Congressional Budget Office

Background Paper

A Model to Project Flows into and Out of Tax-Deferred Retirement Savings Accounts

Previously Published as CBO Technical Paper 2004-15

October 2004

CONGRESS OF THE UNITED STATES

Abstract

This paper provides the complete specification of a model to project contributions to, investment earnings of, and distributions from tax-deferred retirement savings accounts. The model is used to project taxable pension distributions and IRA withdrawals for the Congressional Budget Office's (CBO's) 10-year baseline. A 75-year version of the model was also used to produce the estimates for the CBO Paper "Tax-Deferred Retirement Savings in Long-Term Revenue Projections." The model is presented in two parts: one covering individual retirement accounts and defined-contribution plans, and the other covering defined-benefit plans. The variables used in the model are defined in an appendix.

Table of Contents

Introduction

This paper is a companion to the Congressional Budget Office (CBO) Paper "Tax-Deferred Retirement Savings in Long-Term Revenue Projections." That document focused on estimating how the flows into and out of tax-deferred retirement plans would affect federal revenues as a percentage of gross domestic product (GDP) over the next 75 years. The model used to generate those projections was described in general terms in Appendix A of that paper.

This paper provides a complete specification of that model, beginning with its primary antecedent (which covered only individual retirement accounts, IRAs) and documenting its subsequent development. The model, which is also used to generate projections of taxable pension and IRA distributions for CBO's semiannual 10-year budget baseline, has two distinct modules: one covering defined-contribution (DC) plans and IRAs, the other covering defined-benefit (DB) plans.

The Defined-Contribution Plan/IRA Module

The DC/IRA module expands upon work done by Sabelhaus on IRA withdrawals.[1] Sabelhaus's IRA model has been enhanced by using the newly available Information Returns Master File (IRMF)—a companion to the annual sample of tax returns prepared by the Statistics of Income (SOI) division of the Internal Revenue Service. The IRA model's

---

[1] See Sabelhaus, John, "Modeling IRA Accumulations and Withdrawals," *National Tax Journal*, vol. 53, no. 4 (December 2000), pp. 865-76.

structure has also been applied to DC plans, and the two models have been integrated. Model results have been calibrated to actual assets and withdrawals between 1998 and 2002.

Review of Sabelhaus's Work

Sabelhaus's model focused on withdrawals from IRAs. In the process of estimating withdrawals, however, Sabelhaus also estimated all the major cash flows into IRAs. He then used those flows to simulate the accumulation of assets within IRAs, and ultimately estimated withdrawals as a percentage of those assets. The data he used to perform those simulations were from a variety of sources covering the year 1995.

Model Mechanics. The accumulation of assets is a mechanical process. From a given starting point, assets increase from one year to the next by (1) receiving a return on the account's investments, and (2) receiving new contributions (including rollovers from employment-based plans). Assets decrease when withdrawals are made.

To make a model of asset accumulation operational, one must be able to estimate investment returns, contributions, rollovers, and withdrawals. Sabelhaus's model was based on the notion that age-specific behavior with respect to contributions and withdrawals is constant over time. Hence, if one knows the percentage of taxpayers of a certain age who contributed to an IRA and how much they contributed in a base year, one can estimate how much taxpayers of that age will contribute in future years. The same principle applies to rollovers and withdrawals. With that in mind, Sabelhaus generated distributions of the following items by one-year age cohort for tax year 1995:[2]

- IRA assets at the beginning of the year (A),
- Total number of taxpayers (T),
- Amount of deductible contributions (C),
- Amount of rollovers received (R), and
- Amount of taxable withdrawals (W).

The model also incorporated several growth parameters. One such parameter, representing growth in the number of tax returns (ω), varied by age. The others, which did not vary by age, included the rate of return on investments (ρ), the growth in average contributions over time (γ), and the growth in average rollovers over time (λ). For each age group (a) in each year (t) after 1995 (t=1), assets were calculated as follows:

1.1 $A(a,t) = A(a\text{-}1,t\text{-}1) + C(a\text{-}1,t\text{-}1) + R(a\text{-}1,t\text{-}1) + I(a\text{-}1,t\text{-}1) - W(a\text{-}1,t\text{-}1)$,

where

1.2 $C(a,t) = T(a,t) * [C(a,1)/T(a,1)] * [1+\gamma(t)]$

1.3 $R(a,t) = T(a,t) * [R(a,1)/T(a,1)] * [1+\lambda(t)]$

1.4 $I(a,t) = A(a,t)*[1+\rho(t)]$

1.5 $W(a,t)= A(a,t) * [W(a,1)/A(a,1)]$, and

---

[2] The basic unit of analysis in Sabelhaus's model was the tax return. For joint returns, the age used to classify returns was that of the primary taxpayer.

1.6 $T(a,t) = T(a-1,t-1) * \omega(a-1,t-1)$

For purposes of measuring the long-term revenue effects of IRAs, equations 1.2, 1.4, and 1.5 generate the flows that affect taxable income.

Data Sources. Whenever possible, Sabelhaus used data from actual 1995 tax returns. Specifically, he used the age distribution from the SOI sample of the number of taxpayers (T), the amount of deductible contributions (C), and the amount of withdrawals (W). He also estimated rollovers received (R) using SOI data on total and taxable pension distributions. He assumed that if taxable pensions were less than 25 percent of gross pensions, then the nontaxable portion represented a rollover to an IRA. In those cases, he set R equal to the nontaxable portion of the pension distribution.

The most conspicuous piece missing from the SOI data was the amount and age distribution of IRA assets (A). Sabelhaus therefore derived his initial distribution of IRA assets from the 1995 Survey of Consumer Finances (SCF). Because the SCF sample was relatively small and was not drawn using the same criteria as the SOI file, he employed a smoothing technique to the data. By doing so, he rendered the two data sets more compatible and avoided having the ripple effects of sampling anomalies affect results throughout the projection period.

The remaining parameters were selected to be consistent with other CBO baseline projections. A projection of the number of returns, for example, is part of the normal process of projecting individual income taxes. At this stage of the model's development, the process was disaggregated to generate growth rates (ω) by filing status for each of four 15-year cohorts. The growth in contributions (γ) and rollovers (λ) was assumed to be driven by wages, which is part of the both the 10-year and long-term economic forecasts generated by CBO.[3] The rate of return on investments (ρ) was assumed to be 7 percent—the interest rate on corporate AAA bonds at the time.

Enhancements to the Sabelhaus Model

CBO enhanced Sabelhaus's model in a variety of ways, by doing the following:

- Using data from information returns (such as Forms W-2, 5498, and 1099-R) that were not available to Sabelhaus,
- Fully accounting for nondeductible contributions and the recovery of basis,
- Using the individual rather than the tax return as the basic unit of analysis, and
- Accounting for IRAs held by nonfilers.

Use of Information Returns. The 1997 IRMF contains useful data from Forms 5498 that were not available to Sabelhaus. For example, the form includes a box reporting rollovers received (R), allowing that aspect of the system to be modeled more precisely. More important, however, are the data on asset levels (A). Those data make it possible to model IRA accumulation using only tax data, thereby eliminating any inconsistency between tax

---

[3] In CBO's 10-year baseline, the growth in IRA contributions is driven by the statutory limit on contributions. Through 2001, the contribution limit was fixed at $2,000 and contributions were assumed to remain flat. Legislation increased the contribution limits to $3,000 in 2002, $4,000 in 2005, and $5,000 in 2008 (higher for those age 50 and above). Consequently, when the contribution limit is scheduled to increase, contributions of those previously contributing the limit are assumed to grow at a rate that reflects the growth in the limit, and all other contributions are assumed to grow with average wages. When the contribution limit is indexed to inflation after 2008, contributions are assumed to increase with inflation.

data and the SCF. The tax data also provide a much larger sample size from which to construct an age distribution than does the SCF. In spite of the increased sample size and greater internal consistency between data sources, however, the raw data still show enough variability from one age group to the next to suggest that the smoothing techniques employed by Sabelhaus remain appropriate.

The use of asset data from Form 5498 requires a structural change to the model, because the form captures end-of-year asset levels instead of the beginning-of-year levels found in the SCF. This requires modifying equation 1.1 to recognize that the end-of-year assets are determined by the current year's contributions, rollovers, and withdrawals instead of the previous year's. Such a modification has the advantage of allowing current-year investment returns to affect current-year withdrawals. Unfortunately, modifying equation 1.1 in that way also creates a simultaneity problem with equation 1.5—current-year assets are a function of current-year withdrawals in the former, while current-year withdrawals are a function of current-year assets in the latter. To avoid this, the enhanced model uses an asset measure ($\tilde{A}$) that disregards current-year withdrawals (that is, $\tilde{A}=A+W$). Equation 1.1 is then rendered as:[4]

$$\text{1.1a} \qquad \tilde{A}(a,t) = \tilde{A}(a-1,t-1) + C(a,t) + R(a,t) + I(a,t) - W(a-1,t-1)$$

In other words, $\tilde{A}$ for the current year takes $\tilde{A}$ from the previous year as its starting point, adds contributions and rollovers from the current year, but subtracts withdrawals from the

---

[4] Throughout the paper, modifications of earlier equations retain the same equation number, but take a lower-case letter indicating the version. Versions a and b specify only those equations that differ substantially from the original. Version c represents a full specification of the final DC/IRA module, including equations that do not differ substantially from the original.

previous year. Thus, assets are no longer dependent on current-year withdrawals. $\tilde{A}$ also replaces A in equation 1.4.

Accounting for Nondeductible Contributions and Basis Recovery. Form 5498 provides data on the total amount of IRA contributions, not just the deductible portion reported on Form 1040. Because nondeductible contributions produce tax-deferred investment returns just as deductible contributions do, it should improve the precision of the model to include those contributions. Furthermore, data from Form 1099-R make it possible to estimate the portion of withdrawals that represent the tax-free recovery of basis (that is, those nondeductible contributions).

To incorporate those items into the model, two new parameters must be estimated from base year values: the nondeductible share of contributions (ϕ), and the portion of rollovers that represents basis (ψ).[5] Instead of C and W (which represent only deductible contributions and taxable distributions), equations 1.1, 1.2, and 1.5 must now use $\hat{C}$ and $\hat{W}$, which represent total contributions and withdrawals.[6] Deductible contributions (C)—the flow that affects taxable income—is thus rendered as:

1.7 $C(a,t) = \hat{C}(a,t)*(1-\phi)$.

To estimate the taxable share of withdrawals, basis (B) must be tracked separately from total assets. That requires the addition of another equation to the model:

---

[5] These parameters were calculated for all taxpayers rather than by age group.

[6] R always represented total rollovers, with the implicit assumption that no basis was included in them. That assumption must now be abandoned.

1.8 $B(a,t) = B(a-1,t-1) + [\hat{C}(a,t)*\phi] + [R(\mathrm{a},t)*\psi] - [\hat{W}(a-1,t-1)-W(a-1,t-1)]$

The taxable share of withdrawals is then assumed to be proportional to the share of assets that were not contributed after tax:

1.9 $W(a,t) = \hat{W}(a,t) * \{1 - [B(a,t)/(\tilde{A}(a,t)]\}$.

Use of Individual Rather than Return Data. A supplemental file provided by SOI contains the age and sex of all taxpayers in an SOI sample between 1987 and 2000. That information allows age distributions to be constructed on an individual basis rather than on a tax-return basis. Thus, secondary taxpayers are classified according to their own age rather than that of the primary taxpayer. The new information also allows parameters to be calculated separately for men and women, which adds an additional dimension to the model (designated with an m for male or an f for female) Furthermore, those two changes allow the redistribution of assets upon death to be incorporated into the model.

To simulate redistributions at death, the model uses age-and-sex-specific mortality tables that are consistent with the Social Security Administration's (SSA's) intermediate population forecast. In each year, the mortality tables are applied to the assets in each age-sex cohort to determine the percentage that should be associated with participants who died during the year (and should therefore be redistributed) and the percentage that should be associated with those who did not die (θ). How assets are redistributed by the model depends on some stylized rules designed to reflect that assets can be redistributed both within a generation and between generations.

- Intragenerational redistributions are represented by married decedents. Assets of such decedents (designated in the model with the subscript j) are assumed to be redistributed to the surviving spouse. Because the average married man is roughly three years older than his wife, assets of married decedents were shifted either three years up or down the age distribution (and to the opposite sex), depending on whether the decedent was male or female.

- Intergenerational redistributions are represented by single decedents (based on the assumption that most such decedents were widowed or divorced with children rather than never married). Assets of such decedents (designated in the model with the subscript s) are assumed to be redistributed to surviving children. The average age of a woman at childbirth is roughly 26, so assets of single female decedents were shifted 26 years down the age distribution, and assets of single male decedents were shifted 29 years down the age distribution.

Implicitly, these rules assume that the intergenerational redistributions of married decedents and the intragenerational redistributions of single decedents (both ignored by the model) offset one another. The same rules hold for redistributing basis.

Assets and basis received by heirs are designated in the model as *D* and *E*, respectively. Because of the different redistribution rules for married and unmarried decedents, the equations representing *D* and *E* must be rendered separately for each marital status. Note that marital status refers to that of the decedent, not of the heir. To implement the separation by marital status, the share of assets held by married taxpayers in 1997 (μ) was calculated and added as an age-and-sex-specific parameter of the model. Equations 1.10 to 1.13 represent the female rendition of those equations.[7]

1.10 $$D(\mathrm{j},\mathrm{f},a,t) = [\tilde{A}(\mathrm{m},a\text{-}1+3,t\text{-}1)\text{-}\hat{\mathrm{W}}(\mathrm{m},a\text{-}1+3,t\text{-}1)] *$$

$$[1\text{-}\theta(m,a\text{-}1+3,t\text{-}1)] * \mu(\mathrm{m},a\text{-}1+3,1)$$

1.11 $$E(\mathrm{j},\mathrm{f},a,t) = [B(\mathrm{m},a\text{-}1+3,t\text{-}1)\text{-}\hat{\mathrm{W}}(\mathrm{m},a\text{-}1+3,t\text{-}1)*B(\mathrm{m},a\text{-}1+3,t\text{-}1)/\tilde{A}(\mathrm{m},a\text{-}1+3,t\text{-}1)] *$$

$$[1\text{-}\theta(\mathrm{m},a\text{-}1+3,t\text{-}1)] * \mu(\mathrm{m},a\text{-}1+3,1)$$

1.12 $$D(\mathrm{s},\mathrm{f},a,t) = [\tilde{A}(\mathrm{f},a\text{-}1+26,t\text{-}1)\text{-}\hat{\mathrm{W}}(\mathrm{f},a\text{-}1+26,t\text{-}1)] *$$

$$[1\text{-}\theta(\mathrm{f},a\text{-}1+26,t\text{-}1)] * [1\text{-}\mu(\mathrm{f},a\text{-}1+26,1)]$$

1.13 $$E(\mathrm{s},\mathrm{f},a,t) = [B(\mathrm{f},a\text{-}1+26,t\text{-}1)\text{-}\hat{\mathrm{W}}(\mathrm{f},a\text{-}1+26,t\text{-}1)*B(\mathrm{f},a\text{-}1+26,t\text{-}1)/\tilde{A}(\mathrm{f},a\text{-}1+26,t\text{-}1)] *$$

$$[1\text{-}\theta(\mathrm{f},a\text{-}1+26,t\text{-}1)] * [1\text{-}\mu(\mathrm{f},a\text{-}1+26,1)]$$

Redistribution of assets to children forces another wrinkle into the model. Unlike retirement assets passed to a spouse, all other retirement assets of decedents must be distributed within five years. Modeling that wrinkle required accounting for assets bequeathed to children (and the associated withdrawals) separately from other assets. $\hat{\text{\c{W}}}$ and $\text{\c{\textit{W}}}$ were therefore created to represent total and taxable withdrawals made under the five-year mandate. The assets and basis from which $\hat{\text{\c{W}}}$ is determined are denoted as $\tilde{\text{\c{\textit{A}}}}$ and $\text{\c{\textit{B}}}$, both of which start out with values of zero in 1997. The two equations representing assets can then be rendered as follows to incorporate redistributions upon death:

1.1b $$\tilde{A}(\mathrm{f},a,t) = \theta(\mathrm{f},a\text{-}1,t\text{-}1) * [\tilde{A}(\mathrm{f},a\text{-}1,t\text{-}1) - \hat{\mathrm{W}}(\mathrm{f},a\text{-}1,t\text{-}1)] +$$

$$\hat{\mathrm{C}}(\mathrm{f},a,t) + R(\mathrm{f},a,t) + I(\mathrm{f},a,t) + \mathrm{D}(\mathrm{j},\mathrm{f},a,t)$$

---

[7] The parallel model for men would, in addition to changing the value of the sex subscripts, change the age shifts in equations 1.10 and 1.11 from +3 to -3, and in equations 1.12 and 1.13 from 26 to 29.

1.14 $\tilde{\textit{\c{A}}}(\mathrm{f},a,t) = \tilde{\textit{\c{A}}}(\mathrm{f},a-1,t-1) - \hat{\text{\c{W}}}(\mathrm{f},a-1,t-1) + D(\mathrm{s},\mathrm{f},a,t).$

Similarly, the two equations representing basis can be represented as:

1.8b $B(\mathrm{f},a,t) = \theta(\mathrm{f},a-1,t-1) * \{B(\mathrm{f},a-1,t-1) - [\hat{\mathrm{W}}(\mathrm{f},a-1,t-1)-W(\mathrm{f},a-1,t-1)]\} +$

$[\hat{\mathrm{C}}(\mathrm{f},a,t)*\phi] + [R(\mathrm{f},a,t)*\alpha] + E(\mathrm{j},\mathrm{f},a,t)$

1.15 $\textit{\c{B}}(\mathrm{f},a,t) = \textit{\c{B}}(\mathrm{f},a-1,t-1) - [\hat{\text{\c{W}}}(\mathrm{f},a-1,t-1)-\textit{\c{W}}(\mathrm{f},a-1,t-1)] + E(\mathrm{s},\mathrm{f},a,t).$

Pulling $\hat{\text{\c{W}}}$ out of $\hat{\mathrm{W}}$ means that the right-hand side of equation 1.5 applies only to the difference between $\hat{\mathrm{W}}$ and $\hat{\text{\c{W}}}$. To fully account for $\hat{\mathrm{W}}$, equation 1.5 must be modified as follows:

1.5b $\hat{\mathrm{W}}(\mathrm{f},a,t) = \tilde{A}(\mathrm{f},a,t)*\{[\hat{\mathrm{W}}(\mathrm{f},a,1)-\hat{\text{\c{W}}}(\mathrm{f},a,1)]/\tilde{A}(\mathrm{f},a,1)\} + \hat{\text{\c{W}}}(\mathrm{f},a,t),$

where

1.16 $\hat{\text{\c{W}}}(\mathrm{f},a,t) = \tilde{\textit{\c{A}}}(\mathrm{f},a,t) * 0.2$

The taxable portion of mandated withdrawals ($\textit{\c{W}}$) mirrors equation 1.9:

1.17 $\textit{\c{W}}(\mathrm{f},a,t) = \hat{\text{\c{W}}}(\mathrm{f},a,t) * \{1 - [\textit{\c{B}}(\mathrm{f},a,t)/\tilde{\textit{\c{A}}}(\mathrm{f},a,t)]\}$

Equations for contributions, rollovers, investment income, and tax returns change only in that a sex subscript is added to the terms that already have an age subscript.

Accounting for Nonfilers. Using tax data for the IRA model provides a much larger sample size than does use of the SCF data, but at the cost of excluding nonfilers from the model. Over the short term, that omission makes virtually no difference. But, in later years, some of those nonfilers could become filers (due to real income growth), and their withdrawals would be overlooked by the model. CBO estimated the number of nonfilers in each age group by subtracting the number of taxpayers from the total population. It then derived an initial distribution of the assets of nonfilers by multiplying the average balance per filer in each age group by the number of nonfilers in that age group. Nonfilers, however, have fewer resources and hence much smaller IRA balances than do filers. Therefore, the initial estimate was scaled back to account only for the difference between assets reported by filers and total IRA assets (as reported in the Federal Reserve Board's Flow of Funds report).

After this change has been made, T should be replaced in equations 1.2, 1.4, and 1.6 by $\check{T}$, which represents the entire population in an age group, not just the filers. Likewise, ω should be replaced in equation 1.5 by $\hat{\omega}$, which represents the growth in population (by age and sex) instead of the growth in returns. Those annual growth factors were fully disaggregated by one-year age cohort instead of by 15-year cohort as before.

Adding Defined-Contribution Plans

Like IRAs, defined-contribution plans involve the accumulation of assets. Withdrawals from such plans are just as likely to be influenced by the amount of those assets as are withdrawals from IRAs. To extend this modeling scheme to DC plans, however, requires drawing on data sources beyond those generated from tax forms. The IRA and DC modules can be integrated by projecting rollovers from one to the other once and then using the projected amount as a distribution in the DC module and a contribution in the IRA module.

Sources of Data. Unlike IRAs, for which all needed information could be extracted from tax and information returns, different sources of data were required for the assets in, contributions to, and withdrawals from DC plans.

Assets. There is no tax form for DC plans analogous to Form 5498. Hence, like the original IRA model, it is necessary to rely on the Survey of Consumer Finances for data on asset levels. The 1998 SCF contains data on beginning-of-year assets, which are comparable to the end-of-year IRA assets available on Form 5498 for 1997. Separate fields cover employer-sponsored DC plans and Keogh plans. Unfortunately, only employed individuals were asked about assets in employer-sponsored plans. The assets of retired people were not included. It was necessary, therefore, to impute those assets.

The imputation relies on the assumption that, for any age group, the ratio of total DC plan assets to the DC plan assets of employed individuals is the same as the corresponding ratio for IRA assets. The age-specific ratios for IRA assets were calculated by dividing assets on Form 5498 between those who are currently employed (that is, reporting wage or self-employment income) and those who are not. The resulting ratios were then applied to the reported DC assets on the SCF. No imputation was made to Keogh assets, because that question was asked of all Keogh owners.

Contributions. Employee contributions to DC plans are reported on Form W-2. Contributions by the self-employed to Keogh plans are reported on Form 1040. Contributions by employers to DC plans, however, are not reported to the IRS in any form that would allow them to be linked to individuals. CBO therefore used other data sources to impute employer contributions.

The data used in this case were from Form 5500, one of which is filed annually by every private tax-deferred retirement plan with more than 100 participants.[8] That form identifies the plan type (DB or DC), and contains data on aggregate contributions by employers and employees separately. All DC plans sponsored by a particular employer (identified by the Employer Identification Number, or EIN) were aggregated, and the ratio of employer contributions to employee contributions was calculated for each. The payer ID number on each W-2 was then matched, when possible, to an EIN associated with one or more 5500s. The ratio of employer contributions to employee contributions calculated from those 5500s was applied to the employee contributions on the W-2 to estimate employer contributions. If no match among the 5500s could be found for a W-2 showing employee contributions, the W-2 was dropped from the sample, and the remaining W-2s with employee contributions were reweighted to compensate.

Some DC plans provide for employer contributions but not employee contributions. In such cases, a ratio of employer to employee contributions was of no use. Instead, a more complicated procedure was used. First, W-2s indicating pension coverage but no employee contributions were linked, whenever possible, to 5500s filed by DC plans with no employee

---

[8] Smaller plans also filed forms, but only one-third of them were required to report financial data in 1997. Another third reported financial data in 1996, and the final third reported financial data in 1995.

contributions. Employer contributions from all the identified 5500s were aggregated and divided by aggregate wages from all the identified W-2s. That ratio was then applied to the wages on each identified W-2 to get an initial estimate of employer contributions. To account for the relative generosity of different employers, the initial estimate for each W-2 was then multiplied by the average contribution per participant from the linked 5500, divided by the average contribution per participant from all identified 5500s. Again, W-2s showing pension coverage but no employee contributions that could not be linked to any 5500s were dropped, and the remaining W-2s with those characteristics were reweighted to compensate.

Withdrawals. Withdrawals from DC plans are reported on Form 1099-R. Unfortunately, they are identified only as coming from a "pension or annuity" without specifying the plan type. A variety of strategies were employed to identify the plan type for as many 1099-Rs as possible.[9] Very generally, government plans (almost all of which are of the defined-benefit type) were identified based on the payer name, which usually contained the name of the sponsoring government. For private plans, some plan types were identified based on key words (such as "defined benefit" or "401k") in the plan name. In cases where the plan type was not clear from the plan name, the payer ID was linked, whenever possible, to the EIN of a 5500. The plan type was then extracted from the 5500. Unfortunately, most payer IDs on Form 1099-R are for plan administrators (usually an investment bank or life insurance company) rather than for the sponsoring employer. Those 1099-Rs could not be linked to any 5500. They were dropped, and the remaining 1099-Rs were reweighted to compensate.

---

[9] For more details on the refining of the 1099-R data, see Congressional Budget Office, *The Taxation of Distributions from Retirement Savings Plans* (April 2004), Appendix A.

Treatment of Rollovers From DC Plans to IRAs. To integrate the DC model with the IRA model, the calculation of rollovers from the former to the latter should be performed only once and used in both models. Such rollovers (referred to hereafter as "internal" rollovers) are considered distributions from DC plans, so the model handles them just as it does taxable distributions from DC plans. At the same time, both IRAs and DC plans receive rollovers from other types of qualified plans ("external" rollovers). Those are treated the same as rollovers were treated in the Sabelhaus model; that is, as contributions. Accommodating both treatments requires adding two subscripts to R: q (representing the source plan type), and p (representing the destination plan type). Likewise, a subscript identifying the source plan type must be added to the portion of rollovers that is basis (ψ). The model recognizes values b (for private defined-benefit plans), g (for state and local government plans), k (for defined-contribution plans) and i (for individual retirement accounts) for p and q. The internal rollover equations corresponding to the female half of the model for DC plans and IRAs are as follows:

1.18 $R(\mathrm{k},\mathrm{i},\mathrm{f},a,t) = \tilde{A}(\mathrm{k},\mathrm{f},a,t) * [R(\mathrm{k},\mathrm{i},\mathrm{f},a,1)/\tilde{A}(\mathrm{k},\mathrm{f},a,1)]$, and

1.19 $R(\mathrm{i},\mathrm{k},\mathrm{f},a,t) = -R(\mathrm{k},\mathrm{i},\mathrm{f},a,t)$.

The full specification of the DC/IRA module for females is completed as follows, where q' has the same value (i or k) as p, and p' has the same value as q (which will always be the opposite of the value of p).

1.1c $\tilde{A}(p,\mathrm{f},a,t) = \theta(\mathrm{f},a-1,t-1) * [\tilde{A}(p,\mathrm{f},a-1,t-1) - \hat{W}(p,\mathrm{f},a-1,t-1) - R(q',p',\mathrm{f},a-1,t-1)] +$

$$\hat{C}(p,\mathrm{f},a,t) + R(\mathrm{b},p,\mathrm{f},a,t) + R(\mathrm{g},p,\mathrm{f},a,t) + I(p,\mathrm{f},a,t) + D(p,\mathrm{j},\mathrm{f},a,t)$$

1.2c $\hat{\mathrm{C}}(p,\mathrm{f},a,t) = \check{\mathrm{T}}(p,\mathrm{f},a,t) * [\hat{\mathrm{C}}(p,\mathrm{f},a,1)/\check{\mathrm{T}}(p,\mathrm{f},a,1)] * [1+\gamma(t)]$

1.3c $R(\mathrm{b},p,\mathrm{f},a,t)+R(\mathrm{g},p,\mathrm{f},a,t)=\check{\mathrm{T}}(p,\mathrm{f},a,t)*\{[R(\mathrm{b},p,\mathrm{f},a,1)+R(\mathrm{g},p,\mathrm{f},a,1)]/\check{\mathrm{T}}(p,\mathrm{f},a,1)\}*[1+\lambda(t)]$

1.4c $I(p,\mathrm{f},a,t) = \tilde{A}(p,\mathrm{f},a,t)*[1+\rho(p,t)]$

1.5c $\hat{\mathrm{W}}(p,\mathrm{f},a,t)= \tilde{A}(p,\mathrm{f},a,t) * \{[\hat{\mathrm{W}}(p,\mathrm{f},a,1)-\hat{\text{\c{W}}}(p,\mathrm{f},a,1)]/\tilde{A}(p,\mathrm{f},a,1)\} + \hat{\text{\c{W}}}(p,\mathrm{f},a,t)$

1.6c $\check{\mathrm{T}}(p,\mathrm{f},a,t) = \check{\mathrm{T}}(p,\mathrm{f},a-1,t-1) * \hat{\omega}(\mathrm{f},a-1,t-1)$

1.7c $C(p,\mathrm{f},a,t) = \hat{\mathrm{C}}(p,\mathrm{f},a,t)*(1-\phi).$

1.8c $B(p,\mathrm{f},a,t) = \theta(\mathrm{f},a-1,t-1) *$

$\{B(p,\mathrm{f},a-1,t-1)-[\hat{\mathrm{W}}(p,\mathrm{f},a-1,t-1)-W(p,\mathrm{f},a-1,t-1)]-R[(q',p',\mathrm{f},a-1,t-1)*\psi(q')]\} +$

$[\hat{\mathrm{C}}(p,\mathrm{f},a,t)*\phi] + [R(\mathrm{b},p,\mathrm{f},a,t)+R(\mathrm{g},p,\mathrm{f},a,t)]*\psi(\mathrm{b}) + E(p,\mathrm{j},\mathrm{f},a,t)$

1.9c $W(p,\mathrm{f},a,t) =[\hat{\mathrm{W}}(p,\mathrm{f},a,t) * B(p,\mathrm{f},a,t)/(\tilde{A}(p,\mathrm{f},a,t)]$

1.10c $D(p,\mathrm{j},\mathrm{f},a,t) = [\tilde{A}(p,\mathrm{m},a-1+3,t-1)-\hat{\mathrm{W}}(p,\mathrm{m},a-1+3,t-1)] *$

$[1-\theta(\mathrm{m},a-1+3,t-1)] * \mu(\mathrm{m},a-1+3,1)$

1.11c $D(p,\mathrm{s},\mathrm{f},a,t) = [\tilde{A}(p,\mathrm{f},a-1+26,t-1)-\hat{\mathrm{W}}(p,\mathrm{f},a-1+26,t-1)] *$

$[1-\theta(\mathrm{f},a-1+26,t-1)] * [1-\mu(\mathrm{f},a-1+26,1)]$

1.12c $E(p,\mathrm{j},\mathrm{f},a,t) = \{\mathrm{B}(p,\mathrm{m},a-1+3,t-1)- [\hat{\mathrm{W}}(p,\mathrm{m},a-1+3,t-1)-W(p,\mathrm{m},a-1+3,t-1)]\} *$

$[1-\theta(\mathrm{m},a-1+3,t-1)] * \mu(\mathrm{m},a-1+3,1)$

1.13c $E(p,\mathrm{s},\mathrm{f},a,t) = \{B(p,\mathrm{f},a-1+26,t-1)-[\hat{\mathrm{W}}(p,\mathrm{f},a-1+26,t-1)-W(p,\mathrm{f},a-1+26,t-1)]\}*$

$[1-\theta(\mathrm{f},a-1+26,t-1)] * [1-\mu(\mathrm{f},a-1+26,1)]$

1.14c $\text{\c{\textit{A}}}(p,\mathrm{f},a,t) = \text{\c{\textit{A}}}(p,\mathrm{s},\mathrm{f},a-1,t-1) - \hat{\text{\c{W}}}(p,\mathrm{f},a-1,t-1) + D(p,\mathrm{s},\mathrm{f},a,t)$

1.15c $\text{\c{\textit{B}}}(p,\mathrm{f},a,t) = \text{\c{\textit{B}}}(p,\mathrm{s},\mathrm{f},a-1,t-1) - [\hat{\text{\c{W}}}(p,\mathrm{f},a-1,t-1)-\text{\c{\textit{W}}}(p,\mathrm{f},a-1,t-1)] + E(p,\mathrm{s},\mathrm{f},a,t)$

1.16c $\hat{\text{\c{W}}}(p,\mathrm{f},a,t) = \tilde{\text{\c{\textit{A}}}}(p,\mathrm{f},a,t) * 0.2$

1.17c $\text{\c{\textit{W}}}(p,\mathrm{f},a,t) = \hat{\text{\c{W}}}(p,\mathrm{f},a,t) * \text{\c{\textit{B}}}(p,\mathrm{f},a,t)/\tilde{\text{\c{\textit{A}}}}(p,\mathrm{f},a,t)$

Calibrating the Model

There are three sources of data that can be used to test and calibrate the model over the period 1998-2002: (1) nontaxable pension distributions from the SOI, (2) IRA and DC plan asset levels from the Flow-of-Funds,[10] and (3) taxable IRA withdrawals from the SOI. Nontaxable pension distributions are mostly rollovers to IRAs, so their levels can be used to determine the extent to which rollover behavior in 1997 was repeated in later years. The asset levels make it possible to retroactively infer an average rate of return on those assets and an underlying portfolio mix consistent with that rate. The taxable IRA distribution amounts make it possible to determine the extent to which participants' rollover and withdrawal behaviors in 1997 were repeated in later years.

Calibrating Rollover Behavior. SOI data indicate that the propensity of pension owners to roll their lump-sum distributions over into IRAs increased steadily between 1997 and 2000. Over that period, taxable pension distributions increased by 8 percent per year, while nontaxable distributions increased by 23 percent per year. That difference in growth rate makes it untenable to assume that rollover behavior in 1997 was repeated in later years. To match the actual growth in rollovers, the percentage of DC plan assets in each age class rolled over to IRAs was increased by 20 percent in 1998 and 40 percent in 1999 and 2000. In 2001, rollovers declined, so a 30 percent increase over 1997 levels was sufficient (see Table 1). The best adjustment going forward is not obvious; the adjustment actually used was 20 percent.

---

[10] The DC plan asset levels must be adjusted to account for 403(b) plans, which are covered by the model but are generally not considered private DC plans in the Flow of Funds. Asset levels for 403(b) plans for 1997 were estimated using data from the *1998 Life Insurance Factbook* (American Council of Life Insurance: Washington, D.C., 1998) and the *1999 Mutual Fund Factbook* (available on line at http://www.ici.org/pdf/factbooks2.html#1999%20Fact%20Book). For 1998-2001, 403(b)s were assumed to constitute the same share of DC plan assets as in 1997.

Table 1.
Calibration of Rollover Behavior

	Rollovers from DC Plans to IRAs *(billions of dollars)*		
	Estimated from SOI Data[1]	**Estimated by Model Using 1997 Parameters**	**Estimated by Model Using Adjusted Parameters**[2]
1998	107.8	87.7	105.2
1999	136.6	97.9	136.1
2000	151.5	109.1	149.4
2001	130.1	102.5	128.6

Source: Congressional Budget Office

1. Rollovers estimated to be 67 percent of nontaxable pension distributions
2. Parameters for 1997 increased by 20 percent in 1998, 40 percent in 1999 and 2000, and 30 percent in 2001.

Identifying the Typical Portfolio. Both the IRA and DC models were run with a variety of values for the rate of return on investments (ρ) to determine which value in each year resulted in end-of-year assets that most closely approximated the aggregate targets from the Flow-of-Funds. The results varied from a high of 21.71 percent for IRAs in 1998 to a low of -14.53 percent for DC plans in 2003 (see Table 2).

Table 2.
Inferred Rates of Return for IRAs and DC Plans, 1998-2003

	IRAs *(percent)*	**DC Plans** *(percent)*
1998	21.71	9.37
1999	18.95	8.90
2000	-4.10	-0.68
2001	-5.94	-8.37
2002	-10.36	-14.53
2003	19.05	17.46

Source: Congressional Budget Office

Different stock/bond portfolio mixes were then tested, to determine which mix produced combined rates of return that most closely approximated those inferred from the reported annual asset levels. Returns on stock were received in two forms: as capital gains and as dividends. Returns on bonds were received as interest. Capital gains returns were represented by the change in the S&P 500 (see Table 3). Dividend rates were calculated as dividends in personal income (from the National Income and Product Accounts) divided by stocks held by households (derived from Flow-of-Funds data).[11] Interest rates were calculated as monetary interest in personal income (from the National Income and Product Accounts) divided by interest-bearing assets directly held by households (derived from Flow-of-Funds data).[12] For IRAs, the mix that came closest to matching

---

[11] These data are derived specifically from line 6 from Flow-of-Funds Table B.100.e (equity shares at market value) minus line 12 from the same table (portion held by private defined-benefit plans). Equities held by private defined-benefit plans are excluded because dividends paid to such plans are counted as imputed interest rather than as personal dividends in the National Income and Product Accounts.

[12] The assets may be found in the following locations in the Flow-of-Funds data:

Asset type	Location in the Flow-of-Funds data
Time and savings deposits	Table B.100 line 12
Money market fund shares	Table B.100 line 13
Credit market instruments	Table B.100 line 14

actual asset levels (as reported in the March 2004 release of the Flow-of-Funds) was 75 percent stocks and 25 percent bonds. For DC plans, the most accurate mix was 50 percent stocks and 50 percent bonds. These mixes were then used to project overall rates of return for each type of plan in years beyond 2003.

Table 3.
Actual Rates of Return on Stocks and Interest-Bearing Assets, 1998-2003

	Rate of Return on Stocks			
	as Capital Gains *(percent)*	**as Dividends** *(percent)*	**Total** *(percent)*	**Effective Interest Rate** *(percent)*
1998	23.27	2.76	26.03	5.26
1999	12.70	2.10	14.81	5.64
2000	-10.09	2.65	-7.44	6.03
2001	-11.30	3.05	-8.25	6.40
2002	-24.00	4.32	-19.68	6.26
2003	31.65	3.54	35.19	5.87

Source: Congressional Budget Office, based on the National Income and Product Accounts and Flow-of-Funds

The 50/50 mix for DC plans does not correspond closely to the breakdown of assets by type provided by the Flow-of-Funds, which indicate a higher percentage in stocks.[13] It is possible that many DC plans are invested in relatively conservative stocks, which, in a volatile market, would make their rates of return look more like those of bonds. The Flow-of-Funds does not provide the detail necessary to estimate the percentage of IRA assets in stocks and compare it with the imputed 75/25 mix.

Security credit	Table B.100 line 26
Investment in bank personal trusts, excluding equity portion	Table B.100 line 29 minus Table B.100.e line 9
Mutual funds, excluding equity portion	Table B.100 line 25 minus Table B.100.e line 16

[13] See Flow-of-Funds Table L.119.c.

Constraining Withdrawal Behavior in Extreme Circumstances. Even after the model approximately replicates IRA and DC plan asset levels between 1998 and 2003, the withdrawal percentages calculated from 1997 data do not generate IRA withdrawals for 1998 to 2003 that are particularly close to actual taxable withdrawals reported by SOI.[14] In 1999, when the rate of return was unusually high, withdrawals were less than the model predicted; in 2000, when the rate of return was unusually low, they were more than predicted (see Table 4). It appears that participants engaged in some mental smoothing of their asset accumulation when deciding how much to withdraw. In fact, if assets in IRAs had yielded an 18 percent rate of return in the good years and a 10 percent rate of return in the bad years, the model would have generated approximately the correct level of taxable IRA withdrawals in each year.

Data for 2001, however, indicate that the mental smoothing may have run its course for that cycle and that by then the prolonged slump in stock prices had begun to affect withdrawal rates. In that year, using the inferred rate of return of -5.94 percent, the model came close to generating the correct amount of IRA withdrawals. In contrast, the extremely poor return in 2002 did not have the same effect on withdrawals; a 2.5 percent rate of return in that year would have generated the correct level of taxable IRAs. In 2003, the actual rate of return was once again above 18 percent. No actual data were available to which to calibrate, so the 18 percent that worked well in other good years was applied in 2003 as well.

---

[14] The SOI does not break down taxable pension distributions between DB and DC plans, so there is nothing with which to compare withdrawals generated by the DC model. The SOI figures for taxable IRA distributions for 1998 to 2001 were adjusted to account for Roth conversions. Most participants converting from a traditional IRA to a Roth IRA in 1998 were permitted to spread the resulting tax liability out over four years. Conversions that were immediately taxable also occurred in each of the four years. Because the model does not cover Roth conversions, those amounts were subtracted from the published totals before being compared with model outputs.

In later years, the projected rates of return are all in the range that should require no mental smoothing. In the absence of any control totals for taxable withdrawals from DC plans, the same assumptions regarding mental smoothing were used for DC plans as for IRAs.

Table 4.
Taxable IRA Withdrawals under Different Behavioral Assumptions, 1998-2002

	Actual[1] *(billions of dollars)*	**Modeled** *(billions of dollars)* **No Mental Smoothing**	**Smoothing from 1998-2002**	**Circumstantial Smoothing**
1998	64.77	62.68	60.93	60.93
1999	77.51	80.15	77.65	77.65
2000	89.05	80.41	88.80	88.80
2001	84.08	78.89	100.33	86.56
2002	89.73	72.82	111.11	90.26

Source: Congressional Budget Office

1. From SOI, excluding taxable amounts converted to Roth IRAs

The Defined-Benefit Plan Module

The DB module covers the defined-benefit plans of private firms and of state and local governments. It also includes annuities offered by life insurance companies, most of which are not "qualified" and hence do not receive the same tax benefits as employment-based plans. The DB module does not cover plans administered by the federal government. Instead, CBO uses projections generated by the administering agencies, modified to conform to CBO's economic forecast. All results are calibrated to be consistent with actual taxable pensions and annuities reported since 1997.

Distributions to Retirees of Private Firms and of State and Local Governments

The model to project distributions from selected defined-benefit plans can best be understood in two parts. The first part projects forward the benefits of "incumbent beneficiaries"—in other words, those who were receiving benefits in 1997, the year from which the model's data and parameters are drawn. The second part identifies those who will retire in future years, assigns benefit amounts to them, and projects those amounts forward.

Incumbent Beneficiaries. The simplest aspect of the model to visualize is the treatment of those who were already retired and receiving benefits from a DB plan in 1997. The first step, therefore, is to identify those people. The model used as a starting point the taxpayers in the SOI sample that received one or more Forms 1099-R reporting pension or annuity (as opposed to IRA) distributions. The strategies described above for identifying withdrawals from DC plans were also used to identify distributions from DB plans.

From the reweighted subsample of 1099-Rs identified as coming from DB plans, the number of recipients of distributions (U) from private DB plans and from state and local plans in each age-sex cohort were tabulated separately. A smoothing technique was applied to these results to avoid the ripple effects of sampling anomalies as the model moves the cohort through time. In each year after 1997, the model applies the survival rate (θ) based on age- and-sex-specific mortality tables that are consistent with the SSA's intermediate population forecast. The result is a projection of the number of DB plan beneficiaries in 1997 who would still be receiving benefits in each subsequent year. For each plan type (p), sex (s), and age group (a) in each year (t) after 1997 (t=1), the number of beneficiaries was calculated as follows:

$$2.1 \qquad U(p,s,a,t) = U(p,s,a\text{-}1,t\text{-}1) * \theta(p,s,a\text{-}1,t\text{-}1)$$

The nature of defined-benefit plans is to maintain a certain level of benefits over the life of a recipient. The only deviation from that are plans that are indexed to the consumer price index (Y). Private DB plans are rarely indexed for inflation (χ=0.0), so gross distributions ($\hat{H}$) from such plans were calculated by assigning the 1997 level of benefits to the number of survivors in each year. Approximately half of state and local plans are indexed for inflation (χ=0.5), but, otherwise, distributions from such plans are calculated in the same way:

$$2.2 \qquad \hat{H}(p,s,a,t) = U(p,s,a,t) * \hat{H}(p,s,a,1)/U(p,s,a,1) \; * \{1 + \chi(p)*[Y(t)/Y(1) - 1]\}$$

Amounts rolled over to IRAs and DC plans (R) were then estimated based on age-and-sex-specific parameters that were in turn estimated using the 1997 data:

$$2.3 \qquad R(q,p,s,a,t) = \hat{H}(p',s,a,t) * R(q,p,s,\mathrm{a},1)/\hat{H}(p',s,a,1)$$

That equation (rendered once each for $q=p'=b$ and $q=p'=g$) also replaces equation 1.3 in the DC model.

Taxable distributions (W) can then be calculated by subtracting rollovers and adjusting for basis recovery, using the age-and-sex-specific parameters from 1997 (shown in equation 2.4, where q' has the same value—b or g—as p).

$$2.4 \qquad W(p,s,a,t) = [\hat{H}(p,s,a,t)-R(q',i,s,a,t)-R(q',k,s,a,t)] * \{W(p,s,a,1)/[\hat{H}(p,s,a,1)-R(q',i,s,a,1)-R(q',k,s,a,1)]\}$$

Later Retirees. Of course, those who were receiving benefits in 1997 are not the only ones who will receive benefits in each subsequent year. More people will retire, and they will receive a different level of benefits than is enjoyed by incumbent retirees. Reflecting the past growth in wages, that benefit level will be higher for each class of retirees. Identifying the new retirees and estimating the level of benefits they will receive adds an additional layer of complexity to the model.

Retirees are deemed to belong to a "class" consisting of everybody who retired in the same year. To distinguish each class of retirees, it is necessary to add a dimension (c) to the model. For those who were receiving benefits in 1997, c has a value of 1. A value of 2 represents those who retired in 1998, and so on. By pairing c and t, it becomes possible to denote benefits received by retirees in class c in each subsequent year t. Note that the value of t can never be less than the value of c. For example, a (c,t) pair of (2,3) denotes the benefits received in 1999 by those who retired in 1998. A (c,t) pair of (3,2), however, is not allowable, because it would denote benefits received in 1998 by those who did not retire until 1999. With that in mind, equation 2.1 can be rewritten as follows for every allowable (c,t) pair (except when c=t).[15]

$$2.1a \qquad U(p,s,a,c,t) = U(p,s,a-1,c,t-1) * \theta(p,s,a-1,t-1)$$

---

[15] Equations 2.1a and 2.5 do not work when c=t because that would result in c being paired with t-1 on the right-hand side of the equation, which, in turn, implies the aforementioned anomaly of collecting benefits prior to retirement.

To identify new retirees, it was necessary to track the number of active participants who could potentially retire (V). The process for identifying active participants in 1997 was similar to the process for identifying benefit recipients. Instead of Form 1099-R, however, Form W-2 was used because it contains a box identifying active participants in an employer-sponsored retirement plan. Unfortunately, the type of plan is not specified on Form W-2, so plan names, key words, and links to Forms 5500 were used to create a subsample of W-2s containing identifiable active participants in private DB plans and state and local retirement plans. That subsample was then reweighted to represent all such participants.[16]

At the next stage, the model estimates the number of active participants who will retire in each year. At this point, however, it is sufficient to establish that the same survival rates that apply to plan beneficiaries apply to active participants. Hence, for every allowable (c,t) pair (except when c=t), the number of surviving active participants can be calculated as:

$$V(p,s,a,c,t) = V(p,s,a\text{-}1,c,t\text{-}1) * \theta(p,s,a\text{-}1,t\text{-}1) \tag{2.5}$$

Having estimated the number of active and retired participants associated with each allowable (c,t) pair in which c is greater than t, it becomes possible to estimate the number of retirees associated with the (c',t) pair in which c' is equal to t—in other words, the number of new retirees in each year. To accomplish this in a particular year t, the combined number of active and retired participants in an age-sex cohort (without regard to when they retired or began participating) is first estimated by applying an age-and-sex-specific population growth rate ($\hat{\omega}$) from the SSA intermediate population forecast to the

---

[16] For more details on the refining of the W-2 data, see Congressional Budget Office, *Utilization of Tax Incentives for Retirement Saving* (August 2003), Appendix A.

corresponding number in that cohort in the previous year (t-1). For example, the number of 65-year-old male participants in 1998 was estimated by applying the growth rate for that age-sex cohort to 65-year-old male participants in 1997. The combined number of participants is then split between those who were active and those who were retired by applying the percentage of participants in that age-sex cohort who were retired in 1997. From the resulting number of retired participants, the number determined to have retired in previous years (found using equation 2.1a for all allowable values of c) was subtracted. The difference represents the number of new retirees in that year:

$$2.6 \qquad U(p,s,a,c',t) = \{\sum_{c=1}^{t-1} [V(p,s,a,c,t-1) + U(p,s,a,c,t-1)] * \hat{\omega}(s,a-1,t-1) * U(p,s,a,1,1)/[V(p,s,a,1,1)+U(p,s,a,1,1)]\} - \sum_{c=1}^{t-1} U(p,s,a,c,t)$$

Because they earned higher wages over the course of their careers, new retirees receive higher pension benefits than those already retired in 1997. To adjust for that, the change in average terminal wages (Q) between the year in which the average recipient in 1997 retired (assumed to be 1990) and the year of retirement for new retirees must be added to equation 2.2, as follows:[17]

$$2.2a \qquad \hat{H}(p,s,a,t) = \sum_{c} \{[U(p,s,a,c,t)*\hat{H}(p,s,a,1)/U(p,s,a,1,1)] * Q(c)/Q(1) * \{1 + \chi(p)*[Y(t)/Y(c) - 1]\}\}$$

---

[17] By denoting everybody who was retired in 1997 with a single value of c, the result is a discontinuity between c=1 and c=2 with respect to average wages (Q). The former refers to average wages in 1990, the year in which the average recipient in year 1 (1997) retired. The latter refers to average wages in year 2 (1998).

Terminal wages refer to the wages received just prior to retirement. Each plan can have its own definition of terminal wages. Typically, private plans use a five-year period and state and local plans use a three-year period. To approximate the actual rules, the terminal earnings for those retiring in a given year were assumed to be the average annual wage over the previous three or five years, depending on the type of plan. For example, the terminal wages of somebody retiring from a private plan in 2003 would be adjusted by the average annual wage between 1998 and 2002 (the five years before that person's retirement) divided by the average annual wage between 1985 and 1989 (the five years before the retirement of the average 1997 pension recipient).

Annuity Payments by Life Insurance Companies

In most respects, the model treats annuity payments by life insurance companies the same way as benefits paid by private plans or state and local plans. The key difference between annuity payments by life insurance companies and benefits paid by employment-based plans is the nature of the data on which the projections are based. The life insurance companies making such payments also administer employment-based plans, so it is impossible to tell whether a distribution made by a life insurance company was for an individual annuity or was made on behalf of an employment-based plan. Hence, the reweighted subsample used to create the data and parameters for private plans and state and local plans contained no records representing annuity payments by life insurance companies. Because of that omission, however, the distributions on the reweighted subsample did not add up to the same totals as the original data base. The residual was therefore assigned to represent annuity payments by life insurance companies.

The residual calculation works for dollar amounts, but does not work for either the number of active or retired participants. Because people can participate in more than one type of plan, the number of participants in each type of plan does not sum to the number of taxpayers participating in any plan. Hence, the residual is meaningless. The numbers of active and retired participants, however, are key components of the model, and cannot simply be omitted. Fortunately, it is less important that the absolute number of participants in an age-sex cohort be correct than it is that the distribution of participants among age-sex cohorts be correct. With that in mind, an arbitrary number of active participants was distributed among age-sex cohorts in proportion to the number of filers in each class. An estimate of the number of annuity recipients in each class was derived by dividing gross annuity payments by 80 percent of the average distribution amount (regardless of type of plan) in that class. That derivation assumed that average annuity distributions vary among age-sex cohorts in the same way as average distributions of all other types of plans. Once the number of active and retired participants was in place, the model could operate as described above.

Contributions to and Investment Income within Defined-Benefit Plans

The nature of defined-benefit plans makes it unnecessary to project contributions or investment income in order to project distributions. The contributions and investment income nevertheless have tax consequences, so the model includes a section to estimate those amounts. The fundamental principle followed in generating those projections was to ensure that all of the projected benefits would be fully funded. Hence the model estimates the additional liability incurred (Z) by the plan in each year, and sets the sum of contributions and investment income equal to that amount.

For an actual DB plan, the additional liability incurred is calculated by using a complicated actuarial formula that takes into account (among other things) the ages of participants, their life expectancies, past and projected earnings, lengths of service, vesting statuses, and expected job mobility. In theory, liability would increase each year, even if none of the participants retired. This model cannot duplicate those calculations, however. As a cohort-based model derived from a snapshot of a single year, it contains no information on past earnings or length of service. Therefore, an alternative measure of the annual liability increment was used—specifically, the discounted amount of lifetime benefits that would be paid to participants retiring in that year. The calculation was performed over all age-sex cohorts as shown in equation 2.7, where δ is the discount rate.

2.7 $$Z(p,c) = \sum_{s} \sum_{a} \sum_{t=c}^{t|a=89} \hat{H}(p,s,a,t)/(1+\delta)^{t}$$

Investment income was calculated by applying a rate of return that reflects the typical portfolio of the plan type to beginning-of-year assets:

2.8 $$I(p,t) = A(p,t) * \rho(p,t)$$

The remainder of the liability increment is assumed to be covered by contributions (which are implicitly assumed to occur at the end of the year):

2.9 $$\hat{C}(p,t) = Z(p,c') - I(p,t)$$

Asset levels can then be advanced one year:

$$2.10 \qquad A(p,t) = A(p,t\text{-}1) + \hat{C}(p,t) + I(p,t) - \sum_{s} \sum_{a} \hat{H}(p,s,a,t)$$

Isolating deductible contributions is a bit trickier. First, the nontaxable liability increment (K) must be calculated for each class in a manner similar to the calculation of Z, but excluding taxable distributions and rollovers:

$$2.11 \qquad K(p,c) = \sum_{s} \sum_{a} \sum_{t=c}^{t|a=89} [(1\text{-}\{[R(q',i,s,a,1)+R(q',k,s,a,1)]/\hat{H}(p,s,a,1)\}) * (1\text{-}\{W(p,s,a,1)/[\hat{H}(p,s,a,1)\text{-}R(q',i,s,a,1)\text{-}R(q',k,s,a,1)]\}) * \hat{H}(p,s,a,t)]$$

Because no investment returns accrue to nontaxable assets, the value of K associated with a retirement class (c') becomes the value of the nondeductible contributions that must be made in the same year (t). Those contributions are subtracted from gross contributions (Ĉ) to get deductible contributions:

$$2.12 \qquad C(p,t) = \hat{C}(p,t) - K(p,c') \,.$$

Distributions by Plans Administered by the Federal Government

Projections of distributions from federal civilian and military pensions and from the Railroad Retirement Account were based on those generated by the administering agencies.[18] Each agency has its own forecasting model and uses its own assumptions concerning employment levels, average wage growth, and inflation. The agencies'

---

[18] Plans are administered by the following agencies: The Office of Personnel Management administers federal civilian pensions. The Department of Defense, Defense Finance and Accounting Service, administers military pensions. The Railroad Retirement Board administers the Railroad Retirement Account.

assumptions for employment and average wage growth were retained, but their projections were adjusted to be consistent with the CBO forecast of the consumer price index.

On average, federal government pensions are larger than their private-sector counterparts. Based on CBO tabulations of Forms 1099-R, the average recipient of a federal pension receives a benefit approximately 50 percent larger than the average recipient of a private DB pension does. Some of that is because the largest civilian pension plan, the Civil Service Retirement System (CSRS) operates outside the Social Security system. Its benefits are higher than average in part because they must make up for the absence of Social Security benefits. Furthermore, federal retirement benefits are indexed to inflation, while most private sector benefits are not—a factor that would tend to increase the importance of federal benefits over time. In 1997, pensions paid by federal agencies accounted for nearly 18 percent of taxable retirement plan distributions. Nevertheless, distributions from federal agencies are projected to decline sharply as a percentage of GDP over the next 75 years, dropping from 0.81 percent in 2003 to 0.13 percent in 2078. That reduction is attributable almost entirely to a forecasted decrease in federal and railroad employment. By 2078, federal plan payments will account for less than 2 percent of taxable retirement distributions.

Calibrating the Model

Rates of return for each type of plan were estimated using the same method used for IRAs and DC plans. Stocks were estimated to constitute 62 percent of assets in private plans, 40 percent of assets in state and local plans, and 20 percent of assets in nonqualified annuities.[19]

The only figures with which to compare model results for taxable distributions are those reported by the SOI between 1998 and 2002. Unfortunately, those figures include withdrawals from defined-contribution plans as well as distributions from all of the sources covered by the DB module. Methods of calibrating the DC module results are discussed above. Projections of distributions from plans administered by the federal government are assumed to have been calibrated by the administering agency and are not further adjusted. The final calibration therefore was designed to scale the remaining sources of distributions (i.e., private, state and local, and nonqualified) such that, when added to the calibrated totals for DC plans and federal plans, they equaled the reported amount of taxable pensions and annuities. Between 1998 and 2001, a scale factor of 1.026 (or 2.6 percent) was required. In 2002, however, the model output did not have to be scaled to hit the published totals. Therefore, no scale factor was applied to projected figures.

---

[19] As with DC plans and IRAs, the portfolio splits for private plans and for state and local plans were tested by applying the weighted average rates of return associated with each split to 1997 assets, adjusting for contributions and withdrawals, and comparing the results to actual reported assets in 1998 through 2001. The splits selected were those that resulted in the best approximations of actual assets in all four years. The share of nonqualified annuity assets invested in stocks is based on amounts from the mid-1990s reported in Table 6.9 of the *1998 Life Insurance Factbook* (American Council of Life Insurance: Washington, D.C., 1998).

Appendix: Definitions of Variables

Variables Denoting Amounts

A	Beginning-of-year assets
$\tilde{A}$	End-of-year assets + withdrawals, or Beginning-of-year assets + all sources of income
$\underset{,}{\tilde{A}}$	Portion of $\tilde{A}$ subject to five-year mandatory withdrawal (DC only)
B	End-of-year basis + amount recovered during year, or Beginning-of-year basis + nondeductible contributions
$\underset{,}{B}$	Portion of B subject to five-year mandatory withdrawal (DC only)
C	Deductible contributions
$\hat{\mathrm{C}}$	Total contributions (deductible and nondeductible)
D	Assets inherited (DC only)
E	Basis inherited (DC only)
$\hat{\mathrm{H}}$	Gross distributions (total withdrawals + rollovers out)
I	Investment income
K	Nontaxable annual liability increment (DB only)
Q	Average terminal wages
R	Rollovers
T	Number of taxpayers
$\check{\mathrm{T}}$	Number of people (taxpayers and nontaxpayers)
U	Number of retirees receiving benefits (DB only)
V	Number of workers actively participating in a retirement plan (DB only)
W	Taxable withdrawals
$\underset{,}{W}$	Portion of W subject to five-year mandate (DC only)
$\hat{\mathrm{W}}$	Total withdrawals (taxable and basis recovery)
$\underset{,}{\hat{\mathrm{W}}}$	Portion of $\hat{\mathrm{W}}$ subject to five-year mandate (DC only)
Y	Consumer price index (CPI)
Z	Annual liability increment (DB only, taxable and nontaxable)

Growth Rates and Model Parameters

γ	Rate of growth in average contributions (DC only)
λ	Rate of growth in average rollovers (DC only)
ρ	Rate of return on investments
ω	Rate of growth in number of taxpayers
$\hat{\omega}$	Rate of population growth
θ	Survival rate
ϕ	Nontaxable share of contributions
ψ	Nontaxable share of rollovers
μ	Percentage of the population that is married
χ	Percentage of plans that are indexed for inflation

Subscripts and their Values

a Age

s Sex
m = Male
f = Female

Marital status
s = Single
j = Married (joint filing status)

t Year

c Class (year of retirement, DB only)
$c' = t$

p Plan type (destination plan in the case of rollovers)
q Plan type originating a rollover
i = Individual retirement account
k = Employment-based defined-contribution plan
b = Private defined-benefit plan
g = State or local government defined-benefit plan
$p' = q$
$q' = p$