Mean-Variance Spanning Tests:
The Fiduciary Case in 401(k) Plans

Farid AitSahlia1 Thomas Doellman2 Sabuhi Sardarli3

February 23, 2017

Abstract

We first highlight critical subtleties in the application of the standard Wald test for regression-based mean-variance spanning under short-sales constraints. We address the issue of near-singularity in particular by appealing to a characterization of stochastic discount factors in the presence of a risk-free asset. Next, we apply our approach to assess our proposal for a regulatory prudent default investment selection strategy in 401(k) plans. It consists of a low-volatility subset of funds that results in better spanning relative to the full plan menu. In effect, our study introduces a new dimension to the debate regarding the low-volatility anomaly.

JEL Classification Codes: G11, G20, G23

Keywords: 401(k) plans, Retirement plan menu, Mean-variance spanning, Style analysis, Mutual funds.

---

1 Corresponding author. Warrington College of Business, University of Florida, Gainesville, FL, 32611; email: farid.aitsahlia@warrington.ufl.edu; Tel: +1 (352) 392-5058.

2 John Cook School of Business, Saint Louis University, St. Louis, MO 63108; email: tdoellma@slu.edu, Tel: +1 (314) 977-3815.

3 College of Business Administration, Kansas State University, Manhattan, KS 66506; email: ssardarli@ksu.edu; Tel: +1 (785) 532-5771.
1. Introduction

Starting with the classic paper of Huberman and Kandel (1987), the traditional understanding of testing for mean-variance spanning is whether a given mean-variance efficient portfolio can be improved with the inclusion of additional assets. Applications of such testing have been as varied as assessing the benefits of international diversification, evaluating mutual fund performance and testing linear-factor asset pricing models (see, e.g., Errunza, Hogan, and Hung, 1999; Li, Sarkar, and Wang, 2003; De Roon and Nijman, 2001; and the textbooks of Campbell, Low, and McKinlay, 1997; Cochrane, 2005). A more recent application is in the area of defined contribution retirement plans such as 401(k) plans in the United States, where the idea is to determine the extent to which the mutual funds in a plan span the same efficient frontier as a set of benchmark funds or indices. In this case, however, the presence of short-sales constraints in 401(k) plan investing results in a measurably more complex problem than the classic version.

In this paper, we show that prior attempts by Elton, Gruber, and Blake (2006) (hereafter referred to as EGB) and Tang, Mitchell, Motolla, and Utkus (2010) (TMMU hereafter) to analyze spanning properties of 401(k) plans fail to correctly apply the relevant Wald test developed by De Roon, Nijman, and Werker (2001) (DNW hereafter) to account for the no-short sales constraints in regression-based mean-variance spanning. In addition, we also show that the proper implementation of the Wald test of DNW is in fact subtle and is particularly prone to numerical instability. As a remedy, we develop an approach that exploits the uniqueness of the mean stochastic discount factor in the presence of a risk-free asset. Finally, we use regression-based mean-variance testing to move along another direction in the context of defined contribution plans: to test whether steering plan participants to a specific subset of plan menu offerings that complies with fiduciary regulation does not result in an inferior efficient frontier.
Defined contribution (DC) plans, such as those provided by employers under section 401(k) of the U.S. Internal Revenue Code, figure prominently in the retirement planning of tens of millions of working Americans (Choi, 2015). For the participants, the DC plans provide access to the financial markets through contribution and allocation decisions. In effect, DC plans form a substantial source of capital for institutional investing (Rydqvist, Spizman, and Strebulaev, 2014). Financial assets held in 401(k) plans specifically have grown significantly over the past decade. By the end of the year 2015, $4.7 trillion were invested in 401(k) plans, up from $3.1 trillion in 2010, and compared with $1.6 trillion in 2002.¹ This rising trend is expected to continue at a faster pace given the accelerating shift from defined benefit to defined contribution plans in the corporate world as well as the underfunding of several public pension funds (see, e.g., Munnell, Aubry, Hurwitz, and Quinby, 2011; Munnell, 2014; Novy-Marx and Rauh, 2011).

DC plans present challenges for both sponsoring firms and their employees. As fiduciaries, employers must conduct due diligence regarding 401(k) plan providers (e.g., mutual fund companies) in order, in part, to identify plans that would appeal to employees with a wide range of risk profiles. To illustrate, we refer to Pool, Sialm, and Stefanescu (2016), for a recent evaluation of plan menu offerings, and Bekaert, Hoyem, Hu, and Ravina (2017), for a study contrasting participants’ appetites for international exposure, depending on their age, education, financial literacy, etc. Employees, on the other hand, are overwhelmingly ill-prepared to make allocation decisions in their 401(k) plans (see, e.g., Huberman and Jiang, 2006; Roche, Tompaidis, and Yang, 2013).

¹ 2016 Investment Company Institute Fact Book (Figure 7.9, p. 141), available at https://www.ici.org/pdf/2016_factbook.pdf
One way to assess the adequacy of a DC plan offered by a given employer is through regression-based mean-variance testing. The idea is to compare the efficient frontier generated by the funds in the plan with that of a set of benchmark indices. If the efficient frontier of the former cannot be improved by the inclusion of one or more of the latter, then the set of funds in the plan are said to span, and thus provide a strong indication that the firm’s DC plan is adequate to capture a variety of risk profiles among its employees.

While the seminal paper (Huberman and Kandel, 1987) on regression-based tests of mean-variance spanning allows short-sales in the constitution of portfolios, fund positions in DC plans cannot be held short. DNW introduce a Wald test for regression-based mean-variance spanning under short-sales constraints. Mean-variance spanning tests of 401(k) plans are presented in EGB and TMMU, where both sets of authors refer to the Wald test of DNW. However, we show here that both EGB and TMMU apply this Wald test incorrectly: they only focus on a necessary, but not sufficient, condition for spanning, thus leading them to overestimate the number of spanning plans. We further show that the suggestion made by DNW for the implementation of their test is easily prone to numerical instability, and that in order to address the latter, we instead appeal to the uniqueness of the mean discount factor in the presence of a risk-free rate. Furthermore, while this issue has not been brought up in the extant literature, we show through principal component analysis that collinearity between spanning assets has a significant impact on regression-based mean-variance testing.

In addition to the spanning issues above, we also address allocation decisions faced by employees who, overwhelmingly, do not efficiently exploit the investments opportunities offered in their 401(k) plans. This limitation may be due to a number of factors, such as bias, behavioral
inertia, framing or lack of financial literacy, as evidenced by a number of studies. On the other hand, the push for widespread financial education, as advocated by certain academics and policy makers, has not resulted in tangible effects (Fernandes, Lynch, and Netemeyer, 2014). Until the year 2006, when the U.S. Congress passed the Pension Protection Act, 401(k) plan fiduciaries, such as sponsoring employers and plan providers, had long been reluctant to make fund recommendations or specific allocations due to possible liability exposure. The Pension Protection Act included relief for plan fiduciaries from liability as long as they comply with the Department of Labor safe harbor rule. The latter was issued in October 2007 and provided guidance concerning the so-called Qualified Default Investment Alternative (QDIA) option in 401(k) plans. Specifically, a QDIA is required to be “diversified so as to minimize risk of large losses” and to be “designed to provide varying degrees of long-term appreciation through a mix of equity and fixed-income exposures.”

The question of how to implement a practical QDIA in 401(k) plans remains open, despite the emergence of so-called Target-Date-Funds (TDFs) as a popular option. With a TDF, an investor picks a retirement age (target date) and the fund manager makes allocation decisions that are pre-determined and which change as the investor ages (the so-called glide path). The glide path typically starts with a tilt towards stocks and ends with more bonds by the target date. The passage of the Pension Protection Act in 2006, along with the QDIA designation of target-date funds and the move to automatic enrollment in 401(k) plans, has led a growing proportion of employees to contribute completely to TDFs. In light of the limitations of the average 401(k) plan beneficiary, and

---


investor, this trend has led to some favorable outcomes. For instance, the average 401(k) investor is now less likely to employ extreme portfolio allocations (i.e., all-equity, all-bond, or all-money-market). Certain characteristics of TDFs, however, have also made their increased use problematic. For example, their glide path schedules are not standard and can vary significantly and, as further detailed in Section 3 below, TDFs have higher shortfall risk than constant equal allocation strategies.

Our final objective in this paper is to develop a practical alternative to TDFs. It consists in identifying a QDIA-compliant subset of funds in the plan that either have low beta or enter in the composition of a minimum-variance portfolio. Our approach, which reduces the number of funds to pick from by 2/3 (from 18 to 6, on average,) is consistent with the fact that individuals prefer to choose among fewer alternatives. We should note that our proposal does not specify the amount to invest in the QDIA funds. Thus, it is quite possible investors will choose an even allocation across this subset, which is reasonable. This simple strategy is easily explained and is intuitively attractive to the average person (see, e.g., Benartzi and Thaler, 2001; Iyengar, Huberman, and Jiang, 2004; Iyengar and Kamenica, 2010). And while naïve, this diversification strategy is perfectly acceptable as recent work has shown that it is hard to outperform (see, e.g., De Miguel, Garlappi, and Uppal, 2009). We show that the subset of funds we thus identify satisfy the QDIA safe harbor provision and actually span better than the complete set of funds in the plan.

---

4 Shortfall risk is defined as the probability of running out of savings during retirement.
5 Equal proportion 1/N, where N is the number of assets. Also labelled even-allocation strategy.
7 Huberman and Jiang (2006) find evidence that participants in 401(k) plans indeed tend to allocate evenly across investment funds chosen.
The validity of our alternative to TDFs in 401(k) investing has significant implications for the industry. If the TPA and/or the plan sponsor would offer guidance on the set of funds identified by our strategy, then a plan participant could be more likely to create an investment portfolio that spans a set of benchmark indices. And fiduciaries would have an objective basis for their fund recommendations. Because of the manner with which our subset of funds is identified, driven by low-beta and variance-minimization considerations, our findings also contribute a new dimension to the debate regarding the superior empirical performance of low-volatility strategies known as the low-volatility anomaly (see, e.g., Baker, Bradley, and Wurgler, 2011; Frazzini and Pedersen, 2014).

The remainder of the paper is organized as follows. In Section 2, we review the methodology of mean-variance spanning tests with short-sales constraints in order to show its incorrect use in the current literature on 401(k) plans and also to highlight their numerical instability when correctly implemented. In this section, we also provide a rationalization for and a description of alternative benchmarks for mean-variance spanning tests. Section 3 covers the practical implementation of QDIA, highlighting deficiencies regarding the popular default QDIA option (TDFs) and details our alternative to it. Section 4 describes our data sources and contains sample details, Section 5 discusses our empirical results, and Section 6 concludes.

2. A Re-Examination of Mean-Variance Spanning Tests

In this section we review the regression-based mean-variance spanning test methodology. We show that in their implementation of the Wald test provided by DNW, both EGB and TMMU incorrectly specify their statistical hypotheses. We also re-derive the main test in DNW by avoiding their use of the stochastic discount factor approach for asset pricing and by relying instead on first principles for portfolio optimization. Due to the multitude of stochastic discount
factors, DNW suggest an implementation of their test that makes use of only the smallest and the largest mean discount factors. However, we show that this approach is easily prone to numerical instability. In order to address the latter, we instead appeal to the uniqueness of the mean discount factor in the presence of a risk-free rate.

2.1 Mean-variance spanning tests: Theory

Suppose a 401(k) plan consists of a set of K funds. To assess the efficiency of this plan relative to a benchmark of N index funds, we want to determine whether the mean-variance frontier associated with the K funds coincides with that generated with the augmented set of K+N funds. In other words, the K funds are “sufficient” to span the frontier of the K+N funds. Formally, denote by $\mu_R$ and $\mu_r$ the expected returns vectors of, respectively, the K plan funds and the N benchmark indices. Also define the corresponding covariance matrices $\Sigma_{R,R}$, $\Sigma_{r,r}$, $\Sigma_{R,r}$, where subscripts R and r refer to the K funds and N benchmark indices, respectively. Then the covariance matrix across the K+N assets is defined as

$$
\Sigma = \begin{pmatrix}
\Sigma_{R,R} & \Sigma_{R,r} \\
\Sigma_{r,r} & \Sigma_{r,r}
\end{pmatrix},
$$

(1)

with the superscript ′ used for matrix transposition. Given short-sales constraints, the mean-variance optimization problem across the K+N assets consists in determining the K-dimensional vector $\omega_R \geq 0_K$ and the N-dimensional vector $\omega_r \geq 0_N$ that maximize

$$
(\omega_R', \omega_r') \begin{pmatrix}
\mu_R \\
\mu_r
\end{pmatrix} - \frac{1}{2} \gamma (\omega_R', \omega_r') \begin{pmatrix}
\Sigma_{R,R} & \Sigma_{R,r} \\
\Sigma_{r,R} & \Sigma_{r,r}
\end{pmatrix} \begin{pmatrix}
\omega_R \\
\omega_r
\end{pmatrix},
$$

(2)

subject to $\omega_R' \cdot i_K + \omega_r' \cdot i_N = 1$, with $i_K$ and $i_N$ unit vectors of dimensions K and r, respectively, and the parameter $\gamma$ capturing risk-aversion.

---

8 For ease of notation, the subscript for a zero-vector is henceforth omitted when the dimension is clear.
If the funds in the plan span, then the optimal mean-variance allocation \((\omega^*_R, \omega^*_T)\) is such that \(\omega^*_R = 0_N\), where \(0_N\) is the \(N\)-dimensional zero vector. Incidentally, there is only one solution because the necessary assumption of positive definiteness of \(\Sigma\) leads to the convexity of the objective function (2).

Given return data on both the \(K\) funds and the additional \(N\) benchmarks indices, empirical spanning tests go back to Huberman and Kandel (1987) who, without short-sales constraints, test the null hypothesis

\[
\left( \beta i_K - i_N \right)^\alpha = 0,
\]

where the vector \(0\) is of dimension \(2N\), in the multivariate regression specification

\[
r = \alpha + \beta R + \varepsilon,
\]

where \(\beta\) is an \(N \times K\) matrix. A rejection of the null hypothesis above entails the possibility of improving the efficient frontier of the \(K\) funds with the inclusion of one or more funds matching the \(N\) benchmark indices. If the null is not rejected, then any strategy \(\omega_r\) involving the \(N\)-dimensional benchmark can be re-expressed as a strategy involving the \(K\) funds only. Indeed, for any \(\omega_r\) involving the \(N\) indexes and satisfying \(\omega'_r \cdot i_N = 1\), there exists a strategy \(\omega_R\) involving the \(K\) funds only and such that \(\omega'_R \cdot i_K = 1\), which is obtained as \(\omega'_r \cdot \beta\) by virtue of the second set of equations in (3).

2.2 Effect of short-sales constraints

In the presence of short-sales constraints, both EGB and TMMU use the specification (4) and set the null hypothesis to
\[ \alpha \leq \mathbf{0}, \quad (5) \]

where \( \mathbf{0} \) is, this time, of dimension \( N \). EGB justify their choice by arguing that “\textit{If short sales are forbidden, then only the addition of an asset with positive alpha can improve the efficient frontier…”} (p. 1304). Similarly, TMMU state: “\textit{As short-sales are not allowed for market benchmark index, if none of the } \alpha_i \textit{ are statistically significantly positive, we could conclude that performance of funds under the plan cannot be improved by holding a long position in any of the eight market benchmark indices.”} (p. 1078). We argue next that (5) is in fact incorrect, in the sense that one can still improve the efficient frontier of a given set with the addition of another asset whose alpha is negative relative to the former, even with short-sales constraints. Consider assets 1, 2, and 3 with return covariance matrix

\[
\begin{pmatrix}
0.011 & 0.002 & 0.001 \\
0.002 & 0.012 & 0.003 \\
0.001 & 0.003 & 0.020
\end{pmatrix}
\]

and their associated expected returns 0.043, 0.001, and 0.028, respectively. Then the mean return of asset 2 (call it \( \mu_2 \)) satisfies \( \mu_2 = 0.001 = -0.012 + 0.1689 \times \mu_1 + 0.1416 \times \mu_3 \), where \( \mu_1 = 0.043 \) and \( \mu_3 = 0.028 \) are the mean returns of assets 1 and 3, respectively. For a given expected return of 3\%, the variance minimizing allocation when only assets 1 and 3 are considered is 0.1333 and 0.8667, with a resulting standard deviation of portfolio return equal to 0.1243. On the other hand, with the addition of asset 2, and for the same level of expected return of 3\%, the variance minimizing strategy consists of the weights 0.5410, 0.2265, and 0.2326, for assets 1, 2, and 3, respectively. The standard deviation of the return on this portfolio is smaller, namely, 0.0773, despite the negative alpha of the additional asset 2.

At this point, one might be tempted to refute the above counter-example via the following argument: First, imply, correctly, that the example is compatible with a risk-aversion
coefficient $\gamma = 19.6$, which results in a zero-beta rate $\eta = -8.7\%$. Then, argue that the counter-example, in fact, is associated with $\eta = 0$, which, thanks to the regression specification

$$(r - \eta i_N) = \alpha + \beta (R - \eta i_K) + \epsilon$$

implies the counter-example is consistent with an expected portfolio return of $4.7\%$, not $3\%$. Note that the specification above is tested on the basis of the means equality

$$\mu_r - \eta i_N = \beta (\mu_R - \eta i_K).$$

However, the above equality must be valid for all values of $\eta$ for spanning, our main focus here, and can be re-expressed as

$$\mu_R - \beta \mu_R + \eta (i_N - \beta i_K) = 0,$$

thus leading to the necessary and sufficient conditions:

$$\mu_r - \beta \mu_R = 0$$

$$i_N - \beta i_K = 0.$$

The second condition above is the second of (3) and the first maps with the first of (3), given the regression specification (4). But these are not the spanning conditions to be satisfied in the presence of short-sales constraints. The (null) spanning condition (see (6)-(7) below) is:

$$\mu_r - \beta (\mu_R - \eta i_K + \delta_K) - \eta i_N \leq 0,$$

The left-hand of which in our counter-example translates to

$$0.001 - (0.1689,0.1416) \times ((0.043,0.028)' + 0.0873 \times i_2 + 0.028) + 0.0873 = 0.05 > 0,$$

where $i_2 = (1,1)'$ and $0_2 = (0,0)'$, which shows that asset 2 does indeed enter in the three-asset portfolio with expected return of $3\%$ (no spanning.)
The above counter-example is, in fact, not completely surprising as hypothesis (5) does not account for any covariance information the way (3) does, via the second set involving the matrix $\beta$. In other words, both EGB and TMMU only test a necessary ($\alpha \leq 0$), but not sufficient, condition. As a result, when they fail to reject $\alpha \leq 0$, they may conclude incorrectly in favor of spanning when, in fact, spanning may not be present. As we will show shortly, spanning is rejected based on the necessary and sufficient condition $\alpha + \eta(\beta i_K - i_N) \leq 0$ [see (13) below and (15) on p. 727 of DNW]. In fact, to test (5), both EGB and TMMU refer to the Wald statistic as expressed in DNW but do not account for the necessary and sufficient conditions in their implementation. DNW arrive at their expression (15) via an argument involving stochastic discount factors for asset pricing. We show below how the same result can be obtained through a more direct portfolio optimization approach.

The Karush-Kuhn-Tucker (KKT) conditions to optimize (2) under short-sales constraints are:

$$(\mu_R, \mu_r) - \gamma \left( \frac{\sum_{rr} \sum_{Rr}}{\sum_{rr}} \right) \left( \omega_R^*, \omega_r^* \right) - \eta \left( i_K - i_N \right) + \delta = 0,$$

where the Lagrange multiplier $\eta$ is a scalar with no sign restriction while the $(K+N)$-dimensional vector $\delta$, rewritten as $\left( \delta_K, \delta_N \right)$, is non-negative and satisfies the complementarity slackness conditions

$$\delta_{K,i} \omega_{R,i} = 0, \text{ for } 1 \leq i \leq K, \text{ and } \delta_{N,i} \omega_{r,i} = 0, \text{ for } 1 \leq i \leq N.$$

If the funds in the plan span, then the optimal mean-variance allocation $(\omega_R^*, \omega_r^*)$ is such that $\omega_r^* = 0_N$. As a result, the KKT conditions become:
\[ \mu_R - \gamma \Sigma_{RR} \omega^*_R - \eta i_K + \delta_K = 0 \]  
\[ \mu_r - \gamma \Sigma_{rR} \omega^*_R - \eta i_N + \delta_N = 0 \]  
\[ \delta_{K,i} \omega^*_{R,i} = 0, \text{ for } 1 \leq i \leq K; \]
\[ \omega^*_R \geq 0_K, \quad \delta_K \geq 0_K \]
\[ \omega^*_R \cdot i_K = 1 \]

Under specification (4), we can rewrite (7) as
\[ \alpha + \beta \mu_R - \gamma \Sigma_{rR} \omega^*_R - \eta i_N + \delta_N = 0. \]  

From (6), we have (assuming \( \Sigma_{RR} \) positive definite)
\[ \omega^*_R = \frac{1}{\gamma} \Sigma^{-1}_{RR} (\mu_R - \eta i_K + \delta_K). \]  

Note that \( \delta_{K,i} = 0 \) for \( \omega^*_{R,i} > 0 \). For notational simplicity, and to avoid introducing an additional index (superscript) as in DNW, we may assume that \( \omega^*_R > 0 \); i.e., for all subscripts. As a result, the optimal allocation \( \omega^*_{R,i} > 0 \) can now be written as
\[ \omega^*_R = \frac{1}{\gamma} \Sigma^{-1}_{RR} (\mu_R - \eta i_K). \]  

Substituting (10) back in (8), we get
\[ \alpha + \beta \mu_R - \Sigma_{rR} \Sigma^{-1}_{RR} (\mu_R - \eta i_K) - \eta i_N + \delta_N = 0. \]  

Note that by the linear regression specification (4) above, we have \( \beta = \Sigma_{rR} \Sigma^{-1}_{RR} \) and thus (11) can now be re-written as
\[ \alpha + \beta \mu_R - \beta (\mu_R - \eta i_K) - \eta i_N + \delta_N = 0, \]
which becomes
\[ \alpha + \eta(\beta i_K - i_N) + \delta_N = 0. \quad (12) \]

Since \( \delta_N \geq 0 \), (12) implies the necessary condition

\[ \alpha + \eta(\beta i_K - i_N) \leq 0. \quad (13) \]

We should note that the assumption of optimal weights in the funds being strictly positive (\( \omega_R^* > 0 \)) can be relaxed. For a given risk tolerance parameter \( \gamma \), denote by \( \mathcal{B} \) the set of indices out of the \( K \) funds in the plan so that for the spanning solution \((\omega_R^*, 0_N)\) satisfying the KKT conditions (6)-(7), we have \( \omega_{R,i}^* = 0 \) for \( i \in \mathcal{B} \). Also, let \( \mathcal{B}^c \) be the complement of \( \mathcal{B} \) in \( \{1, 2, ..., K\} \). Then we can easily show, similarly to (13), that spanning is equivalent to

\[ \alpha + \tilde{\beta}(\mu_R - \eta i_K) \leq +\eta(\beta i_K - i_N) \leq 0, \]

where \( \tilde{\beta} \equiv \Sigma_{r\mathcal{B},B} \times \Sigma^{-1}_{B,RR} \), with \( \Sigma_{r\mathcal{B},B} \) and \( \Sigma^{-1}_{B,RR} \) consisting of the columns in \( \mathcal{B} \) and rows in \( \mathcal{B} \) of, respectively, the matrices \( \Sigma_{rB} \) and \( \Sigma^{-1}_{RR} \). The difference between the above expression and (13) is the presence of \( \tilde{\beta}(\mu_R - \eta i_K) \), where it vanishes in (13). One can make the case that \( \tilde{\beta}(\mu_R - \eta i_K) \) is negligible as the inequality above has to apply to all risk factors. As a result, the binding indices would have to be the same for all risk parameters, which for a large number of funds in a plan, would likely mean that \( \mathcal{B} \) is empty. One could further validate this argument through Monte Carlo simulation.

Condition (13) is also sufficient for (12) since \( \delta_N \geq 0 \) is uniquely characterized by (12).

In other words, if (13) is true, then setting

\[ \delta_N = -[\alpha + \eta(\beta i_K - i_N)] \]
will make (12) true (a tautology). Note that (13) is the same as (15) on page 727 in DNW once we divide through our expression by \( \eta \), which is the inverse of their \( \upsilon \). Finally, multiplying both sides of (9) by \( \omega_R^* \) and using \( \omega_R^* \cdot i_K = 1 \) and \( \omega_R^* \cdot \delta_K = 0 \) yields

\[
\omega_R^* \cdot \mu_R - \gamma \omega_R^* \Sigma_{RR} \omega_R^* - \eta = 0.
\]

With the expression (9) above for \( \omega_R^* \), the latter equation becomes

\[
\frac{1}{\gamma} (\mu_R - \eta i_K) \Sigma_{RR}^{-1} \eta i_K - \eta = 0.
\]

Assuming \( \gamma < \mu_R \Sigma_{RR}^{-1} i_K \), or equivalently \( \eta > 0 \), an implicit assumption in DNW, see (14) and (15) below, the solution to the last equation is the Lagrange multiplier \( \eta \) for the portfolio allocation that satisfies:

\[
\eta = \frac{\mu_R \Sigma_{RR}^{-1} i_K - \gamma}{i_K \Sigma_{RR}^{-1} i_K}.
\]  

(14)

2.3 Mean-variance spanning tests: Implementation challenges

For ease of comparison with DNW, we multiply through (13) by \( \upsilon = 1/\eta > 0 \) to rewrite it as

\[
\upsilon \alpha + \beta i_K - i_N \leq 0.
\]  

(15)

For spanning, (15) has to be satisfied for the entire range of values for the mean discount factor \( \upsilon \). DNW suggest (pp. 729-730) that it is enough to jointly test

\[
1 \alpha + \beta i_K - i_N \leq 0
\]

(16)

\[
\upsilon_{\min} \alpha + \beta i_K - i_N \leq 0,
\]

---

9 We again recall here that we do not use the equivalent of their superscript \( (\upsilon) \) for notational simplicity, but we similarly refer here to the dimensions of \( \alpha \) and \( B \) associated with assets with non-zero, i.e. positive, allocation.
where 1 is the upper bound on the values of $v$ and their lower bound is $v_{\text{min}} = \frac{1}{E[R^{GMV}]}$, with $E[R^{GMV}]$ being the mean return of the global minimum variance portfolio.

The inequalities above restrict linear transformations on the $N$ elements of $\alpha$ and the $N \times K$ elements of $\beta$. Following Kodde and Palm (1986), as suggested by DNW, a Wald statistic can be used to test the inequalities in (16), namely

$$\xi = \min_{\gamma \geq 0} (\bar{y} - \gamma)^T \Sigma^{-1} (\bar{y} - \gamma),$$

where

$$\bar{y} = \begin{pmatrix} -\hat{\alpha} - \hat{\beta} \times i_K + i_N \\ -\frac{1}{1+\mu} \hat{\alpha} - \hat{\beta} \times i_K + i_N \end{pmatrix},$$

with $\mu = E[R^{GMV}] - 1,

$$\Sigma = \begin{pmatrix} -I_N & -A \\ \frac{1}{1+\mu} I_N & -A \end{pmatrix} \Omega \begin{pmatrix} -I_N & -A \\ \frac{1}{1+\mu} I_N & -A \end{pmatrix}',$$

with $I_N$ defined as the $N \times N$ identity matrix, $A$ as the Kronecker product $I_N \otimes i_K'$, and $\Omega$ as the $(N + NK) \times (N + NK)$ covariance matrix between the multivariate intercept $\alpha$ and the loading matrix $\beta$ in the multivariate regression (4). Using standard notation, $\hat{\alpha}$ and $\hat{\beta}$ refer to estimates of $\alpha$ and $\beta$, respectively.$^{10}$ Note that estimates for $\mu$ are typically two orders of magnitude smaller than 1. Therefore, when they are indeed very small for global minimum variance portfolios, as typically occurs, the first $N$ rows (resp. columns) in the matrix pre- (resp. post) multiplying $\Omega$ are almost identical to the latter $N$, making $\tilde{\Sigma}$ nearly singular and resulting frequently in non-computable inverses, as we experienced with many 401(k) plans. As a result, we instead appeal

$^{10}$ As a reminder, $\alpha$ and $\beta$ are vectors of dimension $N \times 1$ and $NK \times 1$, respectively.
to the fact that in the presence of a risk-free rate, say $r$, there is only one mean stochastic discount factor, namely $\frac{1}{1+r}$. Consequently, instead of the two sets of inequalities in (16), we only need to deal with one set in (15) with $\nu = \frac{1}{1+r}$, and for (17), we now have:

$$
\bar{\gamma} = -\frac{1}{1+r} \alpha - \hat{\beta} \times \bar{i}_K + \bar{i}_N
$$

$$
\bar{\Sigma} = \left(\frac{1}{1+r} \bar{I}_N - A\right) \Omega \left(\frac{1}{1+r} \bar{I}_N - A\right)'.
$$

(19)

2.4 Style Perspective

An issue that may affect regression-based mean-variance testing in a significant way is that of aggregation, such as merging all international benchmark indexes into a single one, or similarly grouping mid- and small- capitalization funds, as done by EGB and TMMU. Consequently, we perform spanning tests using (i) the same benchmark indices as those of EGB and TMMU and (ii) a different, style-driven benchmark set. This benchmark choice is partly predicated on the preponderance of return-style analysis for fund performance evaluation and also on behavioral attributes of the average investor. Specifically, individual participants are unlikely to engage in optimizing their allocations in their 401(k) plans, or anywhere else for that matter. More likely, during initial portfolio setup they will chase winners or be influenced by fund style descriptions (see, e.g., Karceski, 2002; Barberis and Shleifer, 2003; and Wahal and Yavuz, 2013). Fund descriptions, however, are often vague because of their limitations to just a handful of words, such as growth, value, balanced, income, etc. In addition, their managers tend to regularly readjust their holdings, thus resulting in fund returns that do not necessarily fit the style description provided in the first place.
The style analysis approach of Sharpe (1992) is widely used in practice for the classification and performance evaluation of mutual funds (see, e.g., Chan, Chen, and Lakonishok, 2002; Brown and Goetzmann 1997; and Ter Horst, Nijman, and De Roon, 2004). It regresses a fund’s return on cash and a set of asset classes. Their corresponding regression coefficients are constrained to be non-negative and to sum up to one. They thus represent the effective style mix in the portfolio and the intercept has a Jensen’s alpha interpretation for performance evaluation.

3. Qualified Default Investment Alternative (QDIA) Options

In this section we highlight some concerns regarding the popular QDIA default option of Target-Date-Funds (TDFs) and introduce our own alternative that is QDIA-compliant. Investors largely share many misconceptions regarding TDFs, such as whether these funds offer guaranteed income. For example, Surz and Israelsen (2007) discuss how TDFs should be assessed and find that their risk-adjusted performance falls short while Spitzer and Singh (2008) use simulation to show that TDFs have a higher shortfall risk than a constant equal-allocation between stocks and bonds. In a similar vein, with a focus on shortfall risk, Scott, Sharpe, and Watson (2009) show that glide-path strategies underperform constant mix strategies. They observe that the former, in contrast to the latter, tend to lock-in poor early returns and thus decrease the likelihood of portfolio recovery should returns improve. From a different perspective, Sandhya (2011) finds that TDFs are also subject to agency problems as mutual fund companies use low quality funds to create TDFs. Additionally, as TDFs are funds of funds, their fees can be significant. In an empirical study, Elton, Gruber, De Souza, and Blake (2015) provide evidence that these fees are somewhat offset by the lower fees of the funds into which the TDFs
are invested. On the whole, however, they show that the resulting alphas are generally lower than alternatives for any given fund family.\footnote{Additional issues associated with TDFs are further presented in the review article of Spitzer and Singh (2012).}

When employees do not opt for the default TDF, if offered, there is evidence that they vastly follow an even allocation and they also prefer to deal with fewer investment choices, as noted earlier. Therefore, our main goal now is to help them identify these choices within their plans in a manner that is prudent and efficient, thus complying with the QDIA safe harbor provision. As also remarked earlier, we do not advocate specific amounts to invest in the particular funds selected and the participants may choose to invest evenly in the subset we identify. This so-called “1/N” strategy has the advantage of being intuitive to the average investor while remaining competitive relative to more sophisticated alternatives (see, e.g., De Miguel, Garlappi, and Uppal, 2009; Pflug, Alois Pichler, and Wozabal, 2012). Furthermore, we also show that the subset of funds tend to span similarly to, and often better than, the fuller set of a given plan.

The QDIA-based fund selection approach we prescribe is mainly driven by portfolio efficiency, specifically, mean-variance spanning. Our approach is also motivated by the strong empirical performances of low-volatility and low-beta portfolios (see, e.g., Karceski 2002; Ang, Hodrick, Xing, and Zhang, 2009; Baker, Bradley, and Wurgler, 2011; Frazzini and Pedersen, 2014) which was, in fact, already anticipated all the way back to Black (1972), especially within the context of restricted borrowing that is characteristic of mutual funds. To determine the QDIA subset of a given plan, we first partition the latter into higher risk equity-based funds and the remainder, which typically includes stable value, money market, government bonds, fixed
We then proceed to perform a variance minimization on the former group and only keep equity funds that are part of the optimal solution. Next, we combine these with the safer funds, which were excluded from the variance minimization altogether. This approach has the same flavor as the classical two-fund theorem; however, it differs from it since the two funds (or sets of funds, with a cardinality typically not exceeding three) are not necessarily on the efficient frontier (e.g., minimum variance portfolio in the presence of a risk-free rate). Lastly, a spanning test is performed on the combined set of funds to determine whether they can span returns of index benchmarks.

4. Data

Our primary dataset is provided by Brightscope, Inc., an independent information provider of retirement plan ratings and investment analytics to plan participants, sponsors, asset managers, and advisors. These information services are important because they provide individuals with resources to better educate themselves about investment decisions in a somewhat opaque retirement plan market. Brightscope provides data on retirement plan quality such as overall plan rating, company matching generosity, and plan costs, which ultimately may help potential employees to evaluate retirement benefits before joining a company.

Brightscope’s proprietary dataset currently contains information on over 55,000 defined contribution (DC) plans, such as 401(k) and 403(b) plans. The specific dataset provided to us is a cross-sectional snapshot of plans at the end of 2007, and it contains over 25,000 DC plans. Items included in the data that are important to our analysis include: plan menu investment fund

---

12 Partitioning of funds is based on Morningstar categories. We first determine all unique Morningstar categories in our data and then identify categories that represent equity based funds.
options, plan size (net assets), individual fund balances, fund expense ratios, administrative costs, and plan sponsor and service provider information.

In this study we focus on companies’ primary DC plans as identified by using the Department of Labor codes, thus we eliminate any supplementary plans offered by the same plan sponsor.\textsuperscript{13} This initial sample includes 17,386 DC plans. We further require full return data availability for every mutual fund within a plan in the 2004-2008 period from either CRSP Mutual Fund Database or Morningstar Direct Database. Because of this return data constraint our final sample used in the analysis consists of 7,975 DC plans. Despite the loss of a large number of plans in the original sample, our final sample is significantly larger and richer than previous studies that have analyzed retirement plan menu efficiency. For instance, EGB analyze a relatively small sample of 417 plans and while TMMU analyze a larger sample of 1,003, all of which are administered by one of the top mutual fund companies in the industry – Vanguard.\textsuperscript{14} Our data covers both publicly traded and private companies of all sizes which hire many types of third-party administrators (TPA), the financial institutions in charge of designing and servicing the retirement plan.\textsuperscript{15}

Table 1, Panel A provides comparative descriptive statistics for the initial and the final sample. The average plan in our initial sample has $22.1 million in total net assets and offers

\textsuperscript{13} We do this because we do not typically have complete data for all of a company’s plans; thus, we simply analyze a company’s primary, or largest, DC plan.

\textsuperscript{14} EGB uses a sample provided by Moody’s Investors Services that collects survey data from for-profit firms. The sample used in TMMU is supplied by Vanguard, a company that is well known to provide low cost and well diversified portfolios with heavy preference to index funds.

\textsuperscript{15} We categorize TPAs in our sample into one of seven categories: mutual fund families, large/small (greater than or less than $50 billion in assets) commercial banks, insurance companies, asset management advisory companies, investment banks, and 401(k) services companies.
In contrast, the average plan in our final sample is larger than the average plan in the initial sample with $31.8 million in total assets and contains 18.2 fund options in its investment menu with a balance of $1.9 million per fund. For the majority of our sample we can also identify the TPA. TPA categories used in our analysis can be found in Panel B of Table 1, along with plan size characteristics across these different categories. Mutual fund families represent a heavy majority of the subsample where we can identify the TPA. This is consistent with the overall retirement plan market where mutual fund companies hold a majority of the market share. In our sample, plans administered by investment banks, large commercial banks, asset management advisory firms, and mutual fund companies are considerably larger than plans administered by small/regional commercial banks, 401(k) services companies, and insurance firms. This is not particularly surprising since the clientele of the latter groups are most likely to be smaller firms with fewer participants and lower retirement plan balances.

In Table 2, we provide more details on the types of funds offered in investment menus of plans in the sample. As summarized in Panel A, an average retirement plan in our sample offers 12.3 domestic equity funds, 1.8 domestic bond funds, 1.9 international funds and 0.6 low risk investments such as money market funds, stable value funds, guaranteed investment contracts, or annuities (MSGA). Not surprisingly, almost all retirement plans include at least one domestic equity fund while 97% of plans offer at least one domestic bond fund and one international fund in their investment menus. Additionally, about 60% of plan menus contain at least one MSGA,

---

16 The average plan size in our initial sample is very comparable to the average plan size ($25.2 million) in EBRI’s 2007 dataset, the largest provider of information on 401(k) plan.

17 One exception is the insurance company group. Insurance firms are underrepresented in our sample due to the common use of proprietary funds that do not exist in our return data sources (CRSP or Morningstar).

18 In our analysis we only focus on mutual funds in the plan and exclude company stock and all MSGA options.
while 4% of plans in the sample also include company stock as one of investment choices.\textsuperscript{19} Plan participants in our sample, on average, direct 68\% of their retirement wealth to domestic equity funds, 9\% to domestic bond funds, 14\% to international funds and 9\% to MSGA options. When company stock is offered in the DC plan, the average plan participant also invests 13\% of plan assets in the stock.\textsuperscript{20} Further, Panel B reports the average number of unique Lipper Investment Objective categories covered by plans in the final sample. On average, 13 different objective categories are represented in plan menus.

For purposes of spanning and style analysis we use two sets of benchmark funds. We obtain all return data for these benchmarks from DataStream. Table 3 lists both sets of benchmark indexes and provides descriptive statistics on monthly returns over the analysis period. Despite some changes in management company affiliations, the first set of benchmarks are identical to sets used by the related spanning literature, including EGB and TMMU. In the first set, the Barclays Capital Aggregate Bond Index, Credit Suisse High Yield Bond Fund, and Citigroup World Government Bond Non-US$ Index capture returns of fixed income securities; the Russell 1000 Growth, Russell 1000 Value, Russell 2000 Growth, and Russell 2000 Value indices capture returns of large-, mid- and small-cap equities; and the MSCI EAFE Index provides international exposure. In the second set, each of investment categories are represented with more benchmark indices. This set is well used in the style analysis literature, including Sharpe (1992). In this set, fixed income benchmarks are Barclays Government Intermediate, Barclays United States Aggregate Long Government/Credit, Barclays Investment Grade: Corporates, Barclays US Agency Fixed Rate MBS, Citigroup World Government Bond Index

\textsuperscript{19} Interestingly, retirement participants of about 92\% of plans in our sample have also borrowed against their retirement wealth, on average 8.3\% of their plan balance.

\textsuperscript{20} Although this figure seems quite high, often there are special incentives in investing in company stock for employees.
World 5+ Year Non-USD; equity benchmarks are Standard and Poor’s 500/ Citigroup - Value; Standard and Poor’s Midcap 400/ Citigroup - Value; Standard and Poor’s Midcap 400/ Citigroup - Growth; Standard and Poor’s Smallcap 600/ Citigroup- Value; Standard and Poor’s Smallcap 600/ Citigroup – Growth; the international benchmarks are MSCI Europe, MSCI Pacific, and S&P IFCI Emerging Market Index. This set then differs from that used by EGB and TMMU by disentangling the small and mid-cap groups and by differentiating between the international regions. As summarized in Table 3, the average monthly returns on most benchmark indices were negative in the 2004-2008 period due to the financial crisis of 2008.21

5. QDIA Implementation and Results

Table 2, Panel A shows that the equity funds category that results from our partitioning constitutes a majority of a plan menu’s investment options. We used benchmark indices of two types: (i) the same eight (or their close equivalent since their studies) as in EGB and TMMU, and (ii) an alternative with 13 indices in Sharpe (1992) as described in the last section. In addition, we performed spanning tests where the fund returns are re-expressed in terms of those of all their principal components [see Lai and Xing (2008), p. 41-44, in order to allay the collinearity between funds in plans].

Relative to the same benchmarks used by EGB and TMMU, Table 4 shows that we find that 46% of plans span. On the other hand, with QDIA-based funds, spanning occurs around 49% of the time. Should our alternative benchmark of 13 indices be used instead, spanning occurs at much lower rates: 28% for all funds and 31% for QDIA-based funds, as reported in Table 5.

21 Average monthly returns are positive for all indices in both sets if year 2008 returns are removed from the descriptive statistics. Note that a positive arithmetic average monthly return does not imply that the corresponding average compounded return is positive.
The first important takeaway from these results is that, according to our analysis, spanning rates are much lower than the 53% and the 97% rates found in EGB and TMMU, respectively. As we noted earlier, this can easily be explained by the fact that both sets of authors only test a necessary, but not sufficient condition. As a result, when they fail to reject the null, they may incorrectly conclude in favor of spanning. The second important finding from these results is that the spanning rates markedly differ according to the two benchmark sets. However, as shown in Table 6, the proportion of spanning increases dramatically for both sets of funds (full plan and QDIA set) and both benchmarks when the principal components of the funds are used. These results clearly highlight the significance of the correlation among fund returns. We also note that our selection approach tends to reduce the number of choices by almost two-thirds, with a reduction from a mean (median) of 18.2 (17) funds to a mean of 6.2 (5) in a typical plan.

Finally, the most important takeaway from Tables 4 through 6 is that limiting the plan menu to QDIA-based funds does not impair the spanning opportunities offered by the larger menu options in the plan. However, given the above mentioned issues with offering the full menu to participants, suggestion of QDIA-based funds could greatly alleviate the burden of the investment decision for the participants. If the TPA and/or the plan sponsor would choose to offer guidance on the set of funds in the plan menu that are part of the minimum-variance optimized portfolio, together with the conservative funds (i.e. QDIA-based funds), then a plan participant is more likely to create an investment portfolio that spans a set of benchmark indices. In this fashion, fiduciaries have an objective basis for their fund recommendations, which may otherwise be viewed under a cloud of suspicion.22 Furthermore, if desired, participants can choose to apply the "1/N" strategy to this QDIA-fund subset, as it also has appealing

mathematical properties (see, e.g., De Miguel, Garlappi, and Uppal, 2009; Pflug, Alois Pichler, and Wozabal, 2012). Therefore, we argue that this is a potential and valid third option for retirement plan participants, who would otherwise either choose as they see fit from the entire 401(k) plan menu or follow the current default in the plan.

Our approach is in contrast to TMMU who further assess the impact of the deviation of the risk-adjusted performance across all participants relative to the mean-variance efficient portfolios and conclude that while plans may be spanning, individual portfolio construction is overwhelmingly inefficient. Their recommendation is then to support strategies targeting behavioral change including improved default strategies and educational programs. However, behavioral change is difficult to achieve (Iyengar and Kamenica, 2010) and financial education has a very short “shelf life”, thus severely limiting its efficacy (Fernandes, Lynch, and Netemeyer, 2014). In addition, the estimation issues that arise in the course of mean-variance optimization have led some to question the practicality of mean-variance efficient portfolios, leading some to further wonder whether the “1/N” strategy could really be a viable strategy [see De Miguel, Garlappi, and Uppal (2009) and references therein]. De Miguel, Garlappi, and Uppal (2009) compare Sharpe ratios for strategies on an out-of-sample basis and find that gains from optimal diversification can be more than offset by estimation error. They find that for parameters calibrated to U.S. stock-market data, for a portfolio of 25 assets, for example, the required estimation window is 3,000 months. Though they refrain from explicitly advocating the “1/N” heuristic, they strongly urge using it as a benchmark for comparison purposes.

6. Conclusion

Regression-based mean-variance spanning tests are ubiquitous in empirical finance. Our paper centers on challenges these tests present when implemented in the context of short-sales
constraints. We show that the standard Wald testing methodology is prone to numerical instability and offer a remedy that rests on the uniqueness of the mean stochastic discount factor when a risk-free asset is present.

We illustrate specific issues in household finance, where we focus on employer-sponsored 401(k) plans. These play an important role in the retirement of tens of millions of Americans and affect financial markets in a significant fashion. The main aspect of their effectiveness has centered on (i) whether they include funds that would enable participants to invest in well-diversified portfolios with returns after fees that are not inferior to what could be offered more widely (spanning), and (ii) whether plan participants would indeed take advantage of these opportunities. In this paper, we find that spanning rates are low and vary significantly based on the benchmark used, in contrast with the extant literature, which we show to be faulty. However, when principal component returns of plan funds are used instead, then the spanning rates increase dramatically, which points to the significant effect of collinearity of the independent variables used for testing mean-variance spanning.

Plan participants have two options when making investment decisions in their 401(k) plan: devising their own specific allocations for the full plan menu, or relying on the default option fund chosen by the plan administrators, which are usually target-date funds. Given the well-documented challenges for individual investors when creating efficient portfolios using the full menu and the many performance, together with design and agency issues of target-date funds, we recommend a third option: Plan sponsors or administrators should steer participants toward funds that are included in a set consisting of (i) those selected for the minimum-variance portfolio of equity funds in the plan and (ii) any of the bond-like funds in the plan. This third option would make the challenge of individual investing less daunting for participants and yet,
would fit the U.S. regulatory provision for diversification and equity/fixed-income mix, as contained in the safe harbor rule for a Qualified Default Investment Alternative (QDIA). Critical to the reasonableness of this procedure, we also confirm that limiting the plan menu to the recommended subset does not result in inferior spanning relative to the full plan funds.

We should stress that our practical approach to determining a fund strategy that falls within the confines of QDIA is different than the classical two-fund theorem set-up since the minimum-variance portfolio is not on the efficient frontier in the presence of a risk-free asset. However, it is consistent with the empirical evidence supporting the adoption of low-volatility strategies. Our results imply that plan administrators can take a more active role in providing guidance to plan participants in their investments decisions. Though our recommendation differs from earlier educational and default TDF strategies, it is consistent with individual subjective preference for fewer decision-making choices, allowing them to use even allocations if they so desire, and with recent studies suggesting that some efficient plans have only a handful of options.
References


Table 1. Plan-Level Descriptive Statistics

Panel A reports descriptive statistics for the size of plans and fund balances in the overall sample provided by BrightScope, Inc. These only include 401(k) plans that are deemed a company’s primary defined contribution plan by the Department of Labor. Since data limitations require us to eliminate a large number of plans from the main analysis, Panel A also reports descriptive statistics for the final sample. Panel B reports descriptive statistics on plan size by the type of company administering the 401(k) plan, i.e. third-party administrator (TPA).

Panel A. Plan and Fund Balance Size

<table>
<thead>
<tr>
<th>Beginning Sample</th>
<th>N = 17,386</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>10th percentile</th>
<th>25th percentile</th>
<th>Median</th>
<th>75th percentile</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan Size (000’s)</td>
<td>22,100</td>
<td>111,000</td>
<td>1,157</td>
<td>2,534</td>
<td>5,894</td>
<td>13,700</td>
<td>34,200</td>
<td></td>
</tr>
<tr>
<td>Average Fund Size (000’s)</td>
<td>1,228</td>
<td>8,456</td>
<td>63</td>
<td>133</td>
<td>313</td>
<td>734</td>
<td>1,865</td>
<td></td>
</tr>
<tr>
<td>Number of Fund Options</td>
<td>22.13</td>
<td>15</td>
<td>11</td>
<td>15</td>
<td>19</td>
<td>26</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final Sample</th>
<th>N = 7,975</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>10th percentile</th>
<th>25th percentile</th>
<th>Median</th>
<th>75th percentile</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan Size (000’s)</td>
<td>31,827</td>
<td>136,058</td>
<td>1,141</td>
<td>2,684</td>
<td>6,697</td>
<td>17,172</td>
<td>52,676</td>
<td></td>
</tr>
<tr>
<td>Average Fund Size (000’s)</td>
<td>1,891</td>
<td>8,564</td>
<td>77</td>
<td>180</td>
<td>426</td>
<td>1,059</td>
<td>3,110</td>
<td></td>
</tr>
<tr>
<td>Number of Fund Options</td>
<td>18.21</td>
<td>6.78</td>
<td>10</td>
<td>13</td>
<td>17</td>
<td>23</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Panel B. Plan Size by Category

<table>
<thead>
<tr>
<th>Plan Size (000’s)</th>
<th>Number of Plans</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual Fund Families</td>
<td>2,340</td>
<td>60,089</td>
<td>272,790</td>
<td>37</td>
<td>13,448</td>
<td>9,641,714</td>
</tr>
<tr>
<td>Asset Management Advisory</td>
<td>519</td>
<td>40,030</td>
<td>204,915</td>
<td>49</td>
<td>6,557</td>
<td>2,602,408</td>
</tr>
<tr>
<td>Investment Banks</td>
<td>183</td>
<td>91,400</td>
<td>355,811</td>
<td>14</td>
<td>14,628</td>
<td>4,040,556</td>
</tr>
<tr>
<td>Large Commercial Banks</td>
<td>725</td>
<td>41,197</td>
<td>139,039</td>
<td>70</td>
<td>8,508</td>
<td>2,261,397</td>
</tr>
<tr>
<td>Small/Regional Comm. Banks</td>
<td>203</td>
<td>12,821</td>
<td>33,027</td>
<td>10</td>
<td>4,756</td>
<td>349,804</td>
</tr>
<tr>
<td>Insurance Firms</td>
<td>435</td>
<td>15,002</td>
<td>32,758</td>
<td>147</td>
<td>5,826</td>
<td>470,023</td>
</tr>
<tr>
<td>401(k) Services Companies</td>
<td>362</td>
<td>18,860</td>
<td>62,782</td>
<td>15</td>
<td>4,091</td>
<td>5,825,942</td>
</tr>
<tr>
<td>TPA Unknown</td>
<td>3,224</td>
<td>13,183</td>
<td>52,622</td>
<td>3</td>
<td>4,090</td>
<td>1,146,085</td>
</tr>
</tbody>
</table>
Table 2. Plan Menu Options Descriptive Statistics

The following panels report information related to the final sample used in the analysis sections. Panel A provides frequency data on the different types of funds available in the final sample’s 401(k) plan, unconditional and conditional on being offer in a plan, along with ratio of plan assets directed to the types of funds. Panel B simply reports information on the average plan’s coverage of different Lipper Investment Objective categories.

### Panel A. Fund Type Coverage

<table>
<thead>
<tr>
<th></th>
<th>Number of Fund Options (unconditional)</th>
<th>Number of Fund Options (conditional)</th>
<th>% of Plans Assets Held in (unconditional)</th>
<th>% of Plans Assets Held in (conditional)</th>
<th>% of Plans Containing at least one</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  St. Dev.</td>
<td>Mean  St. Dev.</td>
<td>Mean  St. Dev.</td>
<td>Mean  St. Dev.</td>
<td>Mean  St. Dev.</td>
</tr>
<tr>
<td>Domestic Equity Funds</td>
<td>12.36   6.01</td>
<td>12.37   6.00</td>
<td>68.18%   13.50%</td>
<td>68.24%   13.35%</td>
<td>99.92%</td>
</tr>
<tr>
<td>Domestic Bond Funds</td>
<td>1.82    1.16</td>
<td>1.88    1.13</td>
<td>8.97%    19.72%</td>
<td>9.25%    19.96%</td>
<td>96.98%</td>
</tr>
<tr>
<td>International Funds</td>
<td>1.86    1.16</td>
<td>1.92    1.13</td>
<td>14.36%   9.63%</td>
<td>14.82%   9.43%</td>
<td>96.90%</td>
</tr>
<tr>
<td>Money Market/ Stable Value/ GIC/ Annuity (MSGA)</td>
<td>0.63    0.57</td>
<td>1.06    0.29</td>
<td>8.80%    11.77%</td>
<td>14.88%   11.99%</td>
<td>59.12%</td>
</tr>
<tr>
<td>Company Stock</td>
<td>0.04    0.21</td>
<td>1.06    0.23</td>
<td>0.51%    3.95%</td>
<td>12.97%   15.36%</td>
<td>3.93%</td>
</tr>
<tr>
<td>Participant Loans</td>
<td></td>
<td></td>
<td>7.67%    9.32%</td>
<td>8.32%    9.43%</td>
<td>92.17%</td>
</tr>
</tbody>
</table>

### Panel B. Lipper Objective Category Coverage

Unique Lipper Category Coverage by Plan

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard Deviation</th>
<th>10th percentile</th>
<th>25th percentile</th>
<th>Median</th>
<th>75th percentile</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.21</td>
<td>6.03</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>17</td>
<td>22</td>
</tr>
</tbody>
</table>
Table 3. Benchmark Indices

This table lists the set of benchmark indices used in spanning and style analysis, and it provides descriptive statistics on net of fees monthly returns for the analysis period (2004-2008). Set 1 indices are consistent with those used in Elton, Gruber, and Blake (2006) and Tang, Mitchell, Motolla, and Utkus (2010) while Set 2 indices are more aligned with those used in Sharpe (1992).

<table>
<thead>
<tr>
<th>Monthly Returns</th>
<th>Mean</th>
<th>St.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSCI EAFE Index</td>
<td>0.28%</td>
<td>4.76%</td>
</tr>
<tr>
<td>Barclays Capital Aggregate Bond Index</td>
<td>0.37%</td>
<td>1.34%</td>
</tr>
<tr>
<td>Credit Suisse High Yield Bond Fund</td>
<td>-0.04%</td>
<td>2.96%</td>
</tr>
<tr>
<td>Citigroup World Government Bond Non-US$ Index</td>
<td>0.50%</td>
<td>2.36%</td>
</tr>
<tr>
<td>Russell 1000 Growth</td>
<td>-0.22%</td>
<td>4.09%</td>
</tr>
<tr>
<td>Russell 1000 Value</td>
<td>0.01%</td>
<td>3.78%</td>
</tr>
<tr>
<td>Russell 2000 Growth</td>
<td>-0.05%</td>
<td>5.64%</td>
</tr>
<tr>
<td>Russell 2000 Value</td>
<td>0.14%</td>
<td>4.99%</td>
</tr>
<tr>
<td><strong>Set 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barclays Government Intermediate</td>
<td>0.06%</td>
<td>0.93%</td>
</tr>
<tr>
<td>Barclays United States Aggregate Long Government</td>
<td>-0.02%</td>
<td>2.89%</td>
</tr>
<tr>
<td>Barclays Investment Grade - Corporates</td>
<td>-0.32%</td>
<td>2.10%</td>
</tr>
<tr>
<td>Barclays US Agency Fixed Rate MBS</td>
<td>0.44%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Citigroup World Govn’t Bond Index World 5+Yr Non-US$</td>
<td>0.57%</td>
<td>2.57%</td>
</tr>
<tr>
<td>Standard and Poor's 500 / Citigroup - Value</td>
<td>-0.28%</td>
<td>3.90%</td>
</tr>
<tr>
<td>Standard and Poor's Mid-cap 400 / Citigroup - Value</td>
<td>-0.03%</td>
<td>4.75%</td>
</tr>
<tr>
<td>Standard and Poor's Mid-cap 400 / Citigroup - Growth</td>
<td>0.01%</td>
<td>4.90%</td>
</tr>
<tr>
<td>Standard and Poor's Small-cap 600 / Citigroup - Value</td>
<td>0.05%</td>
<td>4.95%</td>
</tr>
<tr>
<td>Standard and Poor's Small-cap 600 / Citigroup - Growth</td>
<td>0.14%</td>
<td>5.01%</td>
</tr>
<tr>
<td>MSCI Europe Index</td>
<td>0.03%</td>
<td>4.99%</td>
</tr>
<tr>
<td>MSCI Pacific Index</td>
<td>0.13%</td>
<td>4.86%</td>
</tr>
<tr>
<td>S&amp;P IFCI Emerging Market Index</td>
<td>0.77%</td>
<td>7.35%</td>
</tr>
</tbody>
</table>
Table 4. Spanning Tests with 8 Benchmark Indices

This table summarizes spanning test results at 5% significance level with 8 indices: MSCI EAFE Index; Barclays Capital Aggregate Bond Index; Credit Suisse High Yield Bond Fund; Citigroup World Government Bond Non-US$ Index; Russell 1000 Growth; Russell 1000 Value; Russell 2000 Growth; Russell 2000 Value. "Fuzzy" refers to the cases where the Wald statistic falls within the upper and lower bounds of the critical value provided by Kodde and Palm (1986). For these cases the critical value was estimated via Monte-Carlo simulation following Wolak (1989) and 1,000 simulation runs. Results correspond to two variants of our data (i) without performing mean-variance optimization at plan level and allowing all funds within the plan to be included in the spanning test, and (ii) performing minimum-variance optimization on equity-only funds and keeping funds with positive weights in the optimized portfolio together with those excluded from entering the optimization. Spanning in this case is performed with the funds kept out of the optimization (typically bond-like) and those selected by the minimum-variance optimization.

<table>
<thead>
<tr>
<th>Outside of Bounds</th>
<th>Fuzzy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span</td>
<td>No Span</td>
<td>Span</td>
</tr>
<tr>
<td>All Funds</td>
<td>1,439</td>
<td>3,060</td>
</tr>
<tr>
<td></td>
<td>18.04%</td>
<td>38.37%</td>
</tr>
<tr>
<td>QDIA Funds</td>
<td>2,393</td>
<td>3,336</td>
</tr>
<tr>
<td></td>
<td>30.01%</td>
<td>41.83%</td>
</tr>
</tbody>
</table>
Table 5. Spanning Test with 13 Benchmark Indices

This table summarizes spanning test results at 5% significance level with 13 indices: US One-month T-bill; Barclays Government Intermediate; Barclays United States Aggregate Long Government / Credit; Barclays Investment Grade : Corporates; BarCap US Agency Fixed Rate MBS; Citigroup World Government Bond Index World 5 + Year Non-USD; Standard and Poor’s 500 / Citigroup - Value; Standard and Poor’s Midcap 400 / Citigroup - Value; Standard and Poor’s Midcap 400 / Citigroup - Growth; Standard and Poor’s Smallcap 600 / Citigroup - Value; Standard and Poor’s Smallcap 600 / Citigroup - Growth; MSCI Europe US Dollar; MSCI Pacific US Dollar; S&P IFCI. "Fuzzy" refers to the cases where the Wald statistic falls within the upper and lower bounds of the critical value provided by Kodde and Palm (1986). For these cases the critical value was estimated via Monte-Carlo simulation following Wolak (1989) and 1,000 simulation runs. Results correspond to two variants of our data (i) without performing mean-variance optimization at plan level and allowing all funds within the plan to be included in the spanning test, and (ii) performing minimum-variance optimization on equity-only funds and keeping funds with positive weights in the optimized portfolio together with those excluded from entering the optimization. Spanning in this case is performed with the funds kept out of the optimization (typically bond-like) and those selected by the minimum-variance optimization.

<table>
<thead>
<tr>
<th>Outside of Bounds</th>
<th>Fuzzy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Span</td>
<td>No Span</td>
</tr>
<tr>
<td>All Funds</td>
<td>512</td>
<td>4,185</td>
</tr>
<tr>
<td></td>
<td>6.42%</td>
<td>52.48%</td>
</tr>
<tr>
<td>QDIA Funds</td>
<td>791</td>
<td>4,023</td>
</tr>
<tr>
<td></td>
<td>9.92%</td>
<td>50.45%</td>
</tr>
</tbody>
</table>
Table 6. Spanning Tests Based on Principal Component Analysis

This table summarizes spanning test results at 5% significance level with principal component returns constructed from funds of the plan. In Panel A, principal components of all funds and QDIA funds of the plan are the regression variables against the 8 benchmark indices described in Table 4, while in Panel B, principal components of all funds and QDIA funds of the plan are the regression variables against the 13 benchmark indices described in Table 5.

Panel A: 8 benchmark indices and principal components of plan funds

<table>
<thead>
<tr>
<th>Outside of Bounds</th>
<th>Fuzzy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span</td>
<td>No Span</td>
<td>Span</td>
</tr>
<tr>
<td>All Funds</td>
<td>7,903</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>99.10%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QDIA Funds</td>
<td>7,847</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>98.39%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Panel B: 13 benchmark indices and principal components of plan funds

<table>
<thead>
<tr>
<th>Outside of Bounds</th>
<th>Fuzzy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span</td>
<td>No Span</td>
<td>Span</td>
</tr>
<tr>
<td>All Funds</td>
<td>7,858</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>98.53%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QDIA Funds</td>
<td>7,857</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>98.52%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>