

NBER WORKING PAPER SERIES

EFFECTS OF SOCIAL SECURITY POLICIES ON BENEFIT CLAIMING, RETIREMENT
AND SAVING

Alan L. Gustman
Thomas L. Steinmeier

Working Paper 19071
<http://www.nber.org/papers/w19071>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2013

This research was supported by grant Number UM12-04 from the U.S. Social Security Administration (SSA) through the Michigan Retirement Research Center (MRRC) to the NBER. The findings and conclusions are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, MRRC, the NBER Retirement Research Center, or Boston College RRC. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Alan L. Gustman and Thomas L. Steinmeier. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Effects of Social Security Policies on Benefit Claiming, Retirement and Saving
Alan L. Gustman and Thomas L. Steinmeier
NBER Working Paper No. 19071
May 2013
JEL No. C61,D31,D91,E21,H55,J14,J26,J32

ABSTRACT

An enhanced version of a structural model jointly explains benefit claiming, wealth and retirement, including reversals from states of lesser to greater work. The model includes stochastic returns on assets. Estimated with Health and Retirement Study data, it does a better job of predicting claiming than previous versions. Alternative beliefs about the future of Social Security affect predicted outcomes. Effects of three potential policies are also examined: increasing the early entitlement age, increasing the full retirement age, and eliminating the payroll tax for seniors. Predicted responses to increasing the full entitlement age are sensitive to beliefs.

Alan L. Gustman
Department of Economics
Dartmouth College
Hanover, NH 03755-3514
and NBER
alan.l.gustman@dartmouth.edu

Thomas L. Steinmeier
Department of Economics
Texas Tech University
Lubbock, TX 79409
thomas.steinmeier@ttu.edu

Introduction

Social Security continues to be under substantial financial pressure (Social Security Administration Trustees Report, 2012). A number of policy changes have been suggested to encourage a population that is becoming increasingly long lived to work longer, thereby improving the financial status of the Social Security.¹ Increasing the early entitlement age for benefits, increasing the full retirement age, and abolishing the payroll tax on earnings for those over the full retirement age are among the policy changes that are frequently mentioned. These policy changes will affect the timing of retirement (withdrawal from the labor market), the resources available to seniors when they leave the labor market, and the timing of benefit claiming.

To analyze the effects of these potential changes in policy, we specify and estimate a structural model that is designed to simultaneously explain retirement outcomes, saving (wealth accumulation) and benefit claiming. This model updates our previous work on a number of dimensions, most importantly allowing us to explain the full range of retirement dynamics, heterogeneity in wealth accumulation, and benefit claiming, an outcome that researchers have had particular problems in explaining.

We then introduce different beliefs about the future of Social Security and trace the effects of these beliefs on claiming, retirement and saving. Three sets of beliefs are incorporated in alternative simulations and compared to outcomes under a baseline scenario. One set compares outcomes if individuals believe Social Security benefits will be reduced in the future with outcomes under a base case, where, as in most current models, covered workers believe they will receive benefits as called for by current law. In a second set of simulations, potential retirees believe that the return on assets will be higher in the future than in the present. A final set of simulations compares outcomes when survivor benefits are not fully valued with results when they are. Although we do not incorporate each of the alternative beliefs in the model estimation, we do incorporate the alternative beliefs in simulations that allow us to place bounds on their likely influences on key outcomes.

¹ These and other policy changes are discussed by the Senate Committee on Aging (2010) and the Congressional Budget Office (2010).

Our estimates are based on data from the Health and Retirement Study. We include detailed employer pension plan descriptions, allowing identification of incentives embedded in the formulas, and utilize matched earnings histories obtained from the Social Security Administration.

Section II briefly reviews the determination of Social Security benefits, and Section III discusses some outstanding issues in the literature. Section IV considers the actuarial valuation of the Social Security annuity and how it affects claiming. A structural model of retirement, saving and claiming is developed in Section V. The model is estimated and its properties are examined through simulation in Section VI. Section VII explores modifications of the model that would explain the apparently excessive claiming observed at younger ages, focusing on differences in beliefs about the future of own benefits and understanding of the value of spouse benefits. The effects of the three key policies, increasing the early entitlement and full retirement ages, and abolishing the payroll tax for those over the full retirement age, are examined in Section VIII, while Section IX concludes.

II. The Determination of Social Security Benefits

Before turning to the estimation of an enhanced structural model and an examination of its properties, it is useful to review briefly some of the key determinants of Social Security. Each of these determinants of benefits is built into our model.

Eligibility. Ten years of covered earnings are required to attain eligibility for own retirement benefits.

Average Indexed Monthly Earnings (AIME). Benefits are based on the highest 35 years of covered earnings, where actual covered earnings in each year are indexed to age 60 using an index of annual changes in average earnings in the economy. For those working more than 35 years, and after age 60, benefits are recomputed if current earnings exceed indexed earnings from an earlier year.

Primary Insurance Amount (PIA). This is the monthly benefit assuming retirement as of the individual's full retirement age. For example, for a person turning age 62 in 2012, benefits at full retirement age (age 66) are calculated as 90 percent of annual indexed

earnings up to \$9,200, plus 32 percent of indexed earnings between \$9,200 and \$55,500, and 15 percent of earnings over \$55,500 up to maximum covered earnings.

Adjustment of Benefits for Age of Claiming. Full retirement age is the base year for calculating adjustments to benefits. It is based on year of birth (e.g., for those born between 1943 and 1954, full retirement age is 66.) Benefits are reduced from their value at full retirement age for those claiming them between the early entitlement age (62) and the full retirement age. Benefits are increased for those who delay claiming until after full retirement age, through age 70.

Spouse and Survivor Benefits. Spouses who do not qualify for own benefits may receive benefits based on the earnings record of their living, divorced, or deceased spouse. Even if a person is entitled to benefits from own work, these benefits may be topped up if own benefits at full retirement age fall below half the benefits of a spouse. A person whose spouse died is entitled to a top up to the level of the deceased spouse's full benefit. Spouse and survivor benefits are, however, adjusted based on the year they were claimed by each spouse.

Other Determinants of Benefits. Different formulas are used for those who worked in uncovered government or other employment and were not subject to the payroll tax, or whose spouse worked in a job that was not covered by the Social Security system.

Earnings Test. An earnings test is applied to retirement benefits for those who collect benefits before reaching their full retirement age. In the years before a covered worker reaches full retirement age, the earnings test withholds benefits at the rate of fifty cents for each dollar of earnings over the exempt amount. In 2012, for those who had not yet reached full retirement age in that year, benefits were withheld at the rate of 50 percent of earnings in excess of \$14,640. Benefit payments made in future years are adjusted for past reductions due to the earnings test.

In addition, there are other more detailed factors affecting benefits that are included in our benefit estimates but not listed separately here.

III. Gaps in the Literature Relating Social Security to Retirement, Benefit Claiming and Saving

Until the last couple of decades, the most prominent feature of the retirement hazard was the spike in retirements at the Social Security full retirement age. This spike was visible in both national and international data (Gruber and Wise, 1999). It was often attributed to unfair actuarial adjustments for work after full retirement age. When a person delayed retiring, Social Security and pension benefits were not paid. Those benefits to be paid in the future were not increased at an actuarially fair rate to compensate for benefit payments lost to any earnings test while continuing at work. In the extreme, a person might find the net wage for work after qualifying for full retirement benefits to have fallen to the value of the nominal wage paid before retirement less the value of the foregone pension or Social Security benefit. Moreover, mandatory retirement was legal and common. Both factors accounted for much of the spike in retirement at full retirement age.

This all changed as a result of Social Security legislation, other legislation abolishing mandatory retirement and related court decisions mandating fairer actuarial adjustments in pensions for older persons. These changes, along with other trends, substantially reduced the spike in retirements around the full retirement age (Anderson, Gustman and Steinmeier, 1999; Gustman and Steinmeier, 2009).

Now the spike in retirements at age 62 is the more prominent feature of the retirement hazard. That spike is almost certainly related to the Social Security early entitlement age. As would be expected, the spike in exits from the labor force at age 62 is accompanied by a spike in benefit claiming at that age.

There is a puzzle, however. For most individuals when benefit receipt is postponed beyond age 62, future Social Security benefits are increased at better than an actuarially fair rate.² A task of retirement and claiming models is to explain why many who have already

² It has been known for some time that at a 3 percent real interest rate, postponing benefit receipt, especially in the first few years after age 62, increases future benefits for couples at better than an actuarially fair rate (Gordon and Blinder, 1980; Feldstein and Samwick, 1992). More recently, Shoven and Slavov (2012a and b) show that at the very low market interest rates following the Great Recession, delayed claiming is an even better deal.

retired do not defer claiming beyond age 62, and why many others do not defer retirement and claiming for a number of years after they reach age 62.

Heterogeneity in time preference rates, where a substantial subgroup of the population has high discount rates, along with an inability to borrow from future income, may account both for the spike in retirement at age 62, and for what in some models without heterogeneity in time preference appears to be an excess of benefit claiming at age 62. Those with a high time preference rate and little or no pension income would also be expected to have low levels of other assets (perhaps excluding their home). Although they might like to retire before age 62, they do not have sufficient assets to allow them to do so. At 62, when Social Security benefits become available, they retire and claim their benefits en masse. Others, with intermediate time preferences, may have accumulated some assets (Coile, et al, 2002), but for them the prospect of increased future benefits is insufficient to compensate for delayed current benefits. These individuals will also want to collect benefits as soon as possible. Further, if they are forced to delay benefits because of an earnings test, the earnings test will serve to reduce their perceived compensation, making them more likely to retire at age 62 when the earnings test becomes effective. Thus, spikes in retirement and claiming at 62 are not necessarily limited to those who are asset constrained.

Heterogeneity in time preference rates also helps to explain the wide variation in wealth, even for those with similar levels of lifetime earnings. Venti and Wise (1999, 2001) find an extraordinarily wide distribution of wealth holdings. Of greatest interest, the distribution of wealth holding is very wide even among individuals who fall within the same decile of lifetime income. (See also Gustman and Steinmeier, 1999.) Evidence from the HRS also shows that there are a distressing number of age 50+ households with essentially no retirement savings outside of Social Security. Among those with no savings outside of Social Security, many have had substantial earnings in the past.

Scholz, Seshadri and Khitatrakun (2006) claim that the wealth distribution at retirement can be pretty well explained without time preference heterogeneity. In their model, almost no one has wealth that is significantly lower than their optimal amount (see their Figure 2). Yet Venti and Wise find a large number of individuals, even at fairly high income levels, who have very little wealth, generating very wide wealth

distributions for those falling within the same lifetime income decile. Moreover, in their Figure 3, Scholz et al. measure the discrepancy between actual and optimal net worth. That figure indicates that for a given income decile, there is a substantial variation in the difference between actual net worth and their calculated optimal net worth. This suggests that their model with uniform time preference does not do as good a job in generating the variation of observed assets within income deciles as might appear.³

Other possible explanations of low net worth for high income individuals are also suspect. One common explanation is stochastic wages. If current wages are higher than expected, then the savings which were done on the expectation of lower wages may appear small relative to current wages. However, this would not explain assets so low as to be almost negligible, and in any case unless the unexpected wage changes were very recent, savings and asset levels would adjust to the higher wage levels over time. A similar argument is that unexpected negative asset returns can cause low assets relative to savings (think Enron). But unless the negative asset returns were very recent, one would expect savings levels to return to normal levels over time as individuals saved more to make up for the asset loss. Both of these explanations run into the problem that unless the stochastic events are fairly recent and severe, it is difficult to reconcile very many individuals with almost no assets in the years just before retirement with relatively low time preference rates.

This issue of time preference is important because policy prescriptions are very different for a population with a uniform (and fairly low) time preference rate than for a population with a mix of some individuals with a low time preference rate and others with a high time preference rate. For instance, a population with a uniform low time preference rate may not require incentives to save for retirement, while a population with some individuals with high time preference rates may require some mechanism to ensure

³ On page 626, Scholz et al. claim that their model explains 86% of the observed variation in net worth. This result may be related to the concentration of observations in the lower left corner of the scatter plot in their Figure 2. These observations correspond to lower-income individuals with low values of both observed and calculated optimal net worth. For low income individuals, these two values must necessarily be fairly close.

that they set aside something for retirement, at least if we are trying to avoid having a fraction of the retired population with very low incomes.

It has been challenging to construct models that capture the spike in claiming and retirement at age 62 while also accounting for the other related outcomes observed in the data, and in particular the wide variation in asset values observed for individuals with similar lifetime income. As explained, models with a uniformly low time preference rate cannot rely on actuarial unfairness as the cause of the spike in claiming at the Social Security early entitlement age, or as the cause of the spike in retirements at age 62. Social Security does not impose actuarial penalties at age 62. Indeed, as we have mentioned, the actuarial values of Social Security increase for many households as benefits are postponed after the early entitlement age. Nor can structural life-cycle models with uniformly low estimates of time preference account for the spike in retirements at the early entitlement age. Equally troubling, they have trouble explaining why there are individuals with relatively high levels of lifetime earnings with very little net worth.

Before turning to a discussion of our approach to reconciling these disparate outcomes in the context of a consistently specified life cycle model, it is useful to briefly consider the structural models of the retirement and saving process constructed to date.⁴ In particular, it is useful to consider some of the weaknesses of existing structural models for an analysis of retirement, saving and benefit claiming.

Typically, currently available structural models focus closely on a limited set of outcomes. Different models consider retirement, saving or benefit claiming, but none addresses the complexities facing an analysis of all three sets of outcomes at once. Indeed, many of these models focus on only one or another aspect of behavior. Often there is a tendency to simplify those aspects of the model that are not part of the immediate focus of the analysis. For example, while Scholz, Seshadri and Khitatarkin (2006) consider the determinants of wealth accumulation, retirement is taken as exogenous and benefit claiming

⁴ Structural analyses of retirement and saving include Rust and Phelan (1997), Berkovec and Stern (1991), Lumsdaine, Stock and Wise (1990, 1992), Gustman and Steinmeier (1986), French (2005), Bound, Stinebrickner and Waidmann (2010), Van der Klaauw and Wolpin (2006), Scholz, Seshadri and Khitatrakun (2006), and a number of others.

is ignored. In Shoven and Slavov (2012b), claiming is a dependent variable, while retirement and wealth, jointly determined outcomes, are treated as explanatory variables. In a more recent paper, Scholz and Seshadri (2012) introduce retirement into their model of saving, but they define retirement as irreversible and do not allow either for partial retirement or for reversals from states of greater to lesser retirement. These simplifications may or may not affect the reliability of these models to describe the behavior of interest, but they do mean that the models are less useful than they might be for understanding the joint determination of retirement, saving, and claiming, and how Social Security policies might affect those outcomes.

Another common way to simplify the investigation is to limit the sample, restricting the population analyzed to those with the least complex budget constraints who have simpler decisions to make. For example, to avoid having to incorporate the details of individual defined benefit pension plans, van der Klaauw and Wolpin (2006) limit the population they study to those without a DB plan. They eliminate from their sample those who have a DB plan on their current job. They also eliminate those who had a DC plan at any time. Rust and his colleagues (in Benitez-Silva et al., 1999 and Benitez-Silva et al., 2004) limit the sample to exclude those who are covered by a pension. French (2005) uses summary measures of pension incentives rather than the individual accrual profiles reported by the HRS for each observation. Bound, Stinebrickner and Waidmann (2010) limit their sample to single individuals. They also assume, counterfactually, that all defined contribution assets and nonpension wealth are paid out as an annuity. After the restrictions on the scope of the model and exclusions for missing data, pension coverage and work history requirements, Bound et al. were able to fit the model to only 196 observations.

Although these analyses are adequate for the subpopulations analyzed, they are handicapped when it comes to predicting outcomes for members of the population facing more complex situations. They are also unsuited for analyzing how policy changes that have not as yet been implemented would affect the overall population in a particular age group, including the age of claiming of Social Security benefits, retirement dynamics (including the age of labor force withdrawal from full-time or part-time work and the spike in retirement transitions at the Social Security early entitlement age), and important features of the

distribution of wealth among the preretirement population. One cannot be sure that a model that is applied only to those without complex budget constraints, or fails to or inadequately explains one or more key outcomes, can nevertheless provide a reliable basis for predicting the effects of policy changes.

The concerns we have expressed regarding the limitations of existing models of retirement, claiming and wealth accumulation also create concerns about previous analyses of policies affecting each of these outcomes. Each of the policy alternatives we examine in this paper has been considered previously. To provide a few examples, Song and Manchester (2007), Mastrobuoni (2009), and Coe, Kahn and Rutledge (2013) examine the effects of changing the full retirement age using cohort differences. Laitner and Silverman (2012) examine the effects of increasing the payroll tax in a basic life cycle model. Coile, Diamond et al. (2002) examine claiming outcomes. Gustman and Steinmeier (2005) examine the effects of increasing the early entitlement age in a model with heterogeneous time preference rates.

However, all of these analyses have been simplified along one or more dimensions. Gustman and Steinmeier (2005) does not model claiming behavior. The life cycle model in Laitner and Silverman is simplified to the point that it cannot explain the spike in retirements at age 62. Coile, Diamond et al. do not specify a model that can explain associated retirement flows and wealth accumulation. Other studies mentioned use reduced form models that cannot explain various features of the retirement hazard, including the spike at early entitlement age. When natural experiments are used, they isolate the effects of a previous policy change, but do not estimate the deep parameter values required to predict the potential effects of policies that have not yet been adopted.

Consequently, our aim is to continue to broaden the scope of analysis in structural models of retirement and saving. In the current paper, we modify a retirement model we have estimated previously to focus simultaneously on three outcomes, the timing of claiming of Social Security benefits, retirement dynamics and saving, while allowing for stochastic returns. We use the model to analyze several potential changes in the Social Security system. We also examine the likely effects of different beliefs about the future of Social Security, and the valuation of spouse and survivor benefits.

IV. Annuities

The central feature in the claiming decision is the tradeoff between a present lump sum and a future annuity. In this light, it will be helpful to spend a few paragraphs looking at how individuals value annuities such as Social Security.⁵

The value of an annuity to an individual can be broken down into two parts. First, there is the question as to the actuarial fairness of the annuity. That is, how does the present financial value of the stream of annuity payments, discounted at an appropriate interest rate, compare to the cost of the annuity? Second, there is the question as to how the individual values the stream of annuity payments relative to its present financial value. In other words, how much would the individual be willing to pay currently to receive the annuity, and how does this compare to the cost of an actuarially fair annuity? For purposes of analyzing claiming, we are interested in the characteristics of a marginal annuity. Social Security and possible defined benefit pensions already provide a base level annuity, and the result of a delay in claiming adds a marginal amount to this annuity.

The actuarial value of the Social Security annuity has been extensively documented, so the discussion here will be brief. The general conclusion is that while the benefits of delaying claiming are roughly actuarially neutral for single individuals, at least until they reach their mid-60's, there is a strong actuarial advantage for the higher earner of a two earner couple to delay claiming benefits. When the individuals are 62, they are eligible for 80 percent of their full benefits, assuming a full retirement age of 65 (which is appropriate for much of the HRS sample). If they wait until age 63, they are eligible for 86.7 percent of their full benefits, which is an 8.3 percent increase. This increase will apply over their own lifetimes, and for a substantial number the increase will also apply over the lifetime of their surviving spouses.

Table 1 gives the actuarial rates for delaying claiming for several circumstances.⁶ The first column pertains to the higher earner of a couple, where the spouse is two years younger,

⁵ For discussions of the demand for the Social Security annuity, see Brown, Casey and Mitchell (2007).

⁶ For a recent related analysis of the actuarial advantage to delayed claiming and references to the previous literature, see Shoven and Slavov (2012b).

with a real interest rate of 2 percent. The first entry indicates that at age 62, for every dollar of benefits lost because the individual delays claiming, the present value of the increased later benefits will be \$1.67. That is, by claiming at age 62, the individual gives up future benefits that would have an actuarial value that is 67 percent higher. This actuarial advantage declines as the individual delays further and increases slightly at age 65 when the delayed retirement credit becomes available at full retirement age. Even at age 69 the tradeoff of present benefits for future benefits is approximately actuarially fair. As shown in the second column, the advantage declines for a real interest rate of 4 percent, but is still more than actuarially fair for ages prior to the full retirement age. The last column looks at a single individual, and it would also apply to the perceived actuarial calculations for a married individual who gave no weight to the utility of a surviving spouse. These figures indicate that the future increases are roughly actuarially fair until the full retirement age and decline sharply during the late 60's.

The other part of the annuity question is how much individuals would be willing to pay for an actuarially fair annuity. This can be investigated with the aid of a very simple consumption model. Let total utility be given by

$$U = \sum_t s_t \left(\frac{1}{1+\rho} \right)^t u(c_t)$$

Where s_t is the survival rate until time t and ρ is the discount rate. Non-annuitized assets start at some level a_0 and evolve according to the standard formula

$$a_{t+1} = (1+r)a_t + b - c_t, \quad a_{t+1} \geq 0$$

where r is the real interest rate and b is the level of annuities from Social Security and possibly defined benefit pensions. As long as non-annuitized assets are positive, the first order conditions yield

$$u_c(c_t) = \frac{1}{s_t} \left(\frac{1+\rho}{1+r} \right)^t u_c(c_0)$$

Eventually at advanced ages, s_t starts to fall rapidly, causing $u_c(c_t)$ to rise and consumption to fall.

The fall in consumption while non-annuitized assets are positive indicates a two-part solution. In the first part, consumption starts at some level c_0 and evolves according to the first order condition as long as non-annuitized assets are positive. At some point the level of

consumption implied by the first order condition falls below the level of the annuity b . At that point, assets reach zero and consumption then follows b . Two extreme cases are that the individual starts out with almost no assets or that the individual starts out with so many assets that they never reach zero. In the first case, the value of an additional \$1 of annuity is

$\sum \frac{s_t}{(1+\rho)^t}$ the cost of an actuarially fair annuity is $\sum \frac{s_t}{(1+r)^t}$, and the ratio of value to

cost is $\frac{\sum \frac{s_t}{(1+\rho)^t}}{\sum \frac{s_t}{(1+r)^t}}$. In the second case, the value is $\sum \frac{1}{(1+r)^t}$, the cost is the

same, and the ratio of value to cost is $\frac{\sum \frac{1}{(1+r)^t}}{\sum \frac{s_t}{(1+r)^t}}$. In the intermediate case,

where assets are depleted at time B , the ratio of the value to cost is given by

$$\frac{\left(\sum_{t=0}^{B-1} \frac{1}{(1+r)^t} + \frac{1}{(1+r)^B} \sum_{t=B}^D \frac{s_t}{s_B} \frac{1}{(1+\rho)^{t-B}} \right)}{\sum_{t=0}^D \frac{s_t}{(1+r)^t}}$$

where D is the maximum age of survival. This reduces to the first expression when $B = 0$ and to the second expression when $B = D$.

Tabulations of the ratio of annuity value to cost for several scenarios are given in Table 2. For instance, the value of 1.38 in the first row indicates that an individual currently aged 62 whose assets will run out at age 80 along the optimal consumption path would be willing to pay 38 percent more than the cost of an actually fair annuity if the real interest rate were 2 percent and the individual's discount rate was 0. There are several things to note about this table. If the individual's discount rate is relatively low, he or she is likely to have accumulated substantial assets by retirement, and because of the low discount rate these assets are likely to last very late into the life cycle. Therefore, for individuals with a low discount rate, the columns on the right side of the table are most likely to be relevant. These columns suggest that annuities are likely to be of considerable value to these individuals. Not only do these individuals give relatively more weight to later years, but the actuarially fair

cost of providing the annuity in those years is relatively low. Recall also that this value is on top of any actuarial advantage of Social Security delay, so for low discount married individuals with earnings higher than their spouses, the value of the increased future benefits may well be double the value of the current benefits foregone by delaying claiming.

Individuals with high discount rates, on the other hand, are unlikely to have amassed much in the way of assets by the time of retirement, and their high preference for current consumption means that whatever assets they do have will not last for long. This means that for individuals with high discount rates, the columns on the left of the table are more likely to be relevant. As indicated in the third group of figures in the table, an individual with few assets and a relatively high discount rate (0.04) may find that the value of the marginal annuity is actually less than the actuarially fair cost of that annuity. In the limit, an individual with a very high discount rate and almost no assets may find that the value of the marginal annuity arising from delaying claiming would be almost zero, in which case anything that causes a delay in claiming (such as the earnings test) would act effectively like a tax.

This creates something of a dilemma for policymakers. Ideally, one would like individuals with high discount rates to choose to delay claiming benefits to obtain an increased annuity income, since they do not have very many assets to fall back on and are probably least prepared financially for retirement. However, those individuals are precisely the ones who place the least value on the annuity and hence are the least likely to choose to delay claiming benefits. Individuals with low discount rates, who are probably much better prepared financially for retirement, are the ones most likely to value the annuity provided by delaying claiming.

The last three rows of the table indicate that increasing the interest rate has different effects on different groups of people. Compare these rows with the second group of three rows, which have the same discount rate but a lower interest rate. For individuals with a lot of assets whose assets will last until a relatively old age, an increased interest rate will make the marginal annuity less attractive. The reason appears to be that the increased interest rate downplays the later years, when the annuity is most valuable relative to its cost. On the other hand, for individuals with few assets who will deplete those assets relatively quickly, the higher interest rate makes the annuity more valuable relative to its cost. This appears to be

because the higher interest rate reduces the cost of the annuity, whereas the value of the annuity for these individuals is governed more by the discount rate, which has not changed. The interest rate thus has an ambiguous effect on the value of an annuity relative to its cost, even among individuals with the same discount rate.

In summary, the relationship between the amount that individuals would be willing to pay for the annuity provided by delaying claiming and the amount given up by foregoing current benefits varies over a wide range. The higher earning individual in a couple with substantial assets and a low discount rate will find delaying extremely advantageous, while a single individual over the full retirement age with few assets and a high discount rate may find it extremely disadvantageous to delay claiming. The commonly expressed feeling that individuals with relatively few assets should appreciate the opportunity to insure themselves against living too long is erroneous, especially if the cause of the low level of assets is that the individuals have relatively high discount rates. The opposite view, that individuals with low asset levels are likely to regard themselves as over-annuitized and are willing to trade at least some of their future annuity for an actuarially equivalent present lump sum, is oftentimes more nearly correct.

V. The Structural Model.

To help fill the remaining gaps in the analyses of behavior and policy, we extend our earlier model of retirement and saving to include benefit claiming. In the process, we modify the structure of our previous model (Gustman and Steinmeier, 2005, 2008) to also include stochastic returns to assets and to model reverse flows from states of greater to lesser retirement.

In the past several years we have developed a series of models whose main purpose is to jointly explain two of the three key outcomes: the age 62 retirement spike and the dispersion of wealth among households with similar earnings histories (e.g., see Gustman and Steinmeier, 2005.) These models estimate the extent of unobserved heterogeneity in time preference rates, which manifest themselves in the dispersion of wealth households have relative to their previous earnings history. Households with high time preference rates also tend to value future Social Security benefits less than actuarial calculations would suggest.

For them, the earnings test starts to look more like a tax on earnings after the early entitlement age, and this may induce them to retire at that point.⁷

We previously used one of these models (Gustman and Steinmeier, 2004b) to examine the issue of Social Security claiming. At that time, the model assumed a fixed, relatively low interest rate, and retirement was an irreversible decision. In that model, the predicted claiming was less than the observed claiming, with the prediction error being greatest at ages prior to the full retirement age. Since that time we have introduced stochastic returns to assets and the possibility of reentering the labor force into the model, but in the context of an assumption that individuals always claim benefits as soon as possible. It seems worthwhile now to reintroduce endogenous claiming into the model and to determine whether these modifications improve the results regarding claiming outcomes.⁸

In sum, our goal is to integrate the effects of Social Security incentives on benefit claiming, especially at and around the early entitlement age, into a model in which retirement and saving are treated as jointly determined outcomes. We abandon a number of the simplifying assumptions made in our earlier work. We then also introduce the effects of different views on the value of future Social Security benefits, and apply the model to analyze the likely effects of some leading policy recommendations designed to prolong the time spent in the labor market.

Consider now the complete structural model that will be estimated and simulated in this paper. The core of the model is an expected utility function

⁷ A related literature explores the effects of the earnings test on retirement (e.g., Burtless and Moffitt, 1984; Disney and Smith, 2002; Friedberg, 2000; Gustman and Steinmeier, 1985 and 2004; and Song, 2002).

⁸ Stochastic returns to assets can cut two ways in this exercise. The mean rate of return is taken to be determined by a portfolio with half stocks and half bonds. This return is higher than the return on fixed assets we used in our previous work, which was taken as being equal to the rate used in the Social Security actuary's calculations. A higher return should raise the value of claiming early relative to waiting. On the other hand, the variance of the return makes the higher mean return somewhat less attractive, working in the other direction. This last effect would be somewhat mitigated by the possibility of returning to work, since the income loss from bad draws from the distribution of returns on assets could be offset to some degree by returning to work.

$$EU = u(C_0, L_0) + \sum_{t=1}^D \left[\left(\frac{1}{1+\rho} \right)^t \sum_{m=1}^3 s_{m,t} u(C_t, L_t) \right]$$

Where C is consumption, L is leisure, m is an index that indicates the survival state of the household, and $s_{m,t}$ is the probability that the household survives to time t in state m . The three states are that both spouses have survived, only the husband has survived, and only the wife has survived. Due to computational limitations arising from the addition of claiming variables into the model, the wife's labor supply is taken as exogenous, so that the leisure variable in the utility function is the husband's leisure.

The within period utility function is given by

$$u(C_t, L_t) = \frac{1}{\alpha} C_t^\alpha + e^{\beta_0 + \beta_a \text{Age}_t + \beta_h \text{Health}_t + \varepsilon_t} L_t^\gamma$$

The coefficient in front of the leisure term indicates the relative weight of leisure in the utility function. In this formulation, leisure gradually gets more valuable relative to work due to physical and mental exhaustion, and a bout of bad health adds to this relative preference for leisure. The ε term is an individual effect that varies from individual to individual and reflects that individual's relative preference for leisure.

Consumption and leisure are chosen to maximize expected utility subject to the budget constraint

$$A_{t+1} = (1+r_t)A_t + E(L_t) + S_t + P_t + O_t - C_t, \quad A_{t+1} \geq 0$$

where E is earnings, S is Social Security benefits, P is pension benefits, and O is other income. The return on assets r is taken to be stochastic and uncorrelated over time. Earnings depend on the amount of leisure, and the wage rate may vary depending on the amount of work the individual chooses. Social Security and pension benefits are the actual amounts paid in a year, and not a measure of accruals. As such, for instance, S is zero before the Social Security early entitlement age, and P is zero before the individual has retired from the job generating the pension.

The model is estimated for a sample of married households in the original cohort of the Health and Retirement Study (HRS). The original HRS included individuals aged 51 to 61 in 1992 and surveyed those individuals every two years starting in 1992. In addition to a large number of economic and labor force variables, the HRS has two supplements that are

useful for this study. The first is the earnings records from the Social Security Administration, and the second is the pension plan descriptions from the providers of the pensions held by the respondents. These two allow for much more precise measurement of the economic incentives faced by the households. The two main requirements for a household to be included in our sample are that the husband must have been full-time in the labor force most years before age 50 with at least some indication of his wage either from the Social Security earnings records or self-reports, and if he had a defined benefit pension in his current job at the time the survey starts, or the last job held before the start of the survey, the pension must be included in the pension provider survey. The latter exclusion arises because if an individual had a pension, but the details of the pension are not available from the provider survey, it is very unlikely that the incentives that the individual faced to retire at particular ages will be correctly reflected in the budget constraint.

In the model, leisure is restricted to three values: 0 for full time work, $\frac{1}{2}$ for partial retirement work, and 1 for full retirement. In the earnings function E , the wages may differ according to the amount of work. For full-time work before retirement, wages for years not directly observed are imputed using coefficients for experience and tenure from a fixed effects wage regression. Unobserved partial retirement wages are inferred from a regression of observed partial retirement wages on full-time wages and other variables. Wages for full-time jobs which individuals have taken after a period of either full or partial retirement are inferred by resetting the tenure variable to zero.

The Social Security benefit amount, which is central to the issues investigated in this paper, is taken to be determined by three state variables in the model. The first of these is the level of the primary insurance amount (PIA) of the husband, which is determined by the age at which the husband leaves full-time work initially. The presumption is that partial retirement jobs, which tend to have lower wages than full-time work, will not significantly impact the PIA. The other two state variables are the adjustment amounts for both the husband and wife, reflecting both any early retirement penalties and delayed retirement credits. If the individual has the maximum adjustment amount relative to his or her age, this indicates that the individual has not claimed any previous benefits, and that individual may delay claiming further if that is the optimal strategy. Once benefits are claimed, the

adjustment amount may be recalculated if the individual is subject to the earnings test and loses benefits in a particular year. The PIA of the wife enters implicitly as an exogenous variable, since it is determined by the exogenous work history of the wife.

The pension benefit variable takes on a different form depending on whether the pension is defined benefit or defined contribution. In both cases, the pension benefit variable is zero before the individual retires from the pension job. In a defined benefit pension, the amount of the benefit is fixed at the time the individual leaves the job by applying the applicable formula from the pension provider documents to the earnings history and tenure that the individual had compiled in that job. For defined contribution pensions, a state variable is created to reflect the balance in the defined contribution account. Year by year, contributions are made to the plan, and the accumulated balance is augmented by the same rate of return that is applied to the non-pension assets. The entire balance is presumed to be made available to the individual in the year following the last year in the job. At that time, the balance of the defined contribution plan is effectively added to the non-pension assets.

The other income variable is primarily composed in most cases of the earnings of the wife. It also includes any pension amounts due to the wife, calculated on the same basis as the pension benefits due to the husband. A final inclusion in some cases is any inheritances that the household receives.

The model contains three important elements of preference heterogeneity. The most important of these is the discount rate reflecting time preference. This is treated as a fixed effect whose value for every household is calculated as the value for which the asset amount calculated by the model matches the observed level of assets in the initial year of the survey. As suggested by the results of the previous section, variations in the discount rate may play an important role in the differing responses of households to opportunities to gain additional annuity amounts by delaying the claiming of Social Security benefits. The second element of heterogeneity is the initial value of the leisure preference parameter ϵ . This value is treated as a random effect taken from a distribution with mean 0 and standard deviation σ_ϵ .

The third element of heterogeneity relates to the relative desire for partial retirement. Recall that leisure has the three values 0, $\frac{1}{2}$, and 1. For these values of L , L^γ has the three values 0, $(\frac{1}{2})^\gamma$, and 1. Thus, γ effectively determines the utility of partial retirement. For

the marginal utility of leisure to be declining, γ must be between 0 and 1, and $(\frac{1}{2})^\gamma$ must be between $\frac{1}{2}$ and 1. If $(\frac{1}{2})^\gamma$ is closer to $\frac{1}{2}$, the value of leisure will be proportional to the amount of leisure, and individuals will tend to choose either full retirement or full-time work depending on whether the value of leisure exceeds the wage or not. If $(\frac{1}{2})^\gamma$ is closer to 1, then partial retirement leisure is almost as valuable as the leisure in full retirement, and the individual is more likely to go through a period of partial retirement. This model supposes that $x = (\frac{1}{2})^\gamma$ and treats this as a random effect drawn from an exponential distribution $f(x) = g(\delta)e^{\delta x}$ truncated below $\frac{1}{2}$ and above 1, where $g(\delta)$ is a term of proportionality to make $f(x)$ integrate to 1 between $\frac{1}{2}$ and 1. The value of δ changes over time according to $\delta = \delta_0 + \delta_1 \text{ Age}$ to reflect that partial retirement may become more desirable over time as individuals grow older. As the value of δ changes over time, however, we assume that an individual's relative position within the distribution stays the same.

The principal stochastic element in the model is the rate of return on assets, which comes from a normal distribution with a mean and standard deviation reflecting the observed returns for a portfolio of roughly half stocks and half short-term government bonds over an extended period of time. A second stochastic element is that after the individual initially retires, the value of ε may unexpectedly change, reflecting that retirement may be more or less enjoyable than anticipated. This is accomplished by introducing a correlation parameter ρ_ε which measures the correlation of ε in adjacent periods after the individual retires. And, of course, life expectancy is stochastic.

There are eight parameters in the complete model: α (the exponent of consumption), β_0 , β_a , and β_h (which affect the weight of leisure in the utility function), σ_ε and ρ_ε (which influence the variance of leisure preferences and how they change over time), and δ_0 and δ_1 (which determine the distribution of partial retirement preferences). State variables in the model, which mediate how past decisions and stochastic events affect the present decisions and how present decisions will affect future circumstances, include the following: the level of assets; whether or not the individual is still in the career job; the level of defined contribution balances if the individual has a defined contribution pension; the primary insurance amount; the Social Security adjustment amounts for both the husband and wife; the

pension benefit if the individual had a defined benefit pension; and the value of leisure relative to consumption.

Given values for the parameters, a value of the discount rate ρ is calculated using the observed or expected retirement dates and estimating a consumption model for the resulting income. Previous research has shown that this gives a good approximation to the median value of ρ . This approach has the advantage that it guarantees that the distribution of assets from the model approximately matches the observed distribution, conditional on lifetime wages and other measures of economic opportunities. Given this value of ρ , the model is solved by backwards induction in the usual process for dynamic stochastic models.

Estimation is achieved using the method of simulated moments technique. This technique chooses the parameter values to best match the moments generated by the model to the corresponding moments observed in the data. Since there are many more moments than there are parameters, the model will be unable to fit all of the moments perfectly, but if the model is correctly specified there should not be any instances where the moments generated by the model are wildly different from the observed moments. The moments that are used are the retirement percentages for both full and partial retirement by age and by health status and lifetime income level, and the percentages of time that individuals return to work after an initial period of full or partial retirement. Moments related to claiming by individuals in the sample are not used in the estimation, for two reasons. First, one of the objectives of this project is to see how close one can come to generating realistic claiming outcomes as a result of optimizing a model built around a retirement decision. And secondly, there are really no parameters in the model that can substantially influence claiming outcomes in any case.

The identification of the parameters is roughly as follows. β_0 shifts the entire retirement distribution. β_a primarily governs how sensitive retirement is to the economic incentives such as pension and Social Security. β_h reflects how sensitive retirement is to health status. σ_ϵ governs how spread out the retirement distribution is. ρ_ϵ helps to determine how many individuals return to work after retirement. δ_0 governs the overall level of partial retirement, and δ_a indicates how partial retirement moves up or down with age. α ultimately reflects how much the retirement distribution shifts with greater or lesser lifetime earnings.

VI. Estimates and Base Simulations.

The model is estimated for the HRS sample previously described, with the results for the parameters being given in Table 3. With the exception of δ_1 , all of the coefficients are individually significantly different from zero at any reasonable level of significance. For a method of moments estimation, if the model is correctly specified, the q statistic is randomly distributed according to a χ^2 distribution with degrees of freedom equal to the number of moments used minus the number of parameters estimated. The current estimates use 43 moments and the model has 8 parameters, so the χ^2 distribution has 35 degrees of freedom. A χ^2 distribution with 35 degrees of freedom has a 5 percent significance value of 49.8 and a 10 percent significance value of 46.1, as compared to the 43.9 value calculated for the estimates. Since low values of χ^2 are consistent with the model being correct, this result does not reject the hypothesis that the model is correct.

By themselves, though, the parameters of the model are somewhat difficult to interpret. Therefore, we turn to a comparison of the results of simulations with the model to observed distributions. These simulations use 10,000 replications per observation, so there should not be much sampling variance in the simulated results.

The first set of simulations tracks the proportion of married men in work and retirement. Figure 1 depicts the difference between the percentage of full-time workers at one age and the corresponding percentage one year later. This might be called the pseudo-retirement percentage, because it is the net result of individuals newly retired from full-time work and a smaller group who were previously retired but returned to full-time work. The most notable feature of this figure is the large spike at age 62. Over 15 percent of the married men in the sample retired at this age, and this spike is fairly well captured in the simulations. It should be emphasized that there is nothing in the utility function that would produce this result; it is the response to the incentives in the budget constraint. It does not occur simply because some individuals have no savings, since they could have always switched to a partial retirement job prior to age 62 and have had at least some income. Rather, it occurs because individuals with fairly high discount rates and relatively low savings often find that the earnings test serves to reduce their effective compensation at age 62. A second considerably lower spike occurs in the data at age 65, but in the simulations this spike is not much more

than a bump. It is likely that this spike arises in the data because of Medicare eligibility, which is not included in the model.

Figure 2 shows the corresponding transitions into full retirement. The spike at age 62 is not quite so large here, either in the data or in the simulation. A plausible explanation for this is that because of the earnings disregard for the earnings test, the earnings test does not bite into partial retirement earnings quite so much as it does full-time earnings. In this figure, also note that the age 65 spike is muted as well. This probably occurs because fewer partial retirement jobs have health insurance than do full-time jobs, so that eligibility for Medicare is not quite so important.

Figure 3 indicates the percentage of husbands claiming Social Security benefits by age, relative to the observed percentages. The simulated percentages are substantially below the observed amounts, especially at the ages before the full retirement age. It had been hoped that the stochastic additions to this model, including stochastic returns to assets and the possibility that some of the previous retired individuals return to work, would have increased the percentage of individuals claiming benefits in the simulations, but evidently this is not the case. These results do reflect the common feeling of many who have examined this issue: that claiming is higher than would be expected, given the actuarial benefits of delaying claiming. (E.g., most recently, see Henriques, 2012.) The percentage of wives who claim benefits is substantially higher, starting at 60 percent at age 62 and rising to 90 percent by age 66.

VII. Modifications to the Model.

There are several potential explanations as to why claiming in the model falls below the levels of claiming observed. In this section, we will look at three of them, introduce them into the model, and assess the degree to which these explanations can explain the relatively high levels of claiming.

The first explanation to be examined is that individuals do not give full weight to future Social Security benefits because they question the solvency of the system and have some doubts whether they will actually receive the full benefits to which the present rules entitle them. Figure 4 gives the results of a simulation in which individuals assumed that

future Social Security benefits would be a fraction of the currently legislated benefits. The original simulated results and observed results are repeated from the previous figure, and the green line in the graph represents the new simulation. In this simulation, the β_0 parameter is adjusted so that the levels of retirement remain at approximately the same level as before. The simulation strongly suggests that the discrepancy between the original simulated results and the observed results can be accounted for if it is presumed that individuals do not believe that they will receive their legislated benefits. The main problem is with the magnitude of the change required to achieve this effect: the simulation supposes that individuals believe that beginning at age 70, they will receive only 50 percent of their legislated benefits. This number would strike many people as low, particularly for the HRS cohort whose retirement occurred largely in the 1990's.

There is a second problem with this explanation. Over time, it would seem that doubts about the future benefits would have increased as the projected time when the trust fund will be exhausted has crept into the possible lifetimes of individuals currently making claiming decisions. If a greater likelihood of reductions in future benefits would cause individuals to be more likely to claim earlier so as to receive benefits before the reductions occur, the claiming rate should be increasing over time. But this is not what has been happening in recent years.

A second explanation of the underestimate of claiming by the current model is that individuals are more optimistic about asset returns than the model assumes. The mean return used in the model is the historical mean of the Ibbotson series dating back to the 1926, but individuals may be using a different period with a higher mean. The simulation in Figure 5 uses a higher mean return than is used in the base simulation. The result of this simulation almost overlays the observed data. The main difficulty with this simulation is again with the magnitude: the mean rate of return is increased by 5 percentage points over the historical series, a rate that many might regard as a bit high.

This explanation, however, does have the advantage that it is more or less consistent with recent trends. The 1990's, during which much of the retirement of the original HRS cohort occurred, was a period of heady returns. During that time, there was a lot of discussion about the poor returns of Social Security, and individuals were likely to claim as

soon as possible in order to take advantage of those returns. In the last dozen years, however, the picture has been much different. Real returns have been inching down into negative territory, and the discussion about the poor returns to Social Security has largely ceased. The recent downturn in claiming would certainly be consistent with generally lower expectations regarding the returns to assets.

The third explanation for why the model may underestimate claiming has to do with the assumption in the model that the husband provides equal weight to his and his wife's utility, weighted only by the discount rate and survival probabilities. The experience with pensions in past decades certainly gives some reason to suspect this assumption. Before there were any regulations, many if not most husbands with defined benefit pensions took those pensions as single life annuities, even given a roughly actuarially fair opportunity to convert them to annuities that would pay the wife in the case he died first. The fact that so many widows were left destitute by this practice led to regulations that made it substantially more difficult for a married person to take a defined benefit pension as a single life annuity. The same approach may lead the husband to discount the benefits to the wife after his death when making the claiming decision.

Figure 6 shows the results of a simulation wherein the husband does not give any weight to the wife's survivor benefits after he dies. Somewhat surprisingly, the lack of concern for the wife's survivor benefits does not make a great deal of difference in claiming in the simulation, especially in the early years of eligibility for Social Security. To understand this result, recall that the majority of survivor benefits will come only after many years have passed. Any calculation that gives much weight to these years will of necessity have a fairly low discount rate. But at low discount rates, the value of the marginal Social Security annuity relative to its cost is probably fairly high, even if the wife's survivor benefits are ignored. After the full retirement age, this advantage starts to erode, and the simulations predict claiming closer to the observed claiming outcomes if the individual does not take into account the wife's survivor benefits.

Notice that the model is not re-estimated under these assumptions. For reasons we discussed earlier, claiming is not included in the moments used in the estimation. Consequently, re-estimating the model with the moments used in the estimation would not

make much difference. Moreover, the effect of the parameters in the model on claiming are very indirect. The identification of the parameters is largely determined by the relations discussed on p. 20 above, and these relationships are already included in the moments used to estimate the model. In other words, there are no parameters in the model which directly influence claiming, and the changes being considered in the simulations do not affect the moments which identify the parameters.

While all three of these explanations have some weak points, they are not mutually exclusive. There is nothing in the simulations which argues that the changes have to be uniform; for instance some individuals may expect almost no benefits and others expect only a modest change to generate the results. It may be that two of three of them in combination would provide an adequate explanation of the observed claiming outcomes without invoking implausible assumptions. We are not arguing that these changes are the correct explanation for the underestimation of claiming, but merely potential explanations for the underestimation. The simulations are meant to give an idea of the magnitude of the adjustments on average that would be required to bring claiming up to the observed level. There may also be other explanations which we have not considered; again, these simulations are only meant to give an idea of the magnitude of these changes that would be required to bring claiming up to the observed levels.

VIII. Policy Analysis.

Next we ask whether these explanations create any differences in the results of three commonly proposed changes in the Social Security system designed to increase employment and increase the solvency of the system. These policies would increase the early entitlement age, increase in the full retirement age, and eliminate the payroll tax after an individual has reached the full retirement age. In each case, we look at simulations for the base case and also at simulations that reflect the three potential explanations for the underestimation of claiming examined in the previous section.

Figure 7 looks at the results for the simulations involving increasing the early entitlement age to 64. The base case and all three modifications of the model are in remarkable agreement: full-time employment will be little affected before age 62 or after age

64, and at ages 62 and 63 full-time employment will increase by approximately 12 percentage points. Essentially, the spike in retirement from full-time work which presently occurs at age 62 would be shifted to age 64 by this change. The solvency of the system will be increased by the extra two years of payroll taxes paid by those who delay, balanced against the fact that for most of these individuals, delaying benefits will result in more than actuarially fair adjustments to later benefits, which will cost the system in the long run.

Figure 8 examines the results for the simulations involving increasing the full retirement age to 67 from its value of age 65 for most of the individuals in the sample. In the figure, “insolvency” refers to the simulation where individuals expect reduced future benefits due to solvency concerns about the system, “returns” refers to the simulation with the increased returns on assets, and “survivor” refers to the simulation wherein the husband discounts the benefits that would be paid to his surviving spouse were he to die first. These simulations show somewhat greater variability than the simulations in Figure 7, but in general they show two things: the increases in full-time work caused by increasing the full retirement age are substantially less than the increases caused by raising the early entitlement age, and the impact of raising the full retirement age is generally positive. The outlier appears to be the simulation which reduces expectations that individuals will receive their full legislated benefits; the results of the other three simulations are relatively much closer. In terms of relieving the Social Security shortfall, however, this is probably a much more effective change than is increasing only the early entitlement age. Whereas an increase in the early entitlement age mostly postpones benefits, an increase in the full retirement age without an offsetting change in the full benefit serves effectively to reduce considerably the present value of lifetime benefits.

Figure 9 shows the results of simulations for eliminating the payroll tax after the full retirement age. These simulations show the result of eliminating only that part of the tax paid by the employee; if the employer’s side of the tax were eliminated and if that saving were to be passed on to employees, one would expect the resulting magnitudes to be roughly double the amounts shown. The simulations also assume that for ages for which the payroll taxes are eliminated, earnings for those ages are not used in the PIA calculations. These simulations indicate a relative agreement between the base simulation and those simulations

incorporating the various explanations for the claiming underestimates in the base model. They all show a reduction in full-time work of between 0.5 and 1 percent between age 60 and age 64 and an increase in full-time work of between 1 and 2 percent at age 65 and thereafter. Summed over all the ages, the net result seems to be an increase in full-time effort, with some shifting of work effort from the period when compensation is unchanged to the period when effective compensation has increased.

IX. Conclusions

Our estimates reinforce the notion that it is difficult to explain claiming with a straightforward life-cycle model, even one with stochastic asset returns and the possibility of returning to work after retirement. In the analysis, three different sets of beliefs about the future of Social Security are used in an attempt to bring the predicted claiming rates closer to the observed claiming rates. These variants in beliefs about the future of Social Security do permit the model to better match simulated claiming outcomes with observed outcomes without much affecting other positive advantages of the model, including the ability to replicate the spike in retirement at the early entitlement age. Incorporating the variants in beliefs also indicates how much individuals' beliefs would have to change (e.g., how much higher would asset returns need to be) to make the observed pattern of claiming optimal. It turns out that these changes are rather large. The variant that best appears to remedy the problem of underestimating claiming while still being consistent with the reduction in claiming over the last few years is one in which individuals use shorter horizons in forming their expectations about returns to assets. Since expectations regarding asset returns have fallen considerably in recent years, claiming relatively early is becoming less attractive against the alternative of waiting to claim benefits and having future Social Security payments increased.

We find that incorporating the different beliefs as to the future of Social Security and interest rates does improve the ability of the model to jointly explain claiming and retirement outcomes. This improvement is not achieved by sacrificing the model's ability to explain heterogeneity in wealth accumulation.

Of the three potential policy changes considered, changes in the early entitlement age, in the full retirement age, and the elimination of the payroll tax for workers over the full retirement age, an increase in the early entitlement age clearly has the most impact on retirement, while the increase in the full retirement age probably has the most impact on the solvency of the system.

We also have examined the effects of the three potential policy changes in the context of varying beliefs. Consistent with our previous work, raising the early entitlement age to 64 will increase full time work at ages 62 and 63 by 11 to 13 percentage points. Increasing the full retirement age from 65 to 67 (for those in the original HRS cohort) would increase full-time work at age 64 by 1.5 to 2.5 percentage points. Eliminating the payroll tax for work after full retirement age would increase full-time work by 1.0 to 1.7 percentage points between the ages of 65 and 67.

For two of the policy changes, namely an increase in the early entitlement age and the elimination of the payroll tax, the model gives consistent answers regardless of which set of beliefs is used. There is more dispersion in predicted effects for an increase in the full retirement age, but even here most of the dispersion is associated with one set of beliefs, namely that there will be a severe reduction in expected future benefits.

A recurring theme of our research is that in designing policies, it is helpful to be aware of heterogeneity in the population and the possibility that policies may have different impacts on different parts of the population. For instance, current Social Security rules allow individuals to trade current benefits for better than actuarially fair increases in future benefits. Ideally, this would allow individuals with low levels of assets to secure an increased level of steady income for the rest of their lives. But individuals with low levels of assets frequently are in that situation because of a relatively high discount rate, and such individuals are unlikely to take advantage of the offer to trade present benefits for increased future benefits. Thus, policymakers face a situation in which individuals who would benefit most from delaying claiming are the least likely to respond to incentives to delay claiming.

In the same vein, an increase in the early entitlement age may have the greatest effect on those with high discount rates. Individuals with high discount rates may not think that the future benefit increases are worth giving up current benefits, and with few assets they may be

forced to work more than they would otherwise find optimal. Of course, that is the way they feel during the period when the benefits are being delayed; ten years later, they may feel glad that they are receiving higher benefits than they would have received had they been able to claim those benefits earlier. The situation is much different for those with low discount rates. Such individuals will most likely have a relatively higher level of assets relative to earnings. They may well find that it is advantageous to delay claiming anyway, since for them the future benefit increases outweigh the current benefits foregone, and they can live off their assets in the meantime even if they are retired.

Another example would be future medical expenditures. Financial planners have increasingly been sounding the alarm about potential high out-of-pocket expenditures for medical expenses in retirement. Individuals with low discount rates may respond to high and variable potential expenditures for medical expenses by increasing saving, working longer, and taking other measures. Those with higher discount rates may not respond much at all to the prospect of higher medical expenses in what is for them the distant future. If they do wind up with substantial medical expenses, they will simply be forced to tighten their belts and rely on whatever programs exist to keep them out of poverty.

This paper also raises a number of questions for future research. It investigated several explanations of why individuals might want to begin claiming Social Security benefits even though actuarial and even utility calculations might suggest that it would be advantageous to delay, and it has noted at least some recent evidence that might favor one of these explanations. But it would certainly be helpful to have more evidence on this score. Also, the model emphasizes heterogeneity in time preference as an important aspect of modeling individuals' decision making, and to a lesser degree heterogeneity in leisure preferences. In the limit, with preferences differing in unobservable ways, it is important to try to ascertain what unobserved dimensions of heterogeneity are critical to answering the questions at hand, and to try to build those dimensions of unobserved heterogeneity into the appropriate model.

References

- Anderson, Patricia M., Alan L. Gustman, and Thomas L. Steinmeier. 1999. "Trends in Male Labor Force Participation and Retirement: Some Evidence on the Role of Pensions and Social Security in the 1970s and 1980s." *Journal of Labor Economics*, 17(4): 757-783.
- Benitez-Silva, Hugo, Moshe Buchinsky, Hiu-Man Chan, John Rust, and Sophia Sheivasser. 1999. "An Empirical Analysis of the Social Security Disability Application, Appeal and Award Process." *Labour Economics*, 6: 147-178.
- Benitez-Silva, Hugo, Moshe Buchinsky, and John Rust. 2001. "Dynamic Models of Retirement and Disability." Paper presented at the Winter Meetings of the Econometric Society, New Orleans, January.
- Benitez-Silva, Hugo A. and Frank Heiland. 2008. Early Retirement, Labor Supply, and Benefit Withholding: The Role of the Social Security Earnings Test. Michigan Retirement Research Center, Working Paper UM08-09.
- Berkovec, James C., and Steven Stern. 1991. "Job Exit Behavior of Older Men." *Econometrica*, 59(1): 189-210.
- Bound, John, Todd R. Stinebrickner and Timothy A. Waidmann. 2010. "Health, economic resources and the work decisions of older men." *Journal of Econometrics*, 156(1): 106-129.
- Brown, Jeffrey R., Marcus D. Casey and Olivia S. Mitchell. 2007. "Who Values the Social Security Annuity: New Evidence on the Annuitization Puzzle". Unpublished MS.
- Burtless, Gary and Robert Moffitt. 1984. "The Joint Choice of Retirement Age and Postretirement Hours of Work." *Journal of Labor Economics* 3(2): 209-236.
- Coe, Norma B., Mashfiqur Khan and Matthew S. Rutledge. 2013. "Sticky Ages: Why Is Age 65 Still A Retirement Peak?". Boston College, Center for Retirement Research. Working Paper 2013-2.
- Coile, Courtney, Peter Diamond, Jonathan Gruber and Alain Jouten. 2002. "Delays in claiming Social Security benefits". *Journal of Public Economics* 84: 357-385.
- Coile, Courtney, and Jonathan Gruber. "The Effect of Social Security on Retirement in the United States". In *Social Security Programs and Retirement Around the World*. Jonathan Gruber and David A. Wise, Editors. Chicago, University of Chicago Press.

- _____. 2005. "Social Security Incentives for Retirement." In *Themes in the Economics of Aging*. David A. Wise Editor. Chicago, University of Chicago Press.
- Congressional Budget Office. 2010. "Social Security Policy Options".
- Disney, Richard and Sarah Smith. 2002. "The Labor Supply Effects of the Abolition of the Earnings Rule for Older Workers in the United States". *The Economic Journal* 112: C136-C152.
- Feldstein, Martin and Andrew Samwick. 1992. "Social Security Rules and Marginal Tax Rates". *National Tax Journal* 45: 1-22.
- French, Eric. 2005. "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior." *Review of Economic Studies*, 72: 395-427.
- French, Eric, and John Bailey Jones. 2004. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." *Econometrica* 79(3): 693-732.
- Friedberg, Leora. 2000. "The Labor Supply Effects of the Social Security Earnings Test." *The Review of Economics and Statistics* 82 (1): 48-63.
- Gordon, Roger H. and Alan S. Blinder. 1980. "Market Wages, Reservation Wages and Retirement". *Journal of Public Economics* 14: 277-308.
- Gruber, Jonathan and David A. Wise. 1999. "Introduction and Summary". In *Social Security and Retirement Around the World*. Edited by Jonathan Gruber and David A. Wise. Chicago: University of Chicago Press: 1-35.
- Gustman, Alan L. and Thomas L. Steinmeier. 1985. "The 1983 Social Security Reforms and Labor Supply Adjustments of Older Individuals in the Long Run". *Journal of Labor Economics* 3(2): 237-253.
- _____. 1986. "A Structural Retirement Model". *Econometrica* 54(3): 555-584.
- _____. 1999. "Effects of Pensions on Savings: Analysis with Data From the Health and Retirement Study". *Carnegie-Rochester Conference Series*, vol. 50 (July, 1999) pp. 271-326.
- _____. 2000. "Retirement in Dual-Career Families: A Structural Model." *Journal of Labor Economics*, 18(3): 503-545.

_____. 2004a. "Social Security, Pensions and Retirement Behavior within the Family." *Journal of Applied Econometrics*, 19(6): 723–738.

_____. 2004b. "The Social Security Retirement Earnings Test, Retirement and Benefit Claiming." Working Paper 10905. Cambridge, Mass.: National Bureau of Economic Research.

_____. 2005. "The Social Security Early Retirement Age In A Structural Model of Retirement and Wealth," with Thomas Steinmeier. *Journal of Public Economics*. Volume 89, Issues 2-3 (February 2005): 441-463.

_____. 2008. "Projecting Behavioral Responses to the Next Generation of Retirement Policies". *Research in Labor Economics* 28: 141-196.

_____. 2009. "How Changes in Social Security Affect Recent Retirement Trends," with Thomas Steinmeier. *Research on Aging* 31(2): 261-290.

Henriques, Alice M. 2012. "How Does Social Security Claiming Respond to Incentives? Considering Husbands' and Wives' Benefits Separately". Washington, D.C. Federal Reserve Board.

Laitner, John and Daniel Silverman. 2012. "Consumption, Retirement and Social Security: Evaluating the Efficiency of Reform that Encourages Longer Careers". *Journal of Public Economics* 96 (7-8): 615-634.

Lumsdaine, Robin L., James H. Stock, and David A. Wise. 1990. "Efficient Windows and Labor Force Reductions." *Journal of Public Economics*, 43(2): 131–159.

_____. 1992. "Three Models of Retirement: Computational Complexity versus Predictive Validity." In David Wise, ed., *Topics in the Economics of Aging*. Chicago: University of Chicago Press, 19–57.

Mastrobuoni, Giovanni. 2009. "Labor Supply Effects of the Recent Social Security Benefit Cuts: Empirical Estimates Using Cohort Discontinuities". *Journal of Public Economics*, 93(11-12): 1224-1233.

Rust, John and Christopher Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets". *Econometrica* 65(4): 781-832.

Scholz, John Karl and Ananth Seshadri. 2012. "The Interplay of Wealth, Retirement Decisions, Policy and Economic Shocks". MRRC. WP 2012-271.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. "Are Americans Saving 'Optimally' for Retirement?" *Journal of Political Economy*, 114(4): 607–643.

Senate Committee on Aging. 2010. "Social Security Modernization: Options to Address Solvency and Benefit Adequacy". Report of the Special Committee on Aging, United States Senate.

Shoven, John B and Sita Nataraj Slavov. 2012a. "The Decision to Delay Social Security Benefits: Theory and Evidence". NBER Working Paper No. 17866.

_____. 2012b. "When Does it Pay to Delay Social Security? The Impact of Mortality, Interest Rates, and Program Rules". NBER Working Paper No. 18210.

Social Security Administration Trustees Report. 2012. "The 2012 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds". Washington, D.C.

Song, Jae G. 2002. "Evaluating the Effects of the Removal of the Retirement Earnings Test in 2000". Mimeo. Social Security Administration. November.

Song, Jae and Joyce Manchester. 2007. "Have People Delayed Claiming Retirement Benefits?" *Social Security Bulletin* 67(2): 1-23.

Venti, Steven F., And David A. Wise. 1999. "Lifetime Earnings, Saving Choices, and Wealth at Retirement." In James P. Smith and Robert J. Willis, Editors, *Wealth, Work and Health: Innovations in Measurement in the Social Sciences*. Ann Arbor, University of Michigan Press, pp. 87-120.

_____. 2001. "Choice, Chance, and Wealth Dispersion at Retirement". In S. Ogura, T. Tachibanaki, and D. Wise (ed.) *Aging Issues in the United States and Japan*. University of Chicago Press.

Van der Klaauw, Wilbert and Kenneth I. Wolpin. 2006. "Social Security and the Retirement and Savings Behavior of Low Income Households." *Journal of Econometrics*, 145(1-2): 21-42.

Table 1
Actuarial Rates for Claiming at Various Ages

Real Interest Rate	0.02	0.04	0.02
Marital Status	Married ^a	Married ^a	Single
Claiming Age			
62	1.67	1.31	1.18
63	1.50	1.18	1.05
64	1.36	1.08	0.95
65	1.48	1.18	1.02
66	1.33	1.07	0.91
67	1.20	0.97	0.82
68	1.09	0.89	0.74
69	0.99	0.81	0.66

^a Calculations assume that the wife is two years younger and will collect survivor benefits based on her husband's earnings record, but will collect benefits based on her own record while he is alive. The assumed full retirement age is 65, and the assumed delayed retirement credit is 8%.

Table 2
Annuities: Willingness to Pay vs. Actuarial Value

Real Interest Rate	Discount Rate	Current Age	Age When Assets Become Zero				
			Immediate	70	80	90	100
0.02	0.00	62	1.23	1.25	1.38	1.60	1.78
		65	1.21	1.23	1.39	1.65	1.85
		70	1.17	1.17	1.37	1.72	1.99
0.02	0.02	62	1.00	1.14	1.34	1.59	1.78
		65	1.00	1.11	1.35	1.64	1.85
		70	1.00	1.00	1.32	1.70	1.99
0.02	0.04	62	0.83	1.05	1.32	1.58	1.78
		65	0.85	1.01	1.31	1.63	1.85
		70	0.87	0.87	1.28	1.69	1.99
0.04	0.02	62	1.20	1.21	1.32	1.48	1.59
		65	1.18	1.21	1.34	1.53	1.66
		70	1.15	1.15	1.34	1.61	1.80

Table 3
Parameter Estimates

Parameter	Description	Value (Standard Deviation)
α	Exponent of consumption	-0.16 (3.36)
β_0	Constant in leisure value expression	-9.558 (408.81)
β_a	Coefficient of age in leisure value expression	0.067 (6.54)
β_h	Coefficient of health in leisure value expression	5.70 (7.40)
σ_ε	Standard deviation of ε in leisure value expression	5.87 (9.33)
ρ_ε	Correlation of ε after retirement	0.71 (5.72)
δ_0	Constant term in partial retirement distribution	-3.62 (8.00)
δ_1	Coefficient of age in partial retirement distribution	0.16 (1.60)
Number of observations:		2231
q-statistic:		43.94

Figure 1
Retirements from Full-Time Work

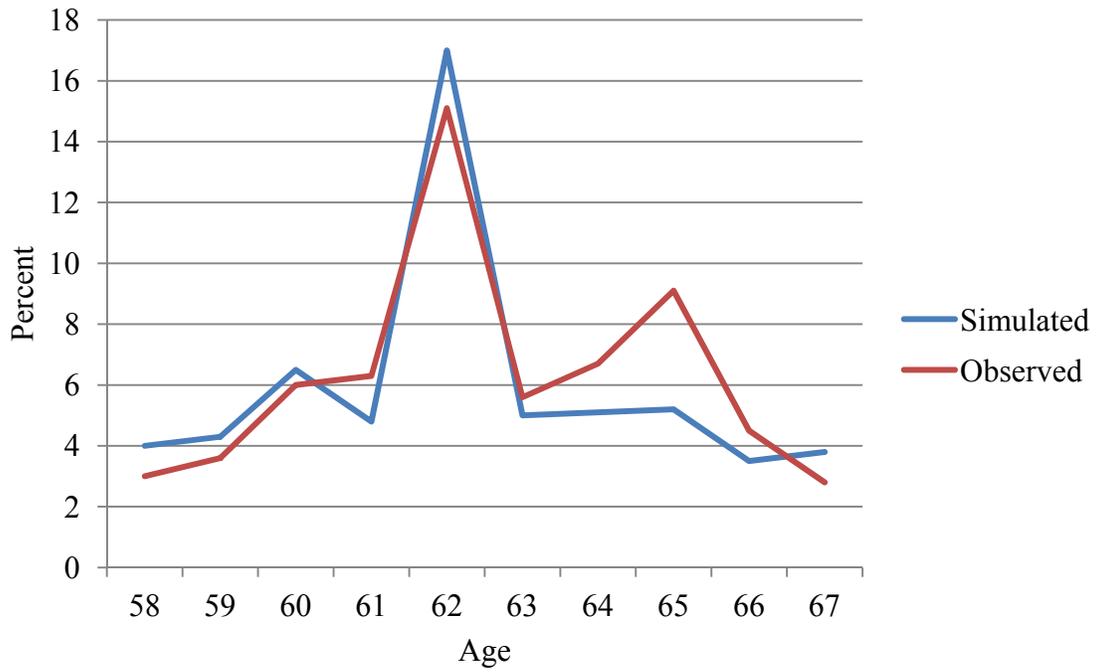


Figure 2
Retirements from All Work

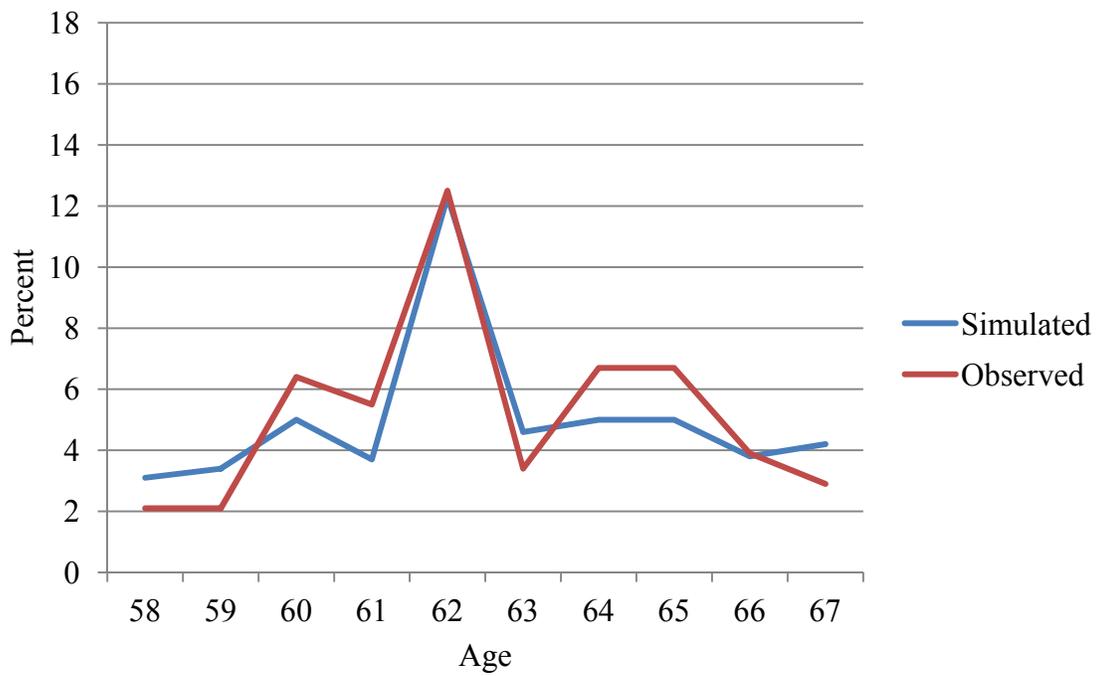


Figure 3
Percent Claiming Social Security Benefits

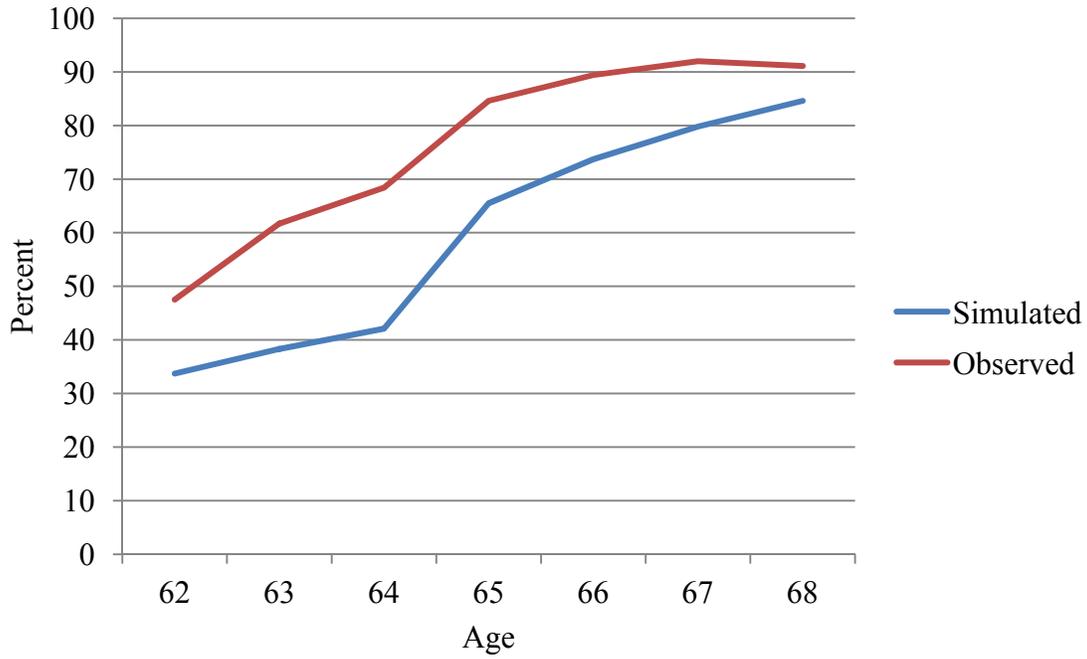


Figure 4
Simulation with Reduced Expectations of Receipt

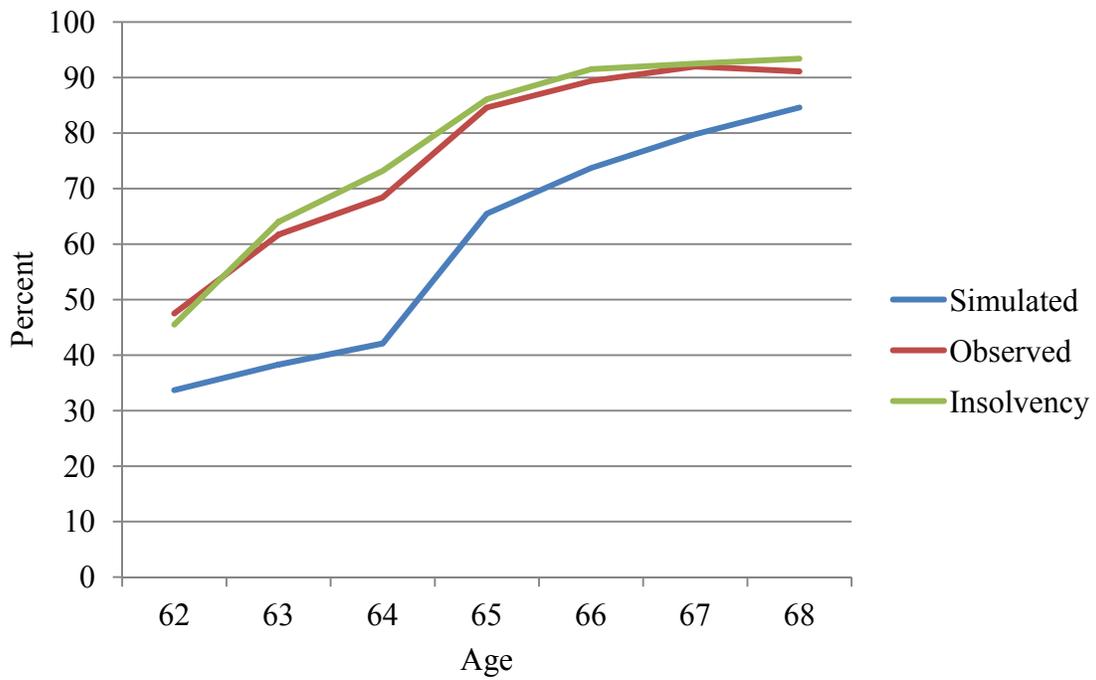


Figure 5
Simulation with Increased Return on Assets

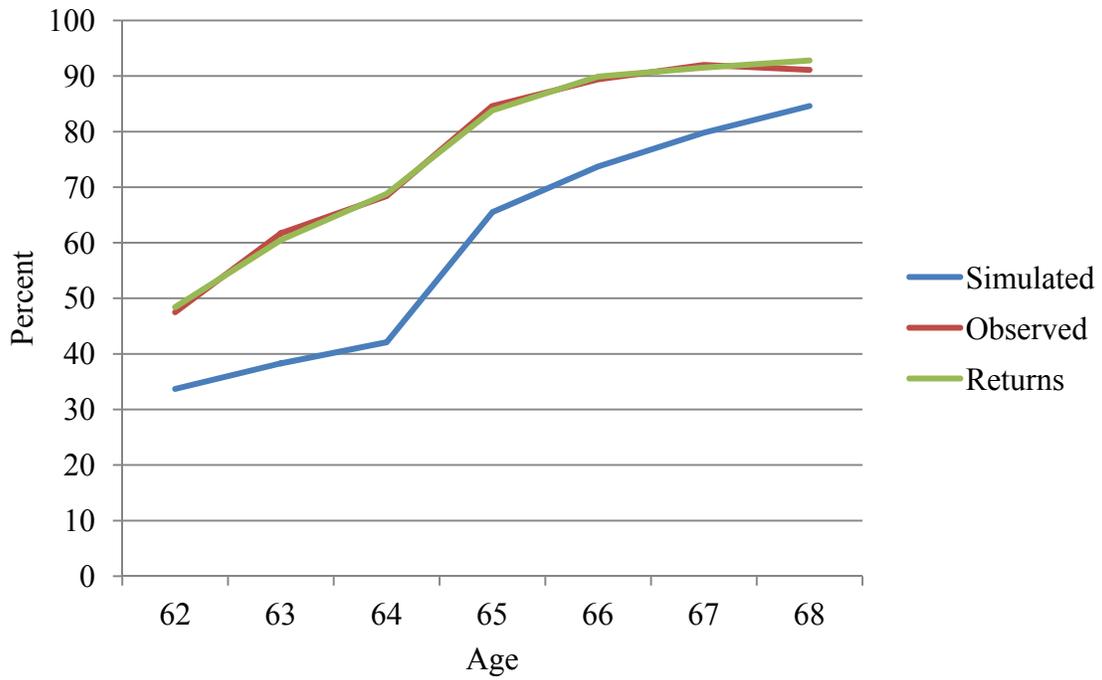


Figure 6
Simulation with Survivor Benefits Ignored by Husband

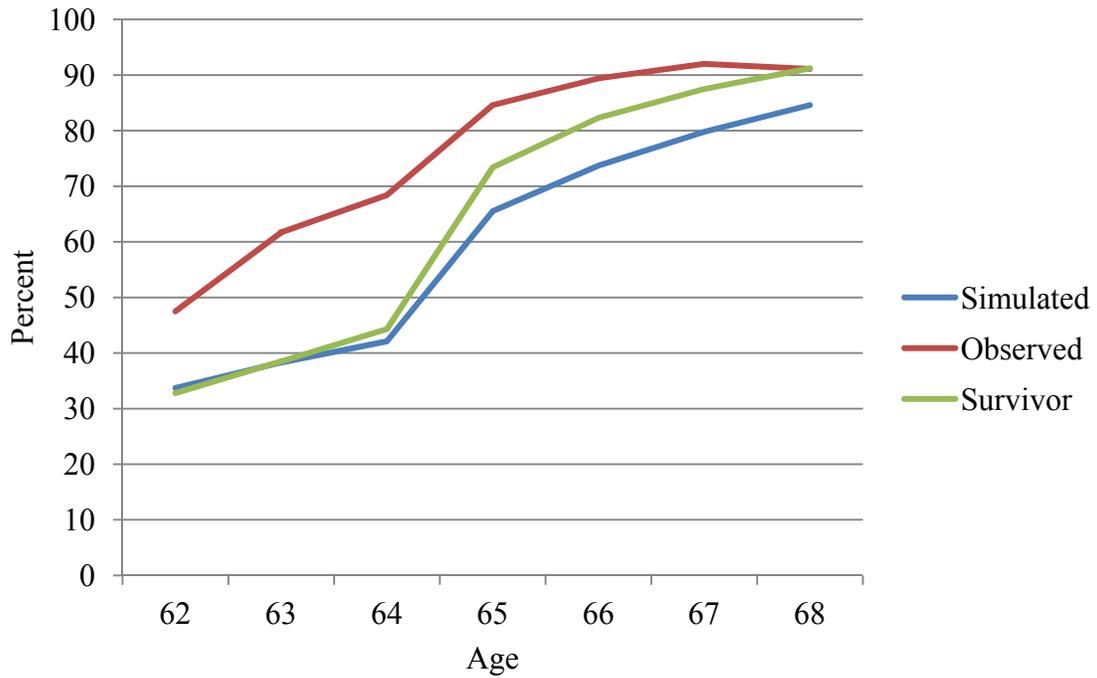


Figure 7
Increase in Full-Time Work with Early Entitlement Age 64

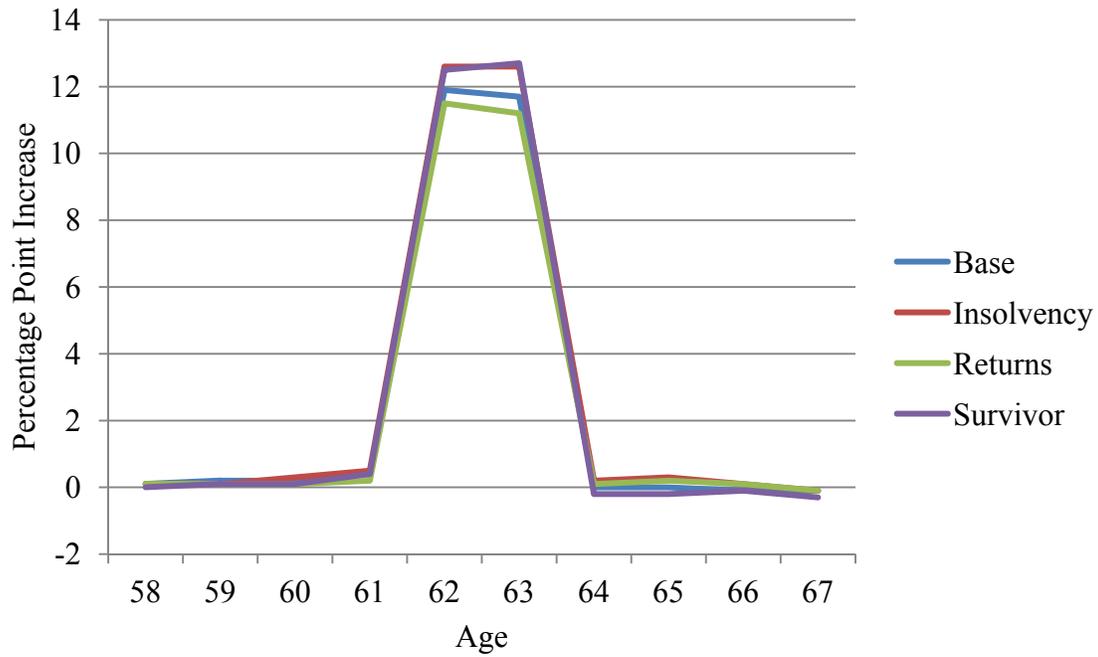


Figure 8
Increase in Full-Time Work with Full Retirement Age 67

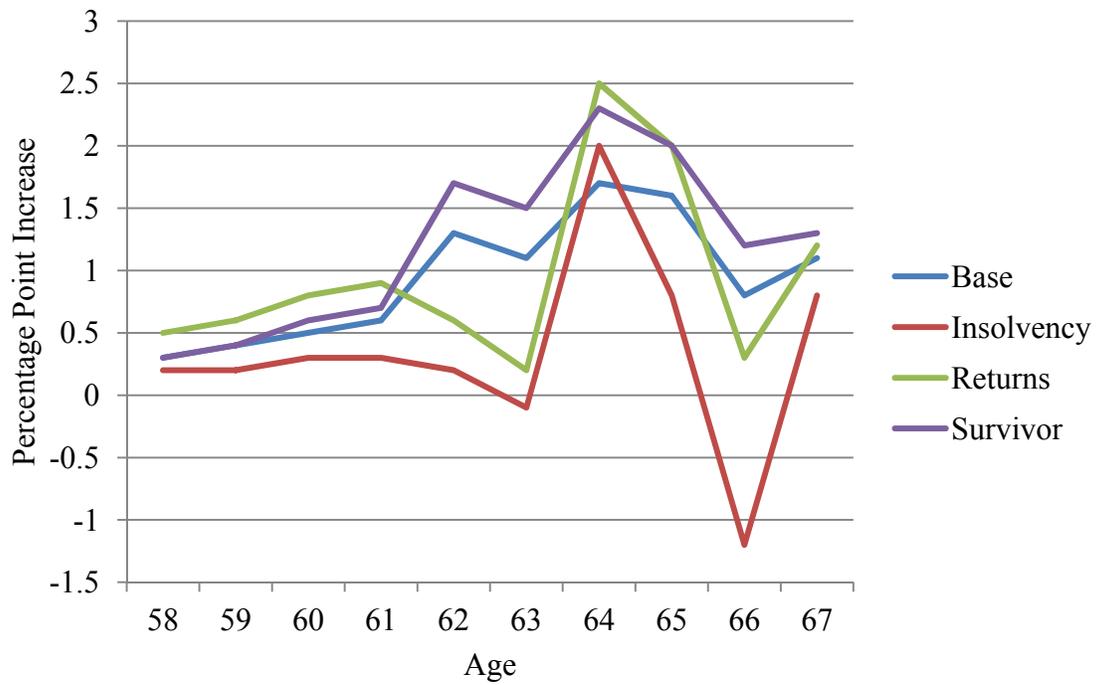


Figure 9
Increase in Full-Time Work with No Payroll Tax After Full Retirement Age

