J Finan Serv Res (2007) 31:1–32
DOI 10.1007/s10693-007-0003-6

Defined Contribution Pension Plans: Determinants of Participation and Contributions Rates

Gur Huberman · Sheena S. Iyengar · Wei Jiang

Received: 21 October 2005 / Revised: 22 August 2006 / Accepted: 5 January 2007 /
Published online: 15 March 2007

Abstract Records of 793,794 employees eligible to participate in 647 defined contribution pension plans are studied. About 71% of them choose to participate in the plans, and of the participants, 12% choose to contribute the maximum allowed, $10,500. The main findings are (other things equal) (1) participation rates, contributions and (most remarkably) savings rates increase with compensation; on average, a $10,000 increase in compensation is associated with a 3.7% higher participation probability and $900 higher contribution; (2) women's participation probability is 6.5% higher than men's and they contribute almost $500 more than men; (3) participation probabilities are similar for employees covered and not covered by DB plans, but those covered by DB plans contribute more to the DC plans; (4) the availability of a match by the employer increases employees' participation and contributions; the effect is strongest for low-income employees; (v) participation rates, especially among low-income employees, are higher when company stock is an investable fund.

Keywords 401(k) plans · defined contribution

1 Introduction

Defined contribution pension plans in general, and 401(k) plans in particular are important vehicles for retirement savings. Although a handful of studies have considered individual and plan-level attributes that affect participation in such plans, or, for participants, levels of contribution, these studies either used only survey data (Papke 2004a; Munnell et al. 2001;

G. Huberman · W. Jiang (✉)
Finance and Economics Division, Columbia Business School, 3022 Broadway,
New York, NY 10027, USA
e-mail: wj2006@columbia.edu

G. Huberman
e-mail: gh16@columbia.edu

S. S. Iyengar
Management Division, Columbia Business School, New York, NY, USA
e-mail: ss957@columbia.edu

Even and Macpherson 2003), or employee records for very few firms (Kusko et al. 1998; Clark and Schieber 1998; Agnew et al. 2003) or used plan-level data (Papke 2004b; Papke and Poterba 1995). With information on almost 800,000 employees eligible to participate in 647 such plans (include those who chose not to participate), this study provides a comprehensive picture of the variables associated with individual participation in and contribution to 401(k) plans.

Individual-level data are important because in general, it is inappropriate to estimate a relation on an aggregate level and then infer that an analogous relation holds at the individual level. In some cases, even the sign of certain sensitivity estimates could be reversed (See a discussion in Freedman 2001). Records of non-participants afford particularly powerful analysis of the participation and contribution decisions.

Some—but not all—of the qualitative relations reported here are straightforward in light of the incentives faced by employees. For instance, the presence of an employer match should increase employees' participation in a 401(k) plan. The data afford going beyond qualitative observations. Specifically, they allow precise estimation of the sensitivities of employee behavior to explanatory variables which are important in their own right, and very useful to designers of retirement savings plans and policy makers at the firm and national levels.

This study goes beyond estimating overall relations between choice variables—participation and level of contribution—and individual and plan attributes. It explores why and how sensitivities of the choice variables to the attributes differ between the participation and contribution decision, and it also considers how the sensitivities of the choice variables to the attributes vary with compensation. The main results can thus be summarized while enumerating the main attributes, both of the plans and individuals.

Although individual characteristics such as gender and age are clearly exogenous, one cannot rule out the possibility that plan design could be catering for the aggregate characteristics and preferences of plan employees (see Mitchell et al. 2006, for an analysis of plan design), or that individual employees self-select into employers who offer plans that suit their retirement savings needs. Without reasonably long panel data it is difficult to tease out such effects using identification based on changes (such as the fixed effects method). All individual-level analyses in the paper control for plan aggregate characteristics (such as plan-average compensation, etc.) to mitigate the endogeneity concern at the individual level. It is possible, however, that some of the effects of plan policies documented here might capture cross-sectional relation rather than a definite causal relation due to unobserved plan-level heterogeneity. In interpreting these results, we also discuss the plausibility of alternative hypothesis.

Plan designs have strong effects on savings outcome (see a recent review by Choi et al. (2004c)). Among plan-level attributes, the first important one is coverage in a defined benefit (DB) plan. Comparing two similar individuals, the one with a DB plan is already saving for retirement, and his propensity to forego current consumption and liquidity in favor of consumption during retirement should be lower. On the other hand, an extreme form of mental accounting will render rights within a DB plan completely irrelevant to choices in a DC plan. Employees covered by a DB plan who have this form of mental accounting will participate in and contribute to their DC plans as if they were not covered by a DB plan. (Shefrin and Thaler 1992, and Thaler 1999, describe and analyze instances of mental accounting.) Moreover, the need to save for retirement may be more salient among those covered by a DB plan, or the savings-prone individuals are more likely attracted to employers that offer both DB and DC plans. The combined effect will lead to the surprising result that those covered by DB plans make stronger usage of 401(k) plans. The data are consistent with this last, and very counterintuitive result: other things equal, participation rates of those covered and not covered by DB plans are similar; contributions of those covered by DB plans are higher.

Many employers–539 plans in the study's sample of 647 plans–offer to match employees' contributions to DC plans. These matches are powerful incentives to participate, and indeed, participation rates are higher in the presence of a match. The incentive effect of the match is strongest for the lowest-income employees, and it decreases with compensation. In fact, at low-income levels (annual compensation between $10,000 and $20,000), a 100% employer match could increase participation probability by nearly 20%; at higher incomes (above $90,000), the incentive effect drops to about 5%.

The data indicate that the presence of a match increases contributions, primarily by increasing participation. In fact, among participants, the presence of a match seems to have no effect on contributions of low-income employees and, surprisingly, negative effect on contributions of those earning between $40,000 and $130,000. Participants' tendency to contribute at the upper limit on employer's match may be responsible for this counterintuitive finding. (Employers typically limit their matches to 5–6% of a participant's salary.)

The strong effect of matching programs on participation, especially of low-income employees, offers an immediate suggestion for a policy that encourages retirement savings in self-directed savings plans such as IRAs: the government could match the savers' contributions. Such a matching program can be more intense for low-income individuals if wealth redistribution is a secondary goal.

The inclusion of company stock in the plan's menu of investable funds guarantees the presence of a familiar option in the menu. Huberman (2001) argues that familiarity breeds investment. In fact, one of his examples is company stock in 401(k) plans. Other studies that consider the impact of company stock on asset allocation in 401(k) plans include Benartzi (2001), Choi et al. (2004a, b, c), Huberman and Sengmuller (2004), Liang and Weisbenner (2002), Mitchell and Utkus (2003), Meulbroek (2002), Ramaswamy (2002), Holden and VanDerhei (2003), and Poterba (2003). One theme common to these studies is that being associated with bad portfolio selection, the presence of company stock in the investable funds is bad for participants.

Overlooked thus far has been the potential salutary effect of including company stock in the investable funds: participation probability may be higher, presumably because eligible employees feel more comfortable participating when a familiar option is available. Empirically, this is the case. Participation probabilities are higher, especially for low-income employees. For employees who earn less than $35,000, the presence of company stocks as an investment option increase participation by more than 5%. The effect diminishes for employees who earn more than $40,000.

Compensation and gender are the more interesting individual attributes. The progressivity of the income tax code entails stronger incentives to participate and contribute to those who earn more. Moreover, low-income employees are more likely to have, or anticipate having liquidity constraints which will deter them from participating or contributing large sums to a 401(k) plan, where the money is locked up until retirement. Additionally, low-income employees expect higher salary replacement rates from social security upon retirement than high-income employees. This anticipation lowers the desire to save for retirement.

The data indeed show that controlling for all other variables, participation probability typically increases by almost 4% and contributions increase by about $900 for an increase of $10,000 in compensation. Moreover, savings rates—the ratios of contributions to compensation—increase with compensation.

Gender matters in saving decisions, adding to prior findings of gender differences in financial decisions (see, e.g., Barber and Odean 2001, and Bajtelsmit and Bernasek 1999). Holding other variables the same (especially compensation!) women's participation probabilities are 6.5% higher and their contributions are close to $500 higher.

This gender difference has at least two explanations, which are not mutually exclusive. One, that women have a stronger preference for saving, perhaps because they typically live longer than men. Two, the unit of decision is the household, and in many cases women are secondary wage earners whose incomes supplement those of their husbands. In these cases the women's recorded incomes are substantially lower than their households' incomes and their behavior is likely to reflect their households' incomes. (Nationally, according to Business Week, in 70% of the married households the husbands earn more than the wives.)

The next section describes the data and the econometric model. Section 3 reports the overall evidence and Section 4 reports how estimates vary with compensation. Section 5 discusses the findings.

2 Data Description and Model Set-up

2.1 Data

The Vanguard Group provided 926,104 participation and contribution employee records (including employees who were eligible but chose to not participate) in defined contribution (DC) pension, mostly 401(k) plans for the year 2001. The data contain 647 plans in 69 industries (by SIC two-digit codes). All plans required eligible employees to opt into the plan. Other concurrent studies using the same dataset including Iyengar et al. (2004, on the effect of offered choices on 401(k) participation), Mitchell et al. (2005 and 2006, on the effect of plan design on plan-level savings behavior; 2006, on the determinants of 401(k) plan design), Huberman and Jiang (2006, on the relation between offerings and choices for individual 401(k) participants), and Iyengar and Kamenica (2006, on choice overload and 401(k) asset allocation).

For the purpose of this research, excluded from the data were observations in at least one of the following categories: (1) The employee was hired after January 1, 2001 (9.6% of the observations). This exclusion criterion ensures that the person is employed for the whole year of 2001; (2) The person is less than 18 years old (0.02% of the observations). (3) The annual compensation is less than $10,000 or above $1 million (7.51% of the observations) to avoid the influence of extreme outliers. 793,794 observations survive. The Appendix offers more details on the construction of variables.

The all-sample participation rate is 71%, and about 76% of the eligible employees have positive balances (comparable to the national average participation rate of 76% reported by the Profit Sharing/401(k) Council of America 2001, 2002). The average individual pre-tax contribution rate for the whole sample and that for the highly compensated employees (defined as those who earned $85,000 or more in 2001) were 4.7 and 6.3%, respectively, compared to the national averages of 5.2 and 6.3% (Council of America 2001, 2002). In summary, the savings behavior of employees in the Vanguard sample seems representative of the overall population of eligible employees.

In the sample, 63% are male, and the mean age is 43. Figures 1 and 2 plot the sample's age and compensation histograms, respectively. Compensation mean and median are $61,150 and $47,430, respectively. In comparison, the same figures from the Survey of Consumer Finance (SCF) are $70,700 and $43,200. The average compensation is $65,900

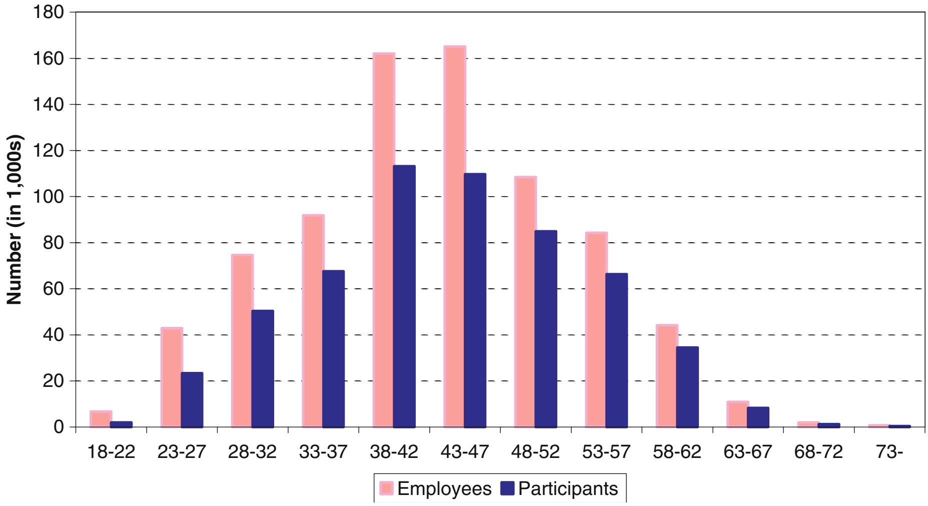


Figure 1 Age distribution of eligible employees and participants

for the 1998 SCF 401(k) eligible employee sample. In 2001, the maximum compensation for defined contribution plan purpose was \$170,000, and therefore the compensation variable used in the regressions is winsorized at \$170,000 (about 3% of the sample). Other information about individual characteristics includes tenure and financial wealth of the nine-digit zip neighborhood the employee lives in. A company called IXI collects retail and IRA asset data from most of the large financial services companies. IXI receives the data from all the companies at the nine-digit zip level, and then divides the total financial assets by the number of households in the relevant nine-digit zip area to determine the average

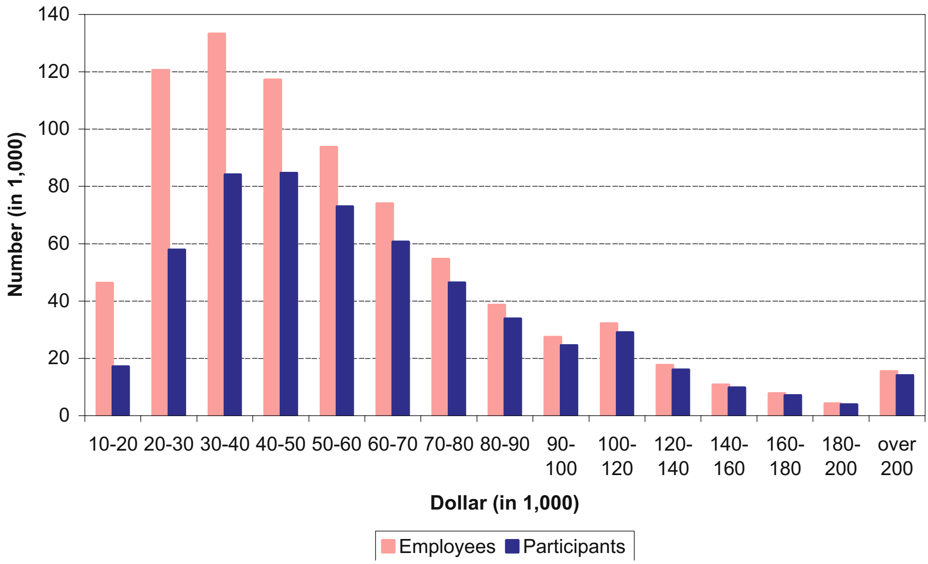


Figure 2 Compensation distribution of eligible employees and of participants

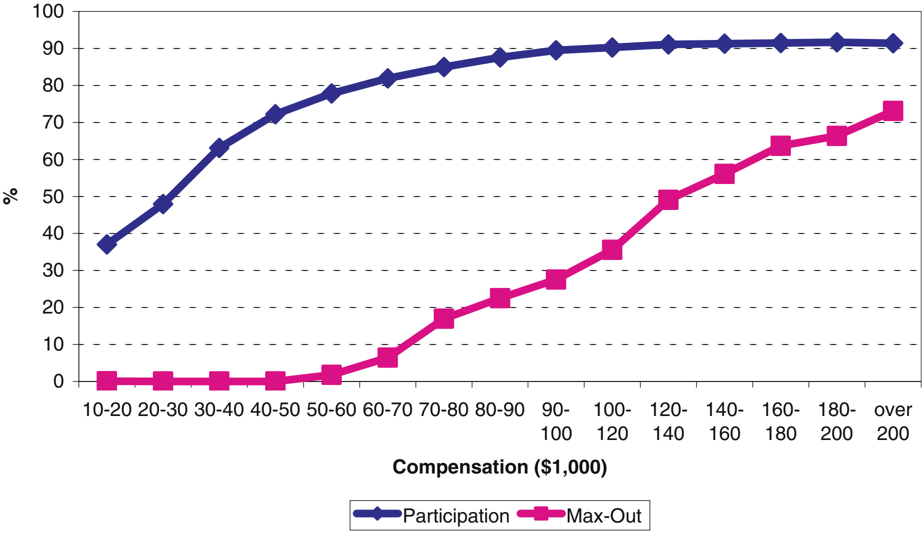


Figure 3 Rates of participation and contributing and maximum at each compensation level

assets for each neighborhood. There are 10–12 households in a nine-digit zip area on average. Subsequently, IXI assigns a wealth rank (from 1 to 24) to the area.

The records break down contributions to DC plans into three parts: employee pre-tax contribution, employee after-tax contribution, and employer contribution (including employer match). All the work reported here uses employee before-tax contributions to be comparable to most other research in 401(k) savings. In this study an employee is considered as a participant in a DC plan in 2001 if she contributes a positive amount before tax. By this criterion, participation rate is 71%, while 75% of the accounts have positive balances in tax-deferred accounts. (The employees who made no contribution in 2001 but had positive balances are probably those who had made contributions in earlier years but not in 2001 or those working for employers who make contributions but they choose not to contribute.)

Most plans ask employees to specify their deferral rates at the beginning of the year. The maximum contribution allowed in 2001 was the lower of $10,500, or 25% of compensation. Some plans impose additional limits on contributions made by highly compensated employees (HCEs, defined as those who earned $85,000 or higher in 2001) to ensure that the DC plans do not overly disproportionately benefit the high-income people (see, e.g., Holden and VanDerhei 2001). The mean deferral rate is 5.2%, and 12% of the participants contributed the maximum amount. Figure 3 plots the relation between participation/maximum contribution[1] and compensation. Both participation probabilities and the probability of contributing the maximum increase with compensation. The majority of those earning $30,000 or above participate. The majority of employees who earn about $130,000 or above contribute the maximum. Nonetheless, about 9% of the high-income employees do not participate at all.

[1] Here we only consider maximum contribution to the IRS limit ($10,500 or 25% of compensation). Section 3.3 discusses potential plan-specific limits that are lower than the IRS limit.

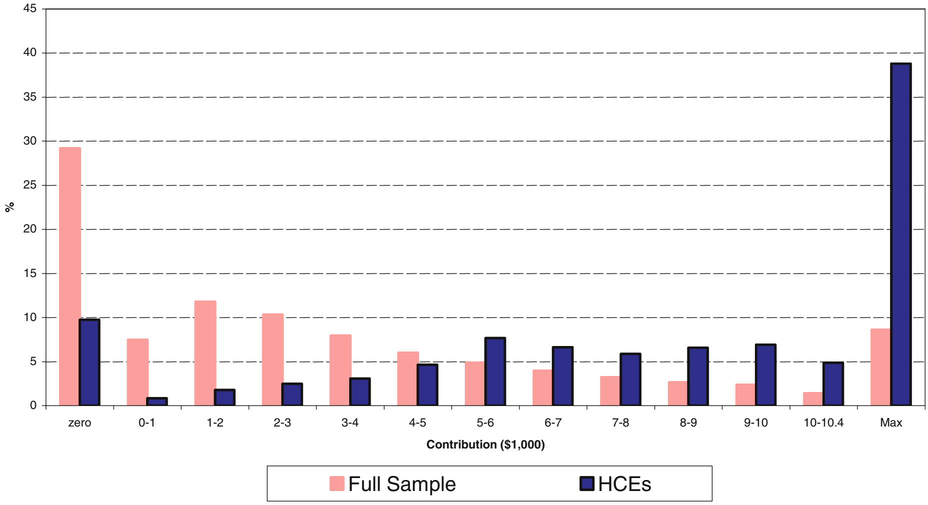


Figure 4 Fraction of employees at each contribution level

Figure 4 plots individual annual contributions for the full sample and the sub-sample with compensation above $85,000 (HCEs). Among all participants, the modal contribution is between $1,000 and $2,000. Those who earn more than $85,000 (16.7% of the sample) contribute more than the typical participants, and about 39% of those earning more than $85,000 contribute the maximum of $10,500. Figure 5 plots the contribution at different percentiles for employees at different levels of compensation. The figure clearly shows that high percentiles respond more intensely to increase in compensation, thereby suggesting that the cross-sectional variance of contributions increases with compensation. Figure 6

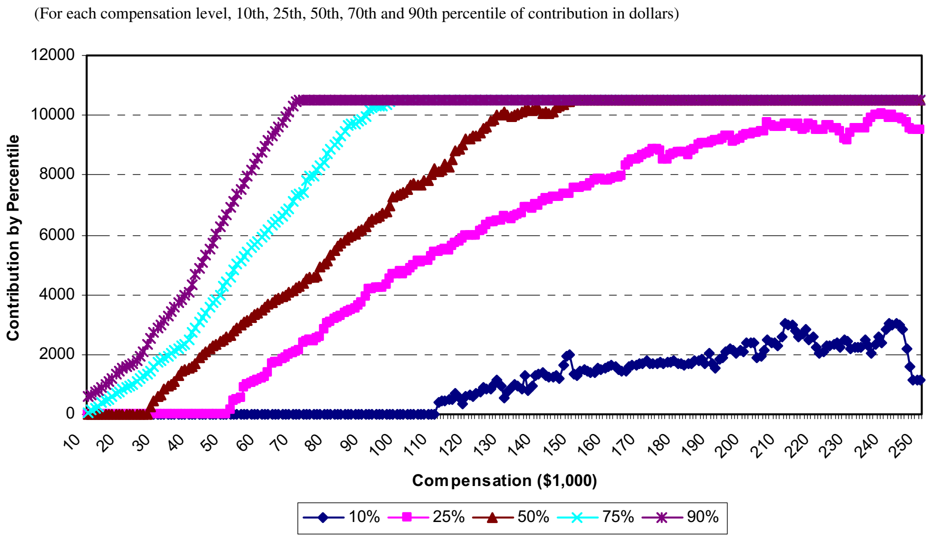


Figure 5 Contribution levels at each percentile

plots the wealth histogram for the general IXI population and for the Vanguard sample. Evidently, the Vanguard sample is somewhat better off than the general population at lower to middle wealth ranks.

The records have information about plan policies, including the presence of defined benefit (DB) plans, the number of investable funds available, employer matching schedule (match range and match rate), the presence of company stock as an investment option, whether the employer's match is in cash or company stock, and if the latter, the restrictions on diversification of the employer's match. 124 plans (covering 58% of the employees in the sample) provide own company stocks as an investment option, among which 47 companies match employee contribution with company stocks only. 216 plans (covering 67% of the employees in the sample) offer defined-benefit plans in addition to the defined contribution plan studied here. The number of funds offered by a plan ranges from 2 to 59 but 90% of the plans offer between 6 and 25 funds. Employers in 539 plans (covering 87% of the employees in the sample) offer some match to their employees' contributions. Most of them offer to match the employee's contribution up to 6% of the employee's salary, and the match rates range from 10 to 250%, mostly between 50 and 100%.

Exploratory data analysis is this study's main goal. Applying probit, one- and two-sided Tobit, and censored median regression analyses, the exploration goes beyond simple tabulation of averages and correlation and linear regression analyses. It affords an understanding of the decisions made by employees regarding their 401(k) savings. However, in the absence of a structural model, there is no single preferred specification.

2.2 Model Specification

The dependent variables studied here are: (1) A dummy variable, *PART*, that equals one if the individual participates, that is, if he contributes a positive amount to his tax-deferred account; (2) A dummy variable, *MAXOUT*, that equals one if the individuals contribute the maximum amount ($10,500 in 2001) to his tax deferred account; (3) Annual contribution, *CONTRIBUTION*, in dollar units or as a percentage of compensation.

The indices i and j represent individuals and plans, respectively. An individual's benefit from participating in a DC plan (net of cost) can be expressed as a function of a set of

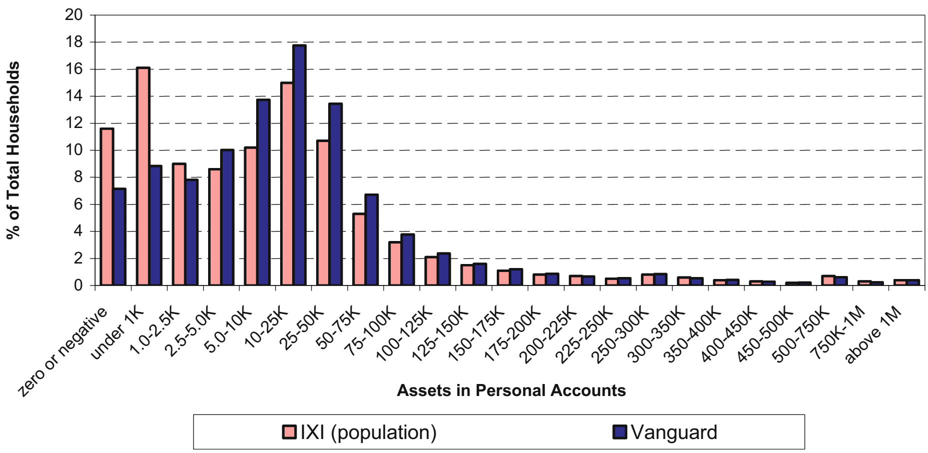


Figure 6 Fractions of households at each wealth level

individual characteristics X_{ij}, plan policies Z_j, and $\overline{X}_j$ represents the plan-level averages of individual characteristics:

$$U_{ij} = \gamma_0 + \gamma_1 X_{ij} + \gamma_2 Z_j + \gamma_3 \overline{X}_j + \varepsilon_{ij}, \varepsilon_{ij} = \eta_j + \varepsilon'_{ij}. \tag{1}$$

The plan-level averages of individual attributes serve as control variables; A later paragraph explains how these variables dampen the influence of endogeneity and peer effect on the coefficient estimation. The disturbance term can be decomposed into a plan-specific unobserved effect, η_j, which is assumed uncorrelated across different plans, and an individual disturbance, ε'_{ij}, assumed independently distributed across individuals. Both η_j and ε'_{ij} could be heteroskedastic across plans or individuals, but are assumed to be independent of the regressors. The individual will participate if $U_{ij}>0$, or

$$\mathrm{PART}_{ij} = \begin{cases} 1, \text{ if } U_{ij} > 0, \\ 0, \text{ otherwise.} \end{cases} \tag{2}$$

Determinants of participation can be analyzed using the linear probability model or maximum likelihood methods such as Probit.

Conditional on participation, the employee's *desired* contribution is described by:

$$y^*_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_j + \beta_3 \overline{X}_j + \delta_{ij}, \delta_{ij} = \varphi_j + \delta'_{ij}. \tag{3}$$

The disturbance δ_{ij} can be decomposed in the same way as ε_{ij}. Due to the max-out restriction, the *observed* contribution for a participant is:

$$y_{ij} = \begin{cases} y^*_{ij}, \text{ if } y^*_{ij} < \overline{v}_{ij}, \\ \overline{v}_{ij}, \text{ otherwise.} \end{cases}. \tag{4}$$

where $\overline{v}_{ij}$ is the maximum allowed contribution. The IRS limit constrains $\overline{v}_{ij}$ to be the lower of \$10,500 and 25% of compensation. If contribution level is expressed as percentage of compensation, $\overline{v}_{ij}$ is then the lower of (\$10,500/compensation) and 25%. Note that the model in Eq. 4 allows for individual truncation levels (as \$10,500/compensation varies with compensation levels). Further, some plans may be subject to plan-specific restrictions on the maximum contribution levels. We discuss the potential effect of such restrictions in Section 3.3.

For non-participants, the observed contribution is zero, and their desired contribution is left unspecified. Note the distinction between corner solution and data censoring: zero observed contribution is due to corner solution and maxed-out contribution is a result of data censoring.

Since contributions are between zero and the maximum limit, it is tempting to analyze them with a two-sided Tobit regression analysis. However, the standard Tobit estimation is not robust to heteroskedasticity, and unfortunately, diagnostic tests (e.g.,

Fig. 5) show that the error terms are highly heteroskedastic which could cause bias in estimation. The estimation tool used here is the censored median regression which is a special case of the censored least absolute deviations (CLAD) proposed by Powell (1984). It is based on the following observation: If y_i is observed uncensored, then its median would be the regression function $x_i'\beta$ under the condition that the errors have a zero median. When y_i is censored, its median is unaffected by the censoring if the regression function $x_i'\beta$ is in the uncensored region. However, if $x_i'\beta$ is on one of the two corners, then more than 50% of the distribution will "pile up" at the corner in which case the median of y_i does not depend on $x_i'\beta$. Thus, the computation of the estimator alternates (till convergence) between deleting observations with $x_i'\widehat{\beta}$ that are outside the uncensored region, and estimating the regression coefficients by applying the median regression to the remaining observations. For this reason, coefficients are not identified for observations with conditional median contribution (given the individual and plan characteristics) outside the non-censored region [0, 10,500]. For the present sample, roughly speaking, the method does not offer sharp predictions on behavior for people who earn below \$20,000 or above \$150,000 (about 10% of the sample). Analysis of this sub-sample is deferred to Section 4.

In Eq. 3, individual characteristics (X_{ij}) include: (1) Annual compensation in \$10,000 or in logarithm (*COMP*); (2) The wealth rank (IXI rank from 1 to 24) of the nine-digit zip neighborhood where the individual lives or the log of average household wealth in that IXI bracket (*WEALTH*). Strictly speaking, the *WEALTH* variable measures the average financial wealth of the neighborhood the employee lives in, which could be a noisy indicator of total personal wealth. On the positive side, this measure is also less susceptible to the endogeneity of personal wealth to savings propensity. This *WEALTH* variable has great explanatory power for participation/contribution, especially participation, but leaving it out of the regressions does not affect other coefficients (except compensation) significantly; (3) A gender dummy (*FEMALE*); (4) Age in years in excess of 18 (*AGE*); (5) The length (in years) of the individual's tenure with the current employer (*TENURE*).

Plan policy variables include the presence of company stocks as an investment option (*COMPSTK*), the presence of DB plans (*DB*), and number of funds offered to employees (*NFUNDS*). Variables for the intensity of employer match very slightly differently in specifications depending on the context. A binary variable, *MATCH*, is equal to one if the employer offers any match. The match rate which appears in the analysis of participation probability is the average match rate for the first 2% of salary, denoted *MATCH_INI*. The match rate used in the contribution analysis is the average match rate for the first 5% of salary, denoted *MATCH_AVG*. About 39% of the employees in the sample face tiered match schedules. The separate measures of match intend to capture the different nature of the strength of the incentive for participation and contribution decisions. In the decision on whether to make *any* positive contribution (participation), the relevant incentive is the match rate for the first dollar contributed. In our sample, this corresponds to match rate for the first 2% of compensation. For the contribution decision, the relevant incentive is the match for the total range, and most plans in our sample match up to 5% of employees.

Finally, plan-level control variables include: (1) average compensation (*COMP_MEAN*); (2) average age (*AGE_MEAN*); (3) average tenure (*TENURE_MEAN*); (4) average wealth (*WEALTH_MEAN*); (5) log number of employees (*NEMPLOYEE*), (6) the rate of web registration among participants within the plan in percentage points (*WEB*). In the absence of information about employee education, *WEB* proxies for the average education level, the sophistication level of the plan participants, or some other firm attribute that is correlated with Internet penetration. *COMP_MEAN* and *WEALTH_MEAN* are in the same units as *COMP* and *WEALTH* in the same regression. To a large extent, these variables can be viewed as exogenous, i.e., out of the control of plan policy makers. Adding these variables in the regressions serves two purposes in addition to the role of conventional control variables. First, they serve as instruments for possible endogeneity of plan policies in response to characteristics and behavior of people within the plan. For example, Mitchell et al. (2005) show that there is a strong tax motivation in employers matching programs subject to the federal non-discrimination rules. As a result, plan-offered match schedules are affected by the average compensation of employees (or those separately of the highly compensated and non-highly compensated employees). If the common component in behavior of employees belonging to the same plan is due to aggregate individual characteristics such as average income, the part of plan-level policies that is orthogonal to plan-level aggregate individual characteristics can be considered as exogenous (see, e.g., Chamberlain, 1985). Second, they serve as instruments to control for potential peer effect, that is, the influence of colleagues participation and contribution choices on an individuals own choices. (Manski 1993, offers detailed analysis of peer effects; Duflo and Saez (2003) examine peer effects in retirement savings decisions.)

3 Participation and Contribution: Full-sample Analyses

3.1 Individual Participation

Table 1 reports participation regressions (Eq. 2 using both the linear probability (Columns 1 and 2) and the probit models (Columns 3 and 4). In linear probability models, *COMP* and *WEALTH* are expressed in log dollars because they are both highly right-skewed variables, and Fig. 3 suggests a concave relationship between participation and compensation. *MATCH_INI* is the average match rate (in percentage points) for up to 2% of salary. (This is the match rate relevant to the participation, as opposed to the contribution decision.) Reported standard errors adjust for heteroskedasticity (both within and across groups, and group-specific disturbances) as well as within group correlation (due to the group-specific disturbance δ_j). A comparison of the two columns indicates that the marginal effect of individual attributes is not much affected by the plan policies. Bear in mind that when standard errors are adjusted for plan random effect (η_j) in Eq. 1, the "effective" sample size for coefficients estimates of individual variables is of the order of the total sample size (about 700,000) while that for coefficients estimates of plan variables is of the order of the number of plans, or just 647. (Wooldridge 2003, offers a general analysis on the asymptotics of cluster samples, especially where number of observations within clusters is large relative to the number of clusters.)

In Probit estimation, *COMP* and *WEALTH* are both in dollar terms and in logarithms (the plan-level average *COMP* and *WEALTH* carry the same units in the same regression.) The marginal probabilities reported (setting all non-dummy variables at their mean values, and all dummy variables at zero) are comparable to the coefficients from linear probability models. Measures of goodness-of-fit are pseudo R-squared and incremental probability of correct prediction. The former is defined as the likelihood ratio $1-\mathrm{Ln(L)}/\mathrm{Ln(L_0)}$, where L_0 is the log-likelihood value with the constant term only. The latter is defined as $\widehat{\Pr}\left(\widehat{y}_i > \frac{1}{2} | y_i > 0\right) + \widehat{\Pr}\left(\widehat{y}_i < \frac{1}{2} | y_i < 0\right) - 1$, where $\widehat{\Pr}$ is the empirical frequency, and $\widehat{y}_i$ is the predicted probability from the estimation. The null value of this probability is zero, and a value of one indicates perfect prediction. The analysis has close to 800,000 observations, and just a handful explanatory variables. Viewed in this context, the explanatory power of

Table 1 Determinants of individual participation. Dependent variable is PART. The all-sample participation rate is 70.8%. All coefficients are multiplied by 100. COMP and WEALTH are expressed in log dollars in Columns (1)–(3). In column (4), COMP is expressed in \$10,000, and WEALTH is expressed in IXI ranks from 1 to 24. Columns (1) and (2) are results from the linear probability model. The *t*-statistics adjust for both heteroskedasticity (both within and across groups, and group-specific disturbances) and within group correlation (due to the group-specific disturbance). Columns (3) and (4) report results from probit estimation. The standard errors are adjusted for correlation within the same plan. Pseudo R-squared is reported for goodness-of-fit. The marginal probabilities are calculated by setting all non-dummy variables at their mean values, and all dummy variables at zero. The sample contains 793,794 eligible employees in 647 plans. The effective sample size on individual (plan) variables is the number of employees (plans)

		Linear probability				Probit					
		(1)		(2)		(3)			(4)		
		COEF	SE	COEF	SE	COEF	SE	Margl. Pr.	COEF	SE	Margl. Pr.
I	CNST	−214.14	36.88	−196.39	40.05	−926.76	81.03	–	−173.06	86.21	–
	COMP	15.27	0.21	15.21	0.23	57.34	4.72	18.12	11.54	0.94	3.71
	WEALTH	5.96	0.06	5.93	0.07	23.16	1.30	7.32	2.14	0.34	0.69
	FEMALE	5.64	0.50	4.66	0.84	18.88	1.29	5.97	20.11	0.93	6.47
	AGE	0.21	0.05	0.22	0.08	0.32	0.44	0.10	0.88	0.51	0.28
	AGE^2	0.00	0.00	0.00	0.00	−0.01	0.01	0.00	−0.01	0.01	0.00
	TENURE	1.30	0.08	1.17	0.14	4.79	0.47	1.51	5.07	0.55	1.63
	TENURE^2	−0.03	0.00	−0.03	0.00	−0.12	0.01	−0.04	−0.12	0.02	−0.04
II	MATCH_INI	0.12	0.02	–	–	0.44	0.04	0.14	0.41	0.04	0.13
	COMPSTK	3.50	1.60	–	–	9.47	4.40	3.01	7.53	3.58	2.41
	DB	−0.28	1.45	–	–	1.01	2.11	0.32	0.26	2.60	0.09
	NFUNDS	−0.23	0.11	–	–	−0.92	0.31	−0.25	−0.78	0.34	−0.25
III	COMP_MEAN	3.30	4.50	3.46	5.39	7.36	6.12	2.32	2.77	0.94	0.89
	WEALTH_MEAN	−1.70	2.62	−1.12	3.13	−5.38	4.58	−1.70	−5.31	2.56	−1.71
	AGE_MEAN	1.49	0.32	0.99	0.59	4.53	1.50	1.43	5.27	2.02	1.70
	TENURE_MEAN	−1.09	0.29	−0.77	0.38	−3.78	0.70	−1.19	−4.17	0.92	−1.34
	WEB	0.07	0.08	0.18	0.09	0.31	0.14	0.10	0.65	0.14	0.21
	NEMPLOYEE	−2.89	0.52	−3.37	0.77	−9.55	2.10	−3.02	−9.80	2.02	−3.15
	Pseudo R-squared	0.19		0.18		0.18			0.13		

the reduced form linear model is remarkable: R^2 is about 19% and incremental probability of correct prediction is about 30%.

Income and wealth are the most important determinants for participation in DC plans. Other things equal and on average, a $10,000 increase in annual compensation is associated with about 3.7% higher probability of participation (unless otherwise stated, reported numbers are the marginal probability estimates from the Probit model in column (4) of Table 1). Females are 6.5% more likely to participate than their male counterparts. The stock phrase “other things equal” is particularly pertinent here. Women’s overall participation rate is 70.0%, less than the 71.3% participation rate of men. However, women typically earn less than men—their median wage is $39,500, whereas men’s’ median wage is $54,000 in this sample—and they have shorter tenure—a median tenure of 9.5 years compared with men’s 10.5 years in this sample. The 6.5% gender difference in participation rates applies after controlling for these and the other variables.

Older and longer tenured employees are more likely to participate. For an average 18-year old who just starts on her job, each year of advance in age (tenure) is associated with an increased 0.2% (1.6%) participation probability, and both marginal effects are decreasing in years. The tax-deferred nature of 401(k) contributions suggests that controlling for income (and the marginal tax rate that goes with it) it is more beneficial to contribute early in one’s career. However, earlier in one’s career is when liquidity constraints are likely to reduce the propensity to save for retirement. Moreover, the salience of retirement (and the need to save for it) may increase with age. Finally, the pattern documented here may arise because employees who join 401(k) plans are very unlikely to leave them. Analysis of a long panel of records can determine the validity of this hypothesis.

With the exception of DB, the plan-level policy variables seem to affect individual participation. Table 1 suggests that participation rate could be about 13% higher in a plan that offers 100% match than in an otherwise equal plan that offers no match. Using *MATCH_AVG*, the sensitivity estimate is about 1 percentage point lower. Further (results not tabulated), the mere existence of a match (regardless of the magnitude) increases participation by 6.3%, and each 1% rise in match rate further increases participation by 0.08%.

When company stock is an investable option, the participation probability increases by 2.4%. One caveat regarding company stock: By and large, firms where company stock is an investable fund are publicly held. It may well be that “company stock” proxies here for “publicly held firm.” Unfortunately, the records available for this study (plan identities are removed) do not allow a further investigation of the issue.

The big surprise is the coefficient on DB, which is small in magnitude and statistically indistinguishable from zero. Controlling for individual and other plan level attributes, it seems that participation rates of those covered and not covered by a DB plan are similar. Moreover, the same result (not tabulated) emerges when the analysis is repeated for the subsample of employees who are at least 40 years old with at least 10 years of tenure. It is those over 40 who are more likely to be conscientious about the status of their retirement savings, and, among them, those with at least 10 years of tenure to have accumulated considerable rights to retirement benefits if their employer offers a DB plan. Their participation rates are similar to comparable employees not covered by a DB plan. Even and Macpherson (2003) report similar findings based on Current Population Survey of 1993. The similarity in behavior suggests that, counter-intuitively, the presence of a DB plan does not affect the participation decision in a 401(k) plan, or that employees who have stronger taste for savings are more likely to work for companies that offer multiple retirement savings vehicles.

Controlling for other variables is crucial to this result and its interpretation because a comparison of the raw data leads to the opposite conclusion. Participation rates for the full sample are 68 and 76%, for employees working for firms that offer or do not offer DB plans, respectively. When attention is confined to the subsample of those over 40 and with at least 10 years of tenure, the corresponding participation rates are 71 and 86%, respectively. However, employers with DB plans tend to be larger employers and the average compensation and wealth levels of those employed by firms that offer DB plans are lower (the correlations between DB and plan size, plan average compensation, plan average wealth are 0.44, −0.12, −0.20, respectively). Excluding those plan-average variables would produce a negative and significant coefficient on DB equal −3.0%; further excluding *WEALTH* from the explanatory variables would yield a coefficient of −3.6%. Therefore, the present result does not contradict Cunningham and Engelhardt (2002). However, DB has no effect on participation only after controlling for these and the other variables in the analysis.

Unfortunately, hazard model-type analysis (used, e.g., in Choi et al. 2002) accommodating changing behavior over time is not feasible here because the data underlying this study consist of a single cross-section. If plan policies change over time and the participation of employees is sensitive to these policies as they evolve, the estimates reported here could be subject to measurement errors. This observation is especially pertinent to *MATCH* which could vary from year to year. The behavior of the 63,043 employees hired in 2001 serves as sensitivity check because their decision to participate was based only the plan policies prevailing in 2001.

Using the same specification as the first column in Table 1 on the new hires subsample, the participation probability is 11% higher for employees who were offered 100% match compared to those without match (significant at less than the 1% level). Still, about 19% of the employees who are offered employer match of at least 50% choose not to participate, and among those who participate, 45% do not contribute up to the match threshold. Such evidence is echoed in Choi et al. (2005) (these authors further estimate that the foregoing matching contributions average 1.3% of the annual pay of the under-contributing employees). Further, the participation probability is 2.2% higher when the company stock is an investable option, and is 2.8% lower when a DB plan is also present, but neither of the effects is statistically significant at less than 10% level after adjusting for the plan random effects.

Interpreting results from this subsample, however, requires some caution. First, some non-participants, especially those who were hired for less than a couple of months, may be simply taking time in making their decisions rather than choosing not to participate. (The subsample participation rate is 45%, as opposed to the all sample participation rate of 71%). Second, the new hires sample is skewed toward the young, inexperienced, and low-income subpopulation of the 401(k) eligible employees, the inference from which may not extend to the general population. Finally, restricting the sample to people who were hired during one particular year may reflect a shock that is particular to that year.

3.2 Individual Contributions

This subsection employs censored median regressions to estimate the relations between individual contribution and individual characteristics as well as plan policies.

Table 2 reports the estimates of three censored median regressions with different dependent variables: contribution in dollar amount; and saving rate in percentage (i.e., the ratio of contribution to compensation). The censoring in the median regressions is designed to account for zero being the lower bound on savings and the lower of $10,500 and 25% of employee compensation being the upper bound. Robustness checks further assess the impact of additional plan-imposed constraints on contributions made by highly compensated employees.

The dollar amount specification suggests that other things equal, contributions increase by $909 for an increase of $10,000 in compensation, and that women contribute $482 more than men. The sensitivities to individual attributes were also estimated separately for each of the 483 plans that had more than 100 employee records. The average estimates from all plans (which assign equal weights on plans regardless of their size) on compensation and gender are $916 and $478, almost identical to the two coefficients from pooled regressions.

Age seems to be negatively associated with contributions for younger employees (below 40) but positively associated with contributions for older employees.

A match increases contributions: an increase in the match from zero to 100% will increase contributions by $457. Further, among companies that offer company stock as an investment option, the effect of 100% match is stronger by $159 when the match is restricted to company stock. (Results not tabulated.) The presence of company stock among the investable funds does not seem to have a consistent impact on contribution. It is slightly negative (but not distinguishable from zero) in the first specification (Column (1) where both contribution and compensation are expressed in dollars), but is positive and significant in the savings rate specification (Columns (2)). The estimation from the latter specification assigns more weight to the low-income employees. The next section shows that they are more responsive to the presence of company stock, which explains the difference in outcome between the first specification and second and third specification.

The presence of a DB plan increases employee contributions by $180. Again, it is important to interpret this observation in the context of controlling for other variables. In the raw data, the median contribution of employees working for firms that have DB plans is $504 lower than their no-DB counterparts; and among employees who are 40 years or older and have at least ten years of tenure, the difference of medians is $1,580. (Unlike average contribution, plan median contribution is not affected by non-participants and maxed-out contributions.) This property in the raw data is consistent with findings of negative relation between the presence of DB plans and contribution rates summarized in Clark and Schieber (1998).

The controls reverse the inference offered by the raw differences because firms that have DB plans tend to have more employees who have longer tenure, but less financial wealth. It is possible that these controls also capture some employees' propensity to save. Such individuals may tend to work for larger companies (that presumably offer safer employment) in which employees stay longer with their employers. The results are even more surprising to the extent that the control variables capture a taste for savings.

Papke (2004a) uses survey data and OLS regressions in which she controls for income, wealth, gender and marital status. She reports that DB coverage is associated with higher contributions to DC plans, but the effect is statistically insignificant. The results reported here are far more reliable because the underlying data are of higher quality and the analysis itself exploits the higher data quality by controlling for plan-level characteristics and allowing for censoring in the contributions. Nonetheless, it is worth remembering that the DB measure reported here (as well as in other studies) is a crude indicator of coverage rather than the more desired measure of the employee's cumulative rights within the plan. Still, to the extent that this measure is correlated with the employees' non-DC benefits upon retirement, the results are valid and the positive correlation between the presence of DB future benefits and current DC contributions is surprising indeed.

Table 2 Determinants of individual contribution: censored median regression analysis. The dependent variables are: contribution in dollars (Columns (1), (3) and (5)), and savings rate (equals contribution/compensation, Columns (2) and (4)). Censored median regression (Powell (1984)) is applied to all specifications. Columns (1) to (4) use the full sample. Column (5) excludes all highly-compensated employees (HCEs, who earned $85,000 or more). Columns (1) and (2) incorporate contribution censoring due to the IRS limit (the lower of $10,500 or 25% of compensation). Columns (3) and (4) further incorporate potential plan-specific limits where an observation is treated upper-censored if (1) the employee contributes to the IRS limit; or (2) the employee is in a "potentially limited plan" and has close to the maximum contribution deferral rate in the plan (within 0.25%). A "potentially limited plan" is defined as a plan that (1) nobody in the plan has total contribution up to 25% of compensation; (2) at least five employees in the plan (who contribute less than $10,500) have contribution deferral rates cluster at the plan maximum (within 0.5%). Pseudo R-squared is the proportion of the sum of absolute deviations in the dependent variable explained by the regression

		IRS limit only				Plan-specific limit adj.				HCE-excluded	
		(1) Linear		(2) Saving rate		(3) Linear		(4) Saving rate		(5)	
		COEF	SE	COEF*100	SE*100	COEF	SE	CEOF*100	Se*100	COEF	SE
I	CNST	−3,958.64	590.84	−591.06	161.61	−3,889.52	580.97	−573.25	168.40	−3,737.80	648.99
	COMP	909.99	11.57	88.66	5.47	911.78	11.36	175.38	14.60	912.86	32.32
	WEALTH	81.22	4.81	16.38	1.27	81.68	4.82	9.72	0.87	73.57	6.69
	FEMALE	481.05	24.84	100.16	7.18	485.70	24.65	100.34	7.35	435.51	38.99
	AGE	−33.37	8.09	−4.88	2.74	−33.63	7.95	−2.48	2.91	−26.29	10.12
	AGE^2	1.11	0.16	0.21	0.05	1.12	0.16	0.14	0.05	0.96	0.21
	TENURE	78.83	7.15	18.40	1.49	79.40	7.00	13.39	1.43	71.56	7.20
	TENURE^2	−2.07	0.24	−0.46	0.05	−2.08	0.24	−0.34	0.04	−1.90	0.24
II	MATCH_AVG	4.60	0.76	1.32	0.20	4.64	0.76	0.95	0.16	4.66	0.83
	COMPSTK	−12.97	54.95	20.74	16.44	−16.61	54.61	18.42	15.91	−0.43	60.55
	DB	177.89	55.91	25.13	12.77	182.06	56.12	7.65	11.24	170.01	63.87
	NFUNDS	−5.45	3.74	−2.15	0.93	−5.74	3.74	−0.96	0.77	−2.45	4.34
III	COMP_MEAN	14.00	15.38	8.67	3.57	13.96	15.25	2.92	2.07	5.31	16.60
	WEALTH_MEAN	−103.58	30.68	−29.52	9.18	−105.25	30.11	−15.31	8.07	−99.24	35.86
	AGE_MEAN	57.47	14.96	19.13	2.94	56.30	14.92	11.95	3.38	54.26	14.96
	TENURE_MEAN	−67.45	17.93	−17.20	3.37	−67.09	17.85	−11.49	3.39	−64.54	19.30
	WEB	19.02	3.25	3.74	0.67	19.32	3.27	2.13	0.63	17.10	3.75
	NEMPLOYEE	−115.15	20.41	−37.72	6.35	−118.66	20.29	−27.97	5.75	−116.18	28.30
	No. individuals & plans	793,794	647	793,794	647	793,794	647	793,794	647	661,104	643
	Pseudo R-squared	0.25		0.10		0.25		0.11		0.19	

The savings rate specification (Column (2)) is quite consistent with the other specification, showing that an increase in compensation form \$40,000–\$50,000 is associated with an almost 1% increase in the saving rate. Women's saving rates are 1.05% higher than those of men. Savings rate increase by 0.18% when company stock is an investable fund and by 0.25% when a DB plan covers the employee.

3.3 Plan-specific Limits on Contribution

Some 401(k) plan sponsors might impose maximum contribution limits on their employees that are lower than the IRS limit (\$10,500 or 25% of compensation). There are two types of such lower limits: uniform plan limits on contribution (usually the total contribution from both employee and employer) as a percentage of employee compensation; and contribution limits for highly compensated employees (HCEs, defined as those who earned \$85,000 or higher in 2001) in compliance with the federal non-discrimination rules. This section analyzes both situations.

First, the plan-specific limits for all employees in a plan. Some of the 401(k) plans in the sample have been historically organized as profit-sharing plans. As such, they were subject to the 15% limit on total employer and employee contributions as a percentage of employee compensation. Other sponsors might have raised the limit to 17–18% or sometimes higher in order to encourage employee contribution. Unfortunately, no explicit information is available. Mitchell et al. (2005) discuss the possible prevalence of these limits in this sample.

We adopt the following algorithm to classify plans that are suspicious of having limits lower than that of the IRS: a plan is classified as a "potentially limited plan" with a limit of c%<25% if: (1) Nobody in the plan has total contribution deferral rate greater than *c%*; and (2) there are five people or more in the plan whose total contribution is more than (c%–0.5%), but lower than \$10,500.[2] The second criterion ensures that there is some clustering at *c%* so that the observed upper bound is not a random incidence. Altogether there are 341 plans (out of 647) that satisfy both criteria above. About 54.5% of our sample eligible employees, and 84.3% of our sample participants are potentially subject to plan-specific limits, and 3.3% of participants are potentially constrained by such limits (that is, they contribute an amount that is lower than \$10,500 but is at the putative plan-specific limits).

Plan-specific limits require modification of Eq. 4 in estimation. An employee's contribution is upper-censored if any of the following holds: (1) He contributes \$10,500 (we allow \$25 for the rounding error); (2) he contributes 25% of his compensation (we allow 0.25% for the rounding error); and (3) he is in one of the "potentially limited plans" described above, and has the highest deferral rate in his plan (we allow 0.25% for the rounding error). If an employee's contribution is upper-censored for any of the three reasons, we record it as $y_{ij} = \bar{v}_{ij}$ (for observed contribution) and $y^{*}_{ij}\bar{v}_{ij}$ (for desired contribution) in the censored regression.

Columns (3) and (4) of Table 2 report the results from the extended censored regression estimation. It should be noted that this specification might over-classify upper-censored

[2] The 0.5% is to allow for rounding error. We err on over-classifying "potentially limited plans" to be on the conservative side.

observations because the "potentially limited plans" may not actually have any mandatory limits.[3] Fortunately results are qualitatively similar to those in (1) and (2) except that in the savings rate specification the marginal effect of compensation is much strengthened (because highly-compensated employees are more likely to be constrained). The similarity arises because only a small portion of the employees are actually constrained although a great majority of them are subject to potential plan-specific limits: about 2.4% of the eligible employees and 3.3% of the participants invest up to the potential plan limits that are lower than the IRS limit.

Second, plan-specific limits for HCEs. Encouraging savings of low- and middle-income American families has been an important mission for policy makers (see recent papers by Bernartzi and Thaler 2004; Duflo et al. 2006). Some plans face additional limits on contributions made by HCEs under the federal non-discrimination rule with the stated goal that the DC plans do not overly disproportionately benefit the high-income people (see, e.g., Holden and VanDerhei 2001; Mitchell et al. 2005). The data set unfortunately does not provide information on such plan-specific restrictions on HCEs. As a sensitivity check, column (5) of Table 2 reports the regression estimates of the main specification (as in column (1)) on the subsample of employees who earned less than $85,000 (about 83% of the sample). Results seem to be consistent with those of the full sample.

3.4 Maximum Contribution

As a by-product of individual contribution analysis, the decision to max out is also considered. Table 3 reports estimates of maxing out using the same model specifications as in the last two columns of Table 1. The first two columns classify maximum contribution according to the IRS limit, while column (3) also includes potentially upper-censored observations due to plan-specific limits. The maxing-out rates among participants is 12.2% in the first two columns, and 15.5% in the third. All individual characteristics affect the probability of maxing out in the same direction as they do the probability of participation.

Not surprisingly, the marginal probability of incremental compensation on maxing-out is higher when potential plan limits are taken into account (semi-elasticity of compensation, in logarithm, on maxing-out increases from 12.7 to 21.2 percentage points). Females are even more likely to max-out than males in the plan-limit adjusted specification: the gender difference increases from 1.4 to 3.2 percentage points (because the more savings-prone gender is more likely to be constrained by plan limits).

The match rate seems to have a negative impact on maxing-out, but the effect goes away once plan specific limits are adjusted for. This difference is explainable by plan specific limits most of which are imposed on total contribution including employer match: for employees who intend to contribute close to the maximum allowable amount by the plans, employer match becomes substitutes for their own contribution. This effect is exacerbated by the positive correlation between the existence of potential plan limits and plan match

[3] It could be that all employees in a plan voluntarily contribute less than certain percentage of their income, and a handful of them (five or more) contribute very close at the top (*c%*). This is more likely to be the case when *c%* is higher, such as those greater than 15%. For example, it is plausible that nobody contributes more than 20% of their compensation, even in the absence of any plan-specific limit. For people with compensation greater than $52,500, the $10,500 IRS limit binds first. For people from the lower compensation group, contributing 20% or more would imply low take-home pay. On the other hand, one can also argue that the classification method may miss-out limited plans, too. Under-classification is quite innocuous for the estimation purpose. Non-detectable plan limits imply that they are basically non-binding (except for maybe less than five people in a plan). A plan-specific limit only affects contribution when it is binding, that is, when the employees are constrained.

Table 3 Determinants of individual maxing out. Dependent variable is MAXOUT, and the estimation method is probit. The maxing-out rate among participants is 12.2% according to the IRS limit (Columns (1) and (2)), and 15.5% taking into account of maximum contribution in "potentially limited plans" (Column (3)). The definition of "potentially limited plans" is the same as in Table 2. All coefficients are multiplied by 100. COMP and WEALTH are expressed in log dollars in columns (1) and (3); and COMP is expressed in \$10,000, and WEALTH is expressed in IXI ranks from 1 to 24 in column (2). Standard errors are adjusted for within plan correlation. Pseudo R-squared is reported for goodness-of-fit. The marginal probabilities are calculated by setting all non-dummy variables at their mean values, and all dummy variables at zero. Only participants are included in estimation. The sample contains 562,013 participants in 647 plans. The effective sample size on individual (plan) variables is the number of employees (plans)

		IRS limit only						Plan-specific limit adj.		
		(1) Log-Comp			(2) Dollar-COMP			(3) Log-Comp		
		COEF	SE	Margl. Pr.	COEF	SE	Margl. Pr.	COEF	SE	Margl. Pr.
I	CNST	−2,879.19	55.03	–	−653.18	53.32	–	−1,421.90	48.92	–
	COMP	207.73	2.41	12.73	19.23	0.26	2.09	118.69	2.29	21.19
	WEALTH	6.84	0.49	0.42	2.20	0.16	0.24	9.53	0.66	1.70
	FEMALE	22.79	1.24	1.40	13.17	1.41	1.43	18.16	2.23	3.24
	AGE	−0.87	0.36	−0.05	0.62	0.41	0.07	−0.71	0.54	−0.13
	AGE^2	0.04	0.01	0.00	0.01	0.01	0.00	0.04	0.01	0.01
	TENURE	0.57	0.23	0.03	0.55	0.28	0.06	1.09	0.35	0.19
	TENURE^2	−0.02	0.01	0.00	−0.01	0.01	0.00	−0.03	0.01	−0.01
II	MATCH_AVG	−0.27	0.04	−0.02	−0.19	0.04	−0.02	−0.10	0.07	−0.02
	COMPSTK	−17.40	3.81	−1.10	−13.88	3.67	−1.54	−7.89	4.52	−1.42
	DB	−9.37	1.85	−0.59	−6.44	2.06	−0.71	−2.40	4.64	−0.43
	NFUNDS	0.55	0.15	0.03	0.70	0.20	0.08	−0.31	0.33	−0.05
III	COMP_MEAN	−11.10	4.95	−0.68	−1.16	0.58	−0.13	−0.80	0.82	−0.14
	WEALTH_MEAN	17.51	3.14	1.07	8.66	1.45	0.94	−1.17	4.35	−0.21
	AGE_MEAN	4.79	0.84	0.29	4.35	0.88	0.47	−2.33	1.08	−0.42
	TENURE_MEAN	−2.29	0.32	−0.14	−2.35	0.37	−0.25	0.57	0.83	0.10
	WEB	0.99	0.10	0.06	1.28	0.10	0.14	0.69	0.21	0.12
	NEMPLOYEE	9.48	1.45	0.58	8.63	1.44	0.94	−1.59	1.33	−0.28
	Pseudo R-squared	0.41			0.22			0.38		

rate (the coefficient of correlation is 0.18). Presence of company stock and DB plans do not seem to have consistent effects on participants' tendency to max out once plan limits are accounted for.

4 Participation and Contribution: The Impact of Variation in Compensation

The evidence so far shows that compensation is a major determinant of participation and contribution. There are a few differences between low- and high-income employees that can lead to this result. One, the tax benefits of saving through a tax-deferred vehicle are more generous to the high-income employees. Two, low-income employees are more likely to face liquidity constraints that will prevent them from putting money away, even in a tax-deferred plan. Three, Social Security benefits offer high salary replacement rates to low-income employees, and render alternative retirement savings less urgent. Four, low-income employees may be less educated and sophisticated about the benefits and costs of participating in a 401(k) plan. Engen and Gale (2000) suggest that the savings behavior varies across earnings groups, and therefore 401(k) plans have different effects on household wealth.

The differences between low- and high-income employees suggest a re-examination of the data separately for various levels of compensation, a luxury easily afforded by almost 800,000 records on hand. This section reports estimates of the probit analysis of the participation and estimates of two sets of Tobit regressions, done at different compensation levels. One is a two-sided Tobit, aimed at estimating a censored linear contribution model for all employees at a given compensation level. Another is a one-sided Tobit aimed at estimating a censored linear contribution model only for participants. The three estimated models produce three sets of slope coefficients. Juxtaposing these coefficients provides a more comprehensive understanding of the employees' decisions.

Technically, estimating the models again for various compensation levels modifies specifications (1)–(2) and (3)–(4) above, by allowing the slope coefficients (i.e., sensitivity of participation and contribution to those factors) to depend on the compensation. Such a modification is reasonable in the absence of a rigid structural model, and enhances the exploration of the rich data set at hand.

The following equation summarizes the relations among the three sets of coefficients (corresponding to probit, one-sided Tobit and two-sided Tobit estimates). Let y^*_{ij} be the *desired* contribution (could be a latent variable) by individual i in plan j, and W_{ij} be a personal or plan characteristic variable. Then:

$$\frac{\partial E\left[y^*_{ij}\middle|W_{ij}\right]}{\partial W_{ij}} = \Pr\left[\mathrm{PART}_{ij}=1\right]\frac{\partial E\left[y^*_{ij}\middle|W_{ij},\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}} + E\left[y^*_{ij}\middle|W_{ij},\mathrm{PART}_{ij}=1\right]\frac{\partial \Pr\left[\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}} \tag{5}$$

In the equation, when the independent variable W is binary (e.g., gender or availability of a DB plan), a partial derivative represents a difference (i.e., the change in the dependent variable when the binary variable changes from zero to one.) The left hand side of the equation, $\frac{\partial E\left[y^*_{ij}|W_{ij}\right]}{\partial W_{ij}}$, is the sensitivity of the desired contribution per employee to a change in a variable W (e.g., compensation, match intensity, gender, etc.) at a given level of compensation. On the right hand side, $\Pr[\mathrm{PART}_{ij}=1]$ is the probability of participation given all attributes, and $\frac{\partial E\left[y_{ij}^*\middle|W_{ij},\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}}$ is the sensitivity of the *desired* contribution by an employee conditional on participation; $E\left[y_{ij}^*\middle|W_{ij},\mathrm{PART}_{ij}=1\right]$ is the expected contribution amount conditioned on participation, and finally, $\frac{\partial \Pr\left[\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}}$ is the marginal change in participation probability to an incremental change in *W*.

If the models are correctly specified, then the two-sided Tobit coefficients are consistent estimates of $\frac{\partial E\left[y_{ij}^*\middle|W_{ij}\right]}{\partial W_{ij}}$; the one-sided Tobit coefficients are consistent estimates of $\frac{\partial E\left[y_{ij}^*\middle|W_{ij},\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}}$, and the marginal probabilities from probit estimation are consistent proxies for $\frac{\partial \Pr\left[\mathrm{PART}_{ij}=1\right]}{\partial W_{ij}}$. Equation 5 implies that the unconditional response to a unit change of an independent variable (two-sided Tobit) is larger (resp., smaller) than the same response conditional on participation (one-sided Tobit) if the variable is positively (resp., negatively) associated with participation probability (Probit regression).

Figures 7, 8, 9, 10 and 11 summarize the estimates of the three regressions for each compensation bin. They are similarly structured with three graphs each. Each graph corresponds to the estimates of one of the three sets of regressions. The horizontal axis, common to the three graphs, indicates the compensation bin. The right vertical axis is the scale of the marginal probability, and the left vertical axis is the scale of the marginal contribution, both for participants and for the whole population in the corresponding compensation bin. All figures depict smoothed coefficients, i.e., weighted averages of actual regression coefficients of the central bin regression (50% weight) and its two neighboring bin regressions (25% each). The dotted lines represent 95% confidence intervals.

The regressions underlying Figs. 7, 8, 9, 10 and 11 are estimated independently. For each, the records used are of the employees (or participants, for the one-sided Tobit) whose compensation ranges from \$5,000 (\$10,000) below the central point of the subsample to \$5,000 (\$10,000) above it, if the central point of the subsample corresponds to compensation below (above) \$100,000. Thus, for instance, the slope coefficients (sensitivities) of the regression labeled \$50,000, are estimates using the records of those earning between \$45,000 and \$55,000.

When considering the evidence sorted by compensation, it is helpful to remember that half the employees in the sample earn less than \$47,000. On the other hand, those who earn more than \$73,500—23% of the sample—contribute half the money in the sample. Therefore, the findings regarding lower income employees should inspire policies that

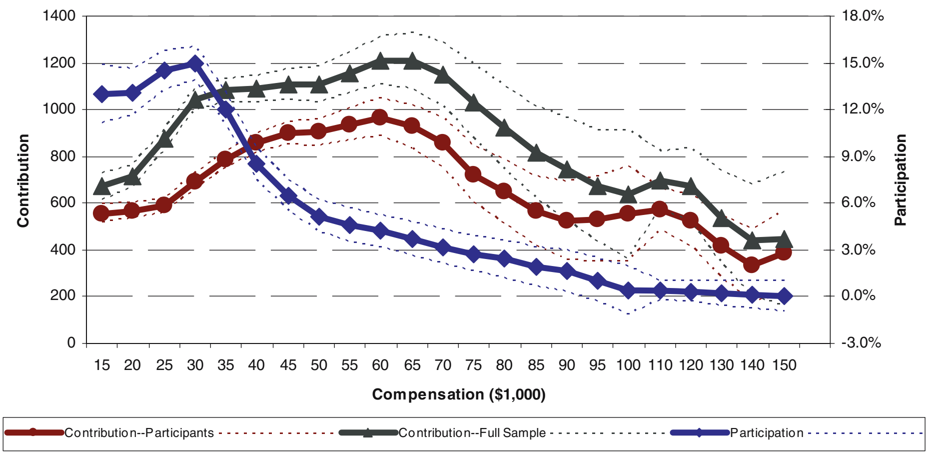


Figure 7 Marginal effect of \$10,000 increase in compensation on participation and contribution

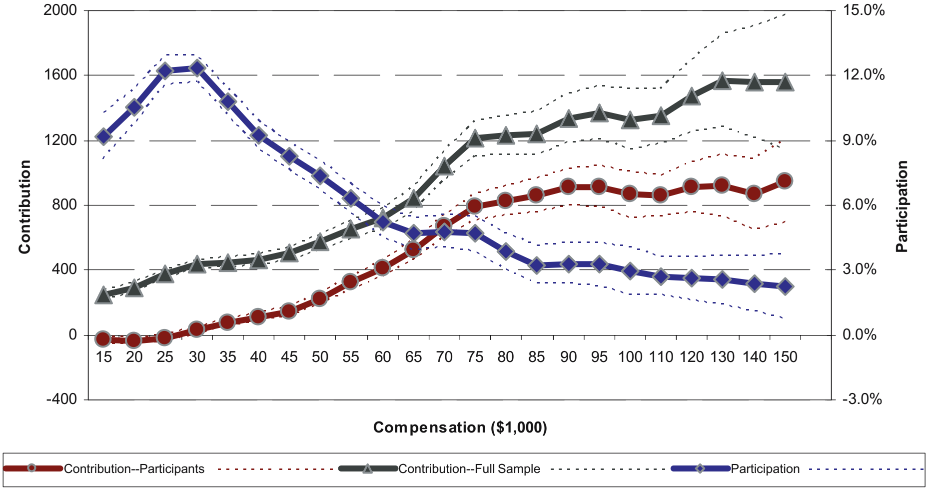


Figure 8 Excess participation rates and contribution levels of women

attempt to affect the most people whereas the findings regarding higher income employees should inspire policies that attempt to affect the most savings.

Figure 7 depicts, for each compensation level, the sensitivity of participation probability and contribution to a $10,000 change in compensation, holding other variables constant. Marginal participation probability peaks at around 15% for those earning $30,000 and declines thereafter. (To appreciate the magnitude, note that the 15% marginal probability is on top of the average participation probability of those earning $30,000 which is 55%.) It is not negative for any compensation level, and for those earning more than $100,000 it is near zero. The marginal effect of compensation on contribution of the eligible employees peaks around a compensation of $60,000 at $1,208. (The $1,208 marginal contribution is on top of the $3,868 average contribution of those earning $60,000.) Even for those earning between $100,000 and $150,000 the marginal contribution ranges between $800 and $500. The

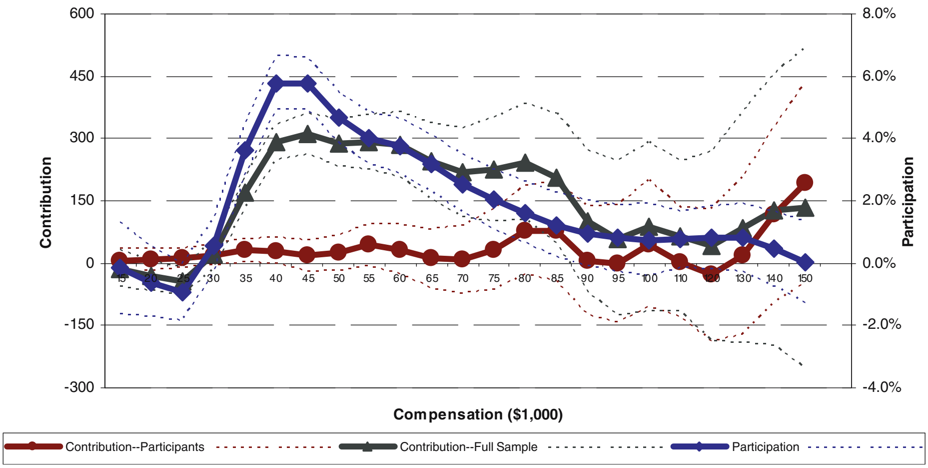


Figure 9 Excess participation rates and contribution levels of those covered by DB

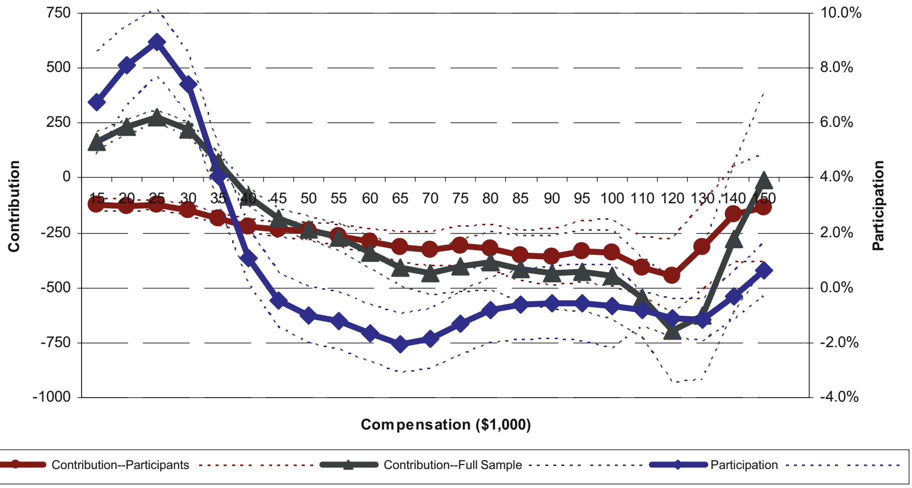


Figure 10 Excess participation rates and contribution levels of company-stock employees

marginal contribution of eligible employees is greater than that of the participants since the marginal participation probability is non-negative.

Figure 8 depicts, for each compensation level, the difference in participation probability and contribution between women and men, holding other variables constant. The average participation probability of men earning \$30,000 is 49% whereas that of women with the same compensation and similar other attributes is 12% higher. The difference declines for higher wages, but even at the highest compensation levels, women's participation probabilities are at least 2% higher than men's. The contributions of women are higher than men's at all compensation levels, and the difference increases (for the whole population) from around \$300 (for those earning around \$20,000) all the way to \$1,500 (for those earning above \$100,000).

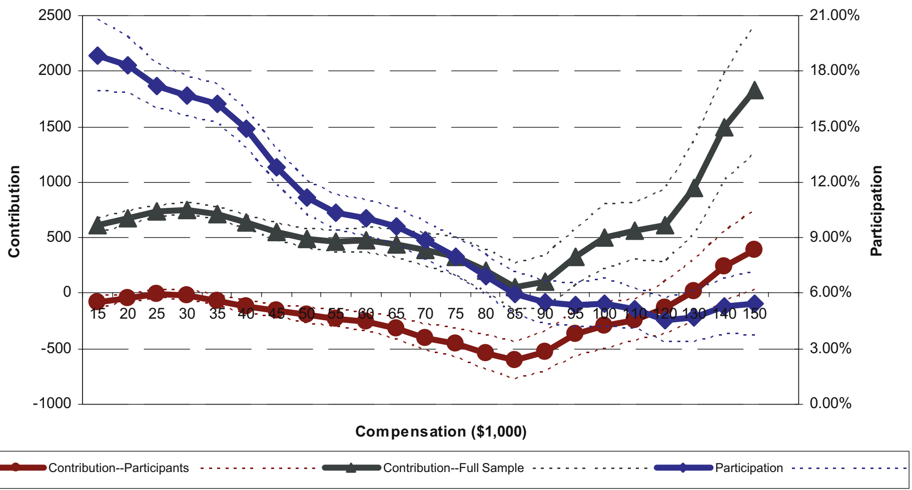


Figure 11 Marginal effect of 100% match on participation and contribution

One explanation for the gender difference is that women are residual income earners in their families: many more low-compensation women than men are married to working spouses and among the majority of working couples, women earn less than men. Consistent with this hypothesis, the data show that women tend to live in wealthier neighborhoods than men of comparable compensation, age and tenure. (Not tabulated.)

If the participation and contribution decision reflects the family's, as opposed to just the employee's needs, then low-income women will participate and contribute as if they had higher incomes than their recorded compensation. This observation may explain the gender gap at the low end of the pay scale, although even here the inclusion of the WEALTH variable should control for the "family" effect. Moreover, if this were the only explanation, women at the higher income levels should behave similarly to men of similar income. But they do not: women earning six digit figures have 2% higher participation probabilities and contribute more than a thousand dollars more than their male counterparts.

Figure 9 depicts the difference in behavior between employees in companies with and without DB plans. The counterintuitive results on higher participation rates and contributions that transpire from Tables 1 and 2 surface here as well, with the additional insight that they are concentrated among those in the middle of the earnings distribution, peaking at incomes of \$40,000–\$45,000.

Remarkably, contributions of participants with and without DB plans are statistically indistinguishable at all income levels. The graph suggests that once an employee decides to participate, the contribution level is unaffected by the presence or absence of the DB plan. The reason that contributions of eligible employees with DB plans are higher than of those without DB plans is that participation probabilities are higher.

Figure 10 depicts the impact of including company stock among the investable funds. For those earning less than \$42,000—41% of the sample—participation probability is higher in the presence of company stock; for those earning \$30,000 or below—30% of the sample, presence of company stock enhances participation probability by about 7%. (Overall participation probability of employees at this income level is 48%.) For those earning below \$35,000 contributions are also higher if company stock is an investable fund, but they are lower for employees earning more than \$35,000. Note that conditional on participation, employees contribute less in the presence of company stock. Presumably, employees attracted to the program by the presence of company stock tend to contribute considerably less than those who would participate regardless of the presence of company stock in the investable funds.

For those earning above \$40,000, the effect of company stock on participation probability is slightly negative (between 0 and −2%), but mostly indistinguishable from zero. More puzzling, and harder to explain is the behavior of contributions: they are lower in the presence of company stock in the investable funds. They can be lower by as much as \$400–750 for those earning between \$65,000 and \$130,000. Why the presence of company stock should adversely affect contributions is unclear, since its presence in the investable funds can be safely ignored by eligible employees and participants, in which case it would leave participation probabilities and contributions unaffected. But this seems not to be the case.

According to Fig. 11 which depicts the relevant sensitivities of participation and contribution to a 100% match, the presence of such a match increases participation at all compensation levels, and such inducement is stronger the lower the compensation. In fact, a 100% match (up to 5% of salary) would lift the average participation rates of those earning \$20,000 by 19%. (Recall that the overall participation rate of employees in this income range is 43%.) Contributions of eligible employees at all income levels are higher when a

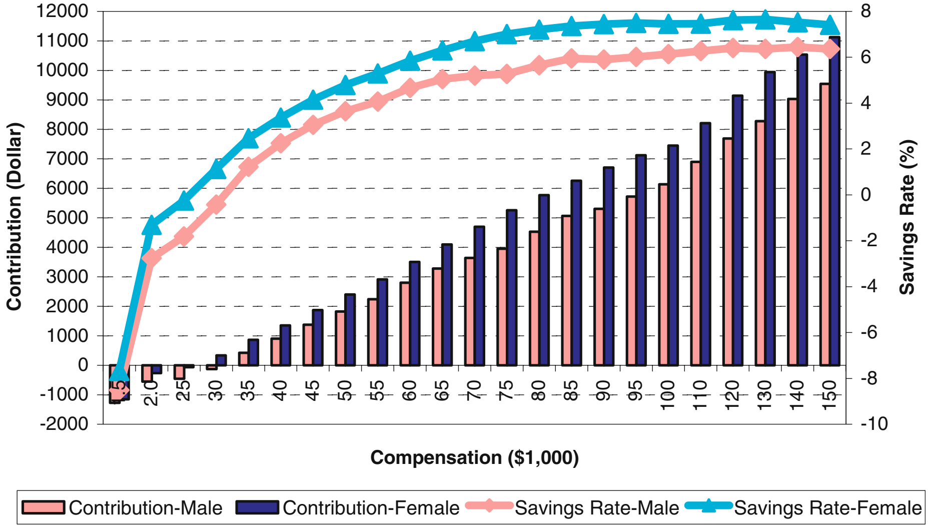


Figure 12 Contribution levels and savings rates for men and women

match is available. In fact, for those earning less than $35,000, a 100% match would increase contributions by about $550–750.[4] (For reference, see Fig. 12 for average contribution levels at various income levels).

The positive impact of a match on all employees' contributions is in contrast with its effect on participants only. It seems that only participants earning above $130,000 increase their contributions in response to the match. Participants earning between $35,000 and $120,000 seem to reduce their contributions in response to a match. To understand this counterintuitive finding requires analysis of the relation between the policy variables—existence of the match, the match rate, and the upper limit on the match—and individuals' choices to participate and, if they participate, how much to contribute. Choi et al. (2002) and Engelhardt and Kumar (2003) summarize how a match affects participants' choices through income and substitution effects.

It is possible that behavioral factors are also necessary to explain the match's influence on employees' choices. A substantial fraction of the contributions are at or near the point where they exhaust the employer's match: the contributions of about 18 (22%) of the participants whose employers offer a match are no more than $100 ($200) away from the upper limit of their employers' match. A similar phenomenon is noted by Choi et al. (2002) on the contribution behavior of participants in one plan that changed the upper limit of match.

A substitution effect could lead participants to cluster their contributions at the point where the employer caps the match. But a substitution effect alone does not cause participants to contribute less when the match rate increases.

A simple explanation for the propensity to contribute an amount which exhausts the employer's match is that the maximal match point is a focal point and interpreted by participants as an implicit suggestion that it is the optimal amount to contribute (Madrian and Shea 2001). This explanation is consistent with the behavior gleaned from

[4] Using an experiment setting, Duflo et al. (2006) find that a 50% match increase the take-up rate of low- and middle-income subjects into IRA contribution by 11 percentage points compared to no match; and increase the contribution by $345. These numbers seem to be in the same order of magnitude as our findings.

Fig. 11, that participants who receive a match contribute less than those who do not receive a match when their incomes are at least \$40,000. The following calculations are consistent with this observation.

A typical upper limit on the match is 5% of salary. It inspires very different contribution levels for low- and high-income participants. The average (median) contribution of participants earning \$40,000 is about \$2,740 (\$2,315), but 5% of \$40,000 is \$2,000. Thus, participants earning around than \$40,000 who contribute just the matched amount will contribute less than their counter-parts who are offered no match. This effect is strengthened when looking at higher earnings, say \$90,000. The average (median) contribution by participants in this income range is \$7,121 (\$7,239), but 5% of \$90,000 is only \$4,500. Thus, participants who use the upper limit on the match as their focal point to choose their contribution level will contribute less than those who do not receive a match at all.

5 Discussion and Concluding Remarks

This is the first study on 401(k) participation and contribution to use non-survey individual-level data that covers a large number of plans (companies) and includes information about non-participants. It offers a few novel and counterintuitive observations on participation in and contributions to 401(k) plans, and provides sharp estimates of sensitivities of these choices to various individual and plan-level variables. The surprising findings are: women are more aggressive users of 401(k) plans; coverage by a DB plan does not adversely affect usage of a 401(k) plan; matching programs positively affect participation rates and contribution of all employees, especially low-income ones, but negatively affect contributions of mid-to-higher income participants.

Other studies have considered some of the issues covered here. However, the results reported here are particularly compelling because of the size and nature of the data used—actual employee records, including non-participants' records, from hundreds of plans. Individual-level data are important because in general, it is inappropriate to estimate a relation on an aggregate level and then infer that an analogous relation holds at the individual level—a problem known as the "aggregation bias" (see, e.g., Freedman 2001; Garrett 2003). For example, at plan level, a \$10,000 increase in average compensation would increase average contribution by \$480, while at individual level the same coefficient is \$907. Further, since individuals choose whether to participate in 401(k) plans and how much to contribute to them, records of non-participants are essential to analyze the participation and contribution decision.

Figure 12 summarizes some of the main findings, plotting contribution and savings rate levels for each gender using predicted contribution imputed from two-sided Tobit coefficients. By the nature of corner solution at zero, the predicted unconditional contribution amount (i.e., accounting for non-participation) could go negative in which case the predicted observed contribution would be zero. The Figure shows that contributions rise with compensation, which is not surprising. It also shows that savings rates rise with compensation, a more remarkable finding. Whether those who earn more also save a larger fraction of their incomes has been a well known question, going back decades prior to Friedman's (1957) classic work on the consumption function. Recently, Dynan et al. (2004) re-visit the issue and conclude that those with higher expected lifetime earnings also have higher savings rates.

In general, the estimation of expected lifetime earnings and of savings is thorny, perhaps intractable. This study uses 2001 compensation records of wage earners, whose incomes fluctuate less than those of the self-employed. The unit of observation is the wage earner, not the household. And savings are narrowly defined as contribution to a 401(k) plan. Keeping these simplifications in mind, Fig. 12 shows quite clearly that savings rate in DC accounts increase with compensation.

The gender difference also transpires from Fig. 12: Relatively more women save, and they save more than men, a result consistent with the findings of Warner and Pleeter (2001) who study US military personnel offered a choice between a lump sum and an annuity upon release from the US armed forces. At the contemporary US government borrowing rate of 7% the present values of the annuities were double those of the corresponding lump sums. Nonetheless, most people took the lump sums. Warner and Pleeter estimate that women's probability of choosing the annuity was 2% higher than men's. Generalizing this finding, one would expect that women are more likely to participate in and contribute to DC plans that defer current consumption into future on favorable terms.

One explanation for the gender difference is that women have a stronger taste for saving, perhaps because they live longer on average. A second explanation starts with the assumption that participation and contribution decisions are made at the household, not the employee level. Women are more likely than men to have working spouses, and women's working spouses earn more than men's working spouses. Thus, comparing a man and a woman with the same income, the woman is likely to live in a household with a higher income, and therefore more likely to participate in a DC plan, and contribute more to it. Without information about marital status, it is difficult to identify such effects. However, the analysis control for the wealth level of the neighborhood in which the household lives (and women do overall live in wealthier neighborhood compared to men of equal earning power), which should to some extent offset the first effect. The second explanation suggests that the gender difference should be strongest for low-income employees and disappear for high-income employees. The differences between women's and men's participation probabilities are indeed highest for those earning less than $40,000, where they exceed 10%. But these differences are still around 2–3% for those earning over $100,000. Also, the gender gap in contribution rates increases rather than decreases with compensation. Thus, the data are consistent with one prediction of the second explanation, but suggest that it is not the only valid explanation.

This paper's findings on DB plans and matching programs are also relevant to those interested in promoting savings. New tax-preferred savings programs can attract new savings (i.e., money that would have been used for consumption) or money from competing savings channels. In the latter case, there would be no increase in aggregate savings. This study shows that surprisingly, other things equal, employees covered by DB plans tend to participate more in, and contribute about the same amount to DC plans once they participate. It is possible that employees have separate mental accounts (Shefrin and Thaler 1992; Thaler 1999) for different accounts of retirement money, and when choosing whether to participate in, and how much to contribute to DC plans, they do not take into account whatever rights they have in their employers' DB plans. Of course, there could be a selection effect at work here: retirement savings-conscientious employees are more likely attracted to firms that provide both DB and DC plans. Moreover, the stronger usage of DC plans by those already covered by DB plans suggests that the presence of a DB plan increases awareness of the need to save for retirement.

The overall impression is that employees save as much in 401(k) plans with or without a DB plan. Thus, the evidence presented here is consistent with the view of Poterba et al.

(1996) ("believing that IRA and 401(k) contributions represent new saving") and not that of Engen et al. (1996) ("we conclude that little, if any, of the overall contributions to existing saving incentives have raised saving.") The implication, then, is that the introduction of new tax-preferred savings programs will likely increase overall savings. This conclusion, however, is not airtight. It is possible that those eligible for DB benefits who are aggressive contributors to 401(k) plans make their contributions with money that would be saved through other channels not covered in the records used in this study.

Match programs increase participation rates, and contributions, primarily of low-income employees. This finding clearly suggests that voluntary participation and contributions in individual retirement accounts are likely to increase if the government were to match the contributions. Moreover, the match will have the strongest impact on the low-income members of society. And, if policy makers find it desirable to limit the subsidy to high-income people, match rates could be set to decline with income.

It does not seem surprising that an employer's match program should increase overall contributions since it increases the compensation, albeit in a deferred form. However, individuals in the medium income range, *conditional on participation*, seem to contribute less when the match is generous. Two explanations come to mind. First, if plan participants have desired levels of total savings, they will contribute less in the presence of employer match than what they would in the absence of such matching. Another influence comes from upper limits on the contribution that is matched, which all plans have. (Usually it is 5–6% of compensation.) The upper limit can serve as a focal point suggesting a desired contribution. Choi et al. (2002) indeed report that many participants contribute exactly to the point where matching ends. To the extent that retirement savings rate increases with compensation, such a focal point will increase contributions of low-income participants and reduce that of mid-to-high-income participants, relative to their counterparts who receive no or little match.

This is a study of choices of employees eligible to participate in 401(k) plans in 2001. Most of the employees' choices were probably made prior to 2001, and a shortcoming of the study is the records' silence on the timing of the employees' choices. Some employee attributes changed over time—certainly age and tenure, and most likely compensation. Moreover, some plan characteristics may have changed between the time an employee made his most recent choice and 2001. A long panel of records can potentially fix some of these issues, but not all. Ameriks and Zeldes (2004) point out that the separation of age, cohort and time effects requires assumptions outside the panel records.

The exploratory data analysis summarized here speaks to a variety of subfields. First, to the growing community of students of retirement plans in general and defined contribution plans in particular. Second, to those interested in savings behavior and especially how it varies across income groups. And third, to those interested in gender differences in decision making. Additionally, this descriptive paper is likely to inform discussions on designs of retirement and savings plans.

Acknowledgement The authors are grateful to Steve Utkus and Gary Mottola from Vanguard® Center for Retirement Research. They made the data available and provided us with constructive comments throughout the process. We also thank Pierre Azoulay, Geert Bakeart, Charlie Calomiris, Sarah Holden, Brigitte Madrian, Jim Powell, Thomas Steinberger, Ed Vytlacil, Elke Weber, Steve Zeldes, an anonymous referee, and seminar participants at Columbia, Rice, Wharton Pension Research Council, New York Federal Reserve Bank, CEPR workshop on Financing Retirement in Europe, and Wharton Workshop on Household Portfolio Choice and Financial Decision Making for their helpful comments.

Appendix

Data Construction

The original data set provided by the Vanguard consists of 926,104 records of 401(k) eligible employees. The following criteria cause elimination of observations: (1) The employee is hired after January 1, 2001. (His recorded annual contribution might be inaccurate.) (2) The employee is less than 18 years old (He might not be the decision maker.) (3) The employee's annual compensation is less than \$10,000 or greater than \$1 million. 793,794 records survive.

The key variables deferral rate, contribution, and compensation appear in all the records. All other individual variables have missing values that are more concentrated in the non-participant sub-sample. 12.8% of the observations do not report gender, among which 62.5% are non-participants; the same percentages for age, tenure and wealth are 12.3% (62.2), 12.2 (62.1%) and 25.6% (76.4). Elimination of all the observations with missing values would cause the study to be based on a partially truncated sample, which is likely to bias the results due to the influence of the selection. Hence the choice to replace them with imputed values.

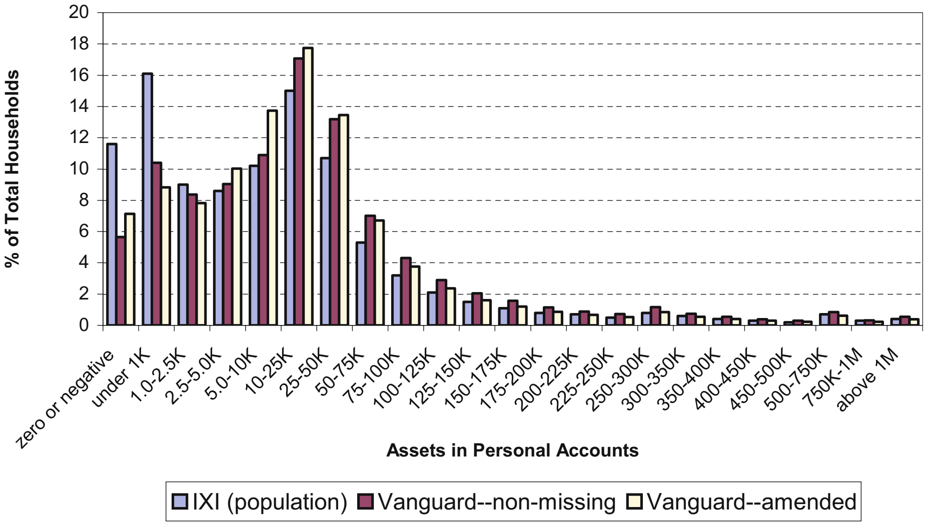
The imputed values are calculated as follows: (1) unidentified gender variables are recorded as the percentage of females in the plan (a record identified as a female being 1 and as a male being 0); (2) missing age and tenure values are replaced with their respective plan mean age or tenure. To fill in the missing values of wealth, the following regression is estimated on non-missing values:

$$\begin{aligned}\text{Log}(\text{WEALTH}) = \beta_0 + \beta_1\text{Log}(\text{COMP}) + \beta_2\text{FEMALE} + \beta_3\text{AGE} + \beta_4\text{AGE}^2 \\ + \beta_5\text{Log}(\text{DCASSETS}) + \varepsilon,\end{aligned}$$

where DCASSETS is the total assets in the defined-contribution accounts. The specification above was chosen among various models for both within-sample goodness-of-fit and out-of-sample robustness. The next step is to predict out-of-sample values and map the predicted values to the closest IXI brackets.

With the exception of the wealth level, IXI, missing values do not account for a significant proportion (about 10%) of observations, and the symmetric trimming method (Honore et al. 1997) is applied to conduct sensitivity check. That is, for each variable create an artificial sample by first taking only records that report that variable (but may miss other variables). The second step eliminates a given number of participants' records at random, so that the participation rate in the subsample matches that of the original sample. (This second step eliminates records of participants, since systematically it is the non-participants whose records are likely to miss values.) Then estimate coefficients for the resulting sub-sample. Repeating this process many times (e.g., 30 or 50), the coefficients estimated on the full sample (with missing values imputed) are close to the average of coefficients estimated on those symmetrically trimmed sub-samples.

For the wealth variable the same sensitivity check cannot be used because only a low proportion of non-participant records have this information. Instead, the comparison is between the inputted wealth variables and the general distribution of wealth in the population. The following figure plots the histogram of wealth distribution for the population (IXI), of the non-missing Vanguard data, and the amended Vanguard data. After filling in the missing values, the sample studied here no long over-represents the wealthy households. The sample still under-represents the very poor households (those with negative or less than $1,000 in balance), and over-represent the lower-middle to middle households (with balances ranging from $5,000 to $100,000), which is consistent with evidences from the Survey of Consumer Finance and the Current Population Survey that 401(k) eligible employees are overall financially better-off than the general population in the lower end.



Further, estimating the main regressions excluding the wealth variable, the coefficients on all variables except compensation show little variation. (The loading on compensation is increased, which is expected.)

References

Agnew J, Balduzzi P, Sunden A (2003) Portfolio choice and trading in a large 401(k) plan. Am Econ Rev 93 (1):193–215 (March)

Ameriks J, Zeldes SP (2004) How do household portfolio shares vary with age? Working Paper, Columbia University

Bajtelsmit, VL, Bernasek, A (1999) Why do women invest differently than men? Financ Couns Plan 7:294–305

Barber B, Odean T (2001) Boys will be boys: gender, overconfidence, and common stock investments. Q J Econ 116(1):261–292 (February)

Benartzi S (2001) Excessive extrapolation and the allocation of 401(k) accounts to company stock? Journal of Finance 56(5):1747–1764

Bernartzi S, Thaler R (2004). Save more tomorrow: using behavioral economics to increase employee saving. J Polit Econ 112(1):S164–S187 (February)

Chamberlain G. Heterogeneity, omitted variables biases, and duration dependence. In: Heckman JJ, Singer B (eds) Longitudinal analysis of labor market data. Cambridge University Press, Cambridge, pp 3–38

Choi J, Laibson D, Madrian B, Metrick A (2002) Defined contribution pensions: plan rules, participation decisions, and the path of lease resistance. In Poterba J (ed) Tax policy and the economy. MIT, pp 67–113

Choi J, Laibson D, Madrian B, Metrick A (2004a) For better and for worse: default effects and 401(k) savings behavior. Perspectives on the Economics of Aging, University of Chicago Press

Choi J, Laibson D, Madrian B, Metrick A (2004b) Employees' investment decisions about company stock. In: Mitchell OS, Utkus SP (eds) Pension design and decision-making under uncertainty. Oxford University Press

Choi J, Laibson D, Madrian B (2004c) Plan design and 401(k) savings outcomes. Natl Tax J 57:275–298, June

Choi J, Laibson D, Madrian B (2005) $100 bills on the sidewalk: suboptimal saving in 401(k) Plans, working paper, Yale School of Management

Clark RL, Schieber SJ (1998) Factors affecting participation rates and contribution levels in 401 (k) plans. In: Mitchell O, Schieber S (ed) Living with defined contribution pensions: remaking responsibility for retirement. University of Pennsylvania Press, Philadelphia, pp 69–97

Cunningham CR, Engelhardt GV (2002) Federal tax policy, employer matching, and 401(k) saving: evidence from HRS W-2 records. Natl Tax J 55:617–645(September)

Duflo E, Saez E (2003) The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. Q J Econ 118:815–842

Duflo E, Gale W, Liebman J, Orszag P, Saez E (2006) Saving incentives for low- and middle-income families: evidence from a field experiment with H&R block. Q J Econ 121(4):1311–1346

Dynan KE, Skinner J, Zeldes SP (2004) Do the rich save more? J Polit Econ 112(2):397–444 (April)

Engelhardt G, Kumar A (2003) Employer matching and 401(k) saving: evidence from the health and retirement study, working paper, Syracuse University

Engen EM, Gale, WG (2000) The Effect of 401(k) Plans on Household Wealth: Differences Across Earnings Groups, NBER Working Paper 8032, December

Engen E, Gale W, Scholz JK (1996) The illusory effects of savings incentives on saving. J Econ Perspect 10 (4):113–138 (Autumn)

Even WE, Macpherson D (2003) Employee participation in 401(K) plans, Working paper, Miami University and Florida State University

Freedman DA (2001) Ecological inference and the ecological fallacy. International Encyclopedia for the Social and Behavioral Sciences 6:4027–4030

Friedman M (1957) A theory of the consumption function. Princeton University Press, Princeton

Garrett TA (2003) Aggregated vs. disaggregated data in regression analysis: implications for inference. Econ Lett 81(1):61–65 (October)

Holden S, VanDerhei J (2001) Contribution behavior of 401(k) plan participants. Investment Company Institute Perspective7(4) (October)

Holden S, VanDerhei J (2003) 401(k) plan asset allocation, account balances, and loan activity in 2001. Investment Company Institute Perspective 9(2) (March)

Honore B, Kyriazido E, Udry C (1997) Estimation of type 3 Tobit models using symmetric trimming and pairwise comparisons. J Econom 76:107–128

Huberman G (2001) Familiarity breeds investment. Rev Financ Stud 14(3):161–178, Fall

Huberman G, Jiang, W (2006) Offering vs. choices in 401(k) plans: equity exposure and number of funds. Journal of Finance XLI(2):763–801

Huberman G, Sengmuller P (2004) Performance predicts asset allocation: company stock in 401(k) plans. Review of Finance 8(3):403–443 (September)

Iyengar S, Kamenica E. (2006) Choice overload and asset allocation: evidence from 401(k) plans, working paper, Columbia Business School and GSB University of Chicago

Iyengar S, Huberman G, Jiang W (2004) How much choice is too much?: contributions to 401(k) retirement plans. In: Mitchell Utkus S (ed) Pension design and structure: new lessons from behavioral finance. Oxford University Press, pp 83–96

Kusko AL, Poterba JM, Wilcox DM (1998) Employee decisions with respect to 401 (k) plans: evidence from individual-level data. In: Mitchell O, Schieber S (eds) Living with defined contribution pensions: remaking responsibility for retirement. University of Pennsylvania Press, Philadelphia, pp 98–112

Liang N, Weisbenner S (2002) Investor behavior and the purchase of company stock in 401 (k) plans—the importance of plan design, NBER Working paper 9131 (August)

Madrian B, Shea D (2001) The power of suggestion: inertia in 401(k) participation and savings behavior. Q J Econ CXVI(4):1149–1187 (November)

Manski CF (1993) Identification of endogenous social effects: the reflection problem. Rev Econ Stud 60:531–542

Meulbroek LK (2002) Company stock in pension plans: how costly is it? Working paper, Harvard Business School

Mitchell OS, Utkus SP (2003) The role of company stock in defined contribution plans. In: Mitchell OS, Smetters K (eds) The pension challenge: risk transfers and retirement income security. Oxford University Press, Oxford, UK pp 33–70

Mitchell OS, Utkus SP, Yang, Tongxuan (Stella) (2005) Turning workers into savers? Incentives, liquidity and choice in 401(k) plan design. NBER Working Paper 11726 and Wharton Pension Research Council Working Paper 2005–18

Mitchell OS, Utkus SP, Yang, Tongxuan (Stella) (2006) Dimensions of 401(k) plan design. In: Blitzstein D, Mitchell OS, Utkus SP (eds) Restructuring retirement risks. Oxford University Press, UK, pp 186–203

Munnell AH, Sunden A, Taylor C (2001) What determines 401(K) participation and contribution? Working paper, Center for Retirement Research, Boston College

Papke L (2004a) Individual financial decisions in retirement saving plans: the role of participant-direction. J Public Econ 88:39–61 (January)

Papke L (2004b) Choice and other determinants of employee contributions to defined contribution plans. Social Secur Bull

Papke LE, Poterba JM (1995) Survey evidence on employer match rates and employee saving behavior in 401(k) plans. Econ Lett 49(3):313–317 (September)

Poterba JM (2003) Employer stock and 401(k) plans. Am Econ Rev 93(2):398–404 (May)

Poterba JM, Venti SF, Wise DA (1996) How retirement savings programs increase saving. J Econ Perspect 10(4):91–112 (Autumn)

Powell JL (1984) Least absolute deviation estimation for the censored regression model. J Econom 25: 303–325

Profit Sharing/401(k) Council of America (2001, 2002) The Profit Sharing/401(k) Survey

Ramaswamy K (2002) Company stock and pension plan diversification, Working Paper, Pension Research Council

Shefrin H, Thaler R (1992) Mental accounting, saving, and self-control. In: Loewenstein G, Elster J (eds) Choice over time. Russel Sage Foundation, New York, pp 98–112

Thaler R (1999) Mental accounting matters. J Behav Decis Mak 12:183–206

Warner JT, Pleeter S (2001) The personal discount rate: evidence from military downsizing programs. Am Econ Rev 91(1):33–53 (March)

Wooldridge JM (2003) Cluster-sample methods in applied econometrics. Am Econ Rev May 2003, 93(2), pp.133–138