ELSEVIER

Journal of Public Economics 87 (2003) 1259–1290

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

Does 401(k) eligibility increase saving? Evidence from propensity score subclassification

Daniel J. Benjamin*

Department of Economics, Harvard University, Cambridge, MA 02138-3001, USA

Received 13 October 2000; received in revised form 12 August 2001; accepted 14 August 2001

Abstract

By comparing 401(k) eligible and ineligible households' wealth, this paper estimates that, on average, about one half of 401(k) balances represent new private savings, and about one quarter of 401(k) balances represent new national savings. Responses to eligibility vary considerably, however, with households who normally save the most largely contributing funds they would have saved anyway. This paper improves on previous research by: (1) employing propensity score subclassification to control more completely for observed household characteristics, (2) controlling for more household characteristics, including several correlated with unobserved savings preferences, and (3) adjusting the observed measure of households' wealth to reduce measurement error.

Keywords: Saving; Retirement; Propensity score

JEL classification: H2; H31; C21

1. Introduction

Despite the enthusiasm of policy makers and the public for tax-advantaged saving incentive programs, economists are still debating whether such programs actually succeed in increasing saving. This paper evaluates the effects on private

*Present address: 12 Mayo Place, Dresher, PA 19025, USA. Tel.: +1-215-646-7988.
E-mail address: daniel_benjamin@post.harvard.edu (D.J. Benjamin).

0047-2727/01/$ – see front matter
PII: S0047-2727(01)00167-0

and national savings and on components of households' saving of 401(k)s, a tax-deferred saving incentive program that has quickly become one of the principal vehicles for retirement saving in the United States (Poterba et al., 1998). The main conclusion is that there is enormous heterogeneity in households' responses to 401(k)s. 401(k)s have induced many households to save more, improving the financial outlook at retirement for those households. In aggregate, however, only about a quarter of 401(k) balances represent new national savings.

401(k) plans are employment-based, tax-deferred savings instruments that have become increasingly popular since 1981. An employee is *eligible* to contribute to a 401(k) account only if the employer sponsors a 401(k) program. Eligible employees choose whether to *participate* — that is, make a contribution — and most employers encourage participation with matching provisions.[1] Contributions are tax-deductible, and interest earned on 401(k) savings accrues tax-free. At time of withdrawal, accumulated 401(k) assets are taxed as income. Employees may withdraw funds without penalty only if over age 59.5, retired and over age 55, or disabled. In exceptional circumstances — financial hardship or worker separation from firm — early withdrawals may be made, but such withdrawals are taxed as income and are subject to a 10% penalty.

One early influential paper (Engen et al., 1994) estimated 401(k) effectiveness by median regression and found that 401(k)s increase private savings little, if at all, because households reduce other asset holdings to fund 401(k) participation. Another important early paper (Poterba et al., 1995) divided households into income intervals and ran median regressions within those intervals, finding that 401(k)s increase private savings dramatically, with little substitution between 401(k) assets and other types of savings. Subsequent work has generally built on these methods and reached conclusions similar to one or the other of these contradictory views.

This paper makes several contributions to this literature. First, regression within propensity score subclasses controls more completely for observed household characteristics and is shown to produce more robust and reliable estimates than regression within income intervals or simple regression. Second, because propensity score methods easily accommodate a large number of control variables, the estimation procedure controls for many observed household characteristics, including several not used in previous research. Third, adjustments to the dependent variable, households' wealth, help to calibrate the magnitude of data deficiencies highlighted by Engen et al. (1996). These three contributions are methodological improvements that reduce biases in the 401(k) effect estimate, and

[1] Highly compensated employees may want other employees to participate in the firm's 401(k) plan so that the highly compensated employees can make large contributions without violating various 'nondiscrimination tests' mandated by Congress (Papke et al., 1996). Plans began matching for this reason (Pare, 1995). Evidence is mixed on whether matching actually does increase employees' participation, and it may depend on the level of matching (see Papke et al., 1996 for discussion).

each has an important impact on the conclusions reached about 401(k) effectiveness. Finally, examining mean effects as well as median effects, and examining the impact of 401(k)s on national savings as well as private savings, enables greater insight into the range of responses to 401(k)s and the implications of those responses for evaluating theory and for achieving policy goals.

The organization of this paper is as follows. Section 2 describes the data used both in previous studies and here and discusses the choice of covariates to control for. Section 3 describes the adjustments made to the measure of household wealth. In Section 4, a tax adjustment distinguishes household contributions to national savings from those to private savings. Because propensity scores are relatively new to economics, Section 5 explains propensity score subclassification in a context that clarifies its relationship to the methods of linear regression and of income subclassification. Section 5 also divides households into subclasses on the basis of estimated propensity scores. Section 6 presents results and systematically explores how each of the methodological improvements introduced in this paper affect the final 401(k) effect estimate. Section 7 summarizes, discusses the main results, relates this work to other recent work, and mentions policy implications. Appendix A estimates the proportion of 401(k) assets converted from pre-existing DC plans, an exercise necessary for the wealth adjustment.

2. The data

As in Engen and Gale (2000), Poterba et al. (1995), and Engen et al. (1994), a sample of households is used from the 1990 Survey of Income and Program Participation (SIPP) wave 4: households in which the reference person is 25–64 years old, in which at least one person is employed, and in which no one is self-employed.[2] All results, especially statements about aggregate 401(k) effects, should be understood as applying to this particular population of households. The sample, which contains 9915 households, has been referred to as the 1991 SIPP because the data were collected between February and May, 1991. Five households were dropped for reporting zero or negative income, leaving $N = 9910$ households. All dollar amounts are in 1991 dollars.

The 1991 SIPP reports, among other things, household financial data across a range of asset categories. In particular, the survey reports whether any member of the household works for a firm that offers a 401(k) program: if so, the household is coded as *eligible*; otherwise, the household is *ineligible*. Thirty-eight percent of the sample households are eligible, and 62% are ineligible. The survey also reports

[2] Analyses have used this sample because 401(k)s are not available to the self-employed or unemployed, because the SIPP only asks 401(k) questions to people 25 and older, and because retirement and saving behavior of the elderly would complicate analyses for people 65 and older. The person in whose name the family's home is rented or owned is the reference person.

401(k) asset balances. Households with positive balances can be considered 401(k) *participants*, while eligible households with a zero balance can be considered *non-participants*. Because eligible employees decide whether or not to participate whereas firms determine eligibility, it is more plausible that eligibility is conditionally random, given observed household characteristics, than that participation is. That is why this analysis focuses on eligibility rather than participation as the independent variable (Poterba et al., 1995 introduced this 'eligibility experiment' approach). In the sample used, 70.5% of eligible households were also participants.

In this analysis, the primary dependent variable is total wealth, which equals net financial assets, including Individual Retirement Account (IRA) and 401(k) assets, plus housing equity plus the value of business, property, and motor vehicles. Using this broadest available measure of wealth reduces possible bias from asset shifting or debt reshuffling. For example, if 401(k) eligible households substitute 401(k) savings for home equity, as Engen et al. (1996) have argued, then net total financial assets for eligible households will exceed net total financial assets for ineligible households by more than the actual increase in savings. The difference in measured total wealth, however, will capture the true difference in savings. Other sources of bias in the eligibility estimate arising from deficiencies in the wealth measure are addressed in the next section.

Other dependent variables are components of total wealth that previous researchers have suspected might substitute for 401(k) savings: IRA assets; non-IRA-401(k) financial assets; and home equity, which equals home value less mortgage. Components of home equity, home value and mortgage, also serve as dependent variables.

The covariates used in this analysis are age, income, family size, education, marital status, and two-earner status, as well as defined benefit (DB) pension status, IRA participation status, and home ownership status. Marital status, two-earner status, DB pension status, IRA participation status, and home ownership status are dichotomous variables, with the two-earner variable indicating whether both household heads contribute to the household's income. The education variable measures the number of years of schooling experienced by the reference person. DB pension status refers to whether the household's employer offers a DB pension plan. IRA participation status refers to whether the household has a positive IRA account balance, and home ownership status refers to whether the household has a positive home value. Log-income is also included as a covariate in the analyses below because, as usual, the distribution of income is skewed. As shown in Table 1, summary statistics indicate that eligible and ineligible households differ: eligibles are somewhat older, more educated, more likely to be married, more likely to have two earners, and earn significantly higher incomes than ineligibles. Eligible households are also more likely to have DB pensions, to participate in IRAs, and to own a house.

Several of the covariates are definitely not exogenous to 401(k) eligibility, most

Table 1
Mean characteristics and assets of 401(k) eligible and ineligible households ($1991)

	Eligible households	Ineligible households	t-statistic[a]
Covariates			
Age	41.5	40.8	3.1***
	(9.6)	(10.7)	
Income	$46 874	$31 514	31.3***
	($25 950)	($22 143)	
Education	13.8	12.9	15.3***
	(2.6)	(2.9)	
Family size	2.9	2.8	1.8*
	(1.5)	(1.6)	
Two-earner indicator	0.48	0.32	$z=16.2$***
	(0.50)	(0.47)	
Marriage indicator	0.67	0.56	$z=10.6$***
	(0.47)	(0.50)	
Defined benefit pension indicator	0.42	0.19	$z=24.5$***
	(0.49)	(0.39)	
IRA indicator	0.32	0.20	$z=13.3$***
	(0.47)	(0.40)	
Home ownership indicator	0.74	0.57	$z=16.8$***
	(0.44)	(0.49)	
Total wealth[b]	$86 251	$50 079	15.6***
	($124 021)	($101 183)	
Assets			
401(k) assets[b]	$10 749	$0	44.1***
	($19 249)	(–)	
Net non-IRA-401(k) assets	$14 709	$7881	5.9***
	($63 486)	($50 973)	
IRA assets among IRA holders[b]	$14 789	$13 837	1.54
	($15 099)	($15 187)	
Home equity among homeowners	$55 710	$50 705	3.5***
	($57 008)	($56 322)	
Home value among homeowners	$110 496	$92 222	10.4***
	($71 009)	($68 050)	
Home mortgage among homeowners	$54 787	$41 517	12.8***
	($42 908)	($39 180)	
Number of observations	3682	6233	

Source: Author's calculations from 1991 SIPP for families whose reference person is 25–64 years old, in which at least one person is employed, and in which no one is self-employed. Standard deviations in parentheses. T-statistics calculated with homoskedastic standard errors. Significance levels: * $P=0.10$, ** $P=0.05$, *** $P=0.01$.

[a] z-statistics are reported for logistic regressions.

[b] Unadjusted measure.

notably DB status, IRA status, and home ownership. One source of endogeneity is that they are all likely correlated positively with unobserved saving preferences. This fact argues for including these covariates in the analysis to mitigate the upward omitted variable bias induced from not controlling for saving preferences. Another source of endogeneity is that DB plans, IRAs, and homes may be substitutes for (and therefore be negatively correlated with) 401(k) eligibility so that controlling for them would introduce a new upward bias. These covariates are included in this analysis because 401(k) eligible households are in fact *more* likely than ineligible households to have DB pensions, IRAs, and homes, and eligibles hold higher IRA balances and home values, indicating that the omitted variable argument is the dominant effect. It is also important to control for the *presence* of a DB pension because DB *assets* are excluded from the wealth measure (the possibility that eligibility causes reduced DB asset accumulation among eligibles with a DB pension is addressed in Section 3). Section 6.4 shows that adding DB status and IRA status to the other covariates makes little difference in the results, but home ownership proves to be a crucial covariate at every stage of the analysis.

3. Adjusting the SIPP's measure of total wealth

The basic problem with the SIPP's measure of households' wealth is that the SIPP does not report all sources of pension wealth. The broadest available measure of wealth in the SIPP, total wealth, includes 401(k) and IRA assets but not assets held in other defined contribution (DC) plans or in DB plans. Assuming the estimation procedure successfully controls for relevant employee characteristics, 401(k) eligible and 401(k) ineligible employees *in the absence of 401(k)s* would behave identically with respect to these unmeasured assets. A firm's introduction of a 401(k) plan, however, may cause the firm or the eligible worker to deposit or transfer funds into the *measured* 401(k) account that would remain in *unmeasured* non-401(k) DC or DB plans for an otherwise similar ineligible worker. As a result of this substitution, non-401(k) DC and DB wealth become negatively correlated with eligibility status, biasing the eligibility effect estimate upward.

Engen et al. (1996) argued forcefully that the biases resulting from substitution out of (unmeasured) non-401(k) DC and DB plans into (measured) 401(k)s are dramatic enough to eradicate positive 401(k) effect estimates. Despite these assertions, researchers have continued to estimate 401(k) effects as though there were no biases. The two adjustments introduced in this section represent an attempt to calibrate the magnitude of this data problem for estimating 401(k) eligibility effects. In much of what follows, the evidence is sketchy, and debatable assumptions are necessary to quantify the adjustments. These assumptions are meant merely as plausible guesses pending better information. Furthermore, this section deals only with the most important potential biases, as judged by the attention accorded them in the literature. Consequently, both of the adjustments discussed here have the effect of reducing the estimated 401(k) effect.[3] The adjustment for converted DC plans reduces the estimated 401(k) effect substantially, which suggests that previous work overstated the 401(k) effect on saving. However, the fact that the average estimated 401(k) effect remains positive despite both adjustments suggests that data deficiencies alone cannot explain away the positive average 401(k) effect.

3.1. Replacement of a pre-existing DC or DB by a 401(k)

One way that a firm's introduction of a 401(k) could cause asset substitution is if the 401(k) replaced outright a previous pension plan. Papke (1999) estimated with US Labor Department's Form 5500 Annual Reports data that for every three 401(k) plans initiated between 1985 and 1992, one DB plan was terminated. Ippolito and Thompson (2000), however, show that most of these 'terminations' were actually plan mergers or changed ID numbers and that 95% of DB plan participants as of 1987 remained covered by their DB plan in 1995. Evidence from Poterba et al. (2001) and Papke et al. (1996) is consistent with the view that 401(k)s only rarely caused DB termination. Because outright replacement of DB plans by 401(k) plans appears to have been rare, no adjustment for this is made.

Conversions of a pre-existing DC plan into a 401(k) plan, however, did occur and may have been common in the 1980s, perhaps partly accounting for the rapid growth in the number of 401(k) plans (Engen et al., 1996). A converted plan's assets, previously unmeasured, are now labeled as 401(k) assets, generating a spurious increase in eligibles' measured assets relative to ineligibles'. New calculations using Form 5500 data, described in Appendix A, suggest that at most 30% of aggregate 401(k) assets in 1991 represent converted assets from pre-existing DC plans.

This estimate, that 70% of 401(k) asset balances are non-converted assets, is a lower bound for the cohort under study here for two reasons. First, this estimate is a lower bound for the aggregate percentage because its calculation conservatively assumes that all 401(k) balances as of 1983 were converted from pre-existing DC plans. Second, the 1991 proportion of non-converted 401(k) assets for households whose heads were aged between 25 and 64 in 1991 is probably greater than the aggregate proportion because this cohort, younger than 65 during all of 1984–

[3]The main data-induced downward bias on the 401(k) effect estimate may be that 401(k) funds kept in a *former* employer's account are not measured because of the SIPP question sequence. I thank a referee for pointing this out to me. Note, however, that if the worker rolls over his 401(k) account balance into an IRA upon separation from the firm, then these funds are measured in the SIPP as part of the worker's IRA, and there is no bias.

1991, made more of their contributions directly into 401(k) plans than the older cohort, who mostly made their contributions before 401(k)s became popular.

Because the evidence indicates that the cohort proportion of non-converted 401(k) assets in 1991 is greater than 70% but less than 100%, it is assumed here to be 80%. Therefore, as a rough adjustment for conversions, sample households' 401(k) assets were multiplied by 0.80 before being incorporated into total wealth. Section 6.4 shows that this adjustment has a relatively large impact on the estimated 401(k) effect.

3.2. Marginal substitution between DB and 401(k) plans

A firm that had a traditional DB pension and added a supplementary 401(k) plan might shift resources from the DB to the 401(k) (Engen et al., 1996). In particular, the firm might pay for 401(k) employer contributions by restricting or reducing average DB pension benefit increases.[4] There is disagreement about the extent to which such marginal substitution occurred. Poterba et al. (2001) argue that it is not generally possible to alter the formula-based DB benefit schedule, but Gale et al. (2000) mention some suggestive and anecdotal evidence that it did happen. If it did occur, the fraction of the employer contribution that represents foregone DB benefits should be subtracted from eligible workers' 401(k) balances because, for ineligible workers, this amount remains in unmeasured DB plans in the form of DB benefits.[5] This adjustment applies only to 401(k) eligible workers who also have a DB pension plan.

Eighty-four percent of firms offering a 401(k) in 1991 made some employer contributions contingent on employee contributions (Hewitt Associates, 1991), so it is assumed for this adjustment that all employer contributions are matching funds. For a given worker's contribution, c, and a given match rate, m, the total (worker plus employer) 401(k) contribution is $c(1+m)$. Suppose that some fraction of the match, ϕ, would have been transferred to the household in the absence of a 401(k). The total foregone transfer is therefore ϕmc. The proportion of the total 401(k) contribution that should not be counted as additional wealth is $\phi m/(1+m)$. The mean employer match in 1991, calculated from Hewitt As-

[4] It is also possible that employers' 401(k) contributions are funded by reducing eligible households' average wage or non-pension benefits. In that case, participating households will actually be receiving a higher true level of compensation than the ostensibly identical ineligible households, and part of the employers' contribution in 1991 should be added to eligibles' measured wage before income is used as a covariate in the statistical analysis. This adjustment, however, would be tiny, amounting to less than 1% of eligibles' income, and would leave the results essentially unchanged.

[5] An eligible worker with a DB pension may also have fewer DB assets than an otherwise similar ineligible worker with a DB pension if the eligibles' 401(k) is from his current job, and the DB plan is from an old job and has stopped accumulating benefits. That scenario would also justify the adjustment in this section.

sociates (1991) (ignoring 401(k) plans uncategorized by match rate and ignoring the fact that most matches are only up to a certain percentage of the employee's income), was 57%. Mean match rates from other studies range from 52% (Papke, 1995) to 62% (Bassett et al., 1998). In this analysis, it is assumed that all household contributions are matched 57%, without regard to the size of the contribution.[6] There is virtually no data from which to calibrate ϕ, so the main results are calculated under the arbitrary assumption that $\phi = 50\%$, although ϕ may well be closer to 0%. For $m = 57\%$ and $\phi = 50\%$, $\phi m/(1 + m) = 18\%$. Section 6.4 also calculates results without adjusting for marginal substitution (i.e. assuming $\phi = 0\%$) and under the assumption of complete offset (i.e. $\phi = 100\%$) and finds that the estimated eligibility effect is relatively insensitive to the assumed value of ϕ.

Within a firm whose benefits package includes both a 401(k) and a DB pension, it is very unlikely that the firm would reduce the DB benefits of 401(k) contributors only. All workers in the firm presumably share the loss in DB benefits. Thus, if a firm funds matches by reducing DB benefits, then total wealth should be adjusted downward to some degree for all its employees. Higher income workers are more likely to be 401(k) eligible and to contribute more on average when eligible, so firms with high-income workers probably would reduce DB benefits by more than firms with lower-income workers. In this analysis, the adjustment for employer matching is made proportional to income.

In the sample, 401(k) eligible households with a DB pension have mean income of \$47 364 and mean 401(k) asset holdings of \$10 486. The mean value of their non-converted 401(k) assets is $\$10\,486 \times 0.80 \approx \$8389$ per eligible household, or \$0.177 per dollar of income per household. Hence, $\$0.177 \times 18\% \approx \$0.032$ per dollar of income represents foregone DB assets for eligible households who have a DB plan. Thus, the adjustment for marginal substitution between DBs and 401(k)s is to subtract \$0.032 per dollar of income from the total wealth of each household who is both 401(k) eligible and has a DB plan.

[6]Fifty-seven percent is almost certainly an overestimate for several reasons. First, it assumes that all employers have matching provisions, but some do not. Some employers do not contribute at all to workers' 401(k)s, and it is plausible that the small proportion of employers who contribute to workers' 401(k)s without matching contribute less to workers' 401(k)s than employers who match. Second, and most importantly, 57% is an estimate of the average match on the first dollar contributed, not on the average dollar. Because most employers match contributions only up to a certain percentage of income, the average match on the last dollar is smaller than the average match on the first dollar. Furthermore, employers with more generous matching provisions tend to have lower matching limits. A small factor in the opposite direction is that, after finding themselves in possible violation of the non-discrimination provisions, some employers may increase their contribution to lower paid employees' 401(k)s *ex post* (Papke, 1995). One other inadequacy with this estimate of the average match (that biases the adjustment in an unknown direction) is that it is calculated from the percentage of firms with various arrangements rather than from the percentage of contributors with various arrangements.

4. Private versus national saving

Two main goals of tax-deferred saving incentives are increasing households' retirement savings and promoting capital accumulation. The effect of 401(k)s on retirement savings lies between their effect on current private savings and their effect on current national savings.[7] The impact of 401(k)s on capital accumulation, although complicated (see Hubbard and Skinner, 1996 for discussion and example calculations), is closely related to the national savings effect. Previous researchers, by not accounting for the fact that 401(k) and IRA assets are pre-tax balances, focused exclusively on current private savings. This section introduces an adjustment to the total wealth measure that converts a household's 401(k) and IRA assets from pre-tax to post-tax balances so that results can be reported for both private and national savings.

Suppose that this year, an eligible employee puts his new retirement savings of \$1000 into his 401(k) account, while an identical but ineligible employee puts his new retirement savings of \$1000 into a non-tax-advantaged mutual fund so that each has made the same contribution to current national savings. The eligible employee pays $\textit{tax rate}_{\text{income}} \times \$1000$ less in taxes than the ineligible employee, where $\textit{tax rate}_{\text{income}}$ is the marginal tax rate on the eligible employee's income. The eligible employee's current private savings will then exceed the ineligible employee's by exactly this amount. Subtracting these foregone taxes from 401(k) contributions is necessary to convert from a measure of private savings to national savings.

The 1991 marginal tax rates for the federal income tax were 15, 28, and 31%. For single individuals, the boundary incomes were \$20 350 and \$49 300, and for married individuals filing jointly, the boundary incomes were \$34 000 and \$82 150. All sample households coded as married were assumed to file jointly. This assumption is probably conservative, tending to put households in higher tax brackets and thereby overestimating the amount to be subtracted from savings measures.[8] On the other hand, because the Tax Reform Act of 1986 gradually reduced federal marginal income tax rates and because many 401(k) contributions were made prior to 1991 (when most households faced higher rates than in 1991),

[7]Because 401(k) assets are taxed at withdrawal, current 401(k) private savings exceeds the present value of 401(k) retirement assets. There are two reasons that the present value of 401(k) retirement assets almost surely exceeds current 401(k) national savings. First, taxation at withdrawal only partially recaptures interest earned by participating households on foregone taxes from time of contribution. Second, the household will probably face a lower marginal tax rate at retirement than at time of contribution.

[8]The cutoff points for married individuals filing separately were \$17 000 and \$41 075. If only one partner earns income, it is advantageous to file jointly. If both partners work, then it is usually (but not in all cases) better to file jointly. Assuming that the household files jointly is conservative when filing jointly would be advantageous. Even when it is disadvantageous financially, however, many households file jointly for legal reasons.

the federal income tax rates assumed here are probably underestimates. For all households, the state and local income tax rates together were assumed to be 5%, a figure above the lowest tax bracket rate for all states and below the highest tax bracket rate for most states.[9]

Transforming private savings to national savings requires multiplying both 401(k) assets and IRA assets by $(1 - \textit{tax rate}_{\text{income}})$ before incorporating them into total wealth, where $\textit{tax rate}_{\text{income}}$ is the sum of the marginal federal and state income tax rates for a household with a given income. This adjustment assumes that the marginal tax rate for a household at time of contribution is the same as the household's current marginal tax rate.

5. Methodology and propensity score estimation

5.1. Motivation for propensity score subclassification

Suppose the true equations determining a household's asset holdings can be written:

$$A_{i0} = f_0(X_i) + \epsilon_{i0} \quad \text{and} \quad A_{i1} = f_1(X_i) + \epsilon_{i1} \tag{1}$$

where A_{i0} is household i's holdings of the asset if i is 401(k) ineligible, A_{i1} is household i's holdings of the asset if i is 401(k) eligible, X_i is a covariate vector of i's observed characteristics, ϵ_{i0} and ϵ_{i1} are unobserved error terms, and f_0 and f_1 are smooth but potentially non-linear functions. Household i's observed asset holdings is $A_i = A_{i1} \times D_i^{\text{elig}} + A_{i0} \times (1 - D_i^{\text{elig}})$, where D_i^{elig} indicates whether i is eligible for a 401(k). The identifying assumption is that eligibility is conditionally random, given the observed covariates X_i. More precisely, (A_{i0}, A_{i1}) is independent of D_i^{elig}, given X_i; and $0 < \text{pr}(D_i^{\text{elig}} = 1|X_i) < 1$.[10]

The effect of 401(k) eligibility on savings, the 'average treatment effect' for the population, is $\tau = \text{E}(A_{i1} - A_{i0}) = \text{E}_X[\text{E}(A_{i1} - A_{i0}|X_i = x_i)]$: the treatment effect for

[9] Pennsylvania is the only state that does not defer taxes on 401(k) contributions, though many local governments also do not (Pare, 1995). Another source of noise affecting the correct household-specific adjustment factor is the fact that part of some 401(k) contributions might not be tax-deferred if the tax deferral would put the employer in violation of the non-discrimination provisions. Because Social Security taxes on 401(k) contributions are not deferred, no adjustment for them is necessary here.

[10] In the standard linear regression framework (3) considered below, this identification assumption implies the usual regression assumption that $\text{Cov}(D_i^{\text{elig}}, \epsilon_i) = 0$. Engen et al. (1996) have challenged this assumption on the grounds that unobserved saving preferences affect both wealth accumulation and 401(k) eligibility. The present paper does not address this concern, although controlling for household characteristics correlated with saving preferences should reduce the omitted variable bias, making the identification assumption more palatable. Holding observed household characteristics constant, sources of exogenous variation in workers' eligibility include industry standards, administrative expenses, and executives' retirement plan preferences. For a somewhat more general framework for propensity score methods, see Rosenbaum and Rubin (1983).

a household with characteristics $X_i = x_i$, integrated with respect to the population distribution of household characteristics. This quantity represents what the average effect of eligibility would be if the entire sample-age US population of households were eligible. Also of interest are the 'average treatment effect on the treated,' $\tau^{\text{elig}} = \text{E}(A_{i1} - A_{i0}|D_i^{\text{elig}} = 1) = \text{E}_{X|D=1}[\text{E}(A_{i1} - A_{i0}|X_i = x_i)]$, and the 'average treatment effect on the untreated,' $\tau^{\text{inelig}} = \text{E}(A_{i1} - A_{i0}|D_i^{\text{elig}} = 0) = \text{E}_{X\,|\,D=0}[\text{E}(A_{i1} - A_{i0}|X_i = x_i)]$. These are the treatment effects for a household with characteristics $X_i = x_i$, integrated with respect to the population distribution of currently eligible households' characteristics or of currently ineligible households' characteristics, respectively. τ^{elig} represents the historical effect that eligibility has had on eligible households. τ^{inelig} represents the counterfactual effect that eligibility would have had on currently ineligible households.

When f_0 and f_1 are not known, one way to proceed is to assume linear relationships:

$$A_{i0} = \alpha_0 + \beta_0' X_i + \epsilon_{i0} \quad \text{and} \quad A_{i1} = \alpha_1 + \beta_1' X_i + \epsilon_{i1} \tag{2}$$

estimate these equations, and then use the fitted values to estimate $\tau = \text{E}_X[\text{E}(A_{i1}|X_i = x_i) - \text{E}(A_{i0}|X_i = x_i)]$. Since $\text{E}(A_{i0}|X_i = x_i - x_0^*) = f_0(x_i - x_0^*)$ and $\text{E}(A_{i1}|X_i = x_i - x_1^*) = f_1(x_i - x_1^*)$, the fitted OLS regression equations, $\hat{\text{E}}(A_{i0}|X_i = x_i - x_0^*) = a_0 + b_0'(x_i - x_0^*)$ and $\hat{\text{E}}(A_{i1}|X_i = x_i - x_1^*) = a_1 + b_1'(x_i - x_1^*)$, can be understood as Taylor approximations to the true functional relationships, f_0 and f_1. These approximations are taken around the mean covariate vectors, $x_0^* = \text{E}(X_i|D_i^{\text{elig}} = 0)$ and $x_1^* = \text{E}(X_i|D_i^{\text{elig}} = 1)$. For median regression, the argument is similar, with expectation operators replaced by median operators and with $x_0^* = \text{med}(X_i|D_i^{\text{elig}} = 0)$ and $x_1^* = \text{med}(X_i|D_i^{\text{elig}} = 1)$. Thus, the fitted equations for (2) will be good approximations to (1) *locally* – that is, for ineligible and eligible households with characteristics similar to x_0^* and x_1^*, respectively.

Two distinct features of the household population can make (2) a poor approximation for (1), making estimates of τ based on (2) unreliable. First, suppose ineligible households (analogously for eligible households) are a heterogeneous group, i.e. their x_is vary considerably. Eq. (2) will not approximate (1) well far from x_0^*, so households with x_is far from x_0^* will generate large regression residuals, making them influential. In that case, the regression estimates will be sensitive to specification. Non-robustness across specifications is symptomatic of this problem. In fact, Table 1 indicates that both eligible and ineligible households are heterogeneous groups: standard deviations of household characteristics are large relative to means.

Second, suppose eligible and ineligible households are very different from each other, i.e. x_0^* is far from x_1^*. In that case, even if (2) generally approximates (1) well for most eligible and ineligible households considered separately, there may be relatively few x_is at which *both* $f_0(x_i - x_0^*) \approx a_0 + b_0'(x_i - x_0^*)$ and $f_1(x_i - x_1^*) \approx a_1 + b_1'(x_i - x_1^*)$ are good approximations. Treatment effect estimators will rely heavily on poor extrapolations. A diagnostic for this problem is covariate *balance*, the degree of similarity between the covariate distributions of eligible and ineligible households (Rosenbaum and Rubin, 1984; Dehijia and Wahba, 1998). The t-tests in Table 1 show that eligible and ineligible households are quite different from each other, indicating that estimation of (2) can provide misleading treatment effect estimates.

The standard estimate of the 401(k) eligibility effect would be the coefficient on the indicator for eligibility in the linear regression:

$$A_i = \alpha + \beta' X_i + \tau D_i^{\text{elig}} + \epsilon_i \tag{3}$$

This approach is a special case of (2), where the main additional assumption is that the treatment effect τ is constant across varying household characteristics: $\tau = \mathrm{E}(A_{i1} - A_{i0}) = \mathrm{E}(A_{i1} - A_{i0} | X_i = x_i)$ for all x_i (hence $\tau = \tau^{\text{elig}} = \tau^{\text{inelig}}$). This assumption would be approximately true in a well-balanced sample (where $x_0^* \approx x_1^*$) that is also homogeneous (so that most of the x_i are near $x_0^* \approx x_1^*$). In an unbalanced sample, the constant treatment effect assumption can make estimates of τ sensitive to specification. Engen et al. (1994) estimate a repeated cross-sections version of (3) by median regression, with $X = \{$age, age^2, income, income^2, age$\times$income, family size, education, two-earner indicator, DB indicator, male indicator, white indicator$\}$. In Section 6.4, it is shown that regression estimates of (3) for the present sample are highly sensitive to specification.

An alternative strategy, employed here, is to stratify the sample into subclasses $s = 1, \ldots, S$, estimate $\tau_s \equiv [\mathrm{E}(A_{i1}) - \mathrm{E}(A_{i0}) | i \in s]$ within each subclass, and then aggregate the subclass-specific estimates into a population-level estimate of τ according to:

$$\tau = \mathrm{E}_s[\mathrm{E}(A_{i1}) - \mathrm{E}(A_{i0}) | i \in s] = \sum\nolimits_s \tau_s \times \mathrm{pr}(i \in s) \tag{4}$$

The sample proportion of households in subclass s estimates $\mathrm{pr}(i \in s)$, and if households are sufficiently balanced and homogeneous within subclasses, then the τ_ss can be reliably estimated by linear regression (3). Note, however, that the coefficients on X in regressions within well-balanced subsamples of an unbalanced population lose their interpretation as the marginal effect of X on A. To understand this, notice that if households are very similar in their observed characteristics, then $\beta \approx 0$ in (3) regardless of which variables are included in X. Because β has no simple interpretation within non-representative samples, these coefficients are not reported in the results.

Poterba et al. (1995) stratified households by income and estimated each τ_s within income intervals ($<$\$10K, \$10–20K, \$20–30K, \$30–40K, \$40–50K, \$50–75K, and $>$\$75K in \$1987) by median regression of (3) with $X = \{$age, income, education, marital status$\}$. Unfortunately, as shown by t-statistics in the top panel of Table 2, dividing households by income interval leaves substantial differences between eligibles and ineligibles within subclasses. They differ especially in DB pension status, and, in several intervals, they remain unbalanced

Table 2
Balance: t-statistics (or z-statistics) for covariates within income intervals and estimated propensity score subclasses

Income intervals	Age	Income	Education	Family size	Two-earner status[a]	Marital status[a]	DB status[a]	IRA status[a]	Home owner[a]
1	1.15	3.45***	2.93***	−1.33	−0.87	−0.32	8.08***	3.85***	1.83*
2	−0.29	5.75***	1.55	−2.34**	−0.47	−2.51**	13.62***	−1.35	1.28
3	−0.50	3.60***	2.55**	−2.77***	−1.68*	−4.24***	10.22***	0.85	1.35
4	1.13	0.22	1.82*	−0.67	−0.08	−1.98**	7.41***	1.83	2.03**
5	0.71	0.72	0.35	−0.44	1.93*	−0.27	6.54***	0.39	3.99***
6	0.02	−0.06	−0.33	1.50	1.70*	0.24	3.74***	0.07	3.09***
7	−0.55	0.09	−0.28	−0.20	3.07***	2.10**	3.67***	2.36**	3.94***
Propensity score subclasses									
1	−0.17	8.03***	0.84	0.61	2.27**	2.69***	2.11**	1.50	3.57***
2	−0.83	−1.88*	−0.09	−0.31	−0.84	−0.46	0.89	−0.78	1.67*
3	−0.53	−0.84	0.25	−0.28	−0.59	−0.92	0.46	−0.35	1.97**
4	0.49	−0.42	0.13	0.03	0.23	−0.02	0.29	0.53	1.15
5	0.77	1.13	−0.05	−0.33	0.04	−0.29	−0.39	0.31	1.49
6	−0.42	0.52	0.42	0.04	1.36	−0.11	0.06	−1.03	1.71
7	−0.10	1.28	1.25	1.00	−0.77	0.98	−2.42**	1.35	0.54
8	−0.89	−0.66	0.67	−0.37	−0.11	−0.12	0.78	−0.46	0.14
9	1.02	0.17	−0.82	0.81	1.64	1.78	0.53	−0.18	2.30**
10	−0.25	0.77	−0.71	−0.20	−0.58	−0.57	–[b]	0.37	2.23**

Source: Author's calculations. Balance for a given covariate within an income interval or propensity score subclass is calculated by regressing the covariate on eligibility and a constant within the subclass. Eicker–White (heteroskedasticity-robust) standard errors are used to calculate t-statistics. The income intervals, calculated in 1987 dollars, are as in Poterba et al. (1995): <$10K, $10–$20K, $20–30K, $30–40K, $40–50K, $50–75K, and >$75K. Two households fall between intervals and are dropped from the analysis. The propensity score specification for which these t-statistics are calculated includes all first and second order covariates, including log-income but excluding home ownership status. Significance levels: * $P=0.10$, ** $P=0.05$, *** $P=0.01$.

[a] z-statistics are reported for logistic regressions.

[b] All ineligible households in Subclass 10 have DB plans.

even on income! These differences imply that Poterba et al.'s identifying assumption — that eligibility is random, given income — is false. Covariate balance could be improved by stratifying along several covariate dimensions at once, such as income and DB pension status. Unfortunately, splitting the sample across more covariate dimensions drastically shrinks the number of households within the subclasses, making inference within subclasses far less precise.

The solution adopted here is to form subclasses on the basis of households' estimated propensity scores. A household's *propensity score* is its probability of treatment, given its observed covariates, $p_i \equiv p(x_i) = \mathrm{pr}(D_i^{\mathrm{elig}} = 1 | X_i = x_i)$. Rosenbaum and Rubin (1983) prove that treated and untreated households matched on their propensity scores are balanced on their multivariate distribution of observed characteristics. That is, X_i is independent of D_i^{elig}, given p_i. Thus, balanced subclasses are a necessary consequence of subclassification on the true propensity score in large samples. Rosenbaum and Rubin (1983) also prove that if (as assumed here) eligibility D_i^{elig} is conditionally random, given X_i, then it is conditionally random, given p_i. Thus, the τ_ss in (4) can be consistently estimated by running linear regression (3) within propensity score subclasses. Moreover, (3) will be a much better approximation within balanced subclasses than for the sample as a whole. Propensity score subclassification does not guarantee homogeneity of x_is within subclasses, but, in practice, subclasses are generally less heterogeneous than the sample as a whole.

5.2. Propensity score estimation

Because propensity scores are not observed (except in randomized experiments), they must be estimated. Recall that the purpose of estimating propensity scores in this context is to generate balanced subclasses, so as long as eligible and ineligible households are similar within subclasses, the propensity score itself does not matter. Thus all choices about the propensity score specification are resolved by trying several specifications chosen as informed guesses and then selecting whichever specification produces the best covariate balance within subclasses (Dehijia and Wahba, 1998). This criterion makes the procedure essentially nonparametric.

In the present analysis, propensity scores were estimated by logistically regressing 401(k) eligibility on the covariates. Interestingly, specifications that included home ownership were no better and often much worse in terms of balance than specifications that excluded home ownership. This means that the factors determining 401(k) eligibility are very different for renters than for homeowners. Separate analyses for homeowners and renters in Section 6 confirm that these groups exhibit quite distinct saving behavior. Because including home ownership did not improve balance within subclasses, it was excluded from the propensity score model.

Several features of the data indicate that second order covariate terms should be included in the specification. For one thing, there is evidence of non-linearities in the relationship between covariates and eligibility. The probability of eligibility increases with income up to about \$170K and decreases thereafter (see Abadie's, 2000, Fig. 1), and the relationship between eligibility and income differs depending on whether the household has a DB pension (Engelhardt, 2000). Age is also non-linearly related to eligibility, with middle-aged households more likely to be eligible than either younger or older households (see Pence's, 2001, Table C). Furthermore, the unequal standard deviations of covariates between eligibles and ineligibles suggests the presence of still more non-linearities in the relationship between covariates and eligibility. Moreover, including second order covariates in the propensity score specification ensures that households within propensity score subclasses will be balanced on both the first and second moments of their covariate distributions.

Thus, the specification that produced the best overall balance included both income and log-income as covariates, excluded home ownership, and included squared terms and all interactions.[11] The main results of this paper are based on this specification, but, for comparison, some results are given in Section 6.4 for other specifications.

To examine covariate balance and to calculate the τ_ss in (4), households are stratified into ten estimated propensity score subclasses. More subclasses creates better homogeneity within subclasses and reduces bias in τ_s estimates but also results in fewer households in each subclass and therefore less precise τ_s estimates (see Du, 1998). Improvement in balance due to subclassification can be measured by 'percent bias reduction'; across a variety of standard distributions for the covariates, the large sample bias reduction for a particular covariate for five subclasses is about 90% and over 95% for ten subclasses (see Cochran, 1968; Rosenbaum and Rubin, 1984). Because there are more ineligible than eligible households in the sample, arbitrary propensity score intervals might contain too few eligible households for reliable inference; therefore, the propensity score intervals defining the subclasses are chosen here to generate a nearly equal number, 369 households, of eligibles within each subclass.

The bottom panel of Table 2 reports *t*-statistics as a measure of balance within propensity score subclasses. Outside Subclass 1 and excluding home ownership, eligible and ineligible households are nearly identical in their observed covariates.[12] Subclassification on the estimated propensity score has simultaneously balanced all covariates included in the propensity score model. Balance on the second order covariates is also excellent. Relative to the imbalances in the sample as a whole, propensity score subclassification reduces bias most in DB status (99%), IRA status (95%), education (94%), and income (94%). It reduces bias least in homeownership status (49%), family size (65%), and marital status (79%). The balance produced by propensity score subclassification is also clearly superior to that produced by income subclassification (the top panel of Table 2).

Eligible households have a mean estimated propensity score of 0.48, with a

[11]The logistic regression results that describe the propensity score model are available at http://www.economics.harvard.edu/~dbenjami/papers.html. Separate logistic regressions for homeowners and renters (excluding the homeownership terms) are also available.

[12]Propensity score subclassification often fails to balance covariates in the first and last subclasses. Here, balance is poor in Subclass 1 because there are a substantial number of ineligible households in that subclass with estimated propensity scores below the lowest eligible household's estimated propensity score. It is possible to improve balance in Subclass 1 by, for example, dropping ineligible households whose estimated propensity score falls below the minimum estimated propensity score among eligible households. Matching methods, though more complicated to implement, can achieve even better balance than subclassification.

standard deviation of 0.17. Ineligible households have a mean estimated propensity score of 0.31, with a standard deviation of 0.19. The fact that mean estimated propensity scores differ by a standard deviation confirms that the two groups are, in a multivariate sense, substantially different and reinforces the desirability of subclassification.

6. Results

6.1. Population-level eligibility effects on savings

Previous researchers have focused on estimating median 401(k) effects because of concerns that outliers would influence mean effect estimates (Poterba et al., 1995). Unless such households are misreported, however, they are important to include for assessing the mean effect of 401(k)s, knowledge of which is essential for evaluating the aggregate impact of 401(k)s. Moreover, the mean 401(k) effect can be compared with the median effect to infer how the distribution of household responses is skewed. Importantly, subclassification reduces the impact of an inappropriately influenced effect estimate by averaging it with effect estimates from other subclasses.

τ_s is estimated for each subclass $s = 1, \ldots, 10$ as the coefficient on eligibility in regression (3) within each subclass. Because some statistically significant differences in household characteristics remain between eligibles and ineligibles within subclasses — especially in home ownership — the independent variables in this regression (besides eligibility itself) are age, income, log-income, family size, marital status, two-earner status, DB status, IRA status, home ownership, and interactions of home ownership with each of the other independent variables (except eligibility). This specification allows household characteristics to affect asset accumulation differently for homeowners and renters but restricts the eligibility effect to be the same for homeowners and renters.

Columns 2 and 5 of Table 3 report results from OLS and median regression estimates of the τ_s values, respectively. These are estimates of the 401(k) eligibility effect on private savings for households with specific, non-representative distributions of characteristics. They have no natural economic interpretation. For example, even though households in higher subclasses have higher average incomes, the τ_ss are not comparable to eligibility effects within income intervals because households vary between subclasses on many characteristics. In fact, because the propensity score model includes squared income terms, many of the highest income households are in lower subclasses. Also, because the propensity score model includes interaction terms, the covariance structure of household characteristics affects the composition of the subclasses.

The τ_s standard error estimates reported in Table 3 would be unbiased if the subclasses were formed using the true propensity scores. Perhaps surprisingly,

Table 3
Regression results within propensity score subclasses

Propensity score subclass	Mean effect: Private savings effect	Mean effect: Mean 401 (k) assets	Mean effect: % NS for avg. HH	Median effect: Private savings effect	Median effect: Median 401 (k) assets	Median effect: % NS for avg. HH
1	−$1852	$4036	−46	$706***	$600	118
	($2922)		(72)	($194)		(32)
2	$3099	$6219	50	$3474***	$1200	289
	($3982)		(64)	($740)		(62)
3	$6899	$8129	85	$1901*	$1400	136
	($6510)		(80)	($1123)		(80)
4	$2859	$8650	33	$2279	$2000	114
	($5114)		(59)	($1914)		(96)
5	$3045	$11 652	26	−$907	$2500	−36
	($8438)		(72)	($1896)		(76)
6	$10 444	$11 072	95	$1200	$3000	40
	($7591)		(69)	($2417)		(81)
7	$3316	$12 043	28	$3181	$3500	91
	($6810)		(57)	($2757)		(79)
8	$5163	$12 658	41	$6851**	$3000	228
	($7184)		(57)	($3306)		(110)
9	$650	$12 625	5	$4994	$4500	111
	($11 456)		(91)	($3358)		(75)
10	$19 892[a]	$20 679	96	$15 124**	$9000	168
	($32 862)		(159)	($6788)		(75)
Overall effect ($\hat{\tau}$)	$3434	$8623	23	$2738***	$2183	131***
	($2444)		(28)	($576)		(22)
Eligibles' effect ($\hat{\tau}^{elig}$)	$5370	$10 772	41	$3895***	$3069	127***
	($3921)		(26)	($953)		(25)
Ineligibles' effect ($\hat{\tau}^{inelig}$)	$2310	$7375	13	$2067***	$1669	134***
	($1931)		(34)	($392)		(22)
Households with total wealth ≤ $100 000 only						
Overall effect ($\hat{\tau}$)	$1795***	$4747	37**	$1819***	$1203	180***
	($640)		(16)	($424)		(42)
Eligibles' effect ($\hat{\tau}^{elig}$)	$1986***	$5404	36***	$2222***	$1586	176***
	($688)		(13)	($658)		(51)
Ineligibles' effect ($\hat{\tau}^{inelig}$)	$1702**	$4429	37**	$1624***	$1017	182***
	($679)		(18)	($334)		(40)

Source: Author's calculations. Mean and median private savings effects are calculated by OLS and median regression, respectively, of benchmark-adjusted private total wealth on the covariates, as well as on interactions of home ownership status with age, income, education, family size, log-income, and IRA status. Mean effects report Eicker–White (heteroskedasticity-robust) standard errors in parentheses, while median effects report bootstrapped standard errors (using 200 repetitions). Mean and median % new saving calculations and population average effect estimates are described in the text. Significance levels: * $P=0.10$, ** $P=0.05$, *** $P=0.01$.

[a] Home ownership is excluded as a right-hand side variable in Subclass 10 due to collinearity.

because households were subclassified on the basis of estimated propensity scores, the reported standard errors are probably *overestimated* (Rubin and Thomas, 1992). To understand this, consider the t-statistics measuring balance reported in Table 2. If households had been subclassified on the basis of their true propensity scores — that is, if households had been randomly assigned to be eligible or ineligible — then about 5% of the balance t-statistics would have been statistically significant at the 5% level (just by chance). In fact, outside Subclass 1 and excluding home ownership, only eight t-statistics were significant out of nine subclasses $\times$ 50 first and second order covariates $=$ 450 calculated t-statistics — about 1.8%, far fewer than 5%. Thus, because the propensity score estimation procedure overfit the data, households within subclasses are more similar on observables (besides home ownership) than they would have been under unconditional random assignment. This reduces the chance that the effect estimates are influenced by random covariate imbalances, raising the estimates' precision. Nonetheless, the standard errors are large relative to the effect estimates, indicating substantial heterogeneity in asset accumulation behavior even among households with similar observed characteristics.

Columns 4 and 7 of Table 3 report the mean and median percentage of 401(k) balances that are new private savings, respectively. These percentages and standard errors are calculated by dividing the private wealth effect and standard error within each subclass by the mean or median eligibles' holding of 401(k) assets (columns 3 and 6) within that subclass, respectively.

The first row in the second panel of Table 3 presents the population-level 'average treatment effect' ($\tau^\wedge$), average 401(k) assets, and new private savings percentage. Recall from (4) that $\tau = \Sigma_s \tau_s \times \mathrm{pr}(i \in s)$. Here, to correct for the SIPP's undersampling of wealthy households, $\mathrm{pr}(i \in s)$ is estimated as the sum of households' population weights in subclass s divided by the sum of households' population weights in the sample. The mean household's new private savings percentage is calculated from the subclass-specific estimates analogously, $\%ns = \Sigma_s \%ns_s \times \mathrm{pr}(i \in s)$.

The standard error for $\tau^\wedge$ is estimated according to $\mathrm{Var}(\tau^\wedge) = \mathrm{Var}[\Sigma_s \tau_s^\wedge \times \mathrm{pr}(i \in s)] \geq \Sigma_s \mathrm{Var}[\tau_s^\wedge \times \mathrm{pr}(i \in s)] = \Sigma_s [\mathrm{pr}(i \in s)]^2 \times \mathrm{Var}(\tau_s^\wedge)$, which would hold with equality if the $\tau_s^\wedge$ values were independent across subclasses. They are not independent because both the $\tau_s^\wedge$ values and the subclassification itself (through the propensity score estimation) depend on the particular sample of households drawn from the population. As a result, the standard error aggregation is biased downward, though not by much (see Du, 1998). Use of regression to control locally for covariates within subclasses substantially mitigates this bias by reducing the dependence of the $\tau_s^\wedge$ values on the sample.

The 'average treatment effect on the treated' estimate ($\tau^{\mathrm{elig}\wedge}$) and standard error are calculated using the fact that $\tau^{\mathrm{elig}} = \Sigma_s \tau_s \times \mathrm{pr}(i \in s | D_i^{\mathrm{elig}} = 1)$; similarly for the 'average treatment effect on the untreated.' The new savings percentage for eligibles and ineligibles are also calculated analogously, $\%ns^{\mathrm{elig}} = \Sigma_s \%ns_s \times$

$\text{pr}(i \in s | D_i^{\text{elig}} = 1)$ and $\%ns^{\text{inelig}} = \Sigma_s \%ns_s \times \text{pr}(i \in s | D_i^{\text{elig}} = 0)$. Estimates of these quantities are given in the last two rows of the second panel of Table 3.

For the population median treatment effect, $\tau^{\text{med}} = \text{med}(A_{i1} - A_{i0})$, it is not correct to calculate the overall effect on median private savings by taking a weighted average of the median subclass effects: $\tau^{\text{med}} \neq \text{E}_s [\tau_s^{\text{med}} | i \in s]$, so these 'median estimates' should not be taken literally. Nonetheless, estimating the 'median' in this way gives a sense for the *relative* magnitude of the population-level median compared with the analogously calculated population-level mean. The 'median estimates' reported here are *not* actually estimates of τ^{med}.

One main result in this paper is that the percentage of 401(k) assets that represent new private savings is substantially smaller for the mean household than for the median household. Since eligibles' mean 401(k) balances are so much larger than their median balances, this result implies that a minority of eligible households — probably those with the most financial sophistication or the greatest taste for saving — accumulate relatively large stocks of 401(k) assets, relatively little of it new savings.

The last panel of Table 3 reports the population-level results when the analysis excludes wealthy households (those with total wealth greater than \$100 000), who are most likely to influence mean effects. As expected, average 401(k) balances and 401(k) eligibility effects measured in dollars are smaller for this sample, and this population's mean effects are estimated much more precisely. The basic pattern persists, however: the percentage of 401(k) assets that represent new private savings is substantially smaller for the mean than for the median household.

Another set of main results addresses the composition of aggregate 401(k) assets. To measure the proportion of the average 401(k) dollar that is new private savings requires weighting the subclass estimates by average subclass 401(k) balances, $\Sigma_s \%ns_s \times \text{pr}(401(\text{k})\$ \in s)$. This figure, 50%, is reported in the first row of Table 4. The analogous figure for post-tax total wealth (applying the adjustment from Section 4) — the proportion of the average 401(k) dollar that is new *national* savings — is 24% (not shown in Table 4). The difference, about a quarter of 401(k) assets, is foregone tax revenue. Furthermore, the analogous figure for unadjusted total wealth (not shown) is 75%, suggesting that about a quarter of aggregate 401(k) assets are conversions (20 percentage points by assumption in Section 3) or marginal firm-level substitution with DB assets (5 percentage points as a result of the benchmark assumption in Section 3). The remaining quarter of aggregate 401(k) balances presumably represents household-level substitution from other assets classes. Particular sources of substitution are examined in Section 6.3.

6.2. Eligibility effects on savings for subpopulations

This section addresses how the average household's new savings percentage differs across different kinds of households. For example, the effect estimates are

Table 4
Overall 401(k) eligibility effect estimates for subpopulations

	Mean effect				Median effect		
	Mean 401(k) assets	% NS for avg. HH	% Net non-IRA-401(k) Fin. assets	% NS for avg. Dollar	Median 401(k) assets	% NS for avg. HH	% Net non-IRA-401(k) Fin. assets
Whole sample							
All households	$8623	23	8	50	$2183	131***[g]	3[g]
		(28)	(14)	(36)		(22)	(5)
Total wealth ≤ $100 000	$4747	38**	−7	38***	$1203	180***	6[h]
		(16)	(8)	(13)		(42)	(9)
Homeownership[a]							
Homeowners	$10 983	0	4	32	$3281	191**	12
		(30)	(16)	(25)		(79)	(8)
Renters	$4076	101***	−17	99**	$702	216***	30
		(39)	(17)	(47)		(64)	(20)
IRA Status[b]							
Households with IRA	$16 253	35	−13	19	$6308	81	37
		(43)	(23)	(36)		(83)	(27)
Households without IRA	$6181	39	24	64***	$1185	220***	−13
		(27)	(14)	(24)		(33)	(8)
Age[c]							
Under age 40	$5245	62	15	55	$1479	130***	6
		(42)	(26)	(38)		(41)	(10)
Age 40 and up	$5457	21	−4	33	$3282	141***	−10[i]
		(31)	(17)	(27)		(41)	(12)
Defined benefit[d]							
Households with DB	$9565	76**	15	80**	$2881	399*	−40
		(31)	(15)	(34)		(239)	(42)
Households without DB	$8269	23	−4	10	$2539	120***	8
		(32)	(19)	(32)		(32)	(8)
Income[d]							
≤$30 000	$4043	66	−10	63	$688	187***[f]	21*[f]
		(49)	(18)	(44)		(53)	(11)
>$30 000	$12 574	47	3	39	$3872	142***	−5
		(28)	(17)	(26)		(42)	(10)
Family size[d]							
Three or more	$8290	−26	−14	18	$2190	134***	0
		(44)	(22)	(34)		(49)	(10)
One or two	$8878	52	17	69**	$2330	151***	8[i]
		(37)	(18)	(29)		(39)	(13)

Table 4. Continued

	Mean effect				Median effect		
	Mean 401(k) assets	% NS for avg. HH	% Net non-IRA-401(k) Fin. assets	% NS for avg. Dollar	Median 401(k) assets	% NS for avg. HH	% Net non-IRA-401(k) Fin. assets
Education[d]							
Finished high school or less	\$7023	34 (36)	12 (13)	54 (43)	\$1504	191***[g] (46)	−5 (10)
More than high school	\$10 396	38 (34)	−1 (21)	49 (31)	\$2485	91** (41)	8 (12)
Marital status[e]							
Married	\$10 093	12 (36)	−4 (17)	36 (27)	\$2841	120*** (37)	−3 (9)
Unmarried	\$6407	17 (33)	20 (26)	41 (41)	\$1441	182*** (64)	−11 (19)
Two-earner[e]							
Two earners	\$9995	52 (33)	−11 (23)	37 (35)	\$2878	170*** (50)	−13 (13)
Single earner	\$7912	28 (33)	16 (19)	56* (30)	\$1880	125*** (26)	9 (7)

Source: Author's calculations. Right-hand side variables in the within-subclass regressions are the first order covariates, including home ownership status and its interactions with age, income, education, family size, log-income, and IRA status, except as noted below. Mean effects report aggregated Eicker–White (heteroskedasticity-robust) standard errors in parentheses, while median effects report aggregated bootstrapped standard errors (using 200 repetitions). 'Median' results are *not* quantitative median effect estimates, but they provide a qualitative basis for comparison with the mean estimates. Significance levels: * $P=0.10$, ** $P=0.05$, *** $P=0.01$.

[a] Right-hand side variables exclude defined benefit pension because of collinearity, as well as all home ownership terms.

[b] Right-hand side variables exclude defined benefit pension status, marital status, two-earner status, and home ownership × family size because of collinearity, as well as IRA status.

[c] Right-hand side variables exclude defined benefit pension status.

[d] Right-hand side variables exclude marital status because of collinearity, as well as defined benefit pension status.

[e] Right-hand side variables exclude defined benefit pension status, two-earner status, and home ownership × IRA status because of collinearity, as well as marital status.

[f] Calculated as overall effect and standard error divided by overall 401(k) assets because eligibles' median 401(k) assets in subclass 9 are zero.

[g] Home ownership is excluded as a right-hand side variable in subclass 10 due to collinearity.

[h] Home ownership × DB status is excluded as a right-hand side variable in subclass 10 due to collinearity.

[i] Two-earner status, IRA status, and home ownership terms are excluded as right-hand side variables in subclass 1 due to collinearity.

calculated separately for homeowners and renters, households with an IRA and households without an IRA, and so on. Because the median sample household is age 40, earns about $30 000, has finished high school, and has a family size of three, the sample is split at these values.

To calculate the subpopulation effects, each subsample is newly divided into ten subclasses based on the estimated propensity scores with an approximately equal number of eligible households in each subclass, just as was done for the combined sample. Average effect estimates, standard errors, and 401(k) assets are calculated by weighted averaging of subclass-specific estimates, as before. Table 4 displays the subpopulation estimates.

Comparing the new savings percentages for the mean and median households, the main differences arise across homeownership status, DB status, IRA status, and educational attainment. The median household's percentage of 401(k) assets that are new savings is greater for less educated households than for more educated households and is also greater for households without an IRA than for households with an IRA. These results support the view that 401(k)s are more effective as saving incentives for less financially sophisticated or less patient households. A high percentage of 401(k) assets are new savings for both the median homeowner and the median renter, but the mean results for these group differ dramatically: on average, all renters' 401(k) assets are new savings, but all homeowners' 401(k) assets would have been saved anyway. Similarly, the median household with a DB pension and the median household without both have high new savings percentages, but mean new savings is higher for households with DBs.

Since 401(k) eligibility presumably affects saving behavior only for households who participate, a particularly high participation rate could explain a subpopulation's large eligibility effect relative to that of its complement even if participation effects are the same. Contrary to this possibility, eligible homeowners, IRA holders, and more educated households have *higher* participation rates than eligible renters, non-IRA holders, and less educated households, respectively, and the rates across DB status are very close. Therefore, the strong eligibility effects for these groups imply large 401(k) participation effects on saving behavior.

6.3. Substitution between assets classes

Repeating the same procedure as above (estimating $\Sigma_s \% ns_s \times \mathrm{pr}(401(\mathrm{k})\$ \in s)$), but using net non-IRA-401(k) financial assets as the dependent variable, estimates the impact of 401(k) eligibility on non-IRA-401(k) financial savings. Multiplying the resulting number by -1 provides an estimate of the percentage of 401(k) assets that would have been saved as non-IRA financial assets in the absence of 401(k) eligibility. Table 4 reports these percentages for the mean and median household. These estimates are not generally statistically different from zero, suggesting little substitution between 401(k)s and non-IRA financial assets.

Among IRA holders, conducting the same type of analysis is possible with IRA assets as the dependent variable (not shown in Table 4). For this group, neither the mean nor the median percentage of 401(k) assets that represent foregone IRA saving is statistically different from zero, suggesting virtually no substitution between these two tax-advantaged retirement saving programs.

Among homeowners, home equity and its components, home value and home mortgage, can function as dependent variables (not shown in Table 4). For homeowners, the percentage of the average 401(k) dollar that represents home equity reduction is about 19%. This home equity reduction appears to be a combination of reduced home value and increased home mortgages.

6.4. Covariates, balance, wealth adjustments, and the 401(k) effect estimate

The empirical strategy pursued here diverges in three main ways from those of previous researchers: households are stratified into balanced subclasses, covariates include additional observed household characteristics, and total wealth is adjusted for conversions and marginal firm-level substitution of pension benefits. Table 5 calculates estimates of the 401(k) eligibility effect on total wealth with a variety of estimation methods in order to understand how each of these factors influences the results presented above. Each panel of Table 5 represents a category of estimation method, and each is discussed individually here. All methods use unadjusted total wealth as the dependent variable until the bottom panel, which compares adjustments. The columns of Table 5 compare estimates as the regressors expand from $X = \{$age, income, education, marital status$\}$ alone (the covariates used by Poterba et al., 1995) to include, by the last column, $X = \{$age, income, education, marital status, log-income, family size, two-earner status, DB pension status, IRA participation status, home ownership status and its interactions with the preceding covariates$\}$ (the full set of covariates used in this paper).

In each row, the estimates in the last column of Table 5 are, practically without exception, smaller than those in all the other columns. This striking observation reinforces the conclusion that controlling fully for home ownership is crucial. Otherwise, because homeowners are both wealthier and more likely to be eligible than renters, 401(k) effect estimates are spuriously inflated.

The first panel of Table 5 shows 401(k) eligibility effect estimates calculated with standard regressions on the whole sample, which is essentially the strategy pursued by Engen et al. (1996). OLS regression is highly sensitive to the regression specification, which is not surprising for an unbalanced and heterogeneous sample. Excluding wealthy households from the sample eliminates outliers, improving precision and robustness to specification, but may bias downward estimates of 401(k)s' aggregate impact. Median regression is less sensitive to specification than OLS and also generates smaller estimates, as expected when the estimates are in dollars rather than percentages.

The second panel displays estimates calculated by (4), where households are

Table 5
Comparison of estimation methods and wealth adjustments

Estimation method	Covariates in total wealth regression: {age, inc, educ, marr}	+{log-inc, fsize, twoearn}	+{db}	+{Dira}	+{homeown interactions}
Whole-sample regression					
OLS regression	\$2233	\$6261**	\$7510**	\$6341**	\$3636*
	(\$2024)	(\$2011)	(\$2053)	(\$1993)	(\$1919)
OLS regression for total wealth ≤\$100 000	\$5873***	\$5650***	\$5535***	\$5367***	\$3659***
	(\$680)	(\$683)	(\$698)	(\$678)	(\$611)
Median regression	\$3810***	\$4705***	\$5086***	\$5075***	\$2648***
	(\$1003)	(\$771)	(\$914)	(\$744)	(\$305)
Median regression for total wealth ≤\$100 000	\$4742***	\$5127***	\$5154***	\$4448***	\$2557***
	(\$534)	(\$550)	(\$553)	(\$571)	(\$177)
Regression within income intervals					
OLS regression	\$6191***	\$7714***	\$8013***	\$6745***	\$4762**
	(\$1970)	(\$1944)	(\$1930)	(\$1854)	(\$1862)
Median regression	\$6874***	\$5472***	\$17692***	–[a]	–[a]
	(\$1206)	(\$1318)	(\$1093)		
Median regression for total wealth ≤\$100 000	\$5242***	\$5179***	\$5829***	\$2886***	–[a]
	(\$905)	(\$935)	(\$931)	(\$883)	
OLS within propensity score subclasses: Comparing propensity score specifications					
{age, inc, educ, marr}	\$5849**	\$7457***	\$8042***	\$6871***	\$4766**
	(\$1942)	(\$1906)	(\$1896)	(\$1844)	(\$1851)
2nd order covariates but DB & IRA status terms	\$8319***	\$8320***	\$8362***	\$8290***	\$5303**
	(\$2053)	(\$2038)	(\$2021)	(\$1967)	(\$1914)
2nd order covariates but IRA status terms	\$8651***	\$8647***	\$8666***	\$8607***	\$5463
	(\$2037)	(\$2014)	(\$2011)	(\$1955)	(\$35546)
All 2nd order covariates	\$8645***	\$8716***	\$8744***	\$8640***	**\$5506****
	(\$2044)	(\$2020)	(\$2015)	(\$1960)	**(\$2455)**
All 2nd order covariates+ {homeownership interactions}	\$8817***	\$8878***	\$8765***	\$8562***	\$5500**
	(\$2013)	(\$1985)	(\$1983)	(\$1927)	(\$1916)
OLS within propensity score subclasses (main specification): Comparing wealth adjustments					
Non-converted assets and $\phi = 0\%$	\$6925***	\$7000***	\$7027***	\$6925***	**\$3807**
	(\$2035)	(\$2010)	(\$2006)	(\$1952)	**(\$2449)**
Non-converted assets and $\phi = 50\%$	\$6551***	\$6628***	\$6655***	\$6552***	**\$3434**
	(\$2035)	(\$2010)	(\$2006)	(\$1951)	**(\$2444)**
Non-converted assets and $\phi = 100\%$	\$6177***	\$6255***	\$6282***	\$6180***	**\$3062**
	(\$2034)	(\$2010)	(\$2006)	(\$1951)	**(\$2440)**

Source: Author's calculations. Results for regression within income intervals are aggregated within-interval OLS estimates ($\tau_s \times$ values), as described in the text; similarly for results for regression within propensity score subclasses. All results use the unadjusted total wealth measure, except for the wealth adjustment comparisons. The main results for the present paper are in bold. Eicker–White (heteroskedasticity-robust) standard errors in parentheses. Significance levels: * $P=0.10$, ** $P=0.05$, *** $P=0.01$.

[a] Median regression failed to converge in at least one income interval.

subclassified by income interval, as in Poterba et al. (1995) and Engen and Gale (2000). These estimates are generally more robust to specification than their whole-sample counterparts and are also substantially larger.

The third panel displays estimates calculated by (4), where households are subclassified by estimated propensity score, and the rows represent various propensity score models. Propensity scores estimated from age, income, education, and marital status only (as in Abadie, 2000) generate results essentially the same as those where income alone determines the subclasses. Including additional covariates as well as second order terms (squares and interactions) in the propensity score model improves balance considerably, generating results that are extremely robust to within-subclass regression specification. These effect estimates are also larger than those produced by a less balanced subclassification scheme (such as on income) or by regressions on the whole sample. The bold estimate, \$5506, is the 401(k) eligibility effect estimate for unadjusted total wealth generated by the propensity score model and within-subclass regressors used in this paper.

The last panel shows how estimates using the main propensity score model change when the total wealth measure is adjusted to varying degrees. The bold estimates focus on the within-subclass regressors used in this paper. All rows include the adjustment for conversions. Poterba et al. (2001) argue that there is little substitution between 401(k) assets and DB pension benefits, which suggests that the $\phi = 0\%$ assumption may be the most reasonable. The first row estimate of \$3807 corresponds to $\phi = 0\%$, adjusting for conversions but not marginal firm-level substitution of pension benefits. Comparing this figure with the unadjusted estimate of \$5506 shows that the conversion adjustment reduces the dollar-measured effect estimate substantially. The second row estimate of \$3434 adjusts for conversions and assumes that $\phi = 50\%$, the benchmark assumption. The third row shows that the conversion adjustment combined with even the most extreme assumption that $\phi = 100\%$ produces a point estimate, \$3062, that remains positive. The adjustment for marginal firm-level substitution from DB benefits into 401(k) contributions has relatively little impact on the 401(k) effect estimate. The estimated percentage of 401(k) assets that are new national savings (not shown) is 24% under the $\phi = 50\%$ assumption, similar to the estimated 29% and 19% under the $\phi = 0\%$ and $\phi = 100\%$ assumptions, respectively.

The benchmark estimate of \$3434 represents the average increase in private savings accumulated by 1991 that would be experienced by a member of the population due to eligibility. Because 401(k) contributions comprised around 12% of 401(k) assets in 1991 (from Table 6), a population member made eligible would have increased current private savings *in 1991* by around $\$3434 \times 12\% \approx \$400$ per household. Since 70.5% of eligible households in the sample participated, a very rough estimate of the 1991 increase in private savings per participant is $\$400 \div 70.5\% \approx \$550$. An analogous calculation puts the figure for national savings at around \$250 per participant.

Table 6
Nominal aggregate assets, contributions, benefit payments, and rate of return for 401(k)-type plans: 1984–1996

Year	Assets (in \$millions)	Contributions (in \$millions)	Benefits (in \$millions)	Rate of return (%) on DC plans[a]	Direct assets (in \$millions)	Direct benefits (in \$millions)	Direct assets ÷ assets (%)
1984	91 754	16 291	10 617	11.0[b]	18 083	0	20
1985	143 939	24 322	16 399	18.4	46 381	3232	32
1986	182 784	29 226	22 098	13.3	77 595	7121	42
1987	215 477	33 185	22 215	4.8	106 214	9431	49
1988	276 995	39 412	25 235	13.1	150 635	12 439	54
1989	357 015	46 081	30 875	9.4	196 838	16 790	55
1990	384 854	48 998	32 028	3.5	236 164	17 658	61
1991	*440 259*	*51 533*	*32 734*	*15.1*	*308 019*	*20 087*	*70*
1992	552 959	64 345	43 166	9.8	375 696	30 200	68
1993	616 316	69 322	44 206	9.8	455 651	30 035	74
1994	674 681	75 878	50 659	3.8	512 851	37 453	76
1995	863 918	87 416	62 163	20.0	663 618	47 253	77
1996	1 061 493	103 973	78 481	15.0	813 402	60 285	77

Source: Columns 2–5 are from the US Department of Labor's Abstract of Form 5500 Annual Reports, Tables E23 and E24, 1999. Columns 6–8 are calculated by the author as described in the text.

[a] This is the rate of return earned by defined contribution plans with 100 or more participants, calculated under the assumption that all cash flows occur in the middle of the plans' reporting period.

[b] This value is not reported by the Department of Labor, so the mean rate of return of 11.0% for the years 1985–1995 is used.[13] The mean rate of return for 1985–1991, 11.1%, is about the same.

7. Summary and discussion

The results in this paper suggest that about one quarter of aggregate 401(k) assets represent new national savings; one quarter, foregone tax revenue; and one quarter, conversions from pre-existing DC plans or foregone DB assets. The remaining quarter presumably represents household-level substitution from other asset classes, largely home equity reduction among homeowners. These aggregate figures, however, mask enormous heterogeneity in saving behavior. Mean percentages of 401(k) balances that are new savings are typically much smaller than median percentages, indicating that the distribution of responses to eligibility is skewed, with a minority of financially savvy or patient households shifting assets they would have saved anyway into tax-advantaged accounts. Observably distinct subpopulations can act quite differently. For example, in the aggregate, all of renters' 401(k) balances are new savings but none of homeowners' are. In fact,

[13]The correlation between the rates of return reported in this table and the prime rate for 1985–1991 is near zero, as is the correlation between these rates of return and the percent change in the S&P 500 index, so these other rates of return cannot be used to estimate the rate of return on DC plans in 1984. The latter correlation of zero is especially surprising, given that a large proportion of DC assets are invested in equities; this finding merits further investigation.

homeowners and renters are such different groups that different factors seem to determine their likelihoods of eligibility. Large standard errors in 401(k) effect estimates also partly reflect the widespread heterogeneity in saving behavior. Previous work has paid insufficient attention to this *observed* heterogeneity.

Findings here on average responses to 401(k)s differ from those of previous research because regression is used within balanced subclasses instead of on the complete unbalanced sample, a large number of covariates are controlled for (with homeownership being especially important), and the measure of households' wealth is adjusted to ameliorate deficiencies in the SIPP's measure. Nonetheless, Engen and Gale (2000), who subclassify on income and also control for many covariates, find that 401(k)s are more effective for renters and households without an IRA than for homeowners and IRA holders, a similar pattern of results as here.

This paper suffers from several limitations. First, the overriding concern in this literature has been that unobserved preferences for saving may cause both asset accumulation and 401(k) eligibility, thereby biasing upward eligibility effect estimates (see, e.g. Engen et al., 1996). Controlling for a large number of correlates of saving preferences, such as DB pension status and IRA status, as done here, reduces but does not eliminate this concern. Pence (2001) shows that eligibility is correlated with self-reported measures of saving preferences even after controlling for many observed household characteristics. Second, households self-report 401(k) eligibility and total wealth in the SIPP data used here, but Engelhardt (2000) shows that lower income households underreport eligibility whereas higher income households overreport it. This pattern of measurement error biases upward estimates of the 401(k) effect on wealth. Third, because 401(k) participation induces households to increase their holdings of financial assets, the difference in wealth between eligibles and ineligibles may reflect better returns on financial versus non-financial assets rather than increased saving by eligibles.[14] If differential returns wholly explained 401(k) eligibility effects, then subpopulations with larger 401(k) balances would have larger new savings percentages, which is generally opposite to the results here. Nonetheless, the argument undoubtedly has validity. Finally, the wealth adjustments used here rely on skeletal information about conversions and virtually no data about marginal firm-level pension substitution. Mistaken assumptions crucial for the wealth adjustments could bias eligibility effect estimates. Importantly, however, the analysis here suggests that deficiencies in the SIPP's wealth measure cannot by themselves explain away the positive estimated average 401(k) effect.

Some of these limitations could be addressed by applying the methods used in this paper to the Survey of Consumer Finances, which includes data on self-reported measures of saving preferences (Pence, 2001), or to the Health and Retirement Study, which includes firm-reported 401(k) eligibility and balances

[14] I thank Karen Pence for suggesting this idea to me.

(Engelhardt, 2000). It would also be of interest to employ the methods in this paper to more recent data. Finally, since it is presumably 401(k) participation and not mere eligibility that affects saving behavior, it would be especially interesting to estimate directly the effect of 401(k) participation on saving. The difficulty is that households choose whether to participate, and this choice almost certainly depends strongly on unobserved saving preferences. Abadie (2000) proposes a propensity score-based method to estimate the effect of 401(k) participation on participating households in which the key identifying assumption is the same as the one used here: that 401(k) eligibility is conditionally random, given the observed covariates. Abadie in fact applies his method to address the 401(k) participation effect, but his application is primarily illustrative, with his propensity score model including only a few covariates and leaving substantial imbalances between eligibles and ineligibles and with his dependent variable including financial assets only.

The results here indicate that some types of households that typically save the most (homeowners, IRA participants, the better educated) respond least to 401(k)s. Many of these households do not increase and may reduce national savings in response to 401(k) eligibility by shifting funds they would have saved anyway into tax-advantaged 401(k) accounts. Such households increase their own current or retirement consumption by reducing the taxes they pay rather than by saving more, effectively forcing a greater tax burden onto households less able or willing to exploit tax-advantaged saving. This outcome seems inconsistent with the goals of a subsidized saving policy. One way to address these concerns would be to target tax-based saving incentives to the types of households who respond by actually saving more. Paradoxically, many of these households normally save little and are relatively less likely to participate in 401(k)s when eligible. Of course, confident policy recommendations require further investigation of how and why households differ in their responses to saving incentives.

Acknowledgements

I thank my advisors from various stages of this project, Tim Besley, Javier Hidalgo, David Laibson, and Donald Rubin, for their insight, encouragement, and patience. Steven Venti provided me with the data, and I am grateful to Joel Dickson, Gary Engelhardt, Richard Ippolito, and Leslie Papke for helpful e-mail correspondence. I am grateful to John Barnard, Dan Beller, Constantine Frangakis, Bill Gale, Ed Glaeser, Jennifer Hill, Richard Johnson, Michael Larsen, Daniele Paserman, Jack Porter, Jim Poterba, Jeremy Tobacman, David van Dyk, Steven Venti, seminar participants at the LSE and the Brookings Institution, and two anonymous but exceptionally thorough referees for valuable advice and comments on previous drafts. I thank Harvard University, the London School of Economics, and the Suntory and Toyota International Centres for Economics and Related

Disciplines (STICERD) for computing facilities and other resources, and I thank the Harvard College Research Program (Abramson Grant), Harvard's Department of Economics (Duesenberry Grant), and the British Marshall Scholarship for providing financial support. All errors are mine.

Appendix A. Estimating the aggregate proportion of non-converted 401(k) assets

There are two ways that funds can enter an employee's 401(k) account: direct contribution by the employee or employer, or conversion from a pre-existing DC plan. Funds exit 401(k) accounts in the form of benefits if a qualifying contributor chooses to withdraw assets. Suppose there were data for each year t on the dollar value of assets at the end of year t, $assets_t$; the dollar value of direct contributions made during year t, $contributions_t$; the dollar value of benefits paid out of the 401(k) account during year t, $benefits_t$; and the rate of return earned by 401(k) assets over year t, r_t. The evolution of a household's 401(k) asset value is given by the accounting identity:

$$assets_t = (assets_{t-1} + contributions_t + conversions_t - benefits_t) \times (1 + r_t) \tag{A1}$$

where $conversions_t$ is the (unmeasured) dollar value of conversions from a pre-existing non-401(k) DC plan during year t. For any year t, the value of directly contributed 401(k) assets evolves according to:

$$direct\ assets_t = (direct\ assets_{t-1} + contributions_t - direct\ benefits_t) \times (1 + r_t) \tag{A2}$$

Conceptually, *direct benefits* refers to the value of benefit payments that can be categorized as having been funded from direct 401(k) contributions rather than from conversions. In many cases, of course, 401(k) assets represent a mix of converted assets with directly contributed assets. In these cases, the proportion of benefits from directly contributed assets for a particular account is approximately the same as the proportion of directly contributed assets to total assets:

$$direct\ benefits_t = benefits_t \times (direct\ assets_{t-1}/assets_{t-1}) \tag{A3}$$

The US Labor Department's Form 5500 Annual Reports contain data on aggregate 401(k) assets, contributions, and benefits paid out for 1984–1995, as well as aggregate rates of return earned by DC plans with 100 or more participants for 1985–1995. These values are shown in columns 2 to 5 of Table 6.

Columns 6 and 7 of Table 6 calculate aggregate values for *direct benefits* and *direct assets* from the data in columns 2–5 for the years 1985–1991 by iteratively applying Eqs. (A2) and (A3) under the conservative assumption that $direct\ assets_{1983} = \$0$, i.e. that all of the assets in 401(k)s before 1984 had been contributed before the plans became 401(k)s.[15] Column 8, calculated as column 6 (*direct assets*) divided by column 2 (*assets*), is the aggregate proportion of non-converted 401(k) assets. According to these calculations, 30% of aggregate 401(k) assets in 1991 were originally contributed into non-401(k) plans.

References

Abadie, A., 2000. Semiparametric estimation of instrumental variable models for causal effects. Harvard University mimeo, March.

US Department of Labor, 1993. Abstract of Form 5500 Annual Reports, 1993–1999. United States Department of Labor, Pension and Welfare Benefits Administration. Private Pension Plan Bulletin. United States Government Printing Office, Washington, DC, http://www.dol.gov/dol/pwba/programs/opr/bullet1995/intro.htm.

Bassett, W.F., Fleming, M.J., Rodrigues, A.P., 1998. How workers use 401(k) plans: The participation, contribution, and withdrawal decisions. National Tax Journal 51 (2), 263–289.

Beller, D., 2000. Office of Policy and Research, Pension and Welfare Benefits Administration, United States Department of Labor. Personal communication.

Cochran, W.G., 1968. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics 24 (2), 295–313.

Dehijia, R.H., Wahba, S., 1998. Propensity score matching methods for non-experimental causal studies. NBER Working Paper 6829, December.

Du, J., 1998. Valid inferences after propensity score subclassification using maximum number of subclasses as building blocks. PhD dissertation, Harvard University, May.

Engelhardt, G.V., 2000. Have 401(k)s raised household saving? Evidence from the Health and Retirement Survey. Syracuse University mimeo, September.

Engen, E.M., Gale, W.G., 2000. The effects of 401(k) plans on household wealth: differences across earnings groups. Mimeo, August.

Engen, E.M., Gale, W.G., Scholz, J.K., 1994. Do saving incentives work? Brookings Papers on Economic Activity 1, 85–151.

Engen, E.M., Gale, G., Scholz, J.K., 1996. The effects on tax-based saving incentives on saving and wealth. NBER Working Paper 5759, September.

Gale, W.G., Papke, L.E., Van Derhei, J., 2000. The shifting structure of private pensions: Evidence, causes, and consequences. Mimeo, August.

Hewitt Associates, 1991. Survey Findings: 401(k) Plan Design and Administration. Hewitt Associates, Lincolnshire, IL.

Hubbard, R.G., Skinner, J.S., 1996. Assessing the effectiveness of saving incentives. Journal of Economic Perspectives 10 (4), 73–90.

Ippolito, R.A., Thompson, J.W., 2000. The survival rate of defined-benefit plans, 1987–1995. Industrial Relations 39 (2), 228–245.

Papke, L.E., 1995. Participation in and contributions to 401(k) pension plans: Evidence from plan data. Journal of Human Resources 30 (2), 311–325.

[15] Eq. (A1) could be used to solve for the aggregate value of conversions for the years 1984–1997. The change from year to year in the sample of small firms surveyed, among other sources of noise, make such estimates imprecise (Beller, 2000), and a few of the aggregate conversion numbers calculated this way are negative for that reason. There is no reason to expect this noise to bias the conversion estimates, however.

Papke, L.E., 1999. Are 401(k) plans replacing other employer-provided pensions? Evidence from panel data. Journal of Human Resources 34 (2), 346–368.

Papke, L.E., Peterson, M., Poterba, J.M., 1996. Do 401(k) plans replace other employer-provided pensions. In: Wise, D.A. (Ed.), Advances in the Economics of Aging. University of Chicago Press, Chicago.

Pare, T.P., 1995. Everything you ever wanted to know about 401(k)s (but were afraid to ask). Fortune 132 (13), 176–180.

Pence, K.M., 2001. 401(k)s and household saving: New evidence from the Survey of Consumer Finances. Mimeo, September.

Poterba, J.M., Venti, S.F., Wise, D.A., 1995. Do 401(k) contributions crowd out other personal saving? Journal of Public Economics 58, 1–32.

Poterba, J.M., Venti, S.F., Wise, D.A., 1998. Implications of rising personal retirement saving. In: Wise, D.A. (Ed.), Frontiers in the Economics of Aging. University of Chicago Press, Chicago.

Poterba, J.M., Venti, S.F., Wise, D.A., 2001. The transition to personal accounts and increasing retirement wealth: Macro and micro evidence. Mimeo, July.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Rosenbaum, P.R., Rubin, D.B., 1984. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association 79, 516–524.

Rubin, D.B., Thomas, N., 1992. Characterizing the effect of matching using linear propensity methods with normal distributions. Biometrika 79 (4), 797–809.