# Offering vs. Choice in 401(k) Plans: Equity Exposure and Number of Funds

Gur Huberman and Wei Jiang*

ABSTRACT

Records of more than half a million participants in more than six hundred 401(k) pension plans indicate that participants tend to use a small number of funds: The number of participants using a given number of funds peaks at three funds and declines after more than three funds. Participants tend to allocate their contributions evenly across the funds they use, with the tendency weakening with the number of funds used. The median number of funds used is between three and four, and is not sensitive to the number of funds offered by the plans, which ranges from 4 to 59. A participant's propensity to allocate contributions to equity funds is not very sensitive to the fraction of equity funds among those offered by his plan. The paper also comments on limitations on inference available from experiments and from aggregate-level data analysis.

---

*: Both authors are at the Finance and Economics Division, Columbia Business School. An earlier version of this article circulated under the title "The 1/N Heuristic in 401(k) Plans." The authors thank Steve Utkus and Gary Mottola from Vanguard® Center for Retirement Research for making the data available and providing us with constructive comments throughout the process. We are grateful to Tano Santos who gave extensive comments on various versions of the draft. We also thank an anonymous referee, Julie Agnew, Brad Barber, Shlomo Benartzi, Stefano DellaVigna, David Goldreich, Ray Fisman, Garud Iyengar, Sheena Iyengar, Eric Johnson, Nellie Liang, Toby Moskowitz, Jim Powell, Ariel Rubinstein, Rob Stambaugh (editor), Suresh Sundaresan, Ed Vytlacil, Elke Weber, Steve Zeldes, and seminar participants at Amsterdam, Berkeley, CEIBS, CEMFI, Columbia, Duke, the Federal Reserve Board, Hebrew University, HKUST, Peking University, Pompeu Fabra, Tel Aviv, Wharton Workshop on Household Portfolio Choice and Financial Decision Making, European Finance Association 2004 Meeting, Stockholm Institute for Financial Research Conference on Portfolio Choice and Investor Behavior, and NBER Behavioral Finance Meeting for their helpful suggestions. We are particularly grateful to Richard Thaler who pointed out oversights in an early draft. Lihong Zhou and Frank Yu Zhang provided excellent research assistance.

How much and how to save for retirement is one of the most important financial decisions made by most people. Defined contributions (DC) pension plans such as the popular 401(k) plans are important instruments of such savings. By year-end 2001, about 45 million American workers held 401(k) plan accounts with a total of $1.75 trillion in assets (Holden and VanDerhei (2003)). An important characteristic of these plans is that the allocation of the savings among the various funds made available by the plan is the participant's responsibility. How responsibly do the participants behave? In particular, how sensitive are their choices to possible framing effects associated with the menu of choices they are offered?

To explore these questions, this paper analyzes a data set recently provided by the Vanguard group consisting of records of more than half a million participants in about 640 DC plans. These plans offer between four and fifty-nine funds in which participants can invest. All plans offer at least one stock fund. 635 plans offer at least one money market fund, 620 offer at least one bond fund. The Vanguard S&P 500 Index Fund is the most popular fund and is available to participants in 596 plans. The proportion of equity funds tends to be higher in plans that offer the higher number of funds.

This study's main findings are as follows. First, participants choose a small number of funds – typically no more than three or four – regardless of the number of funds they are offered. Second, a substantial fraction of them tend to allocate their contributions evenly among the funds they choose. Third, there is little relation between the proportion of contributions which participants allocate to equity funds (equity allocation) and the proportion of equity funds which their plans offer (equity exposure).

A relation between equity allocation and equity exposure would have theoretical and policy implications. On the theoretical side, it would suggest that two otherwise identical individuals who happen to participate in plans that offer different equity exposures would end up with substantially different portfolios – an indication of irrational behavior. On the policy side, if the plan's menu was important in participants' equity allocations, then menu design would be an important task that should be carefully and thoughtfully undertaken. However, the absence of a relation between equity allocation and equity exposure suggests that menu design is not important and that the data failed to reject the null hypothesis of rationality in the direction of the alternative that plan menus influence participants' equity allocations.

Asset allocation in 401(k) plans is related to, but different from, the classic portfolio selection problem that calls for the allocation of invested money among various assets. The problem may look different when the only assets available are funds of more assets. This is approximately the situation facing participants in 401(k) retirement saving plans, where the assets available for investment are mainly mutual funds, including money market funds. (Company stock and guaranteed investment contracts are often also available.) Two hypotheses can be examined using the data. One is rooted in neoclassical economics and the other is inspired by observations on the tendency to diversify.

Economic theory suggests that an investor should not be concerned with the number of assets in his portfolio, or the composition of the ensemble offered to him. Rather, the investor's focus ought to be the selected portfolio's risk-return profile. Investors with this attitude need not spread their holdings across more than a handful of funds, and the fraction of equity funds among the offered funds should not affect the fraction of their savings allocated to equity funds as long as the set of offered funds is sufficiently diverse. These predictions are in sharp contrast with a behavioral insight derived from studies showing the propensity to diversify, whether rationally justifiable or not. (These studies include Simonson (1990) and Read and Lowenstein (1995).) In particular, Benartzi and Thaler (2001) point out that if DC plan participants apply such naïve diversification to the allocation of their DC savings, they will spread them evenly across the funds made available by their plans, i.e., follow a 1/n rule.

The hypothesis that participants use the 1/n strategy (or the 1/n hypothesis) has a few versions examined in this study. One way of distinguishing among the versions is by considering whether the n's chosen by participants are sensitive to the n's offered by their plans. The basic version of the 1/n hypothesis is that participants tend to allocate their contributions evenly among the funds they choose (which may be a subset, even a small subset of the funds offered). Such allocation could be justified as rational investing, and is different from Benartzi and Thaler's (2001) version of 1/n where the equality of allocation is among the funds offered. The menu-effect (or framing) version is that participants tend to use more funds in plans that offer more funds, and they allocate proportionately more money to equity funds in plans where the proportion of equity funds in the overall offerings are higher.[1] This study explores these hypotheses.

In fact, equally weighted allocation to *chosen* funds is quite prevalent. Consider, for instance, the 20,268 participants in the sample who started their 401(k) plans in 2001 (where information about current-year contribution allocation on a fund-by-fund basis is available) and allocated their contributions to between two and five funds. (For technical reasons which are explained in Section II.C below, this part of the analysis excludes investments in company stock.) About one third of them allocated their contributions approximately evenly among the funds they chose. Another 14,588 participants allocated all their contributions to a single fund. Thus, the 1/n intuition seems valid when it comes to the allocation of contributions among the funds chosen by participants. Such an allocation need not be inconsistent with rationality of the decision makers. On the other hand, the framing-effect version of the intuition is inconsistent with rationality because it implies that very similar individuals make very different choices in a very important context.

In contrast, the data are less supportive of the framing-effect version of the 1/n intuition. The median number of funds used by individuals ranges between three and four, regardless of the number of funds offered. In fact, only a negligible minority of participants has positive balances in all the plans available to them: slightly less than 0.5% of the over half million participants studied here. Even among plans that offer 10 funds or fewer, the same proportion is about 1%.

In a similar vein, the proportion participants invest in equity is not very sensitive to the proportion of equity funds offered to them. All the plans allow their participants to choose equity allocation between zero and 100%; 13% of the participants chose to allocate their contributions to funds which entail no equity exposure, whereas 34% of them chose to invest only in equity funds. The ratio of the number of equity funds to the total number of funds a plan offers to participants varies from 25% to 87.5% in the present sample, but about 99% of the participants have a more narrow equity exposure of between 50% and 80%. Among plans that offer 10 funds or fewer, equity allocation and exposure are statistically significantly correlated, which is consistent with both a framing effect and a constrained choice hypotheses. Once plans offer an abundance of choices (above 10 funds), there is no correlation between equity allocation and exposure. In both cases, variation in equity exposure hardly explains the variation in participants' chosen allocations. Further, the probability that a participant facing higher equity exposure allocates proportionally more to equity funds than another participants with similar attributes (compensation, gender, and age) but facing lower equity exposure is indistinguishable from half.

To summarize, overall the offered fund mix and number of funds offered hardly influence participants' choices of funds. The result is more compelling when the fund mix is sufficiently diverse. A wide range of plan offerings are comparable in that they induce similar choices by participants of similar attributes, and thereby are similar in the welfare they confer on them.

This paper builds on Benartzi and Thaler (2001) who "show that some [401(k)] investors follow the '1/n strategy': they divide their contributions evenly across the funds offered in the plan. Consistent with this naïve notion of diversification, we find that the proportion invested in stocks depends strongly on the proportion of stock funds in the plan." Thus, a remarkably simple behavioral insight would imply (at least potentially) serious financial welfare consequences to unwittingly naïvely diversifying DC plan participants. The inference of Benartzi and Thaler (2001) is based primarily on experiments and plan-level data. The inference of this paper is based on data more suitable to the task at hand: records of actual individual choices. Section IV discusses the differences between aggregate- and individual-level analysis and the limitations of using aggregate data for inferences about individual behavior.

The rest of the paper proceeds as follows: the next section describes the data used here; Section II documents that the number of funds participants typically use is small, and does not vary with the number of funds offered by plans; Section III documents the insensitivity of the fraction of the contributions participants allocate to equity funds to the fraction of equity funds among the funds which are offered to them; Section IV discusses the findings and Section V concludes.

## I. Data Description and Definition of Variables

The data underlying this study, provided by the Vanguard Group, are a cross section of records of eligible employees (including those who choose to not participate) in 647 defined contribution (DC) pension plans, mostly 401(k) plans for the year 2001. The data span 69 SIC two-digit industries. All plans required eligible employees to opt into the plan.

For a more detailed description of the data, see Huberman, Iyengar and Jiang (2004). Table I contains summary statistics of the main variables used in this study.

[Insert Table I approximately here.]

An employee is classified a 401(k) participant if in 2001 he contributed to the plan.[2] 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) Survey by Council of America (2002)). Individual contributions range from zero to the lower of $10,500 and 25% of employee salary, the statutory maximum in 2001. 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 (2002)). In summary, the savings behavior of employees in the Vanguard sample seems representative of the overall population of eligible employees.

Six plans did not provide information about asset allocation by individuals, and three more plans did not provide information about participants' current year contribution allocation, and are thus excluded. The final sample for the main analysis contains records of 572,157 participants in 638 plans. (Some regressions may have slightly different sample sizes because of different information requirement.)

The focus of the first part of the paper are the number of funds in which a participant chooses to invest his balance (*NCHOSEN*) and the number of funds that a participant uses to hold 95% of his balance (*NCHOSEN95*), versus the number of fund options available to employees of the plan (*NCHOICE*). Offered funds in the sample are mostly from, but not limited to, the Vanguard family. Further, most plans offer a small subset of funds from all the funds managed by Vanguard.[3] Participants choose the allocation of their contributions to the available funds when they join a 401(k) plan,[4] and they may modify the initial allocation later. However, such modifications appear to be infrequent (Ameriks and Zeldes (2001), Agnew *et al*. (2003)). The total number of funds chosen by a participant may *overstate* the number of funds to which a participant contributes in the current year. Three funds is both the median and mode of participants' choices. Table I shows that almost all the sample participants – more than 95% of them – use no more than seven funds, although 96% of the participants have access to seven or more funds.

The two key variables used in the second part of the paper are chosen equity allocation, *%EQ*, the proportion of current year contribution that goes to equity funds, and offered equity exposure, *%EqOffered*, the proportion of equity funds among all funds offered by a plan. A balanced fund counts as 0.5 equity fund in both variables. (This choice follows the practice in Benartzi and Thaler (2001). Robustness checks show that results are not sensitive if balance funds are counted as majority-equity or are excluded). The sample includes 50 plans in which employers match employee contributions only in company stock. In such cases contribution to company stock may not reflect employees' desired allocation. For this reason, company stock is excluded from both total contribution and from allocation to equity in the main analysis, but is examined in robustness checks.

The median participant allocates 80% of his current year contribution to equity funds excluding company stocks, while the average is about 69%. 34% of the participants contribute only to equity funds, and 13% do not contribute to equity funds at all. The proportion of equity funds offered tends to increase with the total number of funds offered. For example, the average proportion of equity funds (excluding company stock) out of total funds options is 53% for plans that offer 10 or fewer investment options, is 55% for plans offering between 11 and 20 funds; the same number increases to 64% and 70% for plans that offer between 21 and 30 options and those offering more than 30 funds, respectively.

The full range of equity exposure is 25% to 87.5%, but the equity exposure of more than 90% (95%) of the participants is between 56% (50%) and 77% (78%) – a range in the neighborhood of 21% (28%). Such a range of offers seems to be representative of DC plans; for example, the plans examined by Benartzi and Thaler (2001) had similar exposure (where the range is 37% to 81%). With such a limited range the sensitivity of the fraction of equity chosen to the fraction of equity offered has to be high to have an economically significant effect.

The records contain personal and plan-level attributes, which serve as control variables in the analysis. Personal attributes include annual compensation (*COMP*, in log dollars), the average financial wealth of the nine-digit zip code neighborhood where the participant lives (*WEALTH*, in log dollars),[5] gender (*FEMALE*, dummy variable), age (*AGE*, in years), tenure with the current employer (*TENURE*, in years), and whether the participant registered for web access to his retirement account (*WEB*, a proxy for education and technological savvy). Plan policy variables include the average match rate by employer up to 5% of the employee's salary (*MATCH*, in percentage points, most plans offer at least 50% match); the availability of company stock among the offered funds (*COMPSTK*, a dummy variable; 52% of the participants have this option), the presence of a defined benefit (*DB*, a dummy variable; 62% of the participants are covered by DB plans). Other plan level characteristics include plan size in terms of number of employees (*NEMPLOY*, in log; the median size plan has 282 employees). Using information about *both* participants and non-participants, one can also construct plan averages of individual characteristics.

Participants make three allocation decisions: An active choice of the allocation of the contributions across the funds offered, an active choice of transfer of past balances between the funds, and a passive one to leave the allocation as is. This passive choice, especially for past balances, is common (See Ameriks and Zeldes (2001) and Agnew et al. (2003)). The balances of a participant who does not rebalance his holdings will reflect not only his initial choice but also the cumulative returns of the various funds to which he allocates his contributions.  Therefore it is better to study participants' choices with contributions, rather than balances data.

Unfortunately, Vanguard provides information on participants' fund-by-fund allocation of their *balances*.  For the contributions only allocation by fund category is available, of which there are seven: money market funds, bond funds, balanced funds, active stock funds, indexed stock funds, company stock funds, and other (mainly insurance policies and non-marketable securities, overall representing less than 0.1% of the total balance). Since for participants who joined in 2001 the allocations of the contributions and the balances are close, it is this subsample that is used to study equality of allocation across the funds used.  On the other hand, contributions data at the category level are available for the full sample, and therefore the full sample is used to study the relation between equity exposure and equity allocation.

The data do not offer information about how fund menus changed over time. Arguably, participants who joined earlier than 2001 made their choices based on the funds offered at that time, and due to inertia did not modify them later although more choices may have become available.  Therefore, the sub-sample of individuals who started to contribute to their 401(k) plans *during* 2001 (using information about the entry date of the record) deserves special attention because its members made their choices based on the current menu and are not subject to the status quo biases.  This sub-sample has 37,558 individuals in 548 plans, and most of them were hired during 2001.

Using the new entrants subsample has its own downside because of potential selection biases:  members of this subsample are overall less well compensated and less experienced compared to the typical 401(k) participants, and this subsample's choices may reflect circumstances specific to 2001. Therefore, the main analyses of this paper use the full sample whenever possible (i.e., when only information about contribution at the category level is required), and the same analyses on the subsample of new entrant serve as sensitivity checks.

Throughout the paper the standard errors reported in regression analyses adjust for arbitrary form of heteroskedasticity and correlation of error disturbances clustered at the plan level. Accordingly the effective sample size for an individual attribute variable is of the order of number of individuals in the sample (about half a million), while that for a plan-level variable is of the order of the number of plans (about 640).[6] Unless otherwise noted, the criterion for statistical significance is the 5% (2.5%) level for a two-tailed (one-tailed) test.

## II. The Number of Funds Participants Use

This section looks at the number of funds participants typically use (no more than three or four, regardless of the number of funds offered to them), the extent to which the number of funds used in a plan increases with the number of funds the plan offers (hardly at all), and the tendency of individuals to allocate their contributions evenly among the funds they use (it declines with the number of funds they use).

### *A. Number of Funds Chosen by Individuals*

#### *A.1. Overview*

Figures 1 and 2 summarize the typical number of funds offered to and chosen by participants. Figure 1(a) describes the relevant universe: the number of plans offering a given number of funds and the number of participants in these plans. Relatively few plans offer (and relatively few individuals are exposed to) fewer than 6 or more than 22 funds. Figure 1(b) plots the numbers of participants whose balances are in a given number of funds; in one plot, the given number of funds is for the total balances, whereas in the other the given number of funds is the lowest number which covers 95% of a participant's balance. The latter plot excludes funds in which only negligible fractions of the balances are invested. Three funds is both the median and the mode of participants' choices.

[Insert Figure 1 approximately here]

The following procedure yields Figure 2(a). For each number of funds offered, rank all the participants by the number of funds they are using, from the lowest to the highest. The number of funds corresponding to the bottom 10% forms the lowest graph in the panel; the number of funds corresponding to the bottom 25% forms the next lowest graph, etc. Figure 2(b) is constructed similarly, but the basic number per participant is the minimum number of funds he uses to invest at least 95% of his balance. Figure 2(c) looks at only 2001 new entrants. (Certain ranges of numbers of funds offered are grouped to make sure that there are at least ten participants in that range so that the five percentile values are well defined.)

[Insert Figure 2 approximately here]

Regardless of the number of available choices, the median participant chooses between three and four funds. Even the $90^{th}$ percentile of the number of funds used for 95% of individual retirement money is mostly around six, and does not exceed eight, even when the number of available funds is thirty or more. Figure 2 suggests then that 401(k) plan participants use a stable number of funds, regardless of the choice menu. Therefore the figure offers no evidence that participants diversify naively by applying the strict 1/n rule (i.e., spread funds evenly among all options offered) or even its weaker version (i.e., use more funds when more options are offered).

### A.2. Sensitivity of Funds Used to Funds Offered

With detailed individual- and plan-level attributes on hand, it is interesting to estimate the sensitivity of the number of funds used to the number of funds offered, while controlling for the other attributes; the following specification is used:

$$NChosen_{i,j} = \gamma NChoice_j + \beta Controls_{i,j} + \varepsilon_{i,j}, \qquad (1)$$

where $NChosen_{i,j}$ is number of funds chosen by individual *i* in plan *j*, $Controls_{i,j}$ is a vector of control variables, and $\varepsilon_{i,j}$ is residual disturbance assumed to be uncorrelated with all regressors and $E(\varepsilon_{ij_1}\varepsilon_{kj_2}) = 0$ for $j_1 \neq j_2$. The specification allows $\varepsilon_{i,j}$ and $\varepsilon_{k,j}$ to be correlated with each other due to plan-level random effects. Included in $Controls_{i,j}$ are three sets of control variables: a vector of individual-specific attributes (including *COMP*, *WEALTH*, *FEMALE*, *AGE*, and *TENURE*); a vector of plan policies other than *NChoice* (including *MATCH*, *COMPSTK*, and *DB*), and a vector of plan average of individual attributes (including non-participants). Presumably, participants who contribute larger amounts could spread their money out to more funds. In some specifications, $Controls_{i,j}$ also includes individual annual total contribution, *CONTRIBUTION*, in thousands of dollars. A potential problem with including this variable is that *CONTRIBUTION* is also a choice variable, and could be correlated with the error disturbance in the equation. However, the results are not sensitive to the exclusion of a subset of the control variables listed above, including *CONTRIBUTION*.

Table II summarizes results from estimating (1). Columns 1 to 3 report regression estimates in which all participants' records constitute the underlying sample, and column 4 reports regression estimates using the subsample of 2001 entrants. In the regressions reported in all columns except 3, the dependent variable is the total number of funds chosen by a participant, and that in column 3 is the number of funds used by an individual to cover 95% of his retirement assets. The correlation between the two dependent variables is 93%. 2,735 participants (or slightly short of 0.5% of the sample) used all funds available to them. Even in the subsample of plans that offer 10 funds or fewer, only about 1% of the participants spread contribution over all funds offered. That so few participants use all the funds available to them is another finding inconsistent with the framing-effect 1/n heuristic in its strict sense. Moreover, this is the number produced by the balance records; the number of participants who use all funds for their current year contribution is strictly lower.

[Insert Table II approximately here]

The coefficient of the number of fund choices available (*NChoice*) is small (about 0.01 additional funds used for every fund added to the menu) and indeed indistinguishable from zero, suggesting that controlling for all other variables, the number of funds used is not sensitive to the number of funds offered. It is noteworthy that *NChosen* is not sensitive to *NChoice* without controls or with a subset of the controls used.

Some of the slope coefficients reported in Table II are statistically significantly different from zero, but their magnitudes are mostly economically insignificant. Only one has a noteworthy magnitude: the coefficient of company stock, which suggests that controlling for other attributes, the inclusion of company stock in the offered funds increases the number of funds used by 0.7. Presumably it captures the propensity of participants to invest in company stock when this investment is available. A participant who contributes $1,000 more than his otherwise equal peer tends to invest in about .1 more funds. Although highly significant statistically, the economic magnitude of this effect is modest given that the maximum contribution in 2001 was $10,500. When *CONTRIBUTION* is present, the effect of compensation (*COMP*) on number of funds chosen becomes negative (although it is positive on its own as shown in column 2). This occurs because by keeping *CONTRIBUTION* constant, *COMP* effectively proxies for the inverse savings rate.

The coefficient of *TENURE* is noteworthy, being mostly on the positive side (either slightly positive and significant, or negative and insignificant), meaning that other things equal, participants with longer tenure use *no fewer* funds than their more recently hired colleagues. If more funds are added over the years, if many participants seldom modify their choices (due to status-quo, or inertia bias), and if the number of funds chosen was higher with a higher number of funds offered, that coefficient would be significantly negative, and presumably large in magnitude. It is not, thereby offering indirect evidence that the number of funds chosen is not sensitive to the number of funds offered, or that the particular inertia effect is absent.

Table II shows that the number of funds chosen by a typical person is quite insensitive to the total number of funds offered. However, it does not rule out the possibility that *some* participants' choices could depend on the number of funds available. For instance, it can be that those who choose more funds than a great majority of participants with the same attributes are different from the typical participants with the same attributes, in that their choice is sensitive to the number of funds offered. Re-estimating (1) with quantile regressions (introduced by Bassett and Koenker (1978)) could assess the sensitivity of choices made to choices offered at different conditional quantiles. In particular, using the identifying constraint that sets the $\theta$-th percentile of the error disturbance to zero, i.e., $Quantile^{\theta}(\varepsilon_{i,j}) = 0$, a quantile regression of (1) provides estimates of the sensitivity of $NChosen_{i,j}$ to $NChoice_j$ of individuals who are on the $\theta$-th percentile of number of funds chosen conditional on their personal characteristics and other plan attributes.

At the following quantiles: 10%, 25%, 50%, 75%, and 90%, for every 10 funds added, the estimated sensitivity coefficients (standard errors) are: -0.02 (0.13), -0.08 (0.08), 0.003 (0.08), 0.15 (0.12), 0.36 (0.18). Sensibly, the sensitivities to the number of funds offered increase almost monotonically from low to high percentiles. However, overall the sensitivity is small in magnitude. At the 90th percentile, participants invest in 0.036 more funds with each additional fund, and that represents the sensitivity of the 10% people who choose the largest number of funds in their respective peer groups. Thus, it seems that if some participants increase the number of funds they use with the number of funds offered, there are very few of them indeed.

*B. Plan-Level Number of Funds Used*

Individual participants use a handful of funds even when dozens of funds are available. If, however, participants choose different funds, then the plan as a whole will invest in more than a handful of funds. Both the dollar amount invested and the number of participants investing in a fund - the number of hits – indicate the intensity of usage.

To assess plan-level fund usage, within each plan, the funds are ranked according to both measures of usage intensity. Figure 3 summarizes the relation between the number of funds offered and the number of funds participants use. It shows, per number of funds offered in a plan, (i) the minimum number of funds that hold 75% or 90% of the plan's total assets; (ii) the minimum number of funds that account for 75% or 90% of the total participant hits, where each participant record in a fund is counted as a hit. Figure 3 suggests that at the plan level, the more funds are offered, the more funds are used. But the increase is fairly moderate. For instance, when ten funds are offered, 75% of the money is invested in five funds and seven funds receive 75% of the hits. When three times as many (thirty) funds are offered, the corresponding numbers are eleven and ten, less than doubled.

[Insert Figure 3 approximately here]

Finally, Figure 4 shows the proportion of plan assets and participant hits concentrated in the top 1, 2 or 3 funds. Note that once the number of funds reaches 20, the concentration of money as well as of people in the most popular funds does not decrease as the number of funds increases.

[Insert Figure 4 approximately here]

On the whole, the analysis of plan-level usage of funds suggests that participants' choices are quite similar and that when more funds are offered, more funds go almost unused. Nonetheless, if the costs of offering more funds are miniscule, offering many funds need not be a foolish choice by plan administrators.

### *C. Equal Allocation of Money to the Funds Chosen*

So far the data indicate that most participants use a small number of funds and that the number of funds used is not sensitive to the number of funds offered. These observations are unfavorable to the framing-effect form of naïve diversification. The next question is whether participants tend to allocate their money equally among funds they choose, called here the *conditional* 1/n rule.

To assess the extent to which participants tend to follow the conditional 1/n rule, first consider the 37,798 participants who started to contribute to their 401(k) plans in 2001 and who contributed positive amounts to non-company-stock funds. This sub-sample (rather than the full sample) is suitable for the analysis because for veteran participants, allocation of balances may reflect choices they made years before 2001 which they did not bother to change, transfers across funds, and the relative returns on the various funds in which they invested.[7] The allocation to funds of the balances of the new entrants should reflect their choices of contribution more closely. Company stock is excluded from the analysis because in some plans investments in company stock result from the employer's restrictive match.

Even the balance of a participant who joined in 2001 and who allocates his contributions evenly across the *n* funds he uses need not equal exactly 1/n because the different funds may have had different returns during 2001. A natural indicator to examine adherence to the conditional 1/n rule is the Herfindahl index, defined for each individual *i*, as the sum of the squared fractions of contributions in each fund:

$$H_i = \sum_{j=1}^{n_i} s_{i,j}^2 \;. \tag{2}$$

In (2), $s_{i,j}$ is the share of individual *i*'s contribution in fund *j*, and $n_i$ is the total number of funds chosen by individual *i*; therefore, $\sum_{j=1}^{n_i} s_{i,j} = 1$. The value of $H_i$ is bounded between $1/n_i$ and 1, and is equal to $1/n_i$ for a participant whose balances are exactly equally divided among the $n_i$ funds he uses.

A participant with Herfindahl index close to $1/n_i$ should count as applying the conditional 1/n rule. To this end, classify a participant as a *1/n* investor if his Herfindahl index is bounded (from above) by the index that would result from a portfolio in which the total deviation from a $1/n_i$ allocation is 20% of $1/n_i$ (that is, $\sum_j \left| s_{ij} - \frac{1}{n_i} \right| \le \frac{20\%}{n_i}$). For example, when $n_i = 2$, the upper bound is 0.505, implying portfolio weights of 45%-55%. For each *n*, denote by $\bar{H}(n)$ the upper bound implied by this approximate 1/n rule, that is,

$$\bar{H}(n) = \max \left\{ \sum_{j=1}^{n} s_j^2 : \sum_j \left| s_j - \frac{1}{n} \right| \le \frac{20\%}{n} \right\}. \tag{3}$$

Denote the lower bound of the index by $\underline{H}(n)$; it is equal to $1/n$.

Table III summarizes the extent to which plan participants who joined in 2001 tend to allocate their contributions evenly among the funds they use. $Freq_1$ is the empirical frequency of individuals falling into the first interval $[\underline{H}, \bar{H})$, i.e., investors who resort to the conditional 1/n rule. The numbers in this column show the prevalence of using the conditional 1/n rule, as a function of the number funds chosen, *n*.

[Insert Table III approximately here]

Among the new entrants, using a single fund turns out to be the most prevalent choice (38.6% of them use a single fund), using two or three funds are the next most common choices (17.5% and 15.6% respectively). (These statistics exclude investments in company stock because at least some of the money invested in company stock is the employer's restricted match.) The balances of 64% of those who use two funds are almost evenly distributed between the two funds they use. A weaker, but still strong tendency for the balances to be evenly distributed among the funds used shows for those who use between three and five funds. The case of three funds is somewhat unusual; the propensity to allocate the contributions evenly seems weaker for those who choose three funds (17.9%) than for those who choose four or five funds (37.4% and 26.6%). However, about 10% of those who use three funds make another natural, and arithmetically easy allocation: they put half their contribution in one fund, and divide the rest evenly between the other two funds they use. (The corresponding *H* value is 0.375.)

Assessing the significance of the conditional 1/n rule empirically amounts to assessing the magnitude of $Freq_1$, relative to other possible values of *H* (defined in (2)). To this end, form all possible intervals with length $\left(\bar{H}-\underline{H}\right)$ on the support $[\bar{H},1]$ (i.e., outside the classified conditional 1/n region). Suppose that among these intervals, the highest observed frequency is $\max_{j\neq 1}\left(Freq_j\right)$. If the conditional 1/n rule is the most prevalent allocation rule, the ratio $Freq_1/\max_{j\neq 1}\left(Freq_j\right)$ should be significantly greater than one. The last column of Table III reports this ratio. For individuals who chose 2, 4, 5, and 10 funds, it seems that the conditional 1/n rule dominates any other division of contribution among funds.

If the tendency to follow the 1/n rule is not specific to participants who joined their 401(k) plans in 2001, the balances of veteran participants should display a qualitatively similar pattern, but noisier. (Unless those conditional 1/n participants happened to rebalance their holding shortly before the end of 2001 to reflect their desired allocation, the differential past returns among different funds may blur their intended equal allocation.) The same test, with a looser allowed deviation from the strict 1/n allocation of 25% ( $\sum_j \left| s_{ij} - \frac{1}{n_i} \right| \leq \frac{25\%}{n_i}$ ), performed on the full sample using balances data indicates that the accumulated balances of the 40% of the participants who used two funds were allocated roughly equally between the two funds, and the ratio $Freq_1 / \max_{j \neq 1} \left( Freq_j \right)$ is 5.7. For individuals who used 3, 4, 5, and 10 funds, the same percentages (and frequency ratios) are 8% (1.8), 10% (2.9), 5% (2.6), and 11% (4.0). For other numbers of funds chosen, the 1/n rule does not represent the most popular allocation.

In summary, the data are consistent with a most basic form of diversification, in that the contributions of a substantial number of participants are approximately evenly divided among the funds they use. In particular, for individuals who chose 2, 4, 5 and 10 funds, such a conditional 1/n rule seems to be popular.[8]

The conditional 1/n allocations may be consistent with rational choice. For example, a 50-50 allocation between a stock fund and a bond fund is easily justifiable by preferences with reasonable risk aversion and investment horizon. A more important question is whether the menus of funds presented to participants would bias their choice of funds. The next section examines one such framing effect: whether the mix of asset allocation is affected by mix of the offered funds.

## III. Effects of Equity Exposure on Equity Allocation

### *A. Overview*

To what extent is the chosen allocation to equity ("equity allocation" hereafter) influenced by the intensity of equity in the ensemble of offered funds ("equity exposure" hereafter)? A positive relation between participants' equity allocations and their equity exposures will emerge if the participants spread money evenly across the funds offered to them or ignored the substantive differences among the funds offered to them and picked funds at random. Associating framing effects with the 1/n heuristic can imply that the influence should be positive and strong. More generally, evidence on framing effects (See, e.g., Tversky and Kahneman (1986)) suggests that even the choice of participants who do not apply the 1/n heuristic may well be positively influenced by the intensity of their exposure to equity through the suggestive power of the offered choices. Benartzi and Thaler (2001) also argue that a financially unsophisticated participant might think the menu of funds designed by plan sponsors represents a recommended mix of equity and fixed-income assets, and therefore allocate his contribution in similar proportions.

To preview this section's main finding: the relation between equity allocation by participants and their equity exposure seems positive but small, with marginal statistical significance or robustness. Variation in the offered equity exposure hardly explains the variation in individual equity allocation. There is no relation between the two when the choices are sufficiently diverse.

A rational investor's desired allocation of his 401(k) contribution to equities depends on a set of economic variables such as risk tolerance, demand for tax shelters, investment horizon, etc. As long as this desired allocation is feasible given the funds offered by the plan, a rational person's allocation to equity funds should not depend on the equity exposure of the choice menu. On the other hand, framing effects could lead the participant to invest more in equities when his 401(k) plan offers proportionally more equity funds.

Presumably, cross-plan variations in equity exposure are not related to *individual* participants' preferences, controlling for individual attributes and plan-level effects. Though almost any allocation between equity and non-equities could be justified by some utility function, a finding of a positive relation between equity exposure and equity allocation, controlling for individual and other plan-level attributes, would indicate sub-optimal choice by participants. Given the important context of saving for retirement, designers of the fund offerings of DC plans should keep in mind such a finding, if it is significant.

The two key variables for analysis are equity allocation, $\%EQ_{ij}$, the percentage of current-year contribution of individual *i* in plan *j* that goes to equity funds; and equity exposure, $\%EQOffered_j$, the percentage of equity funds out of all offered funds in plan *j*. Following the practice of Benartzi and Thaler (2001), a balanced fund counts as 0.5 stock fund both in $\%EQ_{ij}$ and $\%EQOffered_j$. In the main analysis, company stocks are excluded from both $\%EQ_{ij}$ and $\%EQOffered_j$ for various reasons. First, company stock is not a universal option for all participants. It is available to about 300,000 participants in 124 plans. Second, in the plans that offer company stock and also match the employees' contributions with company stock – there are 50 such plans – the allocation to that stock reflects the matching formula as much as the employees' choices. Robustness checks show that results are similar if company stock is counted as an equity fund and if the analysis is performed separately on plans that do not offer company-stock-only match (i.e., including all plans that do not offer company stock, and those in which participants can invest in company stock, but the match, if it exists, is in cash.). Finally, plan participants may perceive company stock as belonging to a special category, distinct from other equity funds. (Benartzi and Thaler (2001) elaborate on this idea.)

Figure 5 plots the distribution of plans and participants by equity exposure, or the proportion of equity funds out of all funds offered. Although equity exposure varies from 25% to 87.5% in the sample, 95% (99%) of the plans (participants) face equity exposure between 50% and 80%. The median equity exposure of all plans is exactly 2/3.

[Insert Figure 5 approximately here]

Juxtaposed with the graph of the distribution of offered equity exposure is the graph of median individual equity allocation against equity exposure at different levels. It offers a first look at a possible sensitivity of the chosen allocation of contributions to offered equity exposure. The graph indicates that when equity exposure is extremely low (high), median equity allocation is noticeably lower (higher). Nonetheless, for the equity exposure range that covers more than 95% (99%) of the plans (individuals), the graph does not suggest that equity allocation increases with exposure.

*B. Regression Analysis*

Regression analysis affords an improved assessment of a possible relation between a participant's equity allocation and his equity exposure. It allows control for various other attributes and evaluation of the magnitude of the estimated effect. This magnitude depends both on the estimated coefficient and the range of the equity exposures. Since for most participants and plans that range is fairly narrow – between 50% and 80% -- the estimated coefficient has to be large to be important.

The results are described first for the whole sample and then separately for the subsamples of participants who are offered no more than 10 funds and those who are offered at least 11 funds. For two reasons one can suspect that in the former subsample the sensitivity of chosen equity allocation to offered exposure is higher. One, participants who are offered few funds can be constrained in their choice set in the sense that a desired option is not available, and those whose plans offer relatively more (few) equity funds may therefore invest in more (fewer) equity funds because a desired non-equity (equity) fund is not available.[9] Two, the framing-effect 1/n heuristic may be stronger when the number of funds is smaller. As a robustness check the results are also described separately for the subset of participants who joined the DC plans in 2001.

The following regression specification offers an empirical design for a formal examination of this effect:

$$\%EQ_{i,j} = \gamma \, \%EQOffered_j + \beta Control_{i,j} + \varepsilon_{i,j} \tag{4}$$

where $Control_{i,j}$ is a vector of control variables that include individual and plan attributes. If investors are rational and their desired allocation is feasible within the offered funds menu, then $\gamma$ in (4) should be zero as long as choices are abundant. A significantly positive $\hat{\gamma}$ estimator would suggest an influence of the equity exposure on allocation to equity. Since the dependent variable $\%EQ_{i,j}$ is bounded between [0, 100%], and its distribution is neither unimodal nor normal, the censored least absolute deviation regression (CLAD) introduced by Powell (1984) and Khan and Powell (2001) is an appropriate technique to estimate (4). The CLAD method accommodates the corner allocations of equity without making assumptions about the *desired* allocation of those who choose zero or all equity, and is robust to non-normality in the error disturbance.[10] Further, allowing the error disturbances, $\varepsilon_{i,j}$, in (4) to be correlated if they belong to the same plan, the standard errors of all reported estimates adjust for arbitrary error correlation clustered at the plan level as well as heteroskedasticity.

Table IV reports the sensitivities of equity allocation on exposure using several specifications. Panels A and B use the full sample and Panel C use the subsample of 2001 new entrants. In the whole sample, the median and mean values of contribution allocation to equity funds are 80% and 69% respectively. Only participants who contribute positive amount to non-company stock assets in 2001 constitute the sample (so that $\%EQ_{i,j}$ is well defined). Panel A reports the estimators of regression equation (5) on the full sample. Column (1) reports the coefficient $\hat{\gamma}$ without any control variables. The estimate is 0.18 (*t*-statistics = 0.67). It appears that $\%EQ_{i,j}$ does not significantly respond to $\%EQOffered_j$ at the individual level. The raw correlation of the two variables is 0.01, and the pseudo-$R^2$ (the proportion of variation, as defined by absolute deviation from the median, in $\%EQ_{i,j}$ that is explained by $\%EQOffered_j$) is 0.02%. (In comparison, *COMP*, on its own, explains 2.2% of the variation in the dependent variable; *WEALTH* explains 1.3%, and *MATCH* explains 0.8% of the variation in $\%EQ_{i,j}$.) This is not surprising given the pattern shown in Figures 5.

[Insert Table IV approximately here]

Column (2) of Table IV reports the estimation with a complete set of control variables, including: (i) individual attributes: 401(k) savings rate, compensation, wealth, gender, age, tenure, registration for web access; and (ii) plan policies: match rate, availability of company stock, restricted match in company stock, presence of a DB plan, and number of funds offered. Plan size and plan-average of individual attributes are present as additional controls. The coefficient on $\%EQOffered_j$ remains at 0.18 but is now statistically significant at 5% level (*t*-statistic = 2.04). With the equity exposure of 99% of the participants (and 95% of the plans) ranging from 50% to 80%, the estimate implies that the equity intensity of funds offered could at most, other things equal, lead to a 5.4% distortion in participants' equity allocation (out of a median allocation of 80%).

When there are few options, investors are likely to be constrained, and their allocation could vary with the offering even without any framing effect. For example, suppose investors would like to diversify into a large-cap stock fund, a small cap stock fund, and an international stock fund. Plan A offers only the first two, and plan B offers all three.

Investors in plan B could invest more in equity than those in plan A, not because of naïve diversification, but because in plan A investors are constrained.

To examine separately the behavior of potentially constrained participants and those whose choices are less likely to be constrained by the choice set offered by their funds, the sample of plans (and their participants) is divided into those that offer up to ten funds, and those that offer more than ten funds.[11] Columns (3) and (4) in Table IV report the analyses of these two subsamples. The first sub-sample covers 47% of plans and 28% of individuals. Interestingly, the positive sensitivity of allocation to exposure only shows up in the few-funds sub-sample (coefficient = 0.29, and *t*-statistic = 2.73), and disappears in plans that offer more than ten fund choices (coefficient = 0.06, and *t*-statistic = 0.64). In the few-funds sub-sample, more than 95% of the participants (as well as plans) face equity exposure between 50% and 75%. A sensitivity coefficient of .29 multiplied by a range of 25% results in the effect of equity exposure on equity allocation being 7.3% (relative to the median allocation of 83.3%).

Although the general impression is that equity allocation is not meaningfully sensitive to equity exposure, some nuances can be gleaned by correlating this sensitivity with individual characteristics, which is done by estimating a modification of regression (6) in which individual characteristics are interacted with *%EQOffered*. It turns out that high income or wealth is associated with lower sensitivity, but the interaction effects are not significantly different from zero at the 10% level. Age has virtually no differential effect. Gender seems to make a difference: Everything else equal and for every 1% increase in equity exposure, a man on average increases his equity allocation by 0.22% (t-statistic = 2.72), whereas a woman only increases her equity allocation by 0.12% (t-statistic = 1.36), and the gender difference is significantly different from zero at less than 10% level.

Interacting *%EQOffered* with tenure has additional implications. The data underlying this study are from 2001, and many of the participants joined the plans well before 2001. It is possible that at the time they last modified their allocations – possibly when they joined the plans – the sets of funds they were offered were smaller than in 2001. Two sensitivity checks explore the possible relation between inertia in participants' choices and the influence of the offered equity exposure. First, the regression is re-estimated with the product of *TENURE* and *%EQOffered* as an additional exploratory variable, reported in Panel B of Table IV. Presumably, if inertia was important and mitigated a potential framing effect, the coefficient of this variable should be negative. Overall the coefficients of the interaction term are not significantly negative.

Second, Panel C of Table IV shows the results from the subsample of 2001 new entrants. The sensitivity of equity allocation to exposure of this subsample is almost identical to that of the full sample, though not statistically significant due to reduced sample size and larger dispersion of individual error disturbances. Further dividing the sample according to number of funds offered does not offer qualitatively different results.

Focusing on the 2001 entrants, Section II.C above documents that a substantial fraction of participants divide their contributions evenly among the few funds they choose. These participants are classified as following the conditional 1/n heuristic. The separate sensitivities of equity allocation to equity exposure of those who follow the conditional 1/n heuristic and those who do not are available from regression (4) with an additional interactive term of a dummy variable (for conditional 1/n investors) with *%EQOffered*. These sensitivities are similar (the difference being smaller than 1.5% and the *t*-statistic about 0.24), indicating that following the conditional 1/n is unrelated to a relation between equity allocation and equity exposure, which is weak in the data to begin with.

Several robustness checks complement the tabulated results. First, consider the subsample of participants who chose equity exposure strictly between zero and 100%. The sensitivities of their equity allocation to their plans' offered equity exposures are essentially zero and statistically indistinguishable from zero. Second, the same analyses as those presented in Table IV were performed while including company stock as an equity fund and excluding plans that provide restrictive employer match in company stock. The results are similar to those in Table IV with lower statistical significance.

Finally, measuring the equity proportion of individual balanced funds should match the perception of their equity exposure by the participants. There are several possibilities. First, most participants view different balanced funds as close substitutes in terms of equity exposure when they make allocation decisions. If this is the case, balanced funds should be counted as some uniform mixture of equity and bonds, such as 50-50. This is the specification adopted in this paper as well as in prior studies such as Benartzi and Thaler (2001). Robustness checks indicate insensitivity of the results to alternative specifications that count all balanced funds as majority equity or bond funds. A second possibility is that participants view balanced funds as a different category from either equity or bond funds and do not relate balanced funds to equity investment. Excluding balanced funds (from both equity allocation and exposure) yields a sensitivity coefficient of 0.12 (t-statistic = 1.38) for the full sample. For the subsample where participants have 10 options or fewer, the coefficient is 0.23 (t-statistic = 3.63), and for the complementary subsample, the coefficient is -0.06 (t-statistic = -0.81).

Another possibility is to measure balanced funds to their exact equity-bond composition and re-estimate the regression with *%EQ* and/or *%EQOffered* adjusted accordingly. Appendix A offers more details on the data and qualifies the interpretation of the results. Briefly, the mean equity exposure of these balanced funds is close to half (52%), and three out of the 17 balanced funds explicitly market themselves as over-weighting equities or over-weigthing bonds. If both *%EQ* and *%EQOffered* adjust for the exact equity exposure of balanced funds at the end of the previous year—based on the assumption that some participants take into account the exact equity exposure of balanced funds in their allocation, and expect balanced funds to keep stable equity/bond composition over time, the resulting equity allocation to exposure sensitivity coefficient is 0.24 (t-statistic = 2.61) for the full sample; and 0.38 (t-statistic = 3.79) and 0.16 (t-statistic = 1.70) for the subsamples of plans that offer limited and extensive options. The numbers are higher both in magnitude and statistical significance than those in the baseline regression, but open to various interpretations, including one ascribing the larger coefficients to a mechanical relation rooted in the procedure itself. Indeed, a simulation applying the same procedure to a calibrated artificial data set yields similarly significant results. Therefore, the strengthened results could be entirely due to the mechanical relation. Appendix A offers more details.

Assuming that asset allocation of 401(k) plan participants is affected by awareness of the exact asset allocation of the balanced funds available to them seems extreme. Nonetheless, one can still ask a more pragmatic question: How sensitive is the actual allocation to equity to the relative number of equity funds offered, if a balanced fund counts as a half equity fund among the offered funds, and in "actual equity allocation" one counts each balanced fund as having its actual equity allocation. In that case, the sensitivity coefficient (t-statistic) of the full sample is 0.13 (1.63). Those of the few-choice subsample are 0.26 (2.74); and those of the many-choice subsample are 0.03 (0.29). These results are consistent with, and weaker than, those in the baseline regression.

Since almost half the sample members choose extreme equity allocations (zero or 100%), it is interesting to assess the relation between equity exposure and the propensity to choose a corner allocation. A probit analysis shows that the propensity to avoid equity altogether is insensitive to the offered equity exposure. On the other hand, a 1% increase in equity exposure is associated with a 0.23% increase in the probability of participants' allocating all their contributions to equity funds, which is significantly different from zero at less than 5% level. (An equity exposure range of 50% to 80% is associated with a probability increase in the range of about 7%, out of an all-sample probability of 34%.) Choosing 100% equity funds when being offered also some non-equity funds (which all plans have) is inconsistent with the 1/n heuristic and it is possible that the correlation is endogenous, that is, plans with participants with stronger preference for equity investments actually offer relatively more equity funds, an issue to be taken up in Section IV.A.

The standard interpretation of a regression such as (4) assumes that the residuals are independent of the explanatory variables, and in particular that the composition of the offered funds is random or exogenous to the preferences of the participants. But this need not be the case: plans in which participants prefer high equity allocation may well accommodate them by offering a high number of equity funds. To examine this possibility, consider a plan-level regression of $\%EQOffered$ on plan-average attributes (average income, wealth, gender, age, tenure, web registration, and plan size) and their standard deviations. It turns out that the dependent variables jointly explains slightly less than 10% of the variations in $\%EQOffered$, indicating that equity exposure is not completely random and plan sponsors may indeed accommodate their participants' preferences; such accommodation would bias the estimated sensitivity of equity allocation to equity exposure upwards, thereby working against the finding of little relation between chosen equity allocation and offered equity exposure. Section IV.A further assesses the relevance of endogeneity in this context.

### *C. Nonparametric Analysis*

The regression analysis of the preceding subsection may leave some readers looking for a more straightforward and robust analysis which would rely minimally on the model specification and deliver a less nuanced picture. Such an analysis can be summarized by the following question: If two participants are drawn at random from the sample, and the offered equity exposure of the first is higher than that of the second participant, is the first participant also more likely to have chosen a higher equity allocation?

A brief digression to the statistical problem helps. Formally, let $X_j$ and $Y_{ij}$ be the equity exposure and equity allocation for individual *i* in plan *j*, and let $Z_{ij}$ be control variables such as compensation. Suppose equity exposure does not affect equity allocation, then

$$H_0 : F_{XY}(x, y \mid z) = F_X(x \mid z) F_Y(y \mid z), \forall (x, y), \tag{7}$$

where *F* stands for distribution functions. Note that the null hypothesis as stated in (7) is nonparametric (i.e., it imposes neither a functional form on the distributions of the variables nor a functional relation between the two variables).

The alternative is that there is a (positive) dependence of $Y_{i,j}$ on $X_j$. Kendall (1962) considers testing a simple version of (7) nonparametrically and suggests focusing on the statistic $\left[2\Pr\left(Y_2 > Y_1 \mid X_2 > X_1\right) - 1\right]$, its null value being zero. A straightforward transformation of Kendall's proposed statistic, which also lends itself to immediate interpretation is

$$\tau = \Pr\left(Y_2 > Y_1 \mid X_2 > X_1, Z\right). \tag{8}$$

Under the null hypothesis, $H_0 : \tau = \frac{1}{2}$. A positive dependence of chosen equity allocation on offered exposure amounts to

$$H_1 : \tau > \frac{1}{2} \tag{9}$$

The test statistic is the empirical analog to (8):

$$\hat{\tau} = \frac{1}{\tilde{N}} \sum_{x_i > x_j} I(y_i > y_j \mid x_i > x_j, \mid z_i - z_j \mid < w), \tag{10}$$

where *I* is an indicator function equal to one if the argument is true and zero otherwise, and $\tilde{N}$ is the total number of observations pairs that have different *x* values and for which the control variable *z* (possibly a vector) falls in the same neighborhood of window width *w*.

The statistic $\hat{\tau}$ defined in (10) is a pair-wise U-statistic that is asymptotically normally distributed regardless of the underlying distributions of X and Y. Further, it is the least variance statistics among all unbiased estimates of $\tau$ defined in (8). For its asymptotic property and construction of standard error estimates, see, e.g., Serfling (1980), and Abrevaya and Jiang (2004). In the language of this paper, the null hypothesis is that if $\%EQOffered_{j1} > \%EQOffered_{j2}$, there is no more than a 50-50 chance that $\%EQ_{i1,j1} > \%EQ_{i2,j2}$. Under the alternative hypothesis of the framing-effect heuristic, the same probability is greater than 50%. The calculations of the statistic and its standard errors are described in the Appendix B.

When comparing equity allocations of two participants, one must reflect on the possibility of a tie, especially when both individuals choose to invest either all or none of their contributions in equity. A pair of observations with $\%EQOffered_i > \%EQOffered_j$ but $\%EQ_i = \%EQ_j$ and the equity allocations are interior is indicative of a lack of the framing effect, and is thus treated as favoring the null. However, pairs where $\%EQ_i = \%EQ_j = 100\%$ or $\%EQ_i = \%EQ_j = 0\%$ are discarded from the computation, because it is impossible to compare the two individuals' relative intensity of taste for equity (as the corner solutions do not reveal their desired allocation).

Table V summarizes the results. Panels A and B report the statistics and their standard errors at the individual level. The statistic $\hat{\tau}_1$ compares any two individuals with different equity exposure without control variables, and the statistic $\hat{\tau}_2$ only compares pair observations drawn from similar income (the absolute difference between two participants' income being less than \$20,000), similar age (the difference being less than 5 years), and same gender. These three control variables are chosen because they have the highest explanatory power toward equity allocation among all variables available in the data set.

(Including all control variables is computationally prohibitive and adds little more insight.) Panel C reports similar statistics for the same test performed at the plan level. For instance, the 47.99% in the second row of Panel B answers the following question. Consider two randomly drawn participants who were new entrants in 2001, earned similar incomes, of similar age, of the same gender, and whose plans offer at most 10 funds. Suppose that the plan of the first participant offers higher equity exposure than that of the second participant. What is the probability that the first participant allocates a higher fraction of his contributions to equity?

[Insert Table V approximately here]

The interpretation of the three rows in Panel C of Table V is different. For instance, the 60.20% in the next to last row is the answer to the following question: Consider two plans which offer at most 10 funds, with one plan having proportionally more equity funds out of the total number of funds than the other plan; what is the probability that the fraction of the total contribution allocated to equity in the first plan is bigger than that in the second plan?[12]

The results in Table V have a straightforward interpretation in the language of decision making. Consider two randomly drawn participants to whom their plans offer different equity exposures. A bet that the participant with the higher offered equity exposure has also chosen a higher equity allocation than the second participant is no more attractive than a bet that a coin flip will come up tails. On the other hand, the first bet at the plan level would be highly lucrative. The difference between the results at the individual and plan levels is noteworthy, and demonstrates that inference from plan-level data need not carry over to individual-level decisions. It is discussed further in the next section.

## IV. Discussion of Results

### *A. Differences between Analyses at the Plan and the Individual Level*

Benartzi and Thaler (2001) ("BT" hereafter) suggest that the general human tendency to diversify may carry over to the fund selection problem of participants in 401(k) plans. They recognize that not all participants would divide their contributions evenly across all the funds offered to them, and suspect that different behavior might be observed in plans that offer the full range of funds from a large mutual fund company. They do not report any direct experimental examination of the consequences of variation in the number of funds offered, the range of choices in their field examination is limited,[13] and much of their work is devoted to the relation between equity allocation and equity exposure. The analysis in BT uses primarily experiments and plan-level regressions.

Experiments can be an excellent tool to establish motives or directions of behavior of decision makers. They are effective because in the sterile experimental setup, most parameters are held constant, and the observer can learn how variation in a single condition causes variation in behavior. But this very source of effectiveness makes it difficult to assess whether the effect generalizes; if it generalizes, how big it is; and the extent to which it is important in the field, where the presence of the other parameters may overwhelm the effect that is studied in the laboratory.

BT may have recognized the limitations of inference drawn from experiments: "These experiments *suggest* that the array of funds offered to plan participants *can* have a surprisingly strong influence on the assets they end up owning." (Italics added.) In fact, this potential limitation (and data availability) probably led them to study actual plan-level relations between the fraction of equity offered in the plans' menu and that chosen by participants.

However, results at the aggregate level in general do not carry over to the individual level. An example used by Freedman (2001) was literacy vs. being native born. In the US in 1930, the proportion of population that were foreign born in a state is highly positively correlated with the state's literacy level, while such a positive relationship is non-existent if one used individual level data, a phenomenon that Freedman terms "the ecological bias." Examples include Goodfriend (1992) who shows that information-aggregation bias invalidates some tests of permanent income hypothesis; and Hanushek et al. (1996) who show that the explanatory power of school quality on student achievement increase dramatically with the level of aggregation. (It is almost non-existent at the student level, strong at the school level and strongest at the state level.)

Pure aggregation as well as endogeneity can bias the coefficients in an aggregate-level analysis. To understand aggregation bias, suppose the underlying relation is $y_{ij} = f\left(x_{ij}, z_j\right) + \varepsilon_{ij}$. Then an equivalent relation at the aggregate level $f\left(E_j\left(x_{ij}\right), z_j\right) = E_j\left[f\left(x_{ij}, z_j\right)\right]$ only holds under very restrictive conditions. Hanushek et al. (1996) identifies the conditions under which a relation at the individual level could be exaggerated by aggregation. Further, aggregation in the specific context of equity allocation fails to accommodate the corner allocations by a large number of individuals (those who choose zero or 100% equity).

The following simulation exercise illustrates the nature of the pure aggregation bias. Consider a simulation of equity exposure (*X*) and equity allocation (*Y*) for individuals in 600 plans. First randomly generate two independent *X* and *Y* series from normal distributions with mean 66% and standard deviation of 8%. *X* is censored at 25% and 90% and *Y* is censored at 0% and 100%.[14] Then select randomly 1% of *X* to be equal to *Y* (i.e., 1% of the participant follow the 1/n rule). If the average plan size is 100 people, the average $\hat{\tau}$ statistic (out of 1,000 simulations) is 52.7%, indicating the probability (in excess of 50%) that higher equity exposure leads to higher equity allocation is 2.7%. If the average plan size increases to 500, 1000 (about the average plan size of sample studied in this paper), and 2000, the statistics further go up to 56.0%, 58.8%, and 66.4%, respectively.

At the aggregated level and as the plan size increases, equity allocations of participants which are uncorrelated with equity exposure offset each other and diminish in importance; the subsample of participants whose allocation approximates the 1/n rule, no matter how few they are (or how little money they control), dominates the relation between equity allocation and exposure. In the limit as the group size goes to infinite, the probability approaches one that plans offering higher equity also have higher equity allocation. By aggregation, in general, a very weak relation at the individual level appears considerably amplified.

It is noteworthy that in the data used in the present study, at the *plan* level, $\%EQ_j$ (the percentage of plan total current year contribution allocated to equity funds) and $\%EQOffered_j$ are positively correlated (the correlation coefficient is 0.23). Further, the coefficient on $\%EQOffered_j$ in the plan-level regression is significantly positive at 0.27 (White *t*-statistics = 4.51) and $\%EQOffered_j$, on its own, explains 5% of the variation in $\%EQ_j$. It is possible that the strong effect of mix of funds in the plans and asset allocation of participants reported by BT is a by-product of the analysis being at the plan, not the individual level. The contrast between the plan-level and individual level results in Table V counsels against using plan-level observations to make inferences about individual behavior.

Endogeneity bias can also affect the interpretation of plan-level analysis. It is possible that plans with participants with stronger tastes for equity funds will also offer more equity funds. Such a catering to participants' tastes suggests that the possible positivity of the slope coefficient $\hat{\gamma}$ in (4) need not be interpreted as reflecting employees with similar tastes making different choices under the influence of their plans' different fund ensembles. Aggregation amplifies such endogeneity bias.

To see this, revisit (4): $\%EQ_{i,j} = \gamma\,\%EQOffered_j + \beta Control_{i,j} + \varepsilon_{i,j}$, and think of a missing variable—embedded in $\varepsilon_{i,j}$—which captures the taste for equity. Suppose that the true value of $\gamma$ is zero. $\varepsilon_{i,j}$ can be decomposed as the sum $\varepsilon_{i,j} = \bar{\varepsilon}_j + \left(\varepsilon_{i,j} - \bar{\varepsilon}_j\right)$, where $\bar{\varepsilon}_j$ represents *between-plan* variation of error disturbance with variance $\sigma_b^2$ and $\left(\varepsilon_{i,j} - \bar{\varepsilon}_j\right)$ represents *within-plan* variation with variance $\sigma_w^2$. Form a projection of equity exposure on $\bar{\varepsilon}_j$: $\%EQOffered_j = \lambda\bar{\varepsilon}_j + \upsilon_j$. Then $\lambda > 0$ if plans cater to their participants' aggregate tastes. Accordingly, the spurious explanatory power of equity exposure on equity allocation is $\frac{\lambda^2}{\left(\lambda^2 + \sigma_\upsilon^2/\lambda^2\sigma_b^2\right)\left(1 + \sigma_w^2/\sigma_b^2\right)}$ for individual regressions and $\frac{\lambda^2}{\left(\lambda^2 + \sigma_\upsilon^2/\lambda^2\sigma_b^2\right)}$ for plan-level regressions. The latter is larger than the former, and the difference is bigger when the ratio $\sigma_w^2/\sigma_b^2$ is higher.

The intuition is as follows: consider the choice of equity exposure by plan sponsors. That choice varies across plans according to between-plans variation in taste if plan sponsors indeed try to accommodate their participants' tastes. When the within-plan variation in taste for equity dominates the between-plan variation, the endogeneity bias in individual-level regressions is small compared with that bias in plan-level regressions.

Ignoring the catering-for-taste possibility, the slope coefficients reported in Table IV can be interpreted as reflecting average sensitivities of individuals' choices of equity allocation to the exogenous equity exposures to which they are assigned by their plans. Accounting for catering-for-taste, these slope coefficients are upper bounds on these average sensitivities. The following procedure produces lower bounds for these average sensitivities.

Divide the sample into 2001 new entrants (superscripted by *n*) and old participants (superscripted by *o*). Given that $\%EQOffered_j$ in the sample was set by the beginning of 2001, it is more likely to cater to the aggregate taste of old participants revealed in the past.

(To the extent that plan sponsors anticipate the tastes of their new employees, finding no effect is less likely.) Construct $\overline{\%EQ}^{o}_{j,<t}$, the proportion of old participants in plan *j* who invested 100% in equity funds before 2001 (by counting the old participants whose balances excluding current-year contribution are all in equity), as a proxy for the part of the plan fixed effect that is correlated with the aggregate taste for equity. Estimate regression (4) on the subsample of 2001 new entrants using $\overline{\%EQ}^{o}_{j,<t}$ as an extra control variable, then the sensitivity coefficients on $\%EQOffered_j$ becomes insignificant (0.11, t-statistic = 0.49).[15]

The interpretation is the following: suppose two otherwise identical employees join two companies where existing employees had shown similar tastes for equity. Would higher equity exposure lead to high equity allocation? Such an interpretation relies on the following assumption: old participants who went for all equity in the past were not influenced by their current equity exposure (since all plans offer equity exposure that far away from 100% equity). If otherwise, the analysis above would under-estimate the effect of equity exposure on new entrants' equity allocation. However, the estimates reported in Table IV remain an upper bound for such sensitivity if there is any catering in plan equity exposure offerings.

*B. Discussion*

The point of departure of this paper from previous work is the basic 1/n intuition that participants in 401(k) plans tend to allocate their contributions evenly across the funds they use. This intuition is confirmed: Substantial fractions of those who use between two and five funds and those who use ten funds allocate their contributions approximately evenly across the funds they use.

This study goes further to explore framing effects, i.e., whether the number of funds offered to participants affects the number of funds they use, and whether participants in plans which offer more equity funds (relative to all the funds they offer) show a stronger tendency to invest in equity funds.

Motivated by a strong intuition, Benartzi and Thaler (2001) set forth the 1/n hypothesis which entails two predictions: first, some 401(k) participants tend to allocate their money evenly among the funds offered; and second, their allocations to equities are highly correlated with their exposures to equity funds. Quite a few studies that follow on Benartzi and Thaler (2001) have reemphasized the claim that 401(k) plan participants follow the framing-effect version of the 1/n heuristic. (See, e.g., Camerer *et al* (2003), Chordia, Roll, and Subrahmanyam (2002), Cogan and Mitchell (2003), Daniel, Hirshleifer and Teoh (2002), Gabaix and Laibson (2003), Kahneman (2003), and Langer and Fox (2003).)

This study finds the specific framing effects to be on the tenuous side. In fact, this study concludes that when it comes to the number of funds used and the sensitivity of the fraction of equity used to that offered, one usually does not reject the hypothesis that plan participants act rationally; they are unaffected by the number of funds offered or the weight of equity among the offered funds. In some specifications the estimator of the sensitivity of equity used to equity offered is significantly positive, but small in magnitude. Only a minute fraction, if any, of the individual variation in chosen equity allocation is explained by variation in offered equity exposure in the large data set underlying this work.

An investor should not be concerned with the number of assets in his portfolio but with the portfolio's risk-return profile. Indeed, classic results on K-fund separation establish conditions under which all investors will select portfolios of the same K funds. The investors vary the portfolio weights of these K funds to accommodate their attitudes toward risk. The emphasis of these results is on the economy of portfolio decision rules. The Capital Asset Pricing Model (CAPM, developed by Lintner (1965), Mossin (1965), and Sharpe (1964)) delivers the most prominent fund separation result, providing conditions under which all investors choose portfolios of just two funds: the market portfolio and the safe asset. (For other results on K- fund separation, see Cass and Stiglitz (1970), Black (1972), and Ross (1978)).

K-fund separation theories suggest that when an investor chooses among numerous funds of primitive assets, he is likely to allocate the money to a small number of such funds. He will allocate money to many such funds if he attempts to diversify across the various funds, regardless of the merits of such diversification. The data are consistent with the former behavior.

The results in this paper should not be interpreted as an argument that portfolio choice, within or without 401(k) plans, is always and everywhere rational. A few examples of 401(k) portfolio choices that appear to be irrational are offered by Holden and VanDerhei (2001) and Liang and Weisbenner (2003) who report that employees invest more of their own money in company stocks when the employer match is already in company stock; by Huberman (2001) who argues that familiarity breeds investment, and, in particular, investment of substantial fractions of 401(k) savings in the employer stock is a by-product of this tendency; and by Choi, et al. (2004a) who report that automatic enrollment lead to high percentage of participants' asset allocation in the default fund (often a money market fund). (Choi, et al. (2004b) review the effect of plan design on asset allocation of 401(k) plan participants.)

This paper’s failure to detect irrationality of asset allocation may be attributed to the way the data are presented to the decision makers. In many 401(k) plans participants face a hierarchical presentation: First they see fund categories, and only then, within each category, they see the individual funds. In this case they may apply the 1/n heuristic to categories, but such a procedure is unlikely to show up at the fund level. There is very little variation in the categories offered across the plans studied here. Therefore it is inappropriate to study the 1/n rule with respect to categories with the present data. In fact, it may well be that most 401(k) plans offer funds from the same categories. It may well be, then, that this study illustrates a point made by Glaeser (2003) who argues that since market outcomes are determined by demand and supply forces, the latter may drive the outcomes away from those observed in experiments that vary conditions on the demand side alone.

Finally, a point about policy implications. BT points out that if 401(k) plan participants' choices were strongly influenced by the menu of choices offered to them, the design of the menu would be very important. But within the current varieties of menu design hardly any such influence is detected. It is likely, then, that the menus offered by the plans studied here are equally good, and a choice among them by a plan designer is not important.

## V. Conclusion

Kahneman's Nobel lecture (2003) mentions the study of Benartzi and Thaler (2001) as a member of "[a] growing literature of field research and field experiments [which] documents *large and systematic mistakes* in some of the most consequential financial decisions that people make, including choices of investments…" (Italics added.) Using a large archival dataset of 401(k) plan participants, this study fails to find such large and systematic mistakes resulting from the influence of fund offering menus on investment choices.

One way to assess the overall findings regarding a possible relation between the fraction of equity funds a plan offers and the fraction of equity funds its members use is to compare the following two bets. One bet considers two randomly drawn participants of similar incomes such that the plan of the first participant offers more equity funds (relative to the total number of funds it offers) than the plan of the second participant. The bet is that the first participant will allocate a higher fraction of his 401(k) contribution to equity funds. The second bet is that a coin flip will come up heads. The data indicate that the first bet is no more attractive than the second.

This study can be interpreted as a test of rational choice of funds against two overlapping alternatives: that investors increase the number of funds they use as the number of funds in their plans increases, and that investors increase their allocation to equity funds as the relative weight of equity in the offered menu increases. The rational choice hypothesis is that participants with similar attributes should not be making systematically different choices in the directions implied by the alternative hypotheses. This study fails to reject the rational choice hypothesis in favor of those alternatives.

A failure to reject may be a statement of the low power of the test, or a statement that the data are not suitable for the task, or a statement of the weakness of the alternative hypotheses. Better data may produce other results. At the moment this is the only study that uses records of individuals in a large number of plans which offer different numbers of funds. Therefore, an appropriate current conclusion is that investors do not deviate from rational choice in the directions of the alternatives entertained in this study.

## Appendix

*A. Analysis with the Exact Asset Allocation of the Balanced Funds*

Seventeen of the eighteen balanced funds in the sample were matched with the Morningstar data base and the end of 2000 mix of equity and debt of these funds was hand collected.

Thirteen of the seventeen balanced funds held more equity than debt. The one balanced fund which could not be matched with the Morningstar records was treated as if it invested half its assets in equity. Two estimations are done with this additional information: (i) Re-estimating regression (4) with both plan equity exposures and individual equity allocations adjusted for the exact equity component of the balanced funds; (ii) Re-estimating regression (4) where individual equity allocations, but not plan equity exposure, adjusted for the exact equity component of the balanced funds. Specification (i) delivers an equity allocation-to-exposure sensitivity coefficient of 0.24 (t-statistics = 2.61). Both the estimated sensitivity and its statistical significance are higher than those estimated in the paper's baseline regression. Specification (ii) yields weaker-than-baseline results: the sensitivity coefficient (t-statistic) of the full sample is 0.13 (1.63). Those of the few-choice subsample are 0.26 (2.74); and those of the extensive-choice subsample are 0.03 (0.29).

Two competing explanations for the stronger results of specification (i) come to mind: (1) The exact equity-bond compositions of the balanced funds offered influences the allocations of a significant number of plan participants, who must necessarily be aware of these exact compositions. Further, investors who are able to discern the portfolio compositions of balanced funds are *more* subject to the framing effect (i.e., invest proportionally more in equity when plan-offered equity exposure is higher). Therefore, when the key independent variable is measured more accurately, the result is more significant; (2) Most participants do not know or ignore the exact equity component in the balanced funds going forward and view different balanced funds as close substitutes in terms of equity exposure. Then the sensitivity coefficient is more significant because of a mechanical correlation.

The following simple example illustrates the mechanical correlation. Suppose Mr. A and Mr. B are in Plan 1 and Plan 2 respectively. Plan 1 offers an equity fund and a balanced fund that is believed by Mr. A to usually invest 50-50 in equities and bonds but happened to have 51% in equity toward the end of 2000. Plan 2 offers an equity fund and a balanced fund that is believed by Mr. B to usually invest 50-50 in equities and bonds but happened to have 49% in equity toward the end of 2000. (The accuracy of the participants' beliefs is immaterial to the example.) Intending to invest 75% in equities and 25% in bonds, both Mr. A and Mr. B invest 50% in the equity fund and 50% in the (different) balanced funds offered to them. Thus, Mr. B's equity exposure is 1% higher than that of Mr. A, and so is his equity allocation, although both participants think they behave alike.

The following calibrated simulation helps assess the magnitude of such a mechanical relation:

Step 1: Replace the equity exposure of the balanced funds in the sample with randomly simulated numbers from a normal distribution with mean 52% and standard deviation 14% (both moments calibrate to the sample moments).

Step 2: Match these artificial balanced funds to the sample and adjust equity allocation/ exposure accordingly.

Step 3: Estimate a regression of the resulting equity allocation on equity exposure with control variables (equation 4 in the draft) and get the sensitivity coefficients and their standard errors.

Step 4: Repeat Steps 1 to 3 30 times.

The average coefficient from the 30 simulations[16] is 0.26. All 30 estimates are larger than the coefficient using 0.5 equity exposure uniformly across all balanced funds (0.18 as shown in

Table IV) and are significantly different from zero at the 5% level. Further, they are very close in magnitude to the coefficient (and standard error) estimated counting balanced funds to their actual equity component.

The stylized example and the simulation illustrate a mechanical correlation which is likely to exist in the data: most plan participants are unlikely to be aware of the exact asset allocation of the balanced funds offered to them; even if they wanted to know, it would be difficult: only three (out of 17) of the balanced funds in the data explicitly stated in the prospectus that they were over-weighting equities (one growth-oriented balanced fund) or over-weighting bonds (two income-oriented balanced funds), and have equity exposure consistently above 60% or below 40% over time. The equity exposure of other balanced funds varied around 45-60% over time, and these funds did not market themselves as either equity-heavy or bond-heavy balanced funds.

Therefore, estimating the regression with the exact equity exposures of balanced funds at a point could bias the sensitivity coefficient upwards if these funds are *perceived* by investors as close substitutes in terms of equity exposure. The weakened results of specification (ii) and the results of the simulation are consistent with this possibility.

*B. Construction of the Nonparametric Statistic (Section III C)*

The test statistic is:

$$\hat{\tau} = \frac{1}{\tilde{N}} \sum_{x_i > x_j} I(y_i > y_j \mid x_i > x_j, |z_i - z_j| < w), \tag{11}$$

where $I$ is an indicator function, $y$ is equity allocation, $x$ is equity exposure, $z$ is a vector of other control variables, and $\tilde{N}$ is the total number of observations pairs that have different $x$ values and for which the control variable $z$ falls in the same neighborhood of window width $w$. Two versions of $\hat{\tau}$ are computed: $\hat{\tau}_1$ compares all pairs *(i, j)* that satisfy $x_i > x_j$ (C1);

$\hat{\tau}_2$ compares all pairs *(i, j)* such that $x_i > x_j$ (C1), $\left|COMP_i - COMP_j\right| \le \$20,000$ (C2), $\left|AGE_i - AGE_j\right| \le 5$ (C3), and $FEMALE_i = FEMALE_j$ (C4). The results are not sensitive to the window width chosen. The three chosen control variables are the ones that have the highest explanatory power for equity allocation. Including all control variables is computationally prohibitive and adds little more insight.

To compute $\hat{\tau}$ when the sample is small, one could review all possible observation pairs that satisfy the conditions C1 or C1-C4 specified above. Since the magnitude of $\tilde{N}$ (number of qualified pairs) is of the order of $O(nob^2)$ where *nob* is the total number of individuals in the sample, this complete counting strategy is computationally infeasible when the sample is large (e.g., the sample of half a million participants underlying this study). Instead, a million qualifying pairs are randomly selected to produce the statistic $\hat{\tau}$ using the following procedure: (1) Randomly pick up observation *i* from the full sample; (2) Randomly pick up another observation *j* among all that satisfy C1 or C1-C4 to form a pair with *i*; (3) Repeat (1) and (2) till there are a million unique pairs.

The standard errors of the statistics are obtained through nonparametric bootstrap. To account for the possible correlation of equity allocation of individuals from the same plan, the bootstrap is done by plan blocks. That is, when an individual gets re-sampled, all other individuals in the same plan automatically get re-sampled (see Chernick (1999), chapter 5). The effective sample size for the standard error of $\hat{\tau}$ is of the order of the number of plans.

## References


Abrevaya, Jason, and Wei Jiang, 2004, A Nonparametric Approach to Measuring and Testing Joint Curvature, *Journal of Business and Economic Statistics*, forthcoming.

Agnew, Julie, 2002, Inefficient Choices in 401(k) Plans: Evidence from Individual Level Data, Working Paper, College of William and Mary.

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

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

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

Bassett, Gilbert, Jr, and Roger Koenker, 1978, Regression Quantiles, *Econometrica* 46, 33-50.

Black, Fischer, 1972, Capital Market Equilibrium with Restricted Borrowing, *Journal of Business* 45, 444-55.

Camerer, Colin, Samuel Issacharoff, George Lowenstein, Ted O’Donoghue, and Matthew Rabin, 2003, Regulation for Conservatives: Behavioral Economics and the Case for ‘Asymmetric Paternalism’, *University of Pennsylvania Law Review* 151, 1211-1254.

Cass, David, and Joseph Stiglitz, 1970, The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation, *Journal of Economic Theory* 2, 122-60.

Chernick, Michael, 1999, *Bootstrap Methods*, John Wiley & Sons, Inc. USA.

Choi, James, David Laibson, Brigitte Madrian, and Andrew Metrick, 2004a, For Better or For Worse: Default Effects and 401(k) Savings Behavior, in David Wise, ed., *Perspectives in the Economics of Aging,* forthcoming.

Choi, James, David Laibson, Brigitte Madrian, and Andrew Metrick, 2004b, Plan Design and 401(k) Savings Outcomes, *National Tax Journal*, forthcoming.

Chordia, Tarum, Richard Roll, and Avanidhar Subrahmanyam, 2002, Evidence on the Speed of Convergence to Market Efficiency, UCLA Working Paper.

Cogan, John, and Olivia Mitchell, 2003, Perspectives from the President's Commission on Social Security Reform, *Journal of Economic Perspectives* 17, 149-172.

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

Daniel, Kent, David Hirshleifer, and Siew Hong Teoh, 2002, Investor Psychology in Capital Markets: Evidence and Policy Implications, *Journal of Monetary Economics* 49, 139-209.

Elton, Edwin, Martin Gruber, and Christopher Blake, 2003, The Adequacy of Investment Choices Offered by 401K Plans, Working paper, New York University.

Freedman, David, 2001, Ecological Inference and the Ecological Fallacy, in N. J. Smelser and Paul B. Baltes, ed.: *International Encyclopedia for the Social and Behavioral Sciences*, (Elsevier).

Gabaix, Xavier, and David Laibson, 2003, Industrial Organization with Bounded Rational Consumers, Working paper, MIT and Harvard University.

Glaeser, Edward, 2003, Psychology and the Market, Discussion Paper 2023, Harvard Institute of Economic Research.

Goodfriend, Marvin, 1992, Information-Aggregation Bias, *American Economic Review*, 82(3), 508-519.

Hanushek, Eric, Steven Rivkin, and Taylor Lori, Aggregation and the Estimated Effects of School Resources, *Review of Economics and Statistics*, 78(4), 611-627.

Holden, Sarah and Jack VanDerhei, 2001. 401(k) Plan Asset Allocation, Account Balances, and Loan Activity in 2000. *ICI Perspective* 7(5) (November 2001), 1-27.

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

Huberman, Gur, Sheena Iyengar, and Wei Jiang, 2003, Defined Contribution Pension Plans: Determinants of Participation and Contributions Rates, Working Paper, Columbia Business School.

Iyengar, Sheena, and Wei Jiang. 2003, How Choices are Demotivating: The Impact of More Options on 401(k) Investment, Working Paper, Columbia Business School.

Iyengar, Sheena, and Mark Lepper, 2000, When choice is Demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology* 76, 995-1010.

Kahneman, Daniel, 2003, Maps of Bounded Rationality: Psychology for Behavioral Economics, *American Economic Review* 93, 1449-1475.

Kendall, M. G., 1968. *Rank Correlation Methods* (Griffin, London).

Khan, Shakeeb, and James Powell, 2001, Two-Step Estimation of Semiparametric Censored Regression Models, *Journal of Econometrics* 103, 73-110.

Langer, Thomas, and Craig Fox, 2003, Naïve Diversification in Allocation among Risky Prospects: Partition-Dependence, Unit-Dependence, and Procedure-Dependence, Working Paper, UCLA.

Liang, Nellie, and Scott Weisbenner, 2003, Investor Behavior and the Purchase of Company Stock in 401(k) Plans – the Importance of Plan Design, Working Paper, Board of Governors of the Federal Reserve System and University of Illinois.

Lintner, John, 1965, Security Prices, Risk, and Maximal Gains from Diversification, *Journal of Finance* 20, 587-615.

Mossin, Jan, 1966, Equilibrium in a Capital Asset Market, *Econometrica* 34, 261-76.

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

Read, Daniel and George Lowenstein, 1995, Diversification Bias: Explaining the Discrepancy between Combined and Separate Choices, *Journal of Experimental Psychology* 1, 7-59.

Ross, Stephen, 1978, Mutual Fund Separation and Financial Theory – The Separating Distributions, *Journal of Financial Economics* 17, 254-86.

Serfling, Robert, 1980. *Approximation Theorems of Mathematical Statistics* (John Wiley and Sons, New York.).

Sharpe, William, 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* 19, 425-42.

Simonson, Itamar, 1990, The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior, *Journal of Marketing Research* 27, 150-62.

Tversky, Amos, and Daniel Kahneman, 1986, Rational Choice and the Framing of Decisions, *Journal of Business* 59, 251-278.

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

**Table I. Summary statistics of individual- and plan-level attributes for the 572,157 401(k) participants records in 639 plans in 2001**

*NCHOSEN* (*NCHOSEN95*) is the number of funds in which a participant chooses to invest all (at least 95%) of his balance. *%EQ* is the proportion of current-year contribution that a participant invests in equity funds. (A balanced fund counts as a 0.5 equity fund.) *%EQOffered* is the proportion of equity funds out of all funds offered by a plan. *CONTRIBTUION* is the dollar amount that a participant contributed to his defined contribution plan in 2001. *COMP* is a participant's annual compensation. *WEALTH* is the average financial wealth of the nine-digit zip code neighborhood where a participant lives. *FEMALE* is the gender dummy variable. *AGE* and *TENURE* stand for a participant's age and his tenure with the current employer. *MATCH* is the average match rate by employer up to five percent of a participant's compensation. *COMPSTK* is a dummy variable for the availability of company stock among the offered funds. *DB* is a dummy variable for the presence of a defined benefit plan. *NCHOICE* is the number of funds available to the plan participants. *WEB* is the proportion of participants who register for web access to their DC accounts in a plan. *NEMPLOY* is the number of employees eligible to participate; it proxies for plan size.

| | **Unit** | **Mean** | **Std. Dev** | **Median** |
|---|---|---|---|---|
| NCHOSEN | 1 | 3.48 | 1.99 | 3 |
| NCHOSEN95 | 1 | 3.12 | 1.69 | 3 |
| %EQ | 1% | 66.84 | 35.40 | 78.94 |
| %EQOffered | 1% | 66.42 | 7.73 | 68.18 |
| CONTRIBUTION | $1,000 | 4.32 | 3.38 | 3.34 |
| COMP | $10,000 | 6.44 | 6.67 | 5.25 |
| WEALTH | $10,000 | 6.06 | 17.84 | 1.64 |
| FEMALE | 0-1 | 0.38 | 0.46 | 0 |
| AGE | year | 43.36 | 9.75 | 44 |
| TENURE | year | 11.06 | 9.25 | 9.08 |
| MATCH | 1% | 68.25 | 26.68 | 50 |
| COMPSTK | 0-1 | 0.52 | 0.50 | 1 |
| DB | 0-1 | 0.62 | 0.48 | 1 |
| NCHOICE | 1 | 13.66 | 5.75 | 13 |
| WEB | 1% | 28.68 | 11.73 | 26.21 |
| NEMPLOY | 100 | 169.77 | 222.53 | 56.8 |

**Table II. Determinants of Number of Funds Used: Estimates of**

$$NChosen_{i,j} = \gamma NChoice_j + \beta Controls_{i,j} + \varepsilon_{i,j}$$

*NCHOSEN* (*NCHOSEN95*) is the number of funds in which a participant chooses to invest all (at least 95%) of his balance. *NCHOICE* is the number of fund options available to employees of the plan. Definitions of control variables are the same as those in Table I. The coefficients and standard errors (S.E.) are multiplied by 100. Columns 1-3 use all participant records and columns 4 uses only records of new entrants in 2001. In column 3, the dependent variable is the smallest number of funds in which at least 95% of the participant's retirement assets are invested; in all other columns it is the total number of funds chosen by an individual. All regressions include plan-averages of individual characteristics as control variables. Compensation and wealth variables enter in logs. Standard errors adjust for both heteroskedasticity and arbitrary correlation of error disturbances clustered at the plan level. The effective sample size for the coefficients on individual (plan) attributes is of the order of the number of individuals (plans). * indicates that the coefficient is statistically different from zero at the 5% level.

| | All Participants | | | | | | New Entrants | |
|---|---|---|---|---|---|---|---|---|
| | NCHOSEN | | | | NCHOSEN95 | | NCHOSEN | |
| | (1) | | (2) | | (3) | | (4) | |
| | COEF*100 | S.E.*100 | COEF*100 | S.E.*100 | COEF*100 | S.E.*100 | COEF*100 | S.E.*100 |
| NCHOICE | **0.95** | 0.70 | **1.03** | 0.70 | **0.56** | 0.52 | **-0.89** | 0.78 |
| CONTRIBUTION | **10.54*** | 0.56 | **--** | **--** | **7.96*** | 0.43 | **12.48*** | 1.73 |
| COMP | **-0.02** | 2.30 | **33.05*** | 2.87 | **-0.81** | 1.57 | **-6.14** | 5.18 |
| WEALTH | **1.20*** | 0.51 | **3.90*** | 0.55 | **1.09*** | 0.41 | **1.18** | 0.89 |
| FEMALE | **14.51*** | 1.97 | **14.84*** | 1.95 | **10.71*** | 1.45 | **7.84*** | 3.57 |
| AGE | **-1.66*** | 0.10 | **-1.35*** | 0.09 | **-1.44*** | 0.09 | **-1.46*** | 0.16 |
| TENURE | **0.88*** | 0.26 | **0.95*** | 0.26 | **-0.27** | 0.18 | **--** | **--** |
| MATCH | **0.00** | 0.24 | **0.00** | 0.23 | **-0.01** | 0.20 | **0.10** | 0.32 |
| COMPSTK | **70.67*** | 12.72 | **67.16*** | 12.68 | **48.99*** | 10.74 | **48.34*** | 18.10 |
| DB | **-6.31** | 15.35 | **-6.06** | 15.21 | **-4.93** | 11.83 | **3.36** | 16.50 |
| WEB | **1.17** | 0.71 | **1.39** | 0.71 | **0.79** | 0.51 | **1.04** | 0.82 |
| NEMPLOY | **-10.28*** | 4.79 | **-9.25*** | 4.73 | **-8.83*** | 3.86 | **-14.93*** | 5.22 |
| Intercept | **1036.95** | 284.44 | **664.25** | 290.06 | **750.53** | 173.14 | **793.19** | 262.33 |
| # Individuals & plans | 572157 | 641 | 572157 | 641 | 572157 | 641 | 38029 | 547 |
| $R^2$ | 0.075 | | 0.060 | | 0.059 | | 0.055 | |

**Table III. The Conditional 1/n Rule: Prevalence of equal allocation among all *chosen* funds by 2001 new participants who chose ten funds or fewer**

The Herfindahl index $H_i = \sum_{j=1}^{n_i} s_{i,j}^2$ measures adherence to the conditional 1/n rule; the variables are as follows: $s_{i,j}$ is the share of individual *i*'s contribution in fund *j* out of his total contribution, and $n_i$ is the total number of funds chosen by individual *i*. Company stock is excluded. $\underline{H}$ ($\bar{H}$) represents the lower (upper) bound of the Herfindahl index values classified as conditional 1/n allocation: $\underline{H}(n)$ is $1/n$; and $\bar{H}(n)$ is equal to an index value that would result from a portfolio in which the total deviation from a strict $1/n$ allocation is 20% of $1/n$ (that is, $\bar{H}(n) = \max\left\{\sum_{j=1}^{n} s_j^2 : \sum_j \left|s_j - \frac{1}{n}\right| \le \frac{20\%}{n}\right\}$). $Freq_1$ is the empirical frequency of individuals falling into the interval $[\underline{H}, \bar{H})$. $\max_{j\neq 1}\left(Freq_j\right)$ is the frequency of individuals falling into an interval, with equal length, out of $[\underline{H}, \bar{H})$ that receives most observations. * indicates that the ratio is significantly greater than one at less than 2.5% significance level using 1,000 nonparametric re-sampling bootstraps. There are 37,798 new entrants in 2001 who contribute positive amount to non-company-stock funds.

| (1)<br>*# funds chosen* | (2)<br>*% of new entrants* | (3)<br>$\underline{H}$ | (4)<br>$\bar{H}$ | (5)<br>$Freq_1$ | (6)<br>$Freq_1 / \max_{j\neq 1}\left(Freq_j\right)$ |
|---|---|---|---|---|---|
| 1 | 38.6% | 1 | 1 | -- | -- |
| 2 | 17.5% | 0.5 | 0.505 | 64.0% | 12.81* |
| 3 | 15.6% | 0.3333 | 0.3356 | 17.9% | 1.78* |
| 4 | 13.2% | 0.25 | 0.2513 | 37.4% | 8.89* |
| 5 | 7.3% | 0.2 | 0.2008 | 26.6% | 8.19* |
| 6 | 3.5% | 0.1667 | 0.1672 | 1.3% | 0.25 |
| 7 | 1.8% | 0.1429 | 0.1433 | 1.0% | 0.19 |
| 8 | 1.1% | 0.125 | 0.1253 | 3.9% | 1.14 |
| 9 | 0.6% | 0.1111 | 0.1114 | 5.1% | 1.20 |
| 10 | 0.4% | 0.1 | 0.1002 | 53.3% | 13.50* |

**Table IV. Sensitivity of Equity Allocation to Equity Exposure: Estimates of**

$$\%EQ_{i,j} = \gamma\,\%EQOffered_j + \beta Control_{i,j} + \varepsilon_{i,j}$$

The dependent variable, *%EQ*, is the percentage of current year contribution that goes to equity funds. The key independent variable, *%EQOffered*, is the percentage of equity funds out of all funds offered. Company stock is excluded from both variables. In regressions with controls, the control variables are: (1) individual attributes: savings rate, log compensation, log wealth, gender, age, tenure, registration for web access; (2) plan policies: match rate, availability of company stock, presence of restricted match in company stock, presence of a DB plan, and the number of funds offered; (3) plan average of individual attributes. Estimates are obtained through censored median regression (Powell (1984)) to account for the constraint that *%EQ* falls within [0, 100%]. The standard errors are adjusted for both heteroskedasticity and arbitrary correlation of error disturbances clustered by plan. * indicates that the coefficient is statistically different from zero at 5% level.

| | (1) | | (2) | | (3) | | (4) | |
|---|---|---|---|---|---|---|---|---|
| | All *NFunds* | | | | *Nfunds* <= 10 | | *Nfunds* > 10 | |
| | COEF | S.E. | COEF | S.E. | COEF | S.E. | COEF | S.E. |
| | Panel A: Full Sample—Uniform Sensitivity | | | | | | | |
| *%EQOffered* | **0.175** | 0.274 | **0.177*** | 0.088 | **0.292*** | 0.107 | **0.058** | 0.09 |
| R-squared | 0.000 | | 0.061 | | 0.063 | | 0.068 | |
| | Panel B: Full Sample—Sensitivity Varying with Tenure | | | | | | | |
| *%EQOffered* | **0.141** | 0.154 | **0.222*** | 0.106 | **0.184** | 0.136 | **0.146** | 0.099 |
| *TENURE * %EQOffered* | **-0.005** | 0.002 | **-0.005** | 0.008 | **0.011** | 0.01 | **-0.009** | 0.008 |
| R-squared | 0.000 | | 0.062 | | 0.063 | | 0.068 | |
| Controls? | N | | Y | | Y | | Y | |
| # Individuals & plans | 549,341 | 638 | 549,341 | 638 | 152,283 | 297 | 397,058 | 341 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Panel C: New Entrants | | | | | | | |
| *%EQOffered* | **0.004** | 0.842 | **0.182** | 0.201 | **0.197** | 0.227 | **0.204** | 0.172 |
| R-squared | 0.000 | | 0.065 | | 0.078 | | 0.065 | |
| Controls? | N | | Y | | Y | | Y | |
| # Individuals & plans | 37,558 | 548 | 37,558 | 548 | 10,198 | 234 | 27,360 | 314 |

## Table V: A Nonparametric Test of the Sensitivity of Equity Allocation to Equity Exposure

The first three columns list the range of *NCHOICE* (number of funds offered), and the numbers of plans and individuals in those ranges. $\hat{\tau}$ is calculated according to

$$\hat{\tau} = \frac{1}{\tilde{N}} \sum_{x_i > x_j} I(y_i > y_j \mid x_i > x_j),$$

where $x_i$ and $y_i$ represent the equity exposure and equity allocation of individual *i*, and $\tilde{N}$ is the total number of observation pairs with different equity exposures. $\hat{\tau}_1$ compares any two individuals with different equity exposures, and $\hat{\tau}_2$ only compares pair observations that have similar compensation (difference smaller than $20,000), age (difference smaller than 5 years), and same gender. Panels A and B report the nonparametric statistic using individual observations; standard errors are adjusted for correlations of observations clustered by the plan. Panel C summarizes results of the same test applied to plan-level aggregate data. * indicates that the null hypothesis of equity allocation being independent from equity exposure is rejected in favor of positive dependence at less than 2.5% significant level.

| NChoice | # Plans | # Individuals | $\hat{\tau}_1$ | s.e. | $\hat{\tau}_2$ | s.e. |
|---|---|---|---|---|---|---|
| | | | A: Individuals—Full Sample | | | |
| 4-59 | 638 | 549,341 | 49.98% | 0.91% | 49.84% | 0.87% |
| 4-10 | 297 | 152,283 | 51.59% | 1.00% | 51.64% | 0.65% |
| 11-59 | 341 | 397,058 | 49.31% | 1.26% | 49.20% | 0.84% |
| | | | B: Individuals—New Entrants | | | |
| 4-59 | 548 | 37,558 | 49.52% | 1.72% | 49.75% | 1.78% |
| 4-10 | 234 | 10,198 | 47.97% | 2.21% | 47.99% | 2.00% |
| 11-59 | 314 | 27,360 | 49.72% | 2.12% | 50.25% | 2.18% |
| | | | C: Plans | | | |
| 4-59 | 638 | -- | 57.72%* | 1.23% | -- | |
| 4-10 | 296 | -- | 60.20%* | 2.23% | -- | |
| 11-59 | 342 | -- | 56.35%* | 2.07% | -- | |

**Figure 1(a):** **The number of plans that offer a given number of funds and the total number of participants in these plans**

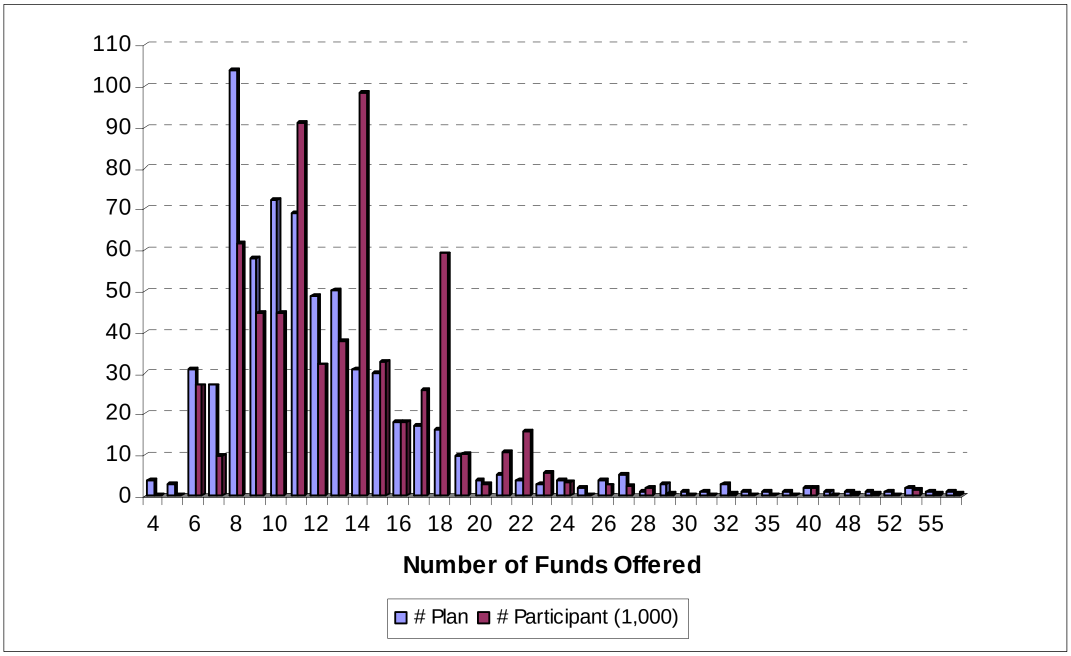



**Figure 1(b):** **The fraction of participants who use a given number of funds (NChosen)), and the fraction of participants with at least 95% of their balances in these many funds (NChosen95)**

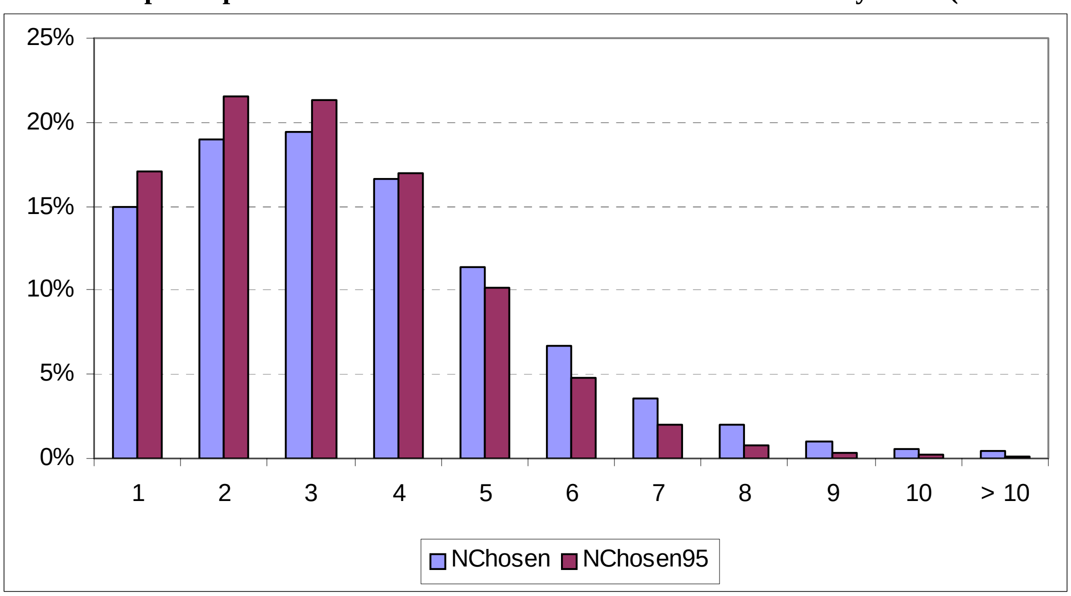


**Figure 2(a): The relation between the Number of Funds Chosen vs. Number of Funds Offered**

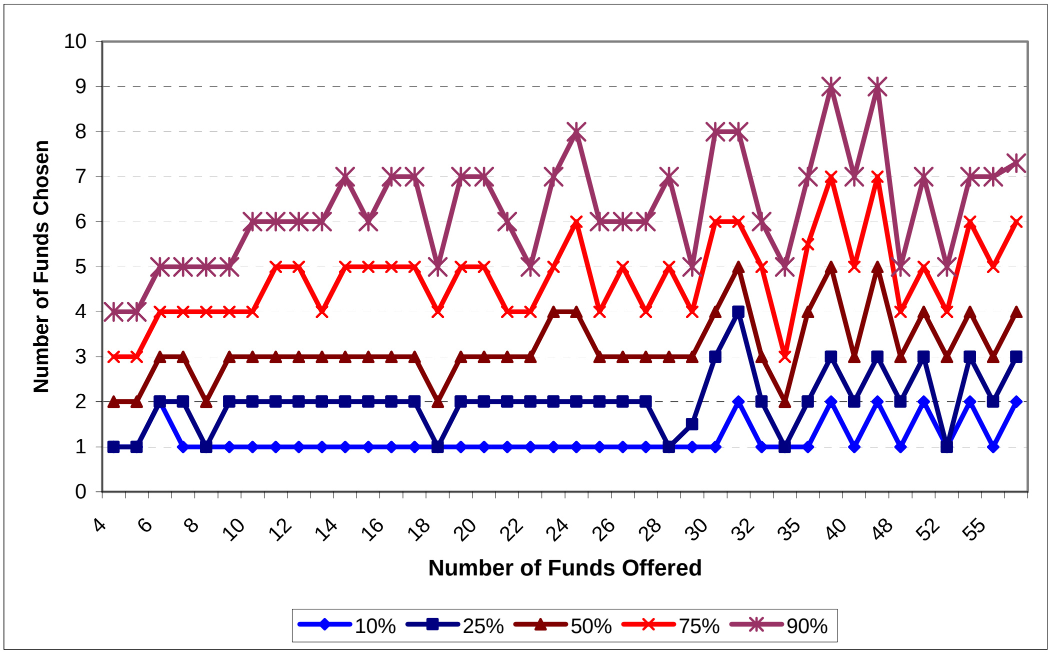



**Figure 2(b): The Relation between the Number of Funds Chosen vs. Number of Funds Offered**

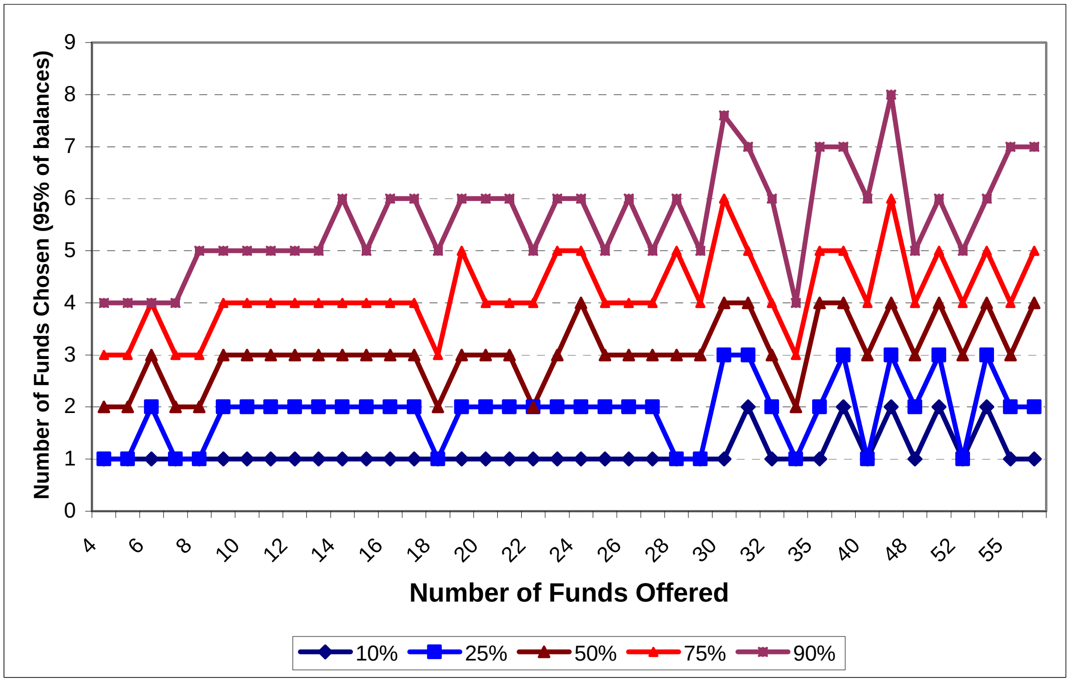


**Figure 2(c): The Relation between the Number of Funds Chosen vs. Number of Funds Offered for 2001 New Entrants**

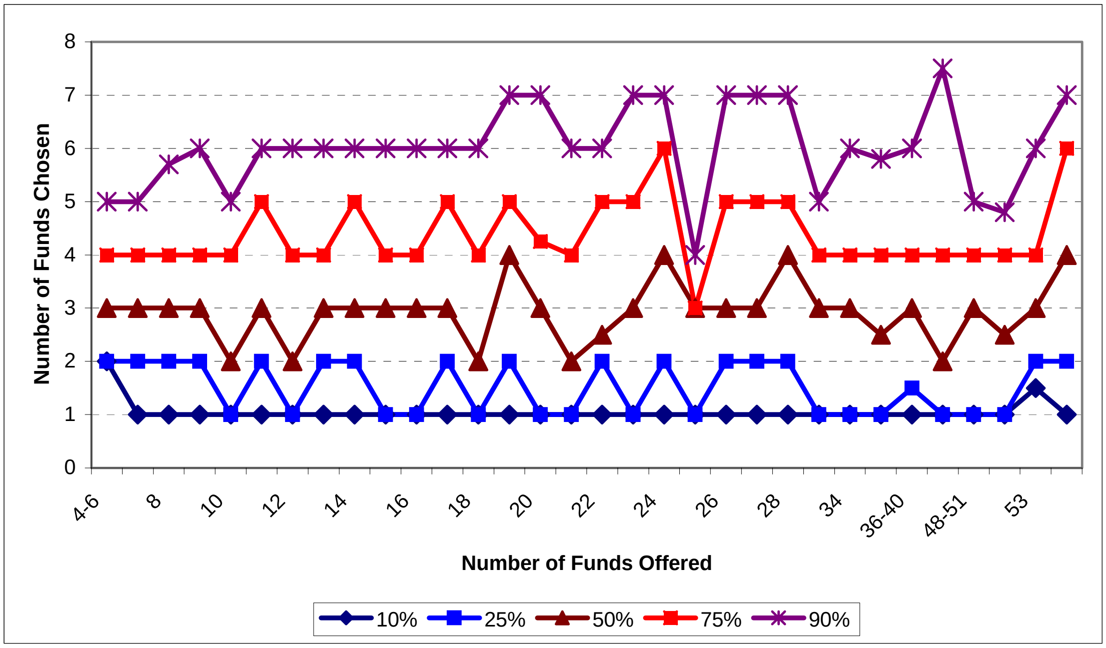



**Figure 3: The Number of Funds Used by Plans vs. Number of Funds Offered**

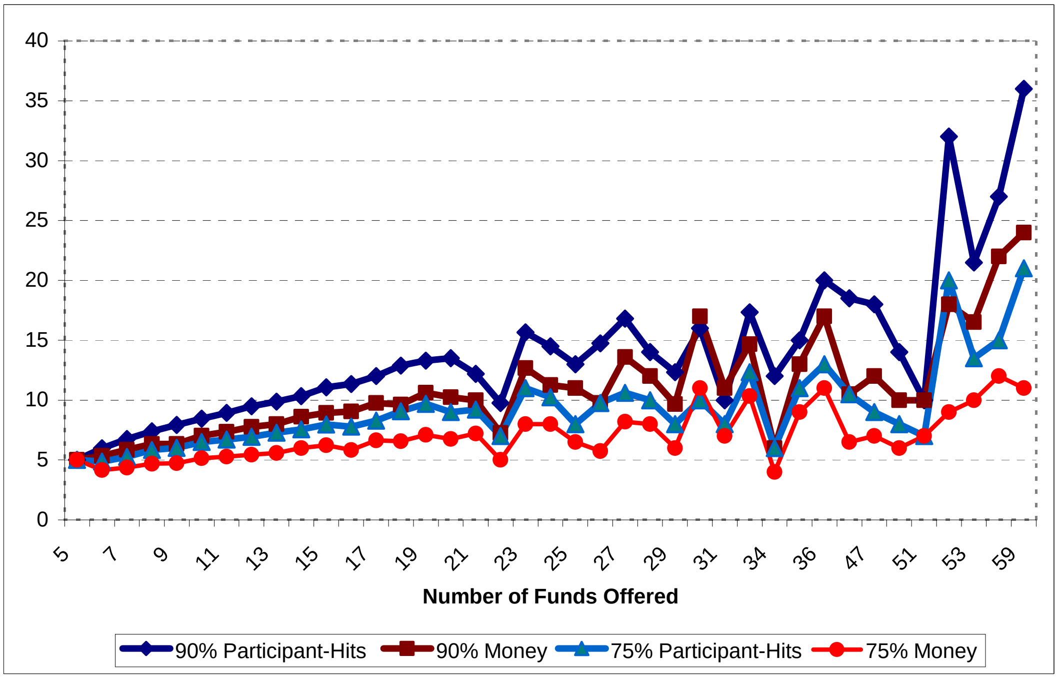


**Figure 4: Concentration of Plan Assets and Participants vs. Number of Funds Offered**

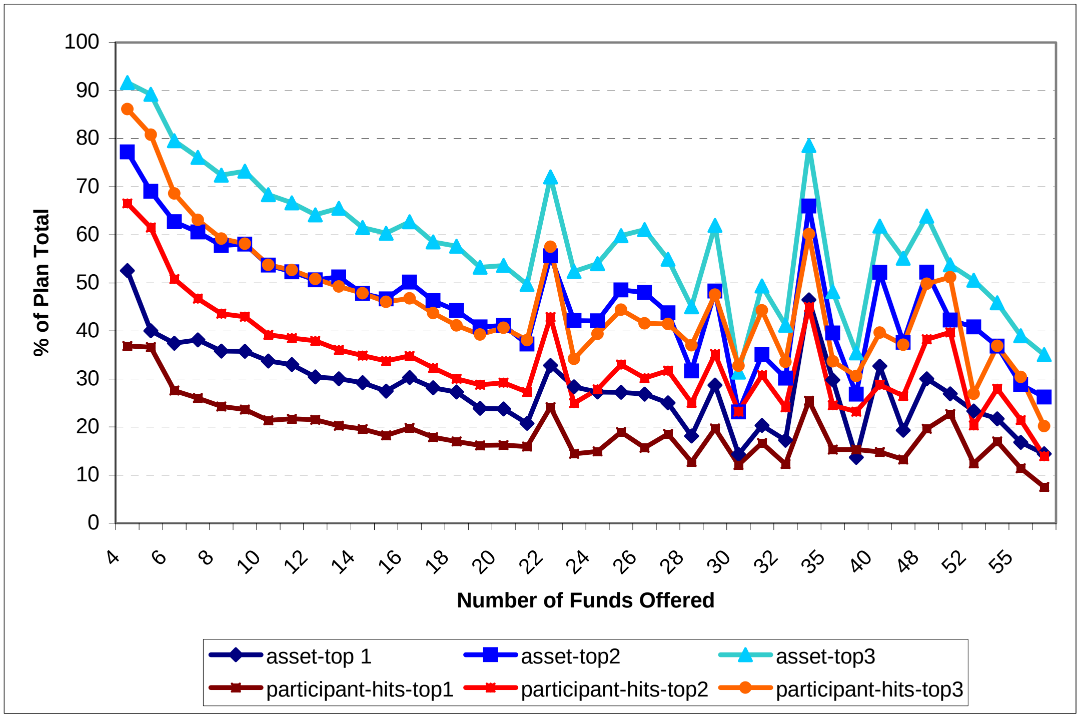
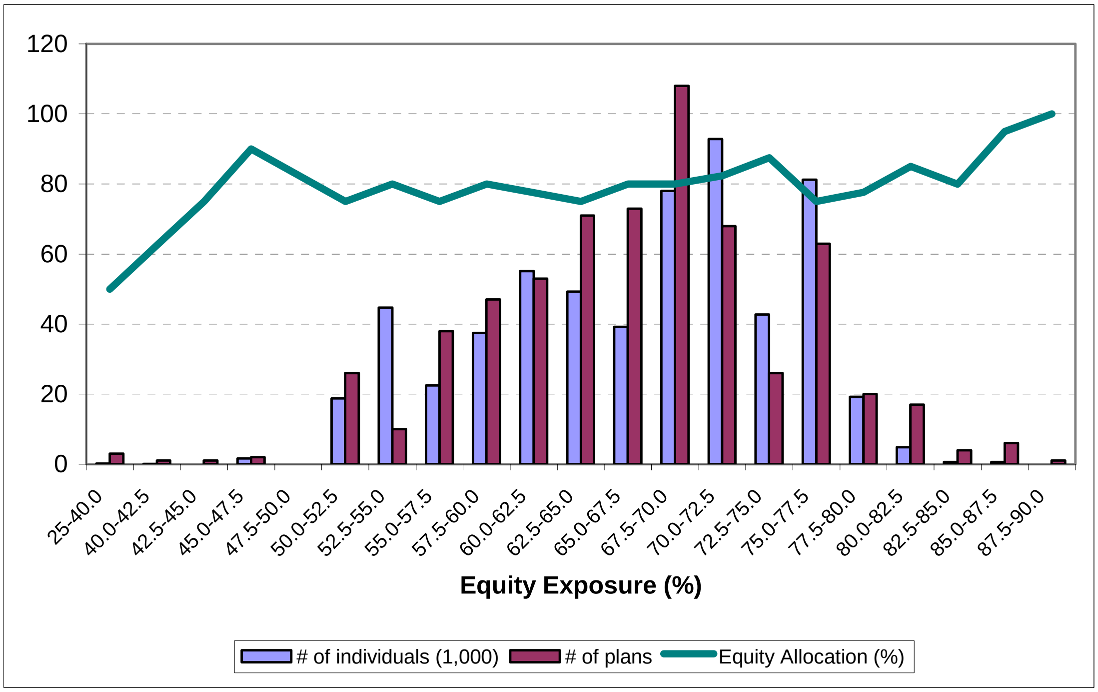



**Figure 5: For a given equity exposure, the number of plans offering that exposure, the number of participants whose plans offer that exposure, and the median equity allocation of these participants.**



## Footnotes:

[1] Framing effects are present in other settings see, e.g., Tversky and Kahneman (1986) for a discussion. In particular, varying the number of choices may lead decision makers to choose differently, including choose not to choose. Iyengar and Lepper (2000) report a clever experiment to this effect.

[2] An employee's total contribution also includes money contributed by his employer.

[3] In 2001, the Vanguard family ran more than 120 different funds. The median number of choices in the present sample is 13.

[4] In our sample, eligible employees are allowed to invest in *all* funds on the menu. Though a few plans impose the minimum of 5% of total contribution to any fund, the constraint (i.e., a maximum of 20 funds chosen), if enforced, does not seem to be binding given that only 118 participants chose more than 15 funds.

[5] 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.

[6] See, e.g., Wooldridge (2003) for a discussion of the relation between effective sample size and cluster-adjusted standard errors.

[7] Unfortunately, the Vanguard data set spells out only the current-year allocations of contributions into *fund categories,* but not to individual funds. They do specify how the balances are allocated across *individual funds*.

[8] Agnew (2002) who uses a different definition of the 1/n heuristic in an examination of a single plan with four funds, reports that "[w]hile the percentage of individuals who follow the 1/n heuristic in this study is lower than that found in previous studies, it still represents 5% of the sample." Liang and Weisbenner (2003) document that plan-level contribution to company stock decreases, on average, at the rate of 1/n as the number of fund choices increases from 2 to 10.

[9] Using a sample of 401(k) plans from 2001 where 69% of the plans offered 10 funds or fewer, Elton et al. (2003) report that 62% of the plans provided an incomplete set of investment alternatives in terms of spanning and achieving Sharpe ratios comparable to the general finance market.

[10] Two-sided Tobit, on the other hand, is consistent only when all distributions are normal and heteroskedastic. Further, the interpretation of its estimates assumes that participants who choose no equity funds would like to take negative positions (short positions) if allowed; and analogously for 100% equity investment. Such an extrapolation of corner solution may not be plausible for typical 401(k) investors.

[11] Splitting the full sample at the median number of funds offered (13) yields similar, but weaker results especially for the fewer-choice subsample. A straight plot (without control) of median equity allocation of participants facing above and below median equity exposure shows that the former lies above the latter when the number of funds offered is fewer than 10. Once there are more than 10 choices, there are no consistent patterns. For example, when number of choices is between 19 and 27, the median allocation of participants facing higher equity exposure lies above that of lower equity exposure, but in the range of 16 to 18 funds, the reverse is true.

[12] The focus of the analysis is the distance of these statistics from the neutral value of ½. Given the standard errors reported in Panels A and C, an upper bound for the standard errors of the differences between individual and plan level estimates is also available: $s.e.(\hat{\tau}_C - \hat{\tau}_A) < s.e.(\hat{\tau}_C) + s.e.(\hat{\tau}_A)$, which implies that all three plan-level estimators in Panel C are significantly greater than their individual counterparts in Panel A at less than 2.5% level.

[13] More than 95% of the plans in their study offer fewer than 12 funds.

[14] Such censoring is meant to calibrate to the 401(k) data. The nature of the simulation does not change if the data are not censored.

[15] Alternatively, one can decompose $\%EQOffered_j$ into two parts: (i) $\%EQ\hat{O}ffered_{j,t}$, the predicted equity exposure from all available exogenous plan-level attributes (such as average compensation, plan size, etc.), and $\overline{\%EQ}^{o}_{j,<t}$; and (ii) $\%EQ\tilde{O}ffered_{j,t}$, the residual component. Again, the sensitivity

of new entrants' equity allocation to $\%EQ\tilde{O}ffered_{j,t}$ is insignificant (coefficient = 0.10, *t*-statistic = 0.18).

[16] There is no need for a larger number of artificial samples in the simulation because the coefficients are extremely tightly distributed: the standard deviation of the 30 estimates is merely 0.003.

## Figure Captions:

**Figure 2(a)**

For a given number of funds offered, the number of funds used by the 10% of the participants who use the fewest funds, by the 25% of the participants who use the fewest funds, etc.

**Figure 2(b)**

For given number of funds offered, the number of funds used by the 10% of the participants who use the fewest funds to hold at least 95% of their balances, by the 25% of the participants who use the fewest funds to hold at least 95% of their balances, etc..

**Figure 2(c)**

For a given number of funds, the number of funds used by the 10% of the participants who use the fewest funds, by the 25% of the participants who use the fewest funds, etc..

**Figure 3**

For each level of number of funds offered, the number of funds needed to hold 90% and 75% of the plan-level balances and total participant hits, where each participant record counts as a hit.

**Figure 4**

The proportion of plan assets and participant hits concentrated in the top 1, 2 or 3 funds.