the HRS collected data on Social Security income). The baseline 2006 benefit (in current dollars, for retirement at age 65) was $1308 per

---

Bound et al. (1995) show that the finite-sample bias of an instrumental variables estimate relative to an OLS estimate of the same parameter is approximately equal to $1/(F)$, where $F$ is the first-stage $F$ statistic. As a rule of thumb, the first-stage $F$ statistic should exceed 10.
month. The CPI was then used to convert these figures to 2009 dollars. Since the CPI is the basis for adjusting Social Security benefits from one year to the next, the benefit amounts are nearly, though not perfectly, identical across years.\(^\text{10}\) This, incidentally, is the reason why an instrumental variables strategy based on the Social Security notch cannot be used in conjunction with fixed-effects estimation methods; so little of the variation in the instrument occurs across years that applying the fixed-effects transformation to the data makes the instrument effectively useless. Thus, it is necessary to use a random effects-based procedure, or, if the random-effects assumptions are too unpalatable, to aggregate the data and focus on cohort-level estimates.\(^{11}\)

Once the retirement benefits for the baseline cohort were calculated, a single measure of retirement benefits for each calendar year was constructed by taking a weighted average of the benefit associated with each retirement age. The weights were given by the proportion of the HRS sample claiming retirement benefits at each possible retirement age. Importantly, the proportions are for the entire sample, not just the 1918 cohort; therefore, no variation in the instrument can be attributed to differences in cohorts’ choices as to when to claim retirement benefits.

For other cohorts, the ANYPIA calculator was used to calculate the 1991–2006 weighted retirement benefits, for workers with the same real earnings histories as the baseline 1918 worker, but born in different years. For this purpose, the CPI was used to construct nominal earnings histories for the non-baseline cohorts. This exercise yielded considerable variation in the instrument across cohorts in the sample; the 2006 weighted retirement benefits for each cohort, in 2009 dollars, are shown in Fig. 5. For birth years prior to 1917, the effects of rising nominal earnings on real benefits are plainly visible. After 1917, real benefits fall with the phaseout of the transition formula, which applied to those born between 1917 and 1921. In later years, real benefits rise (slowly) because of gradual increases in the Social Security earnings ceiling.

In preparation for later empirical results, versions of the instrument were also created for use with the education subsamples. Values of the low-education instrument were obtained by multiplying the earnings histories for each cohort by 0.916 and then recalculating the instrument. This fraction, obtained from the 1960 U.S. census, is the ratio of the median earnings of workers born in 1918 with up to a high school education to the median earnings of all workers born in 1918. Similarly, for the high-education version of the instrument, the earnings histories of each cohort were multiplied by 1.483. Engelhardt et al. (2005) also use this method to adjust their instrument when estimating results for differently educated subsamples of their data, though they use 1962 March CPS data to obtain the necessary ratios. In any event, though using this procedure does improve the fit of the first stage, it does not appreciably change the final estimation results. Fig. 5 also contains values of the instrument for the education subsamples. It would not be appropriate to adopt this procedure for the marital status subsamples because, unlike education, it is not reasonable to suppose marital status is a predetermined variable.

\(^{10}\) In particular, there is a lag in the application of the cost-of-living adjustment, which is based on year-on-year movement in the CPI-W as of the third quarter of the previous year. For instance, the CPI-W increased by 3.3% between the third quarters of 2005 and 2006; however, this increase was not reflected in Social Security beneficiaries’ actual paychecks until January 2007. Therefore, converting the Social Security amounts to real dollars removes much, but not all, of the variation in benefits across years.

\(^{11}\) Several previous, related studies make use of micro-level cross-sectional data (Moran and Simon, 2006; Snyder and Evans, 2006). It is worth noting that, if one truly believes that the individual-level error term – which potentially includes an individual fixed effect, regardless of how many time periods the individual is actually observed – is only correlated with endogenous Social Security income, and no other variable in the model (i.e., if one has no problem with micro-level, cross-sectional 2SLS estimates where the instrument is based on the Social Security notch), then one has, in effect, accepted the critical assumption of the random-effects model.

Finally, the last step was to attach values of the instruments to individuals in the HRS sample according to the birth year of the Social Security beneficiary and the calendar year for which the beneficiary’s Social Security benefits were recorded. For married couples, the instrument was multiplied by 150%, which is the spousal benefit multiplier.

### 3.3. First-stage relationships

Although the Social Security instrument was constructed to reflect benefits received by ‘baseline’ workers from each cohort, it is useful to examine the extent to which the instrument is correlated with Social Security benefits received by actual beneficiaries in the data. To this end, Panel A of Table 3 reports estimates of \(\pi_1\) in the first-stage regression

\[
\text{SocSec}_{it} = \pi_0 + \pi_1 \cdot Z_{it} + \pi_2 \cdot X_{it} + \mu_{it}
\]

where SocSec\(_{it}\) is the annual Social Security retirement benefit, in 2009 dollars, of respondent \(i\) in year \(t\). Similarly, \(Z_0\) is the value of the instrument and \(X_{it}\) is a vector of explanatory variables for respondent \(i\) in year \(t\). When the cohort level data are analyzed, the first stage has

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Estimates of the effects of Social Security income (per thousand 2009 dollars) on older workers’ labor supply.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: First-stage results</strong></td>
<td><strong>Micro sample</strong></td>
</tr>
<tr>
<td>Instrument</td>
<td>0.398 (0.020)</td>
</tr>
<tr>
<td>(F)</td>
<td>401.2</td>
</tr>
<tr>
<td>Partial R(^2)</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Panel B: Instrumental variables (G2SLS/GLS) estimates</strong></td>
<td></td>
</tr>
<tr>
<td>Working for pay ((\times 100))</td>
<td>–0.83 (0.28)</td>
</tr>
<tr>
<td>Hours worked per week</td>
<td>–0.83 (0.09)</td>
</tr>
<tr>
<td>Weeks worked per year</td>
<td>–0.51 (0.13)</td>
</tr>
<tr>
<td><strong>Panel C: G2SLS estimates</strong></td>
<td></td>
</tr>
<tr>
<td>Working for pay ((\times 100))</td>
<td>–0.10 (0.03)</td>
</tr>
<tr>
<td>Hours worked per week</td>
<td>–0.09 (0.01)</td>
</tr>
<tr>
<td>Weeks worked per year</td>
<td>–0.06 (0.01)</td>
</tr>
<tr>
<td>Observations</td>
<td>42,089</td>
</tr>
</tbody>
</table>

Notes: The cohort level results use cell sizes as sample weights. Heteroskedasticity-robust standard errors are in parentheses. Social Security income is measured in thousands of 2009 dollars. The covariates in the model are beneficiary’s years of education, beneficiary’s age, beneficiary’s age squared, spouse’s years of education, spouse’s age, and spouse’s age squared; dummy variables for beneficiary’s sex, beneficiary’s race, spouse’s race, marital status and region of residence; a dummy variable equal to one if the household is a “couple household”; and a full set of year dummy variables.
the same form as in Eq. (1), but the variables are within-cell means, i refers to year of birth, and the results are weighted by cell size. The standard errors are heteroskedasticity-robust and clustered by year of birth. The ‘A’ panels of Tables 4 and 5 report estimates of $r_t$ for subsamples of the data by marital status and years of education.

The variables in $X_e$ are beneficiary’s years of education, beneficiary’s age, beneficiary’s age squared, spouse’s years of education, spouse’s age, and spouse’s age squared; dummy variables for beneficiary’s sex, beneficiary’s race, spouse’s race, marital status and region of residence; a dummy variable equal to one if the household is a “couple household”; and a full set of year dummy variables. When spousal information is not applicable (i.e., when the respondent is not living with a spouse), the spouse variables are set equal to zero. In these cases, and only in these cases, the couple household indicator is equal to zero. Therefore, the couple household variable serves to “dummy out” observations where the spouse variables are irrelevant.

When the microdata are analyzed, the first stage is that obtained from Balestra and Varadharajan-Krishnakumar’s (1987) generalized two-stage least squares (G2SLS) estimator. This estimator is essentially a random-effects generalization of the familiar two-stage least squares (2SLS) estimator. As mentioned earlier, due to the fact that there is very little variation in the instrument over time, it is necessary to use a random-effects model to take full advantage of the panel structure of the data. When the cohort level data are analyzed, the first stage is estimated by weighted least squares, where the weights are cell sizes; the standard errors are heteroskedasticity-robust and clustered by birth year. The cohort level results, though not as precise, do not require the stringent assumptions of the G2SLS estimator.12

In all cases, the magnitudes of the coefficients in the ‘A’ panels are similar to or somewhat greater in magnitude than those of comparable coefficients reported by Engelhardt et al. (2005). One explanation for this difference is that, unlike the CPS data that Engelhardt et al. (2005) analyze, the HRS data contain exact information on respondents’ years of birth. As a result, values of the instrument can be mapped more precisely onto observations in the data, which improves the correlation between the instrument and Social Security retirement income.

Notes: as in Table 2, except that covariates relating to marital status are excluded where there would be no variation in these covariates.

Notes: as in Table 2.

---

**Table 4**
Estimates of the effects of Social Security income (per thousand 2009 dollars) on older workers’ labor supply, by marital status.

<table>
<thead>
<tr>
<th></th>
<th>Micro sample</th>
<th>Cohort level sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single</td>
<td>Married</td>
</tr>
<tr>
<td><strong>Panel A: First-stage results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrument</td>
<td>0.642 (0.037)</td>
<td>0.257 (0.027)</td>
</tr>
<tr>
<td>$F$</td>
<td>294.2</td>
<td>174.7</td>
</tr>
<tr>
<td>Partial $R^2$</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Panel B: Instrumental variables (G2SLS/2SLS) estimates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working for pay ($\times$ 100)</td>
<td>$-1.63 (0.30)$</td>
<td>$-0.62 (0.46)$</td>
</tr>
<tr>
<td>Hours worked per week</td>
<td>$-1.01 (0.10)$</td>
<td>$-0.75 (0.16)$</td>
</tr>
<tr>
<td>Weeks worked per year</td>
<td>$-0.90 (0.14)$</td>
<td>$-0.41 (0.21)$</td>
</tr>
<tr>
<td>Wife working for pay ($\times$ 100)</td>
<td>$-2.35 (0.45)$</td>
<td>$-0.88 (0.15)$</td>
</tr>
<tr>
<td>Wife’s hours worked per week</td>
<td>$-1.30 (0.22)$</td>
<td>$-0.88 (0.15)$</td>
</tr>
<tr>
<td><strong>Panel C: GLS/OLS estimates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working for pay ($\times$ 100)</td>
<td>$-0.05 (0.03)$</td>
<td>$-0.16 (0.05)$</td>
</tr>
<tr>
<td>Hours worked per week</td>
<td>$-0.07 (0.01)$</td>
<td>$-0.13 (0.02)$</td>
</tr>
<tr>
<td>Weeks worked per year</td>
<td>$-0.04 (0.01)$</td>
<td>$-0.10 (0.02)$</td>
</tr>
<tr>
<td>Wife working for pay ($\times$ 100)</td>
<td>$-0.07 (0.04)$</td>
<td>$-0.07 (0.04)$</td>
</tr>
<tr>
<td>Wife’s hours worked per week</td>
<td>$-0.04 (0.01)$</td>
<td>$-0.04 (0.01)$</td>
</tr>
<tr>
<td>Wife’s weeks worked per year</td>
<td>$-0.03 (0.02)$</td>
<td>$-0.03 (0.02)$</td>
</tr>
<tr>
<td>Observations</td>
<td>25,986</td>
<td>16,103</td>
</tr>
</tbody>
</table>

Notes: as in Table 2.
The first-stage $F$ statistics in Panel A of Table 3 indicate that, in both the cohort level and micro-level versions of the HRS sample, the instrument is sufficiently correlated with Social Security retirement income for instrumental-variables estimates based on this instrument to be reliable. The general rule in this case is that the first-stage $F$ statistic should exceed ten (Staiger and Stock, 1997). The partial $R^2$ statistics also indicate that, at the cohort level, the instrument explains a significant portion of the total variation in Social Security benefits ($R^2 = 0.218$). At the individual level, the variation explained by the instrument is much less ($R^2 = 0.004$), largely because there is much more variability in Social Security income at the individual level than there is at the cohort level. Nevertheless, even though the partial correlation at the individual level is relatively low, the value of the corresponding first-stage $F$ statistic is a healthy 401.2, which indicates that there are sufficiently many individuals in the data that potential weak instrument problems are a non-issue.

When the first-stage $F$ statistics are calculated within each subsample, the ‘A’ panels of Tables 4 and 5 show that, with one exception, the instrument is sufficiently correlated with Social Security income to reliably identify the G2SLS and 2SLS estimates. The exception is the cohort level single sample, where the identification is borderline—the first-stage $F$ statistic for this set of estimates is only 6.0. Because of this, the results for the cohort level single subsample should be interpreted with a degree of caution.

4. Empirical results

The ‘B’ panels of Tables 3–5 contain instrumental variables estimates of $\beta$, for the full HRS sample and the subsamples by marital status and education, in the second-stage equation

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{SocSec}_{it} + \beta_2 \cdot X_{it} + e_{it}$$

where SocSec$_{it}$ and $X_{it}$ are defined as in Eq. (1) (the first-stage equation) and $Y_{it}$ is a measure of labor supply. The estimator of $\beta_1$ is Ballestra and Varadharajan-Krishnakumar’s (1987) G2SLS estimator when the microdata are analyzed and weighted 2SLS when the cohort level data are analyzed. The weights in the latter case are cell sizes. The standard errors presented with the weighted 2SLS results are heteroskedasticity-robust and clustered by year of birth.

The findings in Panel B of Table 3 indicate that exogenous variation in Social Security benefits has significant, measurable effects on the labor supply of older workers. The coefficient of $-0.85$ in the first row means that every $1000 in annual Social Security retirement income (in 2009 dollars) reduces the probability that a Social Security beneficiary will participate in the labor force by 0.85 percentage points. Considering that the average probability (from Table 2) that Social Security beneficiaries in the sample participate in the labor force is 12.3%, this effect is large and economically significant. Moreover, since the estimate from the cohort level sample is similar in magnitude ($-1.48$), this result is robust to the choice of estimation procedure. This result contrasts with Krueger and Pischke (1992), who generally find that the notch cohorts did not increase their labor supply in response to legislatively imposed benefit cuts.

The results in the remaining rows of Panel B lend themselves to similar conclusions. In particular, each $1000 of annual Social Security income reduces beneficiaries’ labor supply by an average of 0.83 h per week and 0.51 weeks per year. Considering that respondents in the sample work an average of 3.1 h per week and 5.3 weeks per year in total, these effects are also large in economic terms. Further, for all of the results in Panel B, there is no statistically significant difference between the estimates obtained from the cohort level and micro-level samples. This is encouraging because it suggests that, the more stringent assumptions of the G2SLS estimator notwithstanding, the micro-level results are reasonable.

To put these estimates into perspective, a $1000 cut in annual Social Security income is an 8.6% benefit cut for the average beneficiary in the sample. Based on the relevant actuarial adjustment factor, this is the same magnitude of cut that would be achieved by increasing the normal retirement age by 15 months (on balance, since the system is designed to be actuarially fair, the present value of the benefit should not depend on whether beneficiaries delay claiming retirement benefits in response to such a change). Therefore, if Congress again increases the full retirement age (as it did in 1983, slowly phasing in increases in the full retirement age from 65 to 67 based on year of birth), the effect on elderly labor supply may be quite similar to the generic $1000 cut described above.

One difference between these results and Krueger and Pischke’s (1992) is that, while Table 3 uses a quadratic specification to control for beneficiaries’ age, Krueger and Pischke (1992) use a set of age dummy variables for this purpose. However, robustness checks that make use of age dummy variables do not yield qualitatively different results (see the Appendix A for details of these results). Using age dummy variables does, however, reduce the precision of the estimates. Since the quadratic specification is actually a very good fit to the data (see Fig. 3), it is useful to make judicious use of this specification in order to increase the power of the estimator.

For a variety of reasons, the effects of Social Security income on older workers’ labor supply might depend on marital status. For instance, married couples tend to have more assets and alternative sources of income than single people do. Moreover, even with nominal assets held constant, a married person may be “wealthier” than a single person in real terms simply because he or she can realize household economies of scale. Because of this, married people might respond less to a given change in Social Security income than single individuals would. On the other hand, since married people can realize gains from specialization, they may have more leeway to replace lost retirement income with income from the labor market. A priori, it is not clear which of these factors outweighs the other.

To examine these questions, Panel B of Table 4 contains estimates of the effects of Social Security income on labor supply for single individuals and married couples. Since the joint response of husbands and wives to a given change in Social Security income is greater than the response of a single person, there is little support for the notion that married couples are less responsive to changes in income than single individuals. For instance, each $1000 of Social Security income reduces a single person’s labor supply by an average of 1.01 h per week. However, the same change in Social Security income reduces a couple’s joint labor supply by an average of 1.63 h per week (0.75 + 0.88).\footnote{This difference (between 1.01 and 1.63) is statistically significant at the 10% level ($p = 0.098$).}

On the other hand, within married couples, there is a marked difference between the responses of husbands and wives to variations in Social Security income. In particular, wives’ labor supply is considerably more sensitive to variation in Social Security income than husbands’ labor supply.\footnote{In the micro sample, this difference is of borderline statistical significance where the probability of working is concerned ($p = 0.122$). For hours of work, the difference is significant at the 10% level ($p = 0.058$), and for annual weeks worked the difference is significant at the 5% level ($p = 0.050$).} In many respects, this is consonant with the common empirical finding that married women’s labor supply is more elastic with respect to nonlabor income than married men’s labor supply (Gruber and Orszag (2003) observe an analogous phenomenon when analyzing male and female workers’ responses to the Social Security earnings test). However, what is intriguing about

\[ \text{Y}_{it} = \beta_0 + \beta_1 \cdot \text{SocSec}_{it} + \beta_2 \cdot X_{it} + e_{it} \]
this result is that many of the reasons often given for the greater elasticity of married women's labor supply (for instance, that women bear most of the time costs of raising children, or that men tend to be in careers that demand continuous, full-time employment) are no longer applicable to a couple in retirement. One explanation is that, pre-retirement, women tend to be in jobs with more flexible working hours, which makes it easier to continue their attachment to the labor force (Hurd and McGarry, 1993).

Another interesting way to stratify the data is to estimate the effects of Social Security income on older workers' labor supply by education category. A priori, it is not clear whether more-educated individuals would respond more or less strongly to changes in Social Security income than less-educated individuals. On one hand, the income effect would suggest that the more-educated elderly, who have more pension and asset-based income, would be less responsive to a fixed change in Social Security benefits. However, the more-educated elderly also have greater earning power then the less-educated elderly, and their job skills are less likely to have been affected by diminished physical capacity as a result of the aging process. Therefore, they may be able to replace lost nonlabor income more easily than workers with less education.

Along these lines, estimates of these effects for the education subsamples are given in Panel B of Table 5. The results indicate that the labor supply of less-educated elderly workers is quite responsive to changes in Social Security income. Moreover, this is especially true on the extensive margin, i.e., the decision whether or not to work. For instance, looking at the G2SLS results, $1000 of Social Security income decreases the probability that a less-educated elderly worker will participate in the labor force by 1.12 percentage points, which is quite significant, considering that the overall labor force participation rate for this group (from Table 2) is 10.1%. This effect is also quite statistically significant ($p = 0.003$).

By contrast, the effect of Social Security income on the labor force participation of elderly workers with at least one year of college education is not significantly different from zero ($p = 0.327$). This finding is consistent with Haider and Loughran's (2001) observation that the more-educated elderly tend to work for nonpecuniary reasons, often trading high wages for flexibility in hours so they can maintain some attachment to the labor force. In this context, it is not surprising that levels of Social Security income have little effect on these workers’ decisions to continue participating in the labor force.

On the intensive margin, the effects of changes in Social Security income on hours or weeks worked are more similar for more- and less-educated elderly workers. The effects for more-educated workers are still less than those for less-educated workers, but the difference is not as pronounced as that for the decision to work in the first place. This shows that, while elderly with higher levels of education may well decide whether to work for nonpecuniary reasons, nonlabor income is still a factor when they decide how many hours to work.

Finally, when examining any set of instrumental variables estimates, it is useful to compare them to ordinary least squares estimates (or GLS estimates, when looking at the G2SLS results). The non-instrumental variables estimates of $\beta_i$ in Eq. (2) are given in the ‘C’ panels of Tables 3–5. In general, they are still negative, but smaller in magnitude than the corresponding instrumental variables estimates. This is the result one would expect if, due to a positive correlation between Social Security income (which is related to past employment) and the unobserved propensity to work, the non-instrumental variables estimates are positively biased.

5. Conclusion

Krueger and Pischke (1992), who were the first to examine the labor supply of the notch cohorts, generally concluded that these cohorts were not particularly responsive to changes in Social Security benefit levels. In large part, this was because labor force participation continued to decline even for those born after January 1, 1917, whose benefits were significantly reduced by the 1977 amendments. But this conclusion is worth revisiting, not least because this critical background fact has reversed itself: labor force participation among the elderly bottomed out in the early 1990s and has been increasing ever since (Gendell, 2008).

The empirical results in this paper show that the labor supply of the notch cohorts is now very sensitive to the level of their Social Security benefits, even at advanced ages. A $1000 reduction in nonlabor income (in 2009 dollars), which is roughly equivalent to the cut that would be imposed by a 15-month increase in the full retirement age, increases elderly labor supply by an average of 0.85 hours per week (a 27% increase relative to average hours worked per week). The response is even greater for singles, spouses of beneficiaries, and the less-educated elderly. As a future research direction, it will become increasingly important to understand the work decisions of individuals past the traditional retirement age, who face unique constraints on their income, the nature of their spending, and the working opportunities available.

From a policy standpoint, the results indicate that if concerns about fiscal solvency lead Congress to reduce Social Security benefits, older workers – even those in their 70s and early 80s (the ages of Social Security beneficiaries in the HRS data most affected by the Social Security notch) – are likely to compensate by increasing their participation in the labor force and working more hours. To some extent, this mitigates the welfare losses elderly workers would suffer if such cuts were imposed. Moreover, the least painful cuts would be those targeted at Social Security beneficiaries with the most capacity to work. Examples of recently discussed proposals along these lines are those that make the benefits formula more progressive (i.e., reduce payments to those with relatively high lifetime earnings). By contrast, changes to the indexing formula (for instance, one suggestion of the President's Commission to Strengthen Social Security was to index lifetime average earnings by prices instead of wages) or raising the full retirement age would reduce benefits more uniformly, and have more negative effects on those who can take up employment less easily.

Appendix A. Robustness checks

The purpose of this appendix is to show that the main conclusions of the paper are robust to potentially important variations in the empirical strategy, particularly where they relate to the construction of the sample and the functional form of the regression model. For reasons of efficiency, the main results make use of a sample as large as possible, and assume that the relationship between age and labor supply can be reasonably approximated with a quadratic specification. However, when these assumptions are relaxed, the qualitative conclusions are very similar. This appendix discusses each of these assumptions in turn.

Appendix A.1. Limiting the sample to the AHEAD and CODA cohorts

The empirical results in this paper are based on HRS households from five cohorts: the AHEAD cohort (born before 1924), the CODA cohort (born between 1924 and 1930), the initial HRS cohort (born between 1931 and 1941), the War Baby cohort (born between 1942 and 1947), and the Early Baby Boomer cohort (born between 1948 and 1953). In all cases the Social Security beneficiary is born between 1901 and 1930, but beneficiaries from the latter three cohorts tend to be men with relatively young wives. Although wives' ages and other characteristics are controlled for in the regression analysis, these controls may not be sufficient, and so it is useful to check whether the empirical results are sensitive to the cohorts used to construct the sample.
Although it is not uncommon to use this kind of specification to approximate the relationship between age and elderly labor supply, the paper is that it makes use of a quadratic specification.

Appendix A.3. More flexible age controls

A third characteristic of the estimation strategy in the main body of the paper is that it makes use of a quadratic specification to approximate the relationship between age and elderly labor supply. Although it is not uncommon to use this kind of specification (Engelhardt et al., 2005; Moran and Simon, 2006), and Fig. 3 suggests that this assumption is quite reasonable, it is still useful to check whether the results change when a more flexible specification is adopted. Panel C of Table 6 presents the results obtained from two such specifications: a quartic model (i.e., one that additionally includes an age cubed term and age raised to the fourth power), and a model that uses single-year age dummy variables for the age of the Social Security beneficiary and, where relevant, the age of the beneficiary’s spouse. As in the previous subsection, the sample is limited to respondents from the AHEAD cohort who were born between 1909 and 1924.

The results in Panel C of Table 6 are the most conservative because they incorporate all of the robustness checks that have been discussed. Even so, the qualitative nature of the empirical results remains unchanged; elderly labor supply responds strongly to variations in Social Security income induced by legislative changes to the Social Security benefits formula. Quantitatively, the estimates are also quite similar to comparable estimates in Table 3, particularly when the standard errors of these estimates are taken into account. For example, according to the cohort level results in Table 3, an additional $1000 in Social Security benefits (annually, in 2009 dollars) reduces elderly labor supply by 1.00 h; in Panel C of Table 6, the corresponding estimate is 0.85 h. Therefore, while the estimates in the main body of the paper are more precisely estimated, there is no evidence that the assumptions necessary to obtain this efficiency led to any bias in the empirical results.

References


