



DISCUSSION PAPER PI-0311

Defined Contribution Pension Plans: Determinants of Participation and Contribution Rates

Gur Huberman, Sheena Iyengar and Wei Jiang

June 2003

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University
106 Bunhill Row London
EC1Y 8TZ
UNITED KINGDOM

<http://www.pensions-institute.org/>

Defined Contribution Pension Plans: Determinants of Participation and Contributions Rates¹

Gur Huberman²

Sheena Iyengar³

Wei Jiang⁴

This Draft: June 2003

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) (i) 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; (ii) women's participation probability is 6.5% higher than men's and they contribute almost \$500 more than men; (iii) 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; (iv) 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.

¹ 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, Thomas Steinberger, Elke Weber, Steve Zeldes, and seminar participants at Columbia, Wharton Pension Research Council, and CEPR workshop on Financing Retirement in Europe for their helpful comments.

² Finance and Economics Division, Columbia Business School, e-mail: gh16@columbia.edu.

³ Management Division, Columbia Business School, e-mail: ss957@columbia.edu.

⁴ Finance and Economics Division, Columbia Business School, e-mail: wj2006@columbia.edu.

Defined Contribution Pension Plans: Determinants of Participation and Contributions Rates

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) (i) 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; (ii) women's participation probability is 6.5% higher than men's and they contribute almost \$500 more than men; (iii) 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; (iv) 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.

1. Introduction

Defined contribution pension plans in general, and 401(k) plans in particular are important vehicles for retirement savings. A handful of studies have considered individual and plan-level attributes that affect participation in such plans, or, for participants, levels of contribution. However, these studies either had only survey data (Papke, 2002; Munnell, Sunden, and Taylor, 2000; Even and Macpherson, 2001) or used employee records for very few firms (Kusko, Poterba, and Wilcox, 1998; Clark and Schieber, 1998; Agnew, Balduzzi, and Sunden, 2003) or used plan-level data (Papke, 1995; 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.

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. 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 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.

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 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, 1990, 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.

The observation that employees behave as if there is no substitution between their rights in DB and DC plans suggests that money that goes into new tax-preferred savings vehicles is not likely to come from other tax-preferred vehicles, but from consumption. Thus, the evidence presented here is consistent with the view of Poterba, Venti and Wise (1996) ("believing that IRA and 401(k) contributions represent new saving") and not that of Engen, Gale and Scholtz (1996) ("we conclude that little, if any, of the overall contributions to existing saving incentives have raised saving.")

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 probability to participate by nearly 20%; at higher incomes (above \$90,000), the incentive effect drops to about 5%.

It seems that an employer's match program should increase contributions since it increases the compensation, albeit in a deferred form. However, 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 which is matched, which all plans have. (In most of them 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 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.

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 negative effect on contributions of those earning between \$40,000 and \$130,000.

The strong effect of matching programs on participation, especially of the low-income employees, has 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: employees tilt their portfolios in favor of company stock, at the heavy cost of under-diversification. Other studies that consider the impact of company stock on asset allocation in 401(k) plans include Benartzi (2001), Choi *et al.* (2003), Huberman and Sengmueller (2002), Liang and Weisbenner (2002), Mitchell and Utkus (2002), Muelbroek (2002) and Ramaswamy (2002). 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 to participate 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, 1996). 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 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 decrease with compensation. Thus, the data are consistent with one prediction of the second explanation, but suggest that it is not the only valid explanation.

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. 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). 793,794 observations survive.

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. Other information about individual characteristics includes tenure and financial wealth of the 9-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 9-digit zip level, and then divides the total financial assets by the number of households in the relevant 9-digit zip area to determine the average assets for each neighborhood. There are 10-12 households in a 9-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, and ignores the other two types of contributions. 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.

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. 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 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.

Figure 4 plots individual annual contributions for the full sample and the sub-sample with compensation above \$80,000. Among all participants, the modal contribution is between \$1,000 and \$2,000. Those who earn more than \$80,000 contribute more than the typical participants, and about 36% of those earning more than \$80,000 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 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, what are restrictions on diversification of the employer's match. 124 plans (covering 58% of the employees in the sample) have 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: (i) A dummy variable, *PART*, that equals one if the individual participates, that is, if she contributes a positive amount to her tax-deferred account; (ii) A dummy variable, *MAXOUT*, that equals one if the individuals contributes the maximum amount (\$10,500 in 2001) to her tax deferred account; (iii) 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 individual characteristics X_{ij} , plan policies Z_j , and \bar{X}_j represents the plan-level averages of individual characteristics:

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

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

$$PART_{ij} = \begin{cases} 1, & \text{if } U_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

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:

$$y_{ij}^* = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_j + \beta_3 \bar{X}_j + \delta_{ij}, \delta_{ij} = \varphi_j + \delta'_{ij}. \quad (3)$$

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:

$$y_{ij} = \begin{cases} y_{ij}^*, & \text{if } y_{ij}^* < \bar{v}, \\ \bar{v}, & \text{otherwise.} \end{cases} \quad (4)$$

where $\bar{v} = \$10,500$, the maximum allowed contribution. 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 \$10,500, 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., Figure 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 $x_i' \hat{\beta}$ 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 equation (3), individual characteristics (X_{ij}) include: (i) Annual compensation in \$10,000 or in logarithm ($COMP$); (ii) The wealth rank (IXI rank from 1 to 24) of the 9-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; (iii) A gender dummy (*FEMALE*); (iv) Age in years in excess of 18 (*AGE*); (v) 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 vary 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*.

Finally, plan-level control variables include: (i) average compensation (*COMP_MEAN*); (ii) average age (*AGE_MEAN*); (iii) average tenure (*TENURE_MEAN*); (iv) average wealth (*WEALTH_MEAN*); (v) log number of employees (*NEMPLOYEE*), (vi) 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. (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 individual's own choices. (Manski, 1993, offers detailed analysis of peer effects.)

3. Participation and Contribution: Full-Sample Analyses

3.1 Individual Participation

Table 1 reports participation regressions (equation (2)) using both the linear probability (Columns 1 to 4) and the probit models (Columns 5 to 6). In linear probability models, *COMP* and *WEALTH* are expressed in log dollars because they are both highly right-skewed variables, and Figure 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). Columns (1) to (3) report results from pooled linear probability regressions. In column (2) *WEALTH* is excluded, and in column (3) all policy variables are removed. In column (4), all coefficients on individual characteristics regressors are obtained using the “within” data, that is, data expressed as deviation from their respective plan-level mean values. By construction, these within estimates are not at all affected by plan-level variables. The “between” data, that is, residuals calculated using “within” estimated coefficients, are used to identify coefficients on all plan-specific regressors. The consistency of the estimates across different specifications is remarkable. Bear in mind that when standard errors are adjusted for plan random effect (η_j) in equation (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 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 $1 - \ln(L)/\ln(L_0)$, where L_0 is the log-likelihood value with the constant term only.

The latter is defined as $\widehat{\Pr}\left(\hat{y}_i > \frac{1}{2} \mid y_i > 0\right) + \widehat{\Pr}\left(\hat{y}_i < \frac{1}{2} \mid y_i < 0\right) - 1$, where $\widehat{\Pr}$ is the empirical frequency, and \hat{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 handful explanatory variables. Viewed in this context, the explanatory power of 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. 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%) 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. But 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 panel of records can determine the validity of this hypothesis.

Finally, the proximity of within estimates (displayed in Column 3, estimated from data differenced out their respective plan-level mean values) and the pooled estimates (both from linear probability and probit specifications) in Table 1 suggests that the control variables are indeed adequate, and that the results are not driven by plan-level unobserved heterogeneity.

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. 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 2.5%. 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 who are not covered by a DB plan. The similarity in behavior suggests that, couterintuitively, the presence of a DB plan does not affect the participation decision in a 401(k) plan.

Controlling for other variables is crucial to the result and its interpretation because a comparison of the raw data lead 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%. 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). Thus, DB has no effect on participation only after controlling for these and the other variables in the analysis.

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; contribution in logarithms (accordingly both compensation and wealth are in logarithms as well); and saving rate in percentage (i.e., the ratio of contribution to compensation).

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 (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 log-log and savings rate specifications (Columns (2) and (3)). The estimations of the latter two specifications assign 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 (2002a) 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.

The coefficients in the log-log specification (Column (2)) can be interpreted as elasticity or semi-elasticity. Column (2) in Table 2 indicates that a 1% increase in compensation is associated with a 5% increase in contribution (the coefficients displayed are multiplied by 100, or in basis points). Females are likely to contribute 0.9% more than otherwise equal males. A 100% match will increase contributions by 1.6%. The presence of company stock increases contribution by 0.6%. Finally, the presence of DB plans increases the contribution by those likely DB beneficiaries by .15%.

The savings rate specification (Column (3)) is quite consistent with the other two specifications, showing that an increase in compensation from \$40,000 to \$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 .18% when company stock is an investable fund and by .25% when a DB plan covers the employee.

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 Table 1, except that Column (2) uses only the sub-sample of the 385,260 participants who earned more than \$42,000 in 2001, and therefore were allowed to contribute up to \$10,500. (Only six of the 176,753 participants who earn less than \$42,000 a year choose to contribute the maximum allowed – 25% of their incomes.) All individual characteristics affect the probability of maxing out in the same direction as they do the probability of participation. However, the effects of plan policies are quite different: when company stock is an

investable fund, the probability of maxing out is lower by 1.5%, and when a DB plan covers a participant, his maxing out probability is .7% lower.

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.

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 estimated 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 relation between 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:

$$\frac{\partial E[y_{ij}^* | W_{ij}]}{\partial W_{ij}} = \Pr[PART_{ij} = 1] \frac{\partial E[y_{ij}^* | W_{ij}, PART_{ij} = 1]}{\partial W_{ij}} + E[y_{ij}^* | W_{ij}, PART_{ij} = 1] \frac{\partial \Pr[PART_{ij} = 1]}{\partial W_{ij}} \quad (5)$$

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, $\frac{\partial E[y_{ij}^* | W_{ij}]}{\partial W_{ij}}$, 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, $\Pr[PART_{ij} = 1]$ is the probability of participation given all attributes, and $\frac{\partial E[y_{ij}^* | W_{ij}, PART_{ij} = 1]}{\partial W_{ij}}$ is the sensitivity of the *desired* contribution by an employee conditional on participation; $E[y_{ij}^* | W_{ij}, PART_{ij} = 1]$ is the expected contribution amount conditioned on participation, and finally, $\frac{\partial \Pr[PART_{ij} = 1]}{\partial W_{ij}}$ 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 $\frac{\partial E[y_{ij}^* | W_{ij}]}{\partial W_{ij}}$; the one-sided Tobit coefficients are consistent estimates of $\frac{\partial E[y_{ij}^* | W_{ij}, PART_{ij} = 1]}{\partial W_{ij}}$, and the marginal probabilities from probit estimation are consistent proxies for $\frac{\partial \Pr[PART_{ij} = 1]}{\partial W_{ij}}$. 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 to 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 Figures 7 to 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 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 17% for those earning \$30,000 and declines thereafter. (To appreciate the magnitude, note that the 17% 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 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).

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 to \$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 -1%), 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 Figure 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 match is available. In fact, for those earning less than \$35,000, a 100% match would increase contributions by about \$550-750 (for reference, see Figure 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. Rather than view the match as an opportunity to earn money for increased saving, they seem to substitute the match for their own saving and probably allow for an increased current consumption.

It is likely that the match formula itself implicitly suggests contribution levels and the nature of the suggestion can explain why participants earning more than \$40,000 seem to reduce their contributions in the presence of a match. For instance, if the match is offered for contributions up to 5% of salary, many participants will contribute exactly 5% of their salary. Further, the higher the match rate, the stronger the "power of suggestion." These very same participants, had they not been aware of the match's termination point, may have contributed different amounts. In fact, the 5% of salary upper limit on match is quite popular. 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 less 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 study 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, and 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. 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. 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 analyses of 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, which is more remarkable. 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, Skinner and Zeldes (2003) 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, Figure 12 shows quite clearly that savings rate increase with compensation.

The gender difference also transpires from Figure 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.

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 seems that they have separate mental accounts (Shefrin and Thaler, 1990; 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. 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 to 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.

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 her most recent choice and 2001. A panel of records can potentially fix some of these issues, but not all. Ameriks and Zeldes (2001) 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. Moreover, this descriptive paper is likely to inform discussions on designs of retirement and savings plans.

REFERENCES:

Agnew, Julie; Balduzzi, Pierluigi and Sunden, Annika. “Portfolio Choice and Trading in a Large 401(k) Plan,” *American Economic Review*, March 2003, 93(1), pp. 193-215.

Ameriks, John, and Zeldes, Stephen P., “How Do Household Portfolio Shares Vary With Age?” Working Paper, Columbia University, 2001.

Bajtelsmit, Vickie L. and Bernasek, Alexander. “Why do women Invest Differently than Men?” *Financial Counseling and Planning*, 1999, 7, pp. 294-305.

Barber, Brad and Odean, Terrence. “Boys will be boys: Gender, Overconfidence, and Common Stock Investments,” *Quarterly Journal of Economics*, February 2001, 116(1), pp. 261-292.

Bernartzi, Shlomo and Thaler, Richard. “Naïve Diversification Strategies in Defined Contribution Saving Plans,” *American Economic Review*, March 2001, 91(1), pp. 79-98.

Bernartzi, Shlomo and Thaler, Richard. “How Much Is Investor Autonomy Worth?” Working paper, UCLA and University of Chicago, 2001.

Bernartzi, Shlomo and Thaler, Richard. “Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy*, forthcoming

Buchinsky, Moshe and Hahn, Jinyong. “An Alternative Estimator for the Censored Quantile Regression Model,” *Econometrica*, May 1998, 66(3), pp. 653-71.

Chamberlain, Gary. “Heterogeneity, Omitted Variables Biases, and Duration Dependence,” in *Longitudinal Analysis of Labor Market Data*, ed. J. J. Heckman and B. Singer. Cambridge: Cambridge University Press, pp. 3-38.

Choi, James; Laibson, David and Madrian, Brigitte and Metrick, Andrew. “For Better and for Worse: Default Effects and 401(K) Savings Behavior,” *NBER Working Paper 8651*, 2001.

Choi, James; Laibson, David and Madrian, Brigitte and Metrick, Andrew. “Defined Contribution Pensions: Plan Rules, Participation Decisions, and the Path of Lease Resistance,” in James Poterba, ed. *Tax Policy and the Economy*, MIT Press, 2002, 67-113.

Choi, James; Laibson, David and Madrian, Brigitte and Metrick, Andrew. “Employees’ Investment Decisions About Company Stock,” working paper prepared for presentation at the Pension Research Council Symposium, April 2003.

Clark, Robert L. and Schieber, Sylvester J. “Factors Affecting Participation Rates and Contribution Levels in 401 (k) Plans,” in *Living with Defined Contribution Pensions: Remaking Responsibility for Retirement*, ed. O. Mitchell and S. Schieber. Philadelphia: University of Pennsylvania Press, 1998, pp. 69-97.

Dynan, Karen E., Skinner, Jonathan, and Zeldes, Stephen P., “Do the Rich Save More?” Working Paper, Columbia University, 2003.

Engen, Eric; Gale, William and Scholz, John K. “The Illusory Effects of Savings Incentives on Saving,” *Journal of Economic Perspective*, Autumn 1996, 10(4), pp. 113-38.

Even, William E. and Macpherson, David. “Employee Participation in 401(K) Plans,” Working paper, 1999.

Freedman, David A. “Ecological inference and the ecological fallacy.” *International Encyclopedia for the Social and Behavioral Sciences*, 2001, 6, 4027–30.

Friedman, Milton. *A Theory of the Consumption Function*, Princeton: Princeton University Press, 1957.

Huberman, Gur. “Familiarity Breeds Investment.” *Review of Financial Studies*, Fall 2001, 14(3), pp 161-78.

Huberman, Gur and Sengmuller, Paul. “Company Stock in 401(k) Plans,” Working Paper, Columbia Business School and Universiteit van Amsterdam, 2002.

Khitatrakun, Surachai and Scholz, John K. “Savings Incentives in the U.S.”, Working paper, Center for Research on Pensions and Welfare Policy, University of Turin, Italy, 2001.

Kusko, Andrea L., Poterba, James M., and Wilcox, David M. “Employee Decisions with Respect to 401 (k) Plans: Evidence from Individual-Level Data,” in *Living with Defined Contribution Pensions: Remaking Responsibility for Retirement*, ed. O. Mitchell and S. Schieber. Philadelphia: University of Pennsylvania Press, 1998, pp. 98-112.

Liang, Nellie and Weisbenner, Scott. “Investor Behavior and the Purchase of Company Stock in 401 (k) Plans – The Importance of Plan Design,” *NBER Working paper 9131*, August 2002.

Madrian, Brigitte and Shea Dennis. “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, November 2001, CXVI (4), pp. 1149-87.

Manski, Charles F. “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, 1993, 60, pp. 531-42.

Meulbroek, Lisa K. “Company Stock in Pension Plans: How Costly Is It?” Working paper, Harvard Business School, 2002.

Mitchell, Olivia S. and Utkus, Stephen P. “Company Stock and Retirement Plan Diversification,” Working paper, Pension Research Council, The Wharton School, 2001.

Munnell, Alicia H; Sunden, Annika and Taylor, Catherine. “What Determines 401(K) Participation and Contribution?” Working paper, Center for Retirement Research, Boston College, 2001.

Olsen, Randall J. “Approximating a Truncated Normal Regression with the Method of Moments,” *Econometrica*, September 1978, 46(5), pp. 1099-106.

Papke, Leslie. “Individual Financial Decisions in Retirement Saving Plans: the Role of Participant-Direction,” *Journal of Public Economics*, forthcoming, 2002a.

Papke, Leslie. “Choice and Other Determinants of Employee Contributions to Defined Contribution Plans,” Working paper, Center for Retirement Research, Boston College, 2002b.

Poterba, James M.; Venti, Steven F. and Wise, David A. “How Retirement Savings Programs Increase Saving,” *Journal of Economic Perspective*, Autumn 1996, 10(4), pp. 91-112.

Poterba, James M.; Venti, Steven F. and Wise, David A. “Saver Behavior and 401(k) Retirement Wealth,” *American Economic Review*, May 2000, 90(2), pp. 297-302.

Powell, James L. “Least Absolute Deviation Estimation for the Censored Regression Model,” *Journal of Econometrics*, 1984, 25, pp. 303-25.

Ramaswamy, Krishna. “Company Stock and Pension Plan Diversification,” Working Paper, Pension Research Council, 2002.

Shefrin, Hersh and Thaler, Richard. “Mental Accounting, Saving, and Self-Control,” in *Choice Over Time*, ed. Loewenstein, George and Elster, Jon, 1992, Russel Sage Foundation, New York, pp. 98-112.

Thaler, Richard. “Mental Accounting Matters,” *Journal of Behavioral Decision Making*, 1999, 12, pp. 183-206.

The Vanguard Group. “How America Saves,” 2002.

Warner, John T. and Pleeter, Saul. “The Personal Discount Rate: Evidence from Military Downsizing Programs,” *American Economic Review*, March 2001, 91(1), pp. 33-53.

Wooldridge, Jeffrey M. “Cluster-Sample Methods in Applied Econometrics,” *American Economic Review*, May 2003, 93(2), pp. 133-138.

Table 1. Determinants of Individual Participation

Dependent variable is *PART*. The all-sample participation rate is 70.8%. All coefficients are multiplied by 100. Columns (1) to (4) are results from the linear probability model. COMP and WEALTH are expressed in log dollars. 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 (5) and (6) report results from probit estimation. In column (6), COMP is expressed in \$10,000, and WEALTH is expressed in IXI ranks from 1 to 24. The standard errors are adjusted for correlation within the same plan. Pseudo R-squared and incremental probability of correct prediction are 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 number of observation is 793,794.

		Linear Probability						Probit							
		Pooled Regressions				Within-Between		Log-COMP			Dollar-COMP				
		(1)		(2)		(3)		(4)		(5)		(6)			
		COEF	SE	COEF	SE	COEF	SE	COEF	SE	COEF	SE	Margl. Pr.	COEF	SE	Margl. Pr.
I	CNST	-214.14	36.88	-230.26	29.15	-196.39	40.05	-214.20	43.72	-926.76	81.03	--	-173.06	86.21	--
	COMP	15.27	0.21	22.37	0.15	15.21	0.23	15.31	0.10	57.34	4.72	18.12	11.54	0.94	3.71
	WEALTH	5.96	0.06	--	--	5.93	0.07	5.93	0.03	23.16	1.30	7.32	2.14	0.34	0.69
	FEMALE	5.64	0.50	7.34	0.56	4.66	0.84	6.32	0.11	18.88	1.29	5.97	20.11	0.93	6.47
	AGE	0.21	0.05	0.19	0.04	0.22	0.08	0.26	0.02	0.32	0.44	0.10	0.88	0.51	0.28
	AGE^2	0.00	0.00	0.00	0.00	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	1.36	0.09	1.17	0.14	1.28	0.02	4.79	0.47	1.51	5.07	0.55	1.63
	TENURE^2	-0.03	0.00	-0.03	0.00	-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	0.12	0.02	--	--	0.12	0.02	0.44	0.04	0.14	0.41	0.04	0.13
	COMPSTK	3.50	1.60	3.27	1.41	--	--	3.56	1.67	9.47	4.40	3.01	7.53	3.58	2.41
	DB	-0.28	1.45	-0.17	1.37	--	--	-0.26	1.56	1.01	2.11	0.32	0.26	2.60	0.09
	NFUNDS	-0.23	0.11	-0.22	0.12	--	--	-0.27	0.12	-0.92	0.31	-0.25	-0.78	0.34	-0.25
III	COMP_MEAN	3.30	4.50	1.36	2.59	3.46	5.39	3.41	4.96	7.36	6.12	2.32	2.77	0.94	0.89
	WEALTH_MEAN	-1.70	2.62	--	--	-1.12	3.13	-1.80	2.51	-5.38	4.58	-1.70	-5.31	2.56	-1.71
	AGE_MEAN	1.49	0.32	1.46	0.27	0.99	0.59	1.47	0.40	4.53	1.50	1.43	5.27	2.02	1.70
	TENURE_MEAN	-1.09	0.29	-1.14	0.31	-0.77	0.38	-1.07	0.37	-3.78	0.70	-1.19	-4.17	0.92	-1.34
	WEB	0.07	0.08	0.11	0.07	0.18	0.09	0.07	0.08	0.31	0.14	0.10	0.65	0.14	0.21
	NEMPLOYEE	-2.89	0.52	-3.13	0.47	-3.37	0.77	-2.92	0.53	-9.55	2.10	-3.02	-9.80	2.02	-3.15
Goodness of Fit		0.19		0.16		0.18		0.19		0.18		0.33	0.13		0.26

Table 2: Determinants of Individual Contribution: Censored Median Regression Analysis

The dependent variables are: contribution in dollars (Column (1)), log of contributions (Column (2), where *COMP* is in logarithms as well), and savings rate (equals contribution/compensation, Column (3)). Censored median regression (Powell (1984)) is applied to all specifications. Pseudo R-squared is the proportion of the sum of absolute deviations in the dependent variable explained by the regression. The number of observation is 793,794.

	<u>Linear</u>		<u>Log-Log</u>		<u>Saving Rate*</u>		
	COEF	SE	COEF*100	SE	COEF*100	SE	
I	CNST	-3973.8698	189.92	-6399.18	410.68	-202.15	44.02
	COMP	909.05	3.50	504.46	25.81	99.37	1.00
	WEALTH	80.97	2.33	80.15	5.04	15.90	0.48
	FEMALE	481.50	17.14	87.89	6.35	105.32	3.91
	AGE	-33.12	3.60	-0.99	0.69	-5.39	0.97
	AGE^2	1.11	0.07	0.07	0.01	0.22	0.02
	TENURE	78.87	3.27	16.06	1.09	18.10	0.66
	TENURE^2	-2.07	0.10	-0.40	0.03	-0.46	0.02
II	MATCH_AVG	4.57	0.29	1.64	0.11	1.27	0.06
	COMPSTK	-12.38	18.15	64.10	5.46	17.59	4.01
	DB	180.46	17.91	14.97	2.86	24.95	3.88
	NFUNDS	-3.01	1.62	-2.75	0.41	-1.33	0.29
III	COMP_MEAN	13.11	6.00	27.93	5.06	6.64	1.18
	WEALTH_MEAN	-100.79	10.92	-7.07	2.68	-26.74	2.20
	AGE_MEAN	58.02	4.52	26.21	2.04	18.24	1.02
	TENURE_MEAN	-67.79	3.74	-20.67	1.57	-16.66	0.77
	WEB	19.32	0.94	0.91	0.13	3.65	0.19
	NEMPLOYEE	-115.62	5.81	-47.95	3.52	-35.62	1.32
	Pseudo R-sqr	0.25		0.10		0.10	

*: A quadratic term of *COMP* is included to account for the concavity of savings rate vs. compensation (see Figure 12). The coefficient is -3.61 (in basis point) and the standard error is 0.04.

Table 3. Determinants of Individual Maxing-Out

Dependent variable is *MAXOUT*. The maxing-out rate among participants is 12.2%. All coefficients are multiplied by 100. Columns (1) to (4) are results from the linear probability model. *COMP* and *WEALTH* are expressed in log dollars. 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 (5) and (6) report results from probit estimation. In column (6), *COMP* is expressed in \$10,000, and *WEALTH* is expressed in IXI ranks from 1 to 24. Standard errors are adjusted for within-plan correlation. Pseudo R-squared and incremental probability of correct prediction are 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 number of observation is 562,013.

		Linear Probability								Probit					
		Pooled Regressions				Within-Between				Log-COMP			Dollar-COMP		
		(1)		(2)*		(3)		(4)		(5)		Margl. Pr.	(6)		Margl. Pr.
		COEF	SE	COEF	SE	COEF	SE	COEF	SE	COEF	SE		COEF	SE	
I	CNST	-280.56	8.44	-497.01	34.06	-284.43	9.04	-280.14	9.79	-2879.19	55.03	--	-653.18	53.32	--
	COMP	26.55	0.11	41.09	0.53	26.53	0.11	26.59	0.10	207.73	2.41	12.73	19.23	0.26	2.09
	WEALTH	1.40	0.03	1.41	0.08	1.43	0.04	1.30	0.03	6.84	0.49	0.42	2.20	0.16	0.24
	FEMALE	3.29	0.16	4.00	0.87	3.54	0.28	2.69	0.09	22.79	1.24	1.40	13.17	1.41	1.43
	AGE	-0.35	0.02	-0.35	0.05	-0.34	0.02	-0.39	0.02	-0.87	0.36	-0.05	0.62	0.41	0.07
	AGE^2	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.04	0.01	0.00	0.01	0.01	0.00
	TENURE	0.09	0.03	0.19	0.09	0.09	0.03	0.07	0.02	0.57	0.23	0.03	0.55	0.28	0.06
	TENURE^2	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	-0.02	0.01	0.00	-0.01	0.01	0.00
II	MATCH_AVG	-0.02	0.01	-0.04	0.04	--	--	-0.02	0.01	-0.27	0.04	-0.02	-0.19	0.04	-0.02
	COMPSTK	-1.85	0.64	-2.84	2.96	--	--	-1.93	0.60	-17.40	3.81	-1.10	-13.88	3.67	-1.54
	DB	-0.80	0.67	-0.91	2.29	--	--	-0.82	0.77	-9.37	1.85	-0.59	-6.44	2.06	-0.71
	NFUNDS	0.05	0.03	0.05	0.13	--	--	0.05	0.04	0.55	0.15	0.03	0.70	0.20	0.08
III	COMP_MEAN	-3.85	0.93	-2.27	4.34	-3.60	1.13	-3.94	1.04	-11.10	4.95	-0.68	-1.16	0.58	-0.13
	WEALTH_MEAN	1.17	0.64	1.88	2.49	0.97	0.65	1.32	0.63	17.51	3.14	1.07	8.66	1.45	0.94
	AGE_MEAN	0.36	0.13	0.81	0.49	0.47	0.14	0.36	0.16	4.79	0.84	0.29	4.35	0.88	0.47
	TENURE_MEAN	-0.47	0.11	-0.54	0.28	-0.60	0.10	-0.49	0.11	-2.29	0.32	-0.14	-2.35	0.37	-0.25
	WEB	0.16	0.03	0.21	0.12	0.13	0.03	0.16	0.03	0.99	0.10	0.06	1.28	0.10	0.14
	NEMPLOYEE	0.65	0.24	1.32	0.83	0.52	0.21	0.69	0.22	9.48	1.45	0.58	8.63	1.44	0.94
Goodness of Fit		0.27		0.29		0.27		0.27		0.41		0.33	0.38		0.34

*: (2) only uses observations that have *COMP* greater than \$42,000 (385,249 observations), the max out rate is 17.8% in this subsample.

Figure 1. Age Distribution of Eligible Employees and Participants

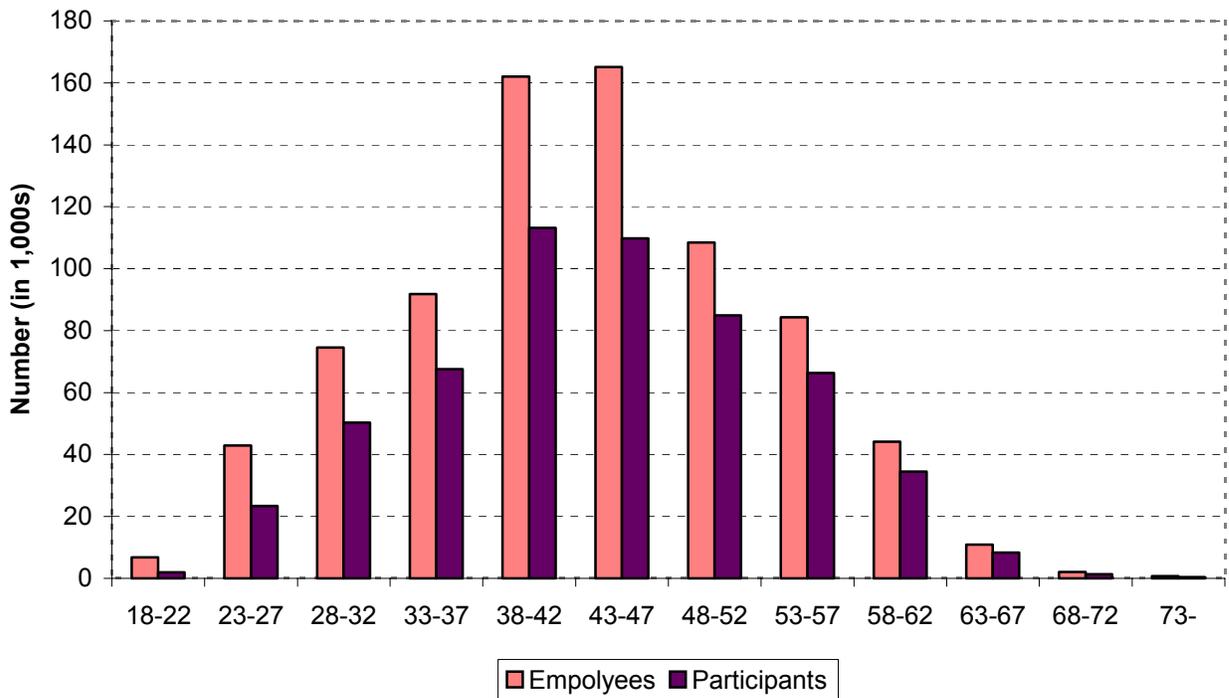


Figure 2. Compensation Distribution of Eligible Employees and of Participants

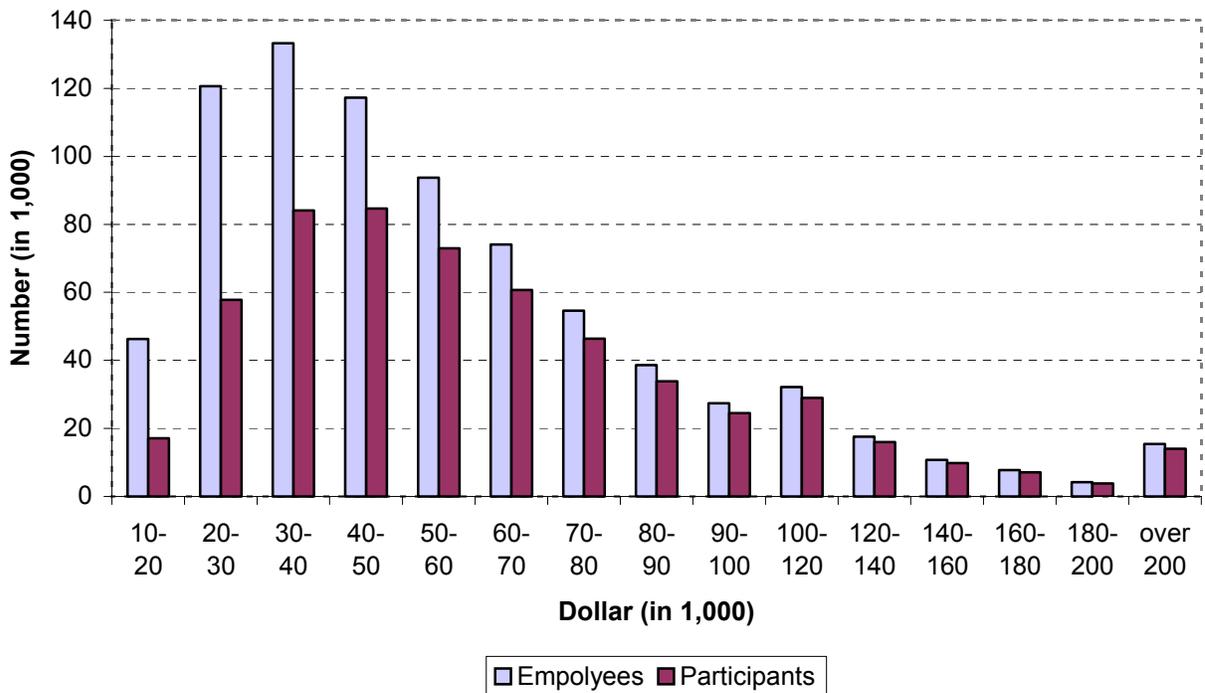


Figure 3: Rates of Participation and Contributing and Maximum at Each Compensation Level

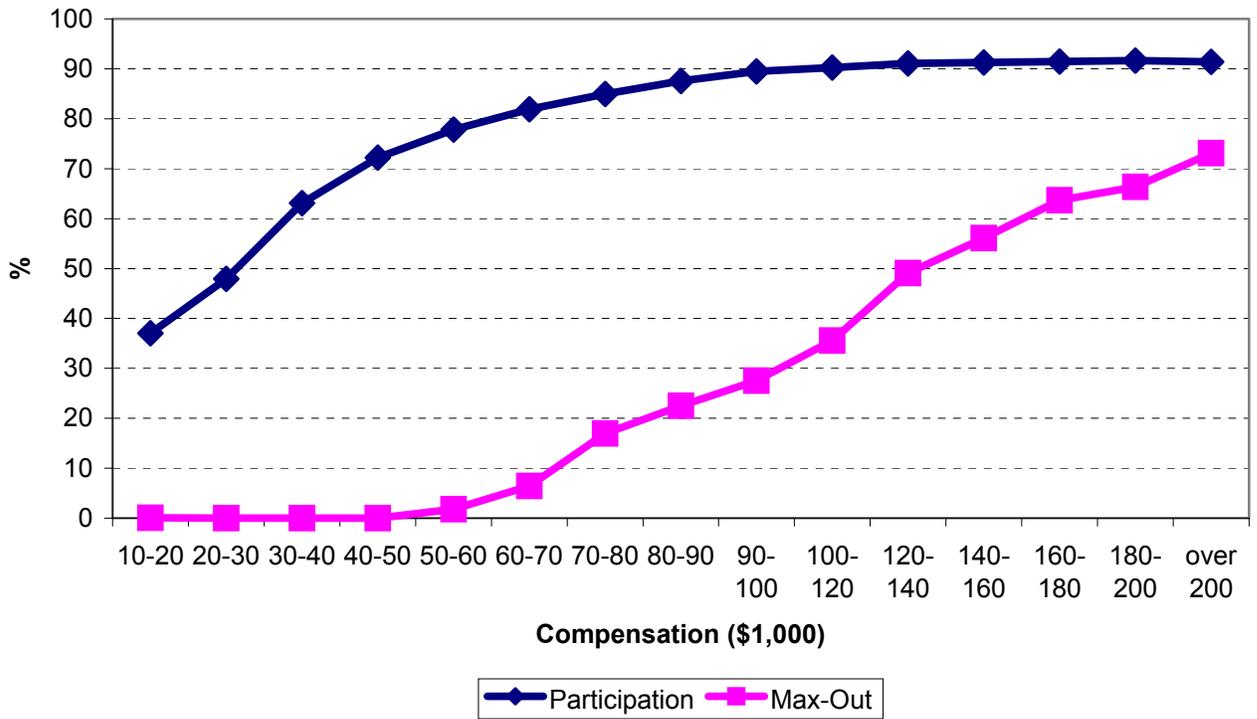


Figure 4: Fraction of Employees at Each Contribution Level

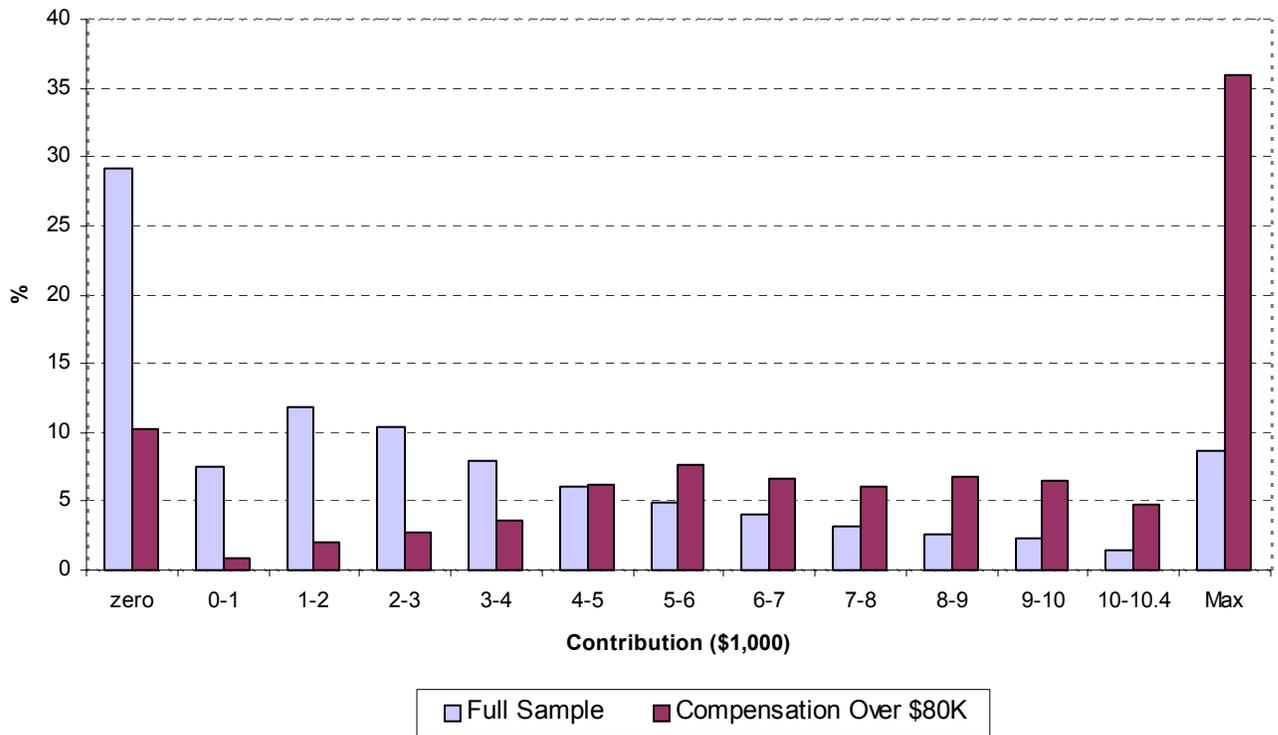


Figure 5. Contribution Levels at Each Percentile

(For each compensation level, the 10th, 25th, 50th, 75th, and 90th percentile of contribution in dollars)

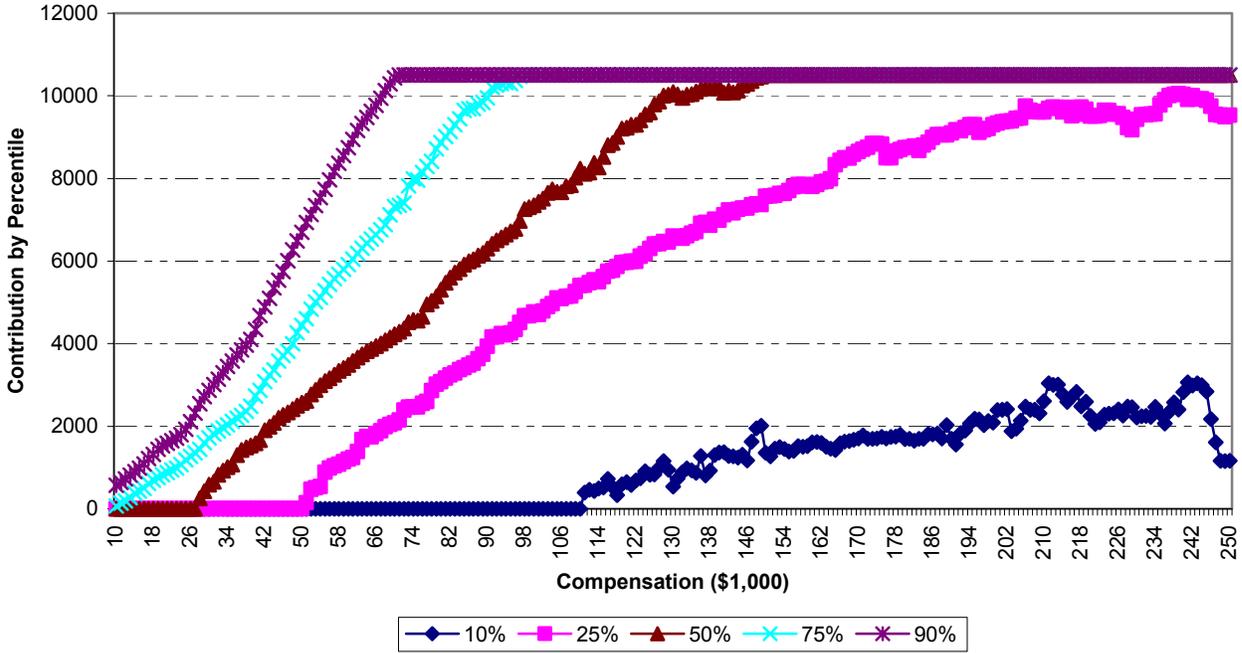


Figure 6. Fractions of Households at Each Wealth Level

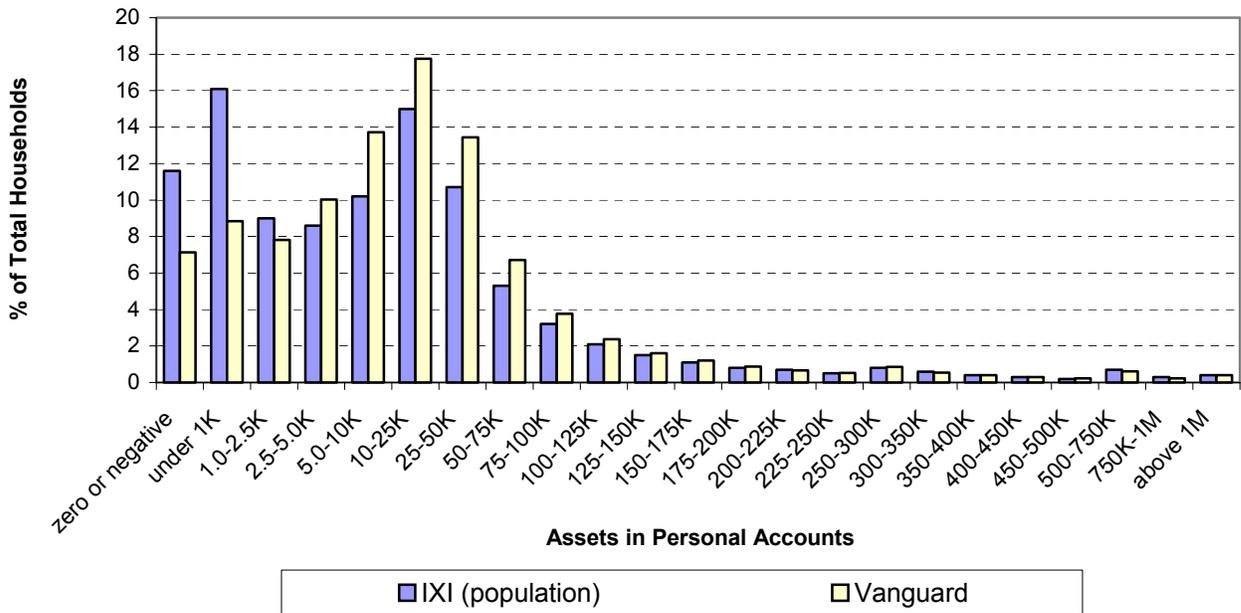


Figure 7: Marginal Effect of \$10,000 Increase in Compensation on Participation and Contribution

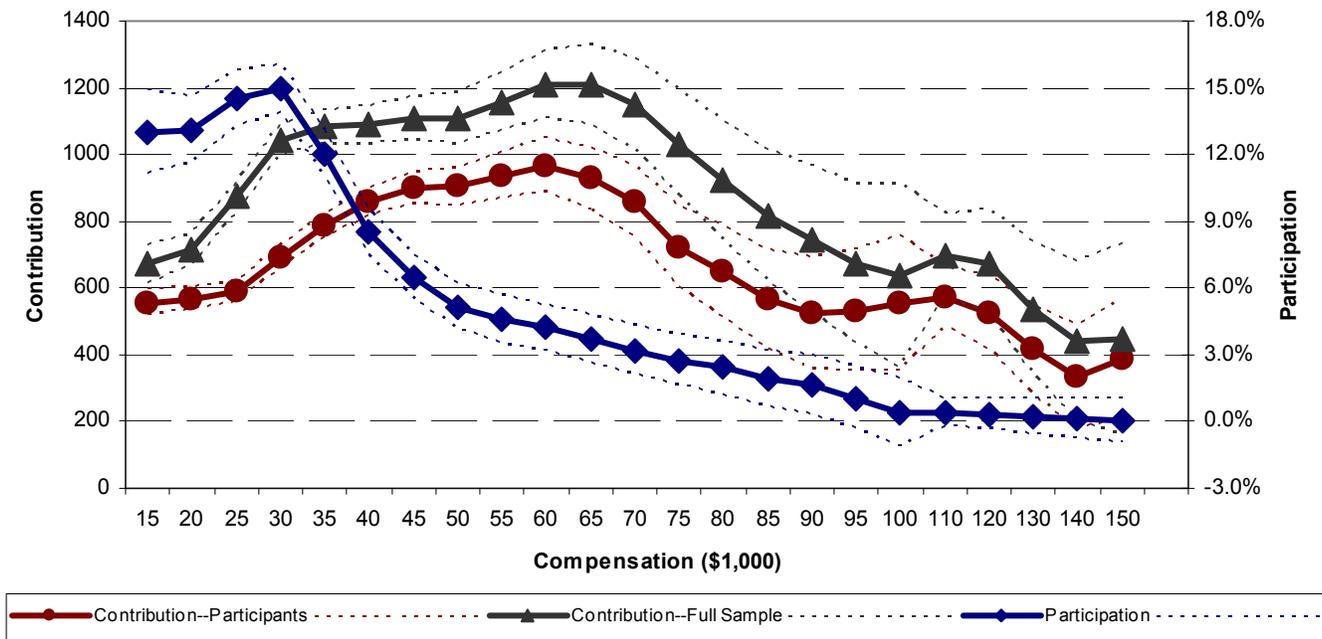


Figure 8: Excess Participation Rates and Contribution Levels of Women

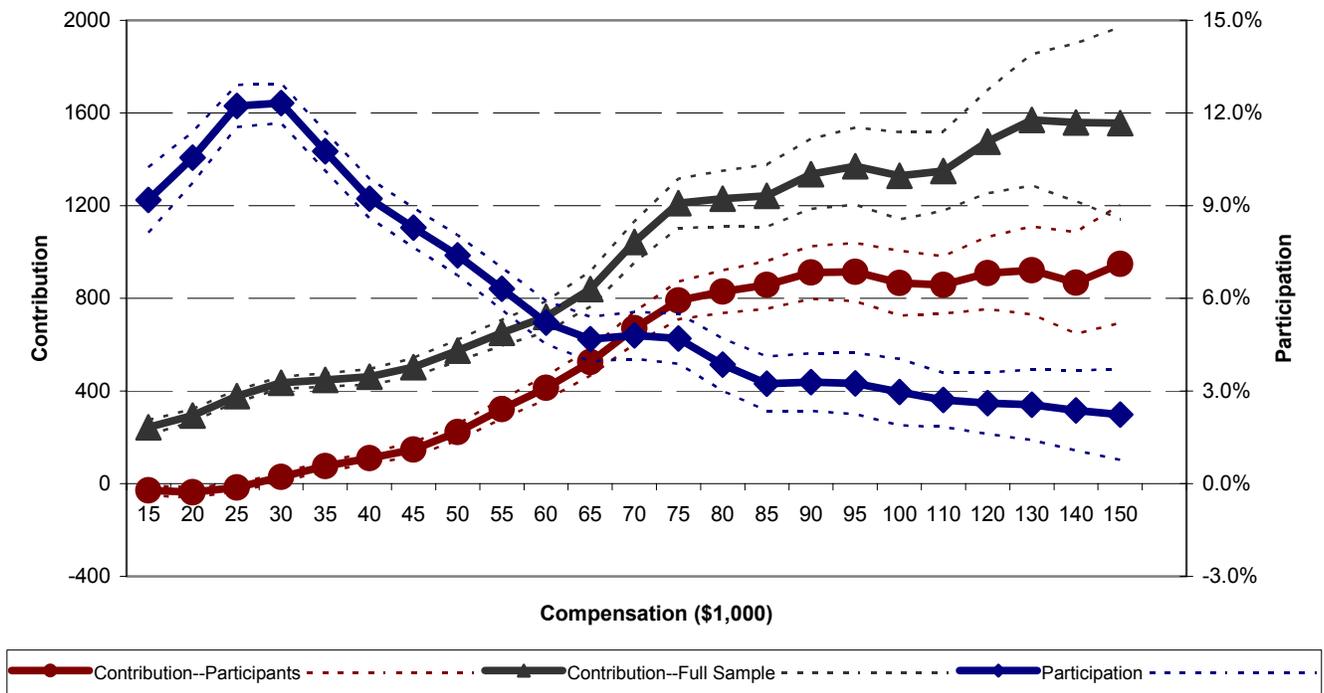


Figure 9: Excess Participation Rates and Contribution Levels of Those Not Covered by DB

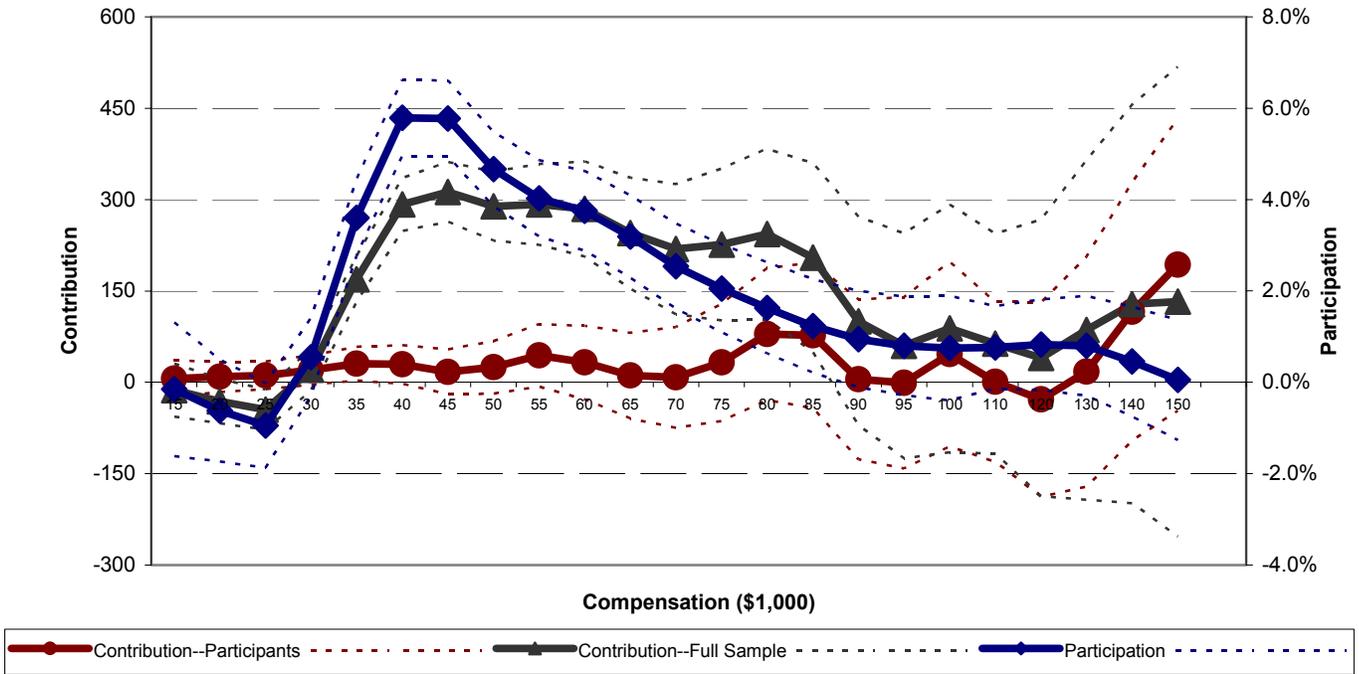


Figure 10: Excess Participation Rates and Contribution Levels of Company-Stock Employees

(Employees in plans with company stock as an investable fund)

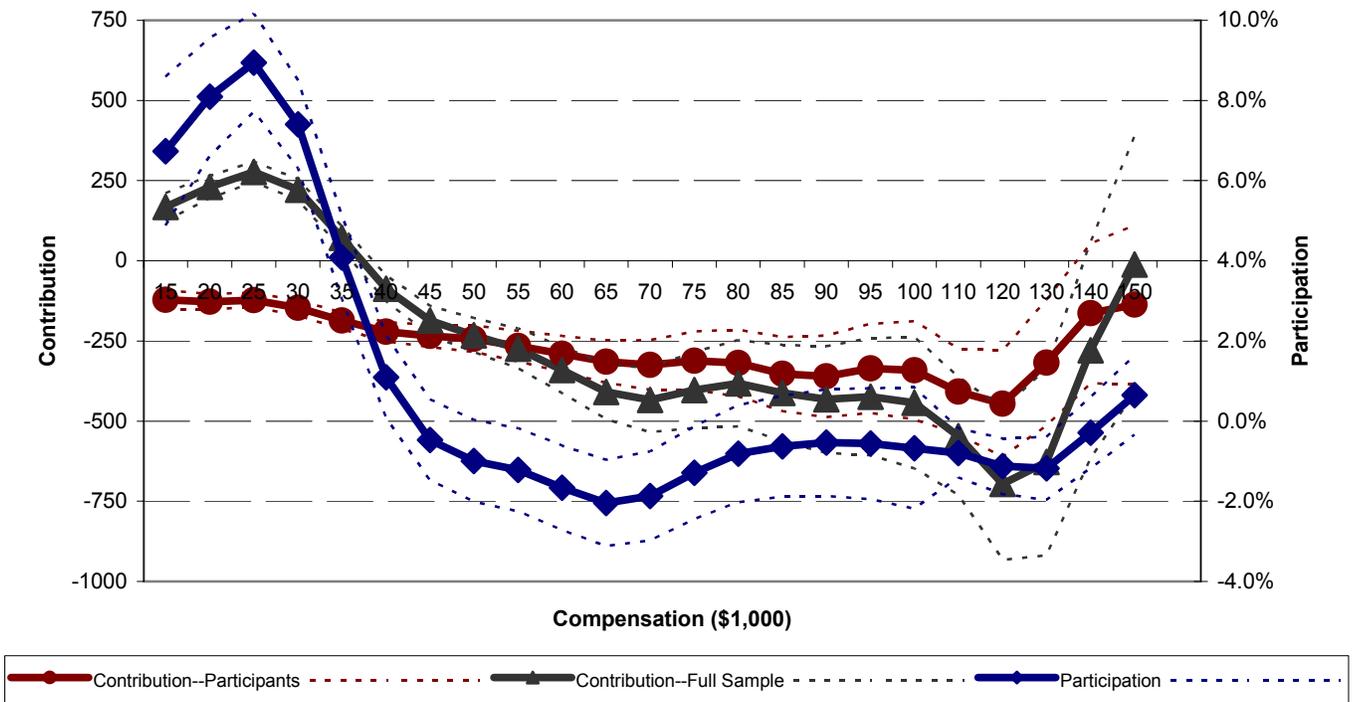


Figure 11: Marginal Effect of 100% Match on Participation and Contribution

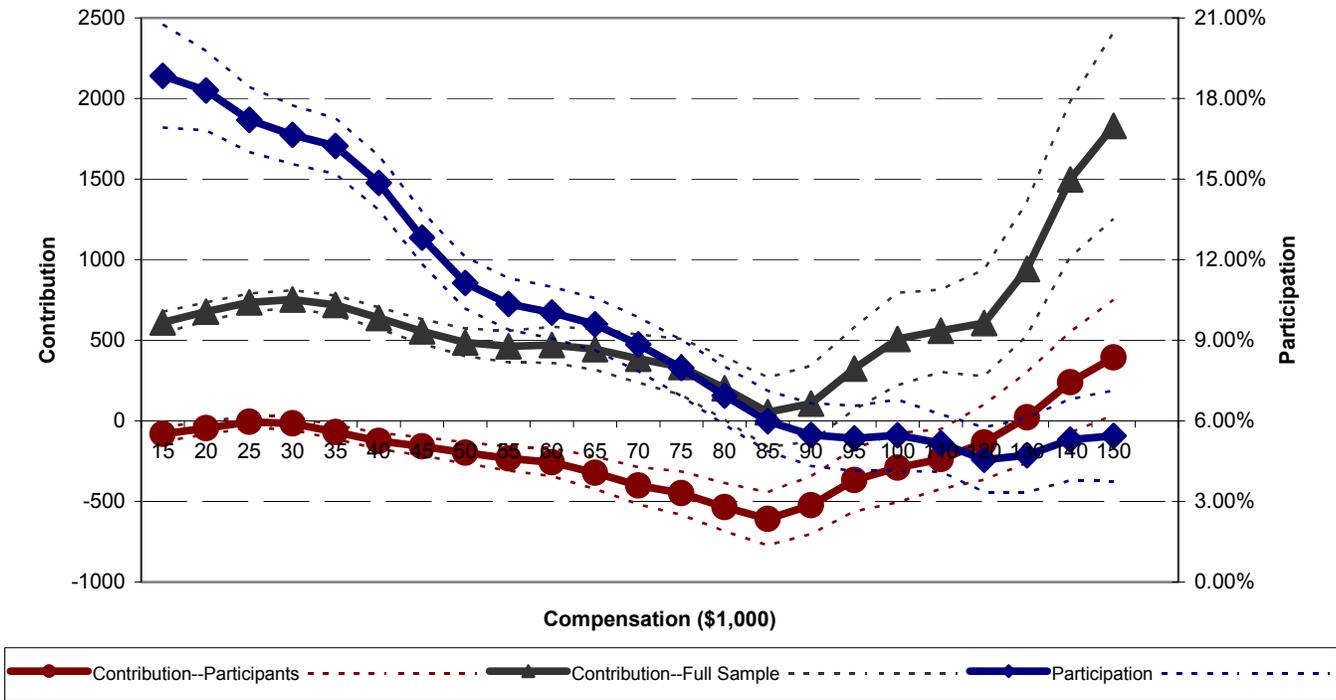


Figure 12: Contribution Levels and Savings Rates for Men and Women

