Economic Brief

November 2020, EB20-12

Lifespan Inequality and Social Security Reform

By John Bailey Jones and John Mullin

What does lifespan inequality mean for Social Security reform? Using a life-cycle model in which the rich tend to outlive the poor, the researchers analyze how various reforms affect the trade-off between distributing lifetime Social Security benefits more equally and encouraging society's most productive members to work longer. They find that social welfare is maximized when benefits are independent of lifetime earnings, the payroll tax cap remains near its current level, and benefits are made less dependent on the age at which they are initially claimed.

A common proposal for stabilizing the U.S. Social Security system is to increase the system's normal retirement age (NRA). An intuitively appealing feature of this proposal is that it counters the trend toward longer lives — a major cause of Social Security's financial difficulties — by delaying the age at which full benefits start. A less appealing aspect of the proposal is that it likely would have a disproportionately negative effect on the poor because the longevity gap between the rich and poor in the United States is large and growing,[1] which means that the poor tend to receive fewer years of Social Security benefits than the rich. Consequently, a one-year increase in the NRA could diminish the lifetime benefits received by the poor by a greater proportion than those received by the rich. While this concern can be mitigated by changing the progressivity of the Social Security benefit formula, a more general question remains: If high-income workers live longer than low-income workers, shouldn't they work longer as well?[2]

Economists have developed an extensive set of tools for analyzing questions about the fairness and efficiency of government tax-and-transfer policies, including Social Security. Standard economic models of optimal fiscal policy start with the premise that policies should be chosen to maximize social welfare, subject to some constraints. Social welfare, in turn, depends on the welfare or "utility" of society's individuals. All other things remaining the same, social welfare in these models increases when income is transferred from the rich (who have relatively low marginal utilities of consumption) to the poor (who have relatively high marginal utilities of consumption). But, of course, all other things do not necessarily stay the same. Except in very special cases (involving "lump-sum" taxes and transfers), government tax-and-transfer programs alter individuals' incentives and thereby affect labor-leisure and spending-saving decisions that can change the size of the overall economic pie. In 1971, James Mirrlees's seminal

paper highlighted the tension between equalizing consumption and encouraging work by the most productive.[3] In a more recent (2010) contribution, Helmuth Cremer, Jean-Marie Lozachmeur, and Pierre Pestieau analyzed the trade-offs associated with public pension programs when beneficiaries have a variety of survival probabilities.[4] With such heterogeneous mortality, "consumption equalization" becomes a tricky concept because equal per-period consumption implies larger lifetime transfers to people who live longer.

In a recent working paper, John Bailey Jones of the Richmond Fed and Yue Li of the University at Albany extend this research by building a heterogeneous-agent, life-cycle model of Social Security to analyze the implications of lifespan inequality for Social Security reform.[5] They find that, relative to the Social Security policies currently in place, policies that would maximize social welfare would reduce work incentives in order to redistribute resources from high earners to low earners. More specifically, quantitative experiments with their calibrated model show that welfare is maximized when benefits are independent of lifetime earnings, the payroll tax cap is kept roughly unchanged from current levels, and when benefits are made less dependent on the age at which they are initially claimed. In their model, additional gains are provided by eliminating taxes on Social Security benefits and by removing the Social Security "earnings test," which reduces the benefits of younger claimants who continue to work.

The Model

The researchers extend the framework used by Selahattin Imrohoroğlu of the University of Southern California and Sagiri Kitao of the University of Tokyo to incorporate a richer treatment of health, mortality, education, and wages.[6] In the extended model, individuals make labor-leisure choices while facing uncertain prospects for their health, wages, medical spending, and mortality — the distributions of which vary by their levels of education. The government collects income, payroll, and consumption taxes and provides Social Security, Medicare, disability insurance, and means-tested social insurance. Individuals can continue working while receiving Social Security benefits, but they may face financial incentives to retire.

In the context of the model, the researchers solve the problem that a social planner would face when attempting to maximize the expected lifetime utility of a newborn, prior to knowing the newborn's endowed characteristics. This optimization criterion, which is standard in the literature, implicitly assigns an equal weighting to all individuals in a society. The criterion also provides a reasonable means of overcoming the problems associated with period-by-period analysis in a setting where agents have heterogeneous mortality.

After calibrating the model to reflect the current state of the U.S. economy and Social Security policies, the researchers evaluate potential reforms. To focus their analysis on policy trade-offs, they evaluate only "parametric" reforms that leave Social Security's overall footprint unchanged. In particular, they limit their analysis to policy changes that do not change aggregate Social Security expenditures, revenues, or general government budget balances. The researchers consider three categories of parametric reforms:

1. Changes to the Social Security payroll tax rate and changes to the cap on taxable payroll earnings;

2. Changes to the formula converting people's earnings histories into their baseline benefits, known as their primary insurance amount (PIA); and

3. Changes to the trade-off between the age at which people first receive their Social Security benefits — and thus the length of the benefit stream — and the size of the annual benefit.

All of the reforms they study have appeared before as proposals — some enacted, some rejected. The first two categories of reforms affect how Social Security payroll taxes and benefits depend on wage realizations and work decisions. The third category affects how benefits depend on the age at which individuals initially claim Social Security benefits. Increases in the NRA can be interpreted as a special

case of this type of reform. This trade-off is formally embodied in the early retirement penalties and delayed retirement credits, which decrease or increase annual benefits, respectively.

Because people can simultaneously work and receive benefits, their work decisions and claiming decisions may appear to be disconnected under current policies. This is not the case, however, for three main reasons. First, the rate at which earnings translate into Social Security benefits is an increasing function of the age at which the benefits are claimed — a mechanism that creates an incentive to work more years before retirement. Second, benefits received by nonretirees can be reduced due to the Social Security earnings test — a mechanism that creates a work disincentive by penalizing people who claim benefits while still working.[7] And third, Social Security benefits tend to push beneficiaries into higher marginal tax brackets — another work disincentive.

Each category of reforms embodies the canonical trade-off between income redistribution and productive efficiency. Under the assumption that Social Security's overall footprint stays the same, raising the payroll tax cap allows for a lower tax rate, but that lower rate applies to a broader range of earnings. This reduces taxes for most workers but raises marginal tax rates for the most productive workers. Linking Social Security benefits to lifetime earnings increases the incentive to work, but it reduces transfers from high earners to low earners. Raising the penalty for claiming benefits at an early age creates a work incentive, but it punishes people with shorter lifespans.

Policy Implications

As a general principle, the researchers find that, relative to the Social Security policies currently in place, the policies that would maximize welfare would reduce work incentives in order to redistribute resources from high to low earners. Under these policies, the PIA would be independent of lifetime earnings, and differences in benefits due to differences in claiming ages would be smaller, while the cap on taxable earnings would remain at more or less its current value. Collectively, this group of reforms would cause both earnings and employment to fall by 1 percent to 2 percent.

However, the researchers show that more than half of the declines in earnings and employment could be reversed by eliminating the earnings test and the taxation of benefits.[8] Adding these reforms also results in larger welfare gains. Because the earnings test and taxation of benefits apply only to older workers, whose elasticity of labor supply is especially high, eliminating them is an especially effective way to encourage work.[9] Removing these provisions uncouples claiming decisions from retirement decisions; under the joint reforms, almost everyone in the model claims benefits at age sixty-two.

The researchers also consider how heterogeneous mortality affects Social Security reform in the face of changing demographic patterns. They construct a hypothetical "2050-demographics" scenario characterized by longer lifespans, lower population growth, and a heightened dependence of mortality on education. They find that the Social Security system that would maximize welfare in the 2050 demographic environment would be quite similar to the one that would maximize welfare today. In both cases, the PIA would be independent of individual earnings; the payroll tax cap would be higher in the 2050 scenario, but claiming adjustments would remain roughly the same in both scenarios. Although increased longevity suggests that larger claiming adjustments are needed to promote longer careers, increased longevity also implies that the adjustments needed to induce claiming delays (and longer careers) are smaller. The net effect is that the optimal claiming adjustments are smaller than in the current demographic environment, but only slightly so.

Conclusion

Within the literature on parametric Social Security reforms, the researchers' main contribution is to consider a broad set of reforms simultaneously and quantitatively while accounting for heterogeneity in income, health, and mortality. While it has

been long recognized that heterogeneous mortality affects the lifetime progressivity of Social Security, and multiple studies have sought to quantify this effect, there has been relatively little progress in quantifying its implications for optimal policies. The researchers believe that there are a number of promising avenues along which to extend their framework. The first would be to add marriage and spousal benefits to their model. A second would be to consider the entire tax and transfer system rather than just Social Security. Jones and Li's results suggest that incorporating a greater degree of heterogeneity into such analyses — in place of some of the demographic simplifications that are commonly used — may have important policy implications. ■

John Bailey Jones is a senior economist and research advisor and John Mullin is an economics writer in the Research Department at the Federal Reserve Bank of Richmond.

Endnotes

[1] Jessie Romero, "The Mortality Gap," *Econ Focus*, Third/Fourth Quarter 2016, vol. 21, no. 3–4, pp. 18–21.

[2] An analysis of the policy implications of changing life expectancies for Social Security can be found in National Academies of Sciences, Engineering, and Medicine, *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses*, Washington, D.C.: The National Academies Press, 2015.

[3] J. A. Mirrlees, "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, April 1971, vol. 38, no. 2, pp. 175–208.

[4] Helmuth Cremer, Jean-Marie Lozachmeur, and Pierre Pestieau, "Collective Annuities and Redistribution," *Journal of Public Economic Theory*, February 2010, vol. 12, no. 1, pp. 23–41.

[5] John Bailey Jones and Yue Li, "Social Security Reform with Heterogeneous Mortality," Federal Reserve Bank of Richmond Working Paper No. 20-09, July 2020.

[6] Selahattin Imrohoroğlu and Sagiri Kitao, "Social Security Reforms: Benefit Claiming, Labor Force Participation, and Long-Run Sustainability," *American Economic Journal: Macroeconomics*, July 2012, vol. 4, no. 3, pp. 96–127.

[7] Social Security benefits withheld because of the earnings test are credited to future benefits. Beneficiaries who are unaware of this provision, however, will view the earnings test as a pure tax, as will individuals who lack liquid assets. See, for example, Gary V. Engelhardt and Anil Kumar. "Taxes and the Labor Supply of Older Americans: Recent Evidence from the Social Security Earnings Test," *National Tax Journal*, June 2014, vol. 67, no. 2, pp. 443–458.

[8] This reform was recommended in John Bailey Jones and Yue Li, "The Effects of Collecting Income Taxes on Social Security Benefits," *Journal of Public Economics*, March 2018, vol. 159, pp. 128–145.

[9] An analysis of the role of labor supply elasticity in Social Security reform can be found in Eric French, "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, April 2005, vol. 72, no. 2, pp. 395–427.