Automatic IRA Reform: How Much Will Retirement Income Go Up?

Martin R. Holmer, Policy Simulation Group (PSG)*

October 18, 2012

The only available answer to this question — from a 2011 study by Butrica and Johnson that uses the Urban Institute DYNASIM microsimulation model[1] — is that after-tax per-adult family income at age seventy will rise on average by three to five percent after a lifetime of experience with automatic IRAs. And the average percentage rise would be from six to twelve percent for those in the bottom half of the pre-reform age-seventy family income distribution (Butrica-Johnson, Table 2).

Are these Butrica-Johnson estimates plausible? It is difficult to answer that question because they do not explain all their assumptions in the paper, because they do not vary most of their assumptions to measure the sensitivity of their estimates, and because the DYNASIM model is not publicly available, which prevents others from using it to determine the sensitivity of their estimates to changes in assumptions.

The plausibility of these DYNASIM estimates can be studied, however, using the PENSIM microsimulation model,[2] which is comprehensively documented, extensively validated, and publicly available at no cost. The analysis described here can be reproduced and checked for accuracy by anyone interested in installing the three PSG models.

*The analysis in this essay has been funded exclusively by PSG; the analysis does make use of PENSIM, a microsimulation model whose development has been funded by the U.S. Department of Labor since 1997, but the views expressed here do not necessarily reflect those of any government agency.

[1]Barbara A. Butrica and Richard W. Johnson, "How Much Might Automatic IRAs Improve Retirement Security for Low- and Moderate-Wage Workers?," Urban Institute Program on Retirement Policy Brief #33, July 2011 (hereafter referred to as Butrica-Johnson).

[2]PENSIM and two related models are documented and available for download at the PSG website ⟨www.polsim.com⟩.

This essay begins by describing how the DYNASIM estimates can be closely replicated using PENSIM when using assumptions similar to those used in the Butrica-Johnson study. Then results are presented from a series of sensitivity tests that change the assumptions required to replicate the DYNASIM estimates to baseline PENSIM assumptions, assumptions that are shown to be more realistic, and hence, more defensible. Two sets of sensitivity tests are conducted. In the first set, just one assumption is changed at a time and all other assumptions remain as in the replication. In the second set, several assumptions are changed in order to get a sense of the cumulative effect on the estimates of changing more than one assumption at a time. The essay concludes with some thoughts on how low the enrollment rate might turn out to be in automatic IRAs and with some evidence on the poor target efficiency of the automatic IRA reform.

PENSIM Replication of DYNASIM Estimates

This section summarizes the DYNASIM estimates and then describes how PENSIM can be used to replicate the DYNASIM estimates.

DYNASIM Estimates

The Butrica-Johnson study uses two different assumptions about the person-weighted (as opposed to earnings-weighted) participation rate among those eligible for automatic IRAs who are 40–49 years old in 2036: the low-enrollment assumption calls for a 36% rate and the high-enrollment assumption calls for a 70% participation rate (Butrica-Johnson, pages 2–4).

Using these two enrollment assumptions in DYNASIM, average after-tax family income (measured per adult and in 2010 dollars) at age seventy for those who are 40–49 years old in 2036 is computed in two situations: pre-reform (without automatic IRAs) and post-reform (with automatic IRAs beginning in 2012). Each of these two average after-tax family income statistics is computed for all individuals and for each of four pre-reform family income quartiles after trimming off the top and bottom one percent of the pre-reform family income distribution. Some of the DYNASIM estimates are reproduced in Table 1.

Given the uncertainties in any such projection it may be prudent to round the simulation estimates as has been done in the Butrica-Johnson study. But aggressive rounding makes the task of replicating the results more difficult because the actual simulation estimates can be anywhere in a fairly broad

Table 1: DYNASIM Estimates of Pre-Reform Level of After-Tax Family Retirement Income and Percentage Reform-Induced Gain in After-Tax Family Income in Retirement

Mean level of pre-reform after-tax family income (thousands $)	Percentage reform-induced gain in after-tax family income	
	low-enrollment	high-enrollment
All individuals	65	3	5
Bottom quartile	16	6	13
Second quartile	33	6	12
Third quartile	59	5	7
Top quartile	150	3	3

Source: Butrica-Johnson, Table 2, where 2010 dollar amounts are rounded to the nearest whole thousand dollars and percentage amounts are rounded to the nearest whole percent.

rounding-error range. For example, the average annual dollar gain induced by the reform under the low-enrollment assumption could be anywhere between \$1619 (0.0251×64501) and \$2286 (0.0349×65499). So, from a replication perspective, the exact dollar gain is a symmetric probability distribution with a standard deviation of about \$184 and an expected value equal to \$1950 (0.03×65000) under the low-enrollment assumption and \$3250 (0.05×65000) under the high-enrollment assumption.[3]

PENSIM Replication

Many of the assumptions used in the Butrica-Johnson study are described in their paper or in publicly available DYNASIM documentation,[4] but some assumptions are not made explicit, and therefore, must be inferred. This means that the replication might not be exact because some inferences may be incorrect.[5]

[3]These two probability distributions are not normal; they have significant negative excess kurtosis because there is less probability mass near the mean and more away from the mean than is the case in the normal distribution.

[4]Karen E. Smith, "Projection Methods Used in the Dynamic Simulation of Income Model (DYNASIM3)" Urban Institute, February 2012 (hereafter referred to as Smith).

[5]PENSIM and the other PSG models can closely replicate social security solvency projections, in part, because the assumptions underlying those projections are made explicit in a comprehensive manner by the Office of the Chief Actuary at the Social Security Administration.

Using the 10/17/12 version of PENSIM and the other PSG models, eleven changes in baseline PENSIM assumptions must be made in order to replicate the Butrica-Johnson estimates.[6] These eleven changes to baseline PENSIM assumptions are as follows:

(a) assume the automatic IRA reform is accompanied by an expanded and refundable saver's credit reform rather than assume current-law saver's credit is unchanged after the automatic IRA reform ⟨explicit at Butrica-Johnson, page 3⟩;

(b) assume there is no small-employer exemption for offering automatic IRAs rather than assume that employers with ten or fewer employees are exempt as in the reform proposals and in the baseline PENSIM assumptions[7] ⟨inferred from the statement that people are "eligible to participate in an automatic IRA ... because they are working in a job that doesn't offer a retirement plan" (Butrica-Johnson, page 4) and because there is no indication that DYNASIM simulates employer size⟩;

(c) assume there are no passive participants at the default contribution rate in automatic IRAs rather than assume that the time pattern of passive participation is the same as observed in employer-sponsored DC plans with automatic enrollment (see section C.16 of *PENSIM Overview* for baseline PENSIM assumptions on time pattern of passive participation)[8] ⟨inferred from "[f]or workers predicted to participate in an automatic IRA, the analysis uses DYNASIM's model of DC contributions to estimate how much they will contribute" (Butrica-Johnson, page 3) and because Butrica-Johnson do not discuss assumptions about how long automatically-enrolled participants will remain passive at the defaults and do not mention what assumptions are made about the default contribution rate (three percent in baseline PENSIM assumptions) and the

[6]Comprehensive documentation of PENSIM and its baseline assumptions can be found in the *PENSIM Overview*, which is available at ⟨www.polsim.com⟩.

[7]J. Mark Iwry, William G. Gale, and Peter R. Orszag, "The Potential Effects of Retirement Security Project Proposals on Private and National Saving: Exploratory Calculations," Retirement Security Project working paper 2006-2 (hereafter referred to as Iwry-Gale-Orszag) assume on page 9 that the small employer automatic IRA offering exemption would imply that only sixty percent of workers without an employer-sponsored pension plan would be offered automatic IRAs. The baseline PENSIM assumption is based on estimates of private-sector employees working for employers with ten or fewer employees.

[8]Iwry-Gale-Orszag (page 9) assume that seventy-one percent of participants in automatic IRAs would be contributing at the default rate of three percent.

default investment (target-date fund in baseline PENSIM assumptions) in automatic IRAs⟩;

(d) assume no contribution skipping among active participants in automatic IRAs rather than assume active participants in automatic IRAs skip contributions with the same frequency as active participants in employer-sponsored DC plans (see section C.17 of *PENSIM Overview* for baseline PENSIM assumptions on contribution skipping in DC plans, which are based on EBRI-ICI and SIPP data) ⟨inferred from lack of discussion of contribution-skipping behavior in Butrica-Johnson paper and DYNASIM documentation⟩;

(e) assume all automatic IRA contributions are new saving out of family income rather than assume that one-third of contributions represent family saving that would have occurred without an automatic IRA reform as in baseline PENSIM assumptions[9] ⟨explicit at Butrica-Johnson, page 3⟩;

(f) assume all automatic IRAs are Roth accounts rather than assume they are all traditional accounts as in baseline PENSIM assumptions ⟨inferred⟩;

(g) assume that pension annuity provider has no costs other than paying promised annuity payments rather than assume that administrative costs are one percent and the required rate of return on capital is seven percent as in baseline PENSIM assumptions ⟨inferred from description of DYNASIM annuity prices as being "actuarially fair" by Smith, page 8⟩;

(h) assume a risk-free equity premium rather than assume that the rate of return on stocks equals the rate of return on government bonds when there is no year-to-year volatility in stock returns as assumed by CBO[10] and by the baseline PENSIM assumptions when macroeconomic variables are simulated in a single time-series projection ⟨explicit at Smith, page 5⟩;

(i) assume all mutual fund fees are 100 basis points rather than assume bond fund fees are 45 basis points and target-date fund fees are 75 basis points as in baseline PENSIM assumptions ⟨explicit at Smith, page 5⟩;

[9]The one-third saving-offset assumption is from Iwry-Gale-Orszag, page 10.

[10]See, for example, the discussion of risk-adjusted equity return on pages 6–7 in CBO [letter from Director Holtz-Eakin to Senator Craig], "Long-Term Analysis of Plan 2 of the President's Commission to Strengthen Social Security," September 30, 2004.

(j) assume that pension limits and income tax brackets are wage indexed rather than assume those amounts are indexed to prices as in current-law policy and baseline PENSIM assumptions ⟨implicit based on Butrica-Johnson, page 3, assumption that parameters of the expanded and refundable saver's credit reform are wage indexed⟩; and

(k) assume contribution rates and rollover rates for employer-sponsored DC plans are somewhat higher than implied by the baseline PENSIM assumptions ⟨implicit based on Butrica-Johnson, page 3, statement that "[a] few workers (mostly young workers and those with small balances) are simulated to cash out balances at job separation" and based on complete lack of documentation about the level of contribution rates⟩.

Replication assumption (k) involves raising the rollover rate among all DC participants. The baseline PENSIM assumptions imply the person-weighted rollover rate is about 76% for DC savings and thrift plan participants who finish a job at ages between 40 and 49. Making a guess, the Butrica-Johnson phrase "a few" is interpreted to mean one out of ten. So replication assumption (k) raises the rollover rate up to roughly 90%. Then replication assumption (k) raises the average contribution rate for all DC participants to about 8.5% for DC savings and thrift plan participants who are between the ages of 40 and 49. This 8.5% is about one percentage point higher than implied by the baseline PENSIM assumptions. The size of this contribution rate increase has been selected so that the pre-reform family income statistic produced in the PENSIM replication equals exactly the DYNASIM estimate of sixty-five thousand dollars.

Replication assumption (k) has been calibrated so that the PENSIM replication can produce the DYNASIM estimate of the pre-reform family income statistic despite the fact that baseline PENSIM assumptions about rollover rates and contribution rates have been extensively validated to ensure they produce simulated statistics that closely match recent historical evidence on DC plans.[11]

The PENSIM estimates using these eleven replication assumptions are compared with the DYNASIM estimates in Table 2. The quartile estimates for the pre-reform family income statistic from the PENSIM replication are close to the DYNASIM estimates. The PENSIM replication estimates of the reform-induced percentage increase in that family income statistic are 2.93% under the

[11]See *PENSIM Overview* sections C.17 and C.19 for documentation of DC contribution and rollover behavior, and chapter 11 for results of validation tests.

Table 2: Estimates of Pre-Reform Level of After-Tax Family Retirement Income and Percentage Reform-Induced Gain in After-Tax Family Income in Retirement from PENSIM Replication (P) and DYNASIM (D)

Mean level of pre-reform after-tax family income (thousands $)			Percentage reform-induced gain in after-tax family income			
			low-enrollment		high-enrollment	
	P	D	P	D	P	D
All individuals	65.0	65	2.93	3	4.94	5
Bottom quartile	19.6	16	6.84	6	13.86	13
Second quartile	35.8	33	6.75	6	11.10	12
Third quartile	59.5	59	3.65	5	5.78	7
Top quartile	145.1	150	1.17	3	1.86	3

Source: PENSIM runs 401, 403, and 405 as specified in the AI-401.rsf, AI-403.rsf, and AI-405.rsf run specification files; and Butrica-Johnson, Table 2, where 2010 dollar amounts are rounded to the nearest whole thousand dollars and percentage amounts are rounded to the nearest whole percent.

low-enrollment assumption and 4.94% under the high-enrollment assumption, in comparison to the rounded 3% and 5% estimates from DYNASIM. The corresponding average dollar increases in the family income statistic of $1907 and $3208 are not significantly different from the expected value of the probability distribution of the DYNASIM estimates under the low- and high-enrollment assumptions.[12]

DYNASIM estimates are also available for the subset of the population who experience a reform-induced gain in this family income statistic of at least two percent. Under the high-enrollment assumption, DYNASIM estimates 53% will experience at least a two percent gain (Butrica-Johnson, Figure 2) and the average of those large gains will be 11% (Butrica-Johnson, Figure 3). The corresponding PENSIM replication estimates are 57.8% and 11.07%, both of which are close to the corresponding DYNASIM estimate.[13]

[12]The dollar difference is $43 $(1950 - 1907)$ under the low-enrollment assumption and $42 $(3250 - 3208)$ under the high-enrollment assumption, both of which are less than one-quarter of the $184 standard deviation of the probability distribution implied by the rounding of the DYNASIM estimates.

[13]The quartile estimates for this large gain percent statistic are 18, 18, 12, and 7, from DYNASIM, versus 17.9, 15.3, 10.5, and 6.0, from the PENSIM replication.

Sensitivity of Estimates to Assumption Changes

Now that it has been demonstrated that the PENSIM replication produces estimates that are quite closely aligned with DYNASIM estimates, PENSIM is used to determine the sensitivity of the reform estimates to changes in the assumptions that were required to replicate the DYNASIM estimates. These sensitivity test results are presented in two parts. In the first part, each of the eleven replication assumptions is eliminated one at a time. In the second part, results are presented for a number of compound sensitivity tests, in which more than one replication assumption is eliminated at a time. These compound sensitivity tests reveal the cumulative effect on the reform estimates of eliminating several replication assumptions. In all cases, when a replication assumption is eliminated, it is replaced with the baseline PENSIM assumption, as described above on pages 4–6. The sensitivity tests are conducted using the low-enrollment assumption, but there is no reason to think the percentage sensitivities would be different for the high-enrollment assumption.

Changing One Assumption at a Time

The results of the solo sensitivity tests are presented in Table 3. The first row of the table shows the reform-induced dollar gain in the family income statistic ($1907) when all eleven PENSIM replication assumptions — from (a) through (k) as described on pages 4–6 — are used along with the low-enrollment assumption. The other rows show how sensitive that estimate is to eliminating one of the replication assumptions and replacing it with the corresponding baseline PENSIM assumption. One of the assumption changes causes the estimated reform-induced dollar gain to increase, but the other ten assumption changes cause a smaller reform-induced dollar gain. For eight of these ten assumption changes, the reform-induced dollar gain declines from eighteen to thirty-six percent. So, the estimates of the impact of the automatic IRA reform are very sensitive to most of the assumptions that are required for PENSIM to replicate the Butrica-Johnson DYNASIM estimates.

Changing More than One Assumption at a Time

The results of the compound sensitivity tests are presented in Table 4. The first row of the table shows the reform-induced dollar gain in the family income statistic ($1907) when all eleven PENSIM replication assumptions — from (a) through (k) as described on pages 4–6 — are used along with the low-

Table 3: Sensitivity of Estimates to Changes in PENSIM Replication Assumptions Under Low-Enrollment Assumption, Part 1

Eliminated replication assumption	Pre-reform family income statistic	Reform-induced dollar gain in family income	Percentage change in dollar gain from the full replication
none	65.0	1907	0.0%
(a)	65.0	1444	–24.3%
(b)	65.0	1552	–18.6%
(c)	65.0	1457	–23.6%
(d)	65.0	1833	–3.9%
(e)	65.0	1414	–25.9%
(f)	65.0	1258	–34.0%
(g)	63.8	1777	–6.8%
(h)	59.4	1212	–36.4%
(i)	65.5	1966	+3.1%
(j)	61.1	1350	–29.2%
(k)	62.9	1549	–18.8%

Source: PENSIM runs 401 and 403 as specified in the AI-401.rsf and AI-403.rsf run specification files; list of replication assumptions (a) through (k) on pages 4–6.

enrollment assumption. The other rows show how sensitive that estimate is to eliminating different combinations of the replication assumptions with each eliminated assumption being replaced with its corresponding baseline PENSIM assumption.

The eight changes in assumptions listed on the line just above the first dividing line in Table 4, which are all highly defensible changes in assumptions, together reduce the size of the estimated reform-induced gain in the family income statistic by about 82 percent. The average gain in the PENSIM replication under the low-enrollment assumption is $1907. But after changing replications assumptions (a) through (e) and (g) through (i) back to the corresponding baseline PENSIM assumptions, the $1907 declines by 82.6% to an average gain of $332. In other words, the DYNASIM estimate of the effect of an automatic IRA reform on retirement income is about five times larger than an estimate based on more realistic assumptions.

Assumption change (a) simply focuses attention on the automatic IRA reform, which is the question posed in the title of the Butrica-Johnson paper: how much might automatic IRAs improve retirement income? While an expanded

Table 4: Sensitivity of Estimates to Changes in PENSIM Replication Assumptions Under Low-Enrollment Assumption, Part 2

Eliminated replication assumptions	Reform-induced dollar gain in family income	Percentage change in dollar gain from the full replication
none	1907	0.0%
(a)	1444	–24.3%
(a)+(b)	1105	–42.1%
(a)+(b)+(c)	783	–58.9%
(a)+(b)+(c)+(d)	767	–59.8%
(a)+(b)+(c)+(d)+(e) [*i.e.*,(a)⇒(e)]	512	–73.2%
(a)⇒(e)+(g)	474	–75.1%
(a)⇒(e)+(g)+(h)	322	–83.1%
(a)⇒(e)+(g)+(h)+(i)	332	–82.6%
(a)⇒(e)+(g)+(h)+(i)+(k)	285	–85.1%
(a)⇒(e)+(g)+(h)+(i)+(k)+(f)	185	–90.3%
(a)⇒(e)+(g)+(h)+(i)+(k)+(j)	221	–88.4%
(a)⇒(e)+(g)+(h)+(i)+(k)+(f)+(j)	144	–92.4%

Source: PENSIM runs 401 and 403 as specified in the AI-401.rsf and AI-403.rsf run specification files; list of replication assumptions (a) through (k) on pages 4–6.

and refundable saver's credit had been included in recent administration budget proposals, it was dropped from the most recent budget proposal. And a saver's credit reform is not logically required by the automatic IRA reform.

Assumption change (b) simply corrects shortcomings in DYNASIM that prevent it from simulating individuals working at a job where the employer has ten or fewer employees.

Assumption change (c) simply corrects shortcomings in DYNASIM that prevent it from simulating the time it takes to move from passive to active participation under automatic enrollment. It is puzzling that DYNASIM has been used to simulate the effects of a pension reform that has automatic enrollment as its central feature when there is no evidence in the documentation that DYNASIM has the ability to simulate the period of time individuals remain passively enrolled at the default contribution rate with investments being made in the default investment. And the Butrica-Johnson paper says nothing about which defaults are assumed in the analysis.

Assumption change (d) simply assumes that active participants in auto-

matic IRAs will occasionally skip a year of contributions on a job just as do other DC plan participants.

Assumption change (e) simply recognizes what even the original proponents of the automatic IRA reform recognized: that some contributions to automatic IRAs would have been placed in taxable savings accounts even without an automatic IRA reform, which means that some automatic IRA contributions are not new saving, and therefore, will not raise retirement income.

As can be seen in Table 4, the cumulative effect of replacing replication assumptions (a) though (e) — that is, focusing on just the automatic IRA reform and correcting what appear to be four structural limitations of DYNASIM — is to reduce the average gain in the family income statistic from $1907 to $512, a 73.2% decline in the size of the reform's effect.

There are at least three additional changes in the replication assumptions that are defensible.

Assumption change (g) simply recognizes that any pension annuity provider has costs other than the cost of making promised annuity payments, namely administrative costs and a competitive rate of return on the capital reserve that must be held to cushion against risks.

Assumption change (h) simply corrects a logically inconsistent assumption that people can earn returns on stocks that are above the risk-free rate without bearing any year-to-year volatility risk, adopting instead an assumption that CBO has made in all its non-stochastic analysis of social security individual accounts.

Assumption change (i) simply changes the Butrica-Johnson assumptions about mutual fund fees back to the baseline PENSIM assumptions about fees, which in some cases are lower.

After adding these three changes to the previous five, the cumulative effect of all eight assumption changes — (a) through (e) plus (g) through (i) — is to reduce the average gain in the family income statistic from $1907 to $332, which is nearly an eighty-three percent decline in the size of the reform's impact on the family income statistic.

And there is one more defensible change in the replication assumptions: assumption change (k) that switches back to the baseline PENSIM assumptions regarding DC pension contribution and rollover behavior, which have been extensively tested to show that they produce contribution, rollover, and on-the-job DC account balance statistics that are close to those observed in recent real-world data. The DYNASIM assumptions regarding these DC behaviors are undocumented and apparently untested. The Urban Institute has not provided

the kind of information that would generate confidence in using DYNASIM assumptions about contribution and rollover behavior in DC pension plans.

The cumulative effect of adding assumption change (k) the prior eight changes can be seen in Table 4 on the line between the two dividing lines. The cumulative effect of those nine assumption changes is to reduce the average gain in the family income statistic from $1907 to $285, which is about an eighty-five percent decline in the size of the reform's impact on the family income statistic.

This is the central finding of this study: the size of the Butrica-Johnson estimates of the effect of an automatic IRA reform on retirement income are driven largely by a series of unrealistic assumptions. And while this sensitivity testing has been conducted under their low-enrollment assumption, there is no reason to think that the percentage declines in the size of the high-enrollment estimate would not be roughly the same.

Making the other two changes in replication assumptions is not very defensible, but they are shown in Table 4 for the sake of completeness. It does seem reasonable to avoid changing replication assumption (f) because it is likely that most participants in automatic IRAs would pay income taxes on their contributions and deposit them in a Roth IRA account, from which withdrawals during retirement are not subject to income taxation. And finally, there are arguments both for and against changing replication assumption (j). It is true that assuming the current-law pricing indexing policy continues over many decades reduces the relative size of the saver's credit as wages rise faster than prices in the long run. In addition, effective income taxes rates rise because of bracket creep. So, assuming for analytical purposes that income tax and saver's credit amounts are wage indexed has its merits. But with all the discussion about the need to reduce tax expenditures — of which the saver's credit and the tax-favored treatment of automatic IRAs are small examples — and the need for more tax revenues to reduce the budget deficit, it is important for any analysis that makes the wage-indexing assumption to show how much difference it makes to the results.

Expectations about Enrollment in Automatic IRAs

The Butrica-Johnson low-enrollment and high-enrollment assumptions, which produce person-weighted participations rates among middle-age eligibles for automatic IRAs of 36% and 70%, respectively, "are intended to represent the lower and upper bounds of possible outcomes in a world with automatic IRAs"

(Butrica-Johnson, page 3). But this is a weak assertion given that Butrica-Johnson offer no evidence that 36% should be considered the low end of the range of expectations about how many eligibles would participate in automatic IRA plans. And, in fact, there are at least two important bits of evidence that suggest the person-weighted participation rate in automatic IRA plans could be lower than 36%.

Because both these bits of evidence refer to participation rates among people of all ages (rather than a rate for those in their forties, which is how Butrica-Johnson compute the 36%), it is essential to know what the all-person participation rate is in the replication of the DYNASIM low-enrollment assumption. The answer is that the person-weighted participation rate for all ages is 25.0% (while being 36.0% for those between the ages of 40 and 49) and the earnings-weighted participation rate for all ages is 49.8%. The earning-weighted participation rate is higher than the person-weighted rate because high earners are more likely to participate than low earners.

The first bit of evidence is that Iwry-Gale-Orszag (page 9) assume a 35% earnings-weighted participation rate in their estimates of the saving impact of an automatic IRA reform. This 35% assumption is well below the 49.8% implied by the Butrica-Johnson low-enrollment assumption.

The second bit of evidence is that enrollment in a (non-automatic) IRA account among a population similar to those who would be eligible for automatic IRA plans is very low according to results from a field experiment conducted by the Retirement Security Project.[14] Dufflo-Gale-Liebman-Orszag-Saez (page 1313) report person-weighted participation rates for three groups: 3% in a control group who received the current-law saver's credit on any IRA contributions, 8% for a group that receive a twenty percent refundable match on IRA contributions up to one thousand dollars, and 14% for a group that receive a fifty percent refundable match on IRA contributions up to one thousand dollars. These person-weighted participations rates are for randomly-selected people filing income taxes in low-income neighborhoods in the St. Louis area during March 2005. Each of the three groups contained at least 4500 people, 66% of whom had at least a $500 income tax refund available to contribute to the IRA, and the average income tax refund was about $1550 in each group.[15]

[14]Esther Dufflo, William Gale, Jeffrey Liebman, Peter Orszag, and Emmanuel Saez, "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block," *Quarterly Journal of Economics*, November 2006 (hereafter referred to as Dufflo-Gale-Liebman-Orszag-Saez).

[15]Dufflo-Gale-Liebman-Orszag-Saez, Table I on pages 1318–1319.

Even if these observed participation rates are scaled up to account for the fact that only 75% of people in each group were offered IRA enrollment (something that Dufflo and her co-authors did not do),[16] the person-weighted participation rates are about 4%, 10%, and 19%. All three of these adjusted all-age rates are below the 25.0% implied by the Butrica-Johnson low-enrollment assumption, with the 4% estimate being the most relevant when considering the automatic IRA reform in isolation.

Both of these bits of evidence about the likely rate of participation in automatic IRA plans come from the Brookings Retirement Security Project that was the original proponent of the automatic IRA reform. Why do Butrica and Johnson fail to mention this evidence that suggests participation rates could be below their low-enrollment assumption?

Target Efficiency of Automatic IRAs

The PENSIM replication of the Butrica-Johnson DYNASIM estimate of the effect on retirement income of an automatic IRA reform permits a reconsideration of the target efficiency of the reform. Target efficiency refers to the extent a policy "direct[s] resources to where they do the most good."[17]

Butrica and Johnson argue that the distributional results on the automatic IRA reform indicate that "automatic IRAs would provide the most help to adults with low and moderate incomes, those with limited plan coverage but strong work histories, and those without private retirement benefits. Thus, the proposal seems to have the largest impact on those it was designed to target."[18]

But the evidence they present to support this conclusion is not entirely convincing. They do find, and the PENSIM replication confirms, that people in the bottom half of the pre-reform family income distribution experience larger percentage increases in family income from the automatic IRA reform than do people in the top half of the pre-reform family income distribution. And if the phrase "provide the most help" means to provide the largest percentage increase in retirement income, then their conclusion is supported by the simulation results.

But if target efficiency is measured by the fraction of reform-induced dollar gains that are received by the bottom half of the distribution, then the

[16]Dufflo-Gale-Liebman-Orszag-Saez, pages 1317–1319.

[17]Peter Schuck and Richard Zeckhauser, *Targeting in Social Programs* (Brookings Institution, 2006), page 1.

[18]Butrica-Johnson, page 4.

Table 5: Target Efficiency Estimates Using All PENSIM Replication Assumptions Under Butrica-Johnson Low-Enrollment or High-Enrollment Assumption

Pre-reform family income quartile	Reform-induced percentage gain in family income		Percentage share of aggregate gain in family income	
	Low	High	Low	High
All individuals	2.93	4.93	100.0	100.0
Quartile 1 (bottom)	6.84	13.86	17.5	21.1
Quartile 2	6.75	11.10	31.7	31.0
Quartile 3	3.65	5.78	28.5	26.8
Quartile 4 (top)	1.17	1.86	22.3	21.1

Source: PENSIM runs 401, 403, and 405 as specified in the AI-401.rsf, AI-403.rsf, and AI-405.rsf run specification files.

automatic IRA reform is not very target efficient. Only about half of the population-wide gains in retirement income flow to those in the bottom half of the income distribution. And the top income quartile receives as many, if not more, gains than does the bottom quartile.

The estimated target efficiency of the reform under the low-enrollment and high-enrollment assumptions combined with all eleven of the replication assumptions (including counting the effects of an expanded and refundable saver's credit as part of the automatic IRA reform effect) is shown in Table 5. Under the low-enrollment assumption, the bottom quartile receives less than 18% of the aggregate gains in retirement income, the bottom two quartiles receive slightly less than 50%, and the top quartile receives 22% of the aggregate gains. Even under the high-enrollment assumption, the bottom quartile receives only about 21% of the aggregate gains in retirement income, the bottom two quartiles receive about 52%, and the top quartile receives about 21% of the aggregate gains.

The estimated target efficiency of the reform under the low-enrollment and high-enrollment assumptions combined with all but one of the eleven replication assumptions (excluding counting the effects of an expanded and refundable saver's credit as part of the automatic IRA reform effect) is shown in Table 6. Under the low-enrollment assumption, the bottom quartile receives slightly more than 15% of the aggregate gains in retirement income, the bottom two quartiles receive about 47%, and the top quartile receives about 23% of the aggregate gains. Even under the high-enrollment assumption, excluding the

Table 6: Target Efficiency Estimates Using All PENSIM Replication Assumptions Except (a), Including Effects of Saver's Credit Reform, Under Butrica-Johnson Low-Enrollment or High-Enrollment Assumption

Pre-reform family income quartile	Reform-induced percentage gain in family income		Percentage share of aggregate gain in family income	
	Low	High	Low	High
All individuals	2.22	4.05	100.0	100.0
Quartile 1 (bottom)	4.54	10.59	15.4	19.7
Quartile 2	5.11	9.11	31.7	31.0
Quartile 3	2.86	4.85	29.5	27.5
Quartile 4 (top)	0.93	1.59	23.4	21.9

Source: PENSIM runs 401, 403, and 405 as specified in the AI-401.rsf, AI-403.rsf, and AI-405.rsf run specification files.

effect of the saver's credit reform implies that the bottom quartile receives slightly less than 20% of the aggregate gains in retirement income, the bottom two quartiles receive almost 51%, and the top quartile receives almost 22% of the aggregate gains.

The automatic IRA reform becomes even less target efficient as more of the replication assumptions are replaced with baseline PENSIM assumptions. So, for example, under the low-enrollment assumption when replication assumptions (a) through (d) are replaced with the corresponding baseline PENSIM assumptions, the top quartile receives $1.86 for each dollar the bottom quartile receives and the top half receives 55% of the aggregate family income gains generated by the reform. Given these results, it is difficult to see how the reform can be characterized as target efficient.

The issue of target efficiency is important because an automatic IRA reform will not be costless: employers will have extra administrative costs and the government will lose income tax revenue on all the automatic IRA contributions that would have been saved in taxable accounts without the reform and because of the favorable income tax treatment of asset returns in automatic IRA accounts. Those who favor an automatic IRA reform must provide an answer to the question: why should employers and taxpayers bear these reform costs when more than half of the reform benefits go to those in the top half of the retirement income distribution? It is a weakness of the Butrica-Johnson paper that this issue is glossed over.

Broader Considerations

The results described above raise a broader issue: what is the rationale for an automatic IRA reform?

In a review of a book by Orszag, Iwry, and Gale, that advocates a number of policy changes including an automatic IRA reform,[19] Olivia Mitchell had the following thoughts in the *Journal of Economic Literature* (December 2008):

> ... some will find it troubling that there is no discussion of how Social Security interacts with pensions for the lower-paid. Social Security rules, for instance, pay lower-wage workers much higher relative benefits than higher-paid workers; the lower-paid also pay less tax and are more likely to receive social insurance disability income and survivor benefits. So why, then, should low-paid workers be required, induced, and/or subsidized to save, when government benefits are already proportionately more generous? Maybe low saving rates are optimal for the poor and middle class .... Accordingly, the authors could have devoted greater effort to supporting their presumption that the poor should save more, before asserting that they should be "induced" to save more out of already low incomes. [page 987]

And on the issue of target efficiency, Mitchell had this to say about an expanded and refundable saver's credit (but the same issues arise with regards to an automatic IRA reform):

> Furthermore, their plan [*i.e.*, a refundable saver's credit reform] may prove to be target-ineffective, since it pays "temporarily" low-wage workers (such as college students) a subsidy, even when they go on to earn much higher future lifetime salaries later in life. [page 988]

The above results on target efficiency suggest strongly that this sort of thing is going on with a reform that requires all but the smallest employers to offer automatic IRA plans if they do not offer any other pension plan to their employees.

[19] Peter R. Orszag, J. Mark Iwry, and William G. Gale (editors), *Aging Gracefully: Ideas to Improve Retirement Security in America*, Century Foundation Press, 2006.