Journal of Investment Strategies **12**(2), 1–12
DOI: 10.21314/JOIS.2023.005

Research Paper

Implementing mean–variance spanning tests with short-sales constraints

Farid AitSahlia,[1] Thomas Doellman[2] and Sabuhi Sardarli[3]

[1]Eugene F. Brigham Department of Finance, Insurance and Real Estate, Warrington College of Business, University of Florida, PO Box 117168, Gainesville, FL 32611-7168, USA; email: farid1@ufl.edu

[2]Richard A. Chaifetz School of Business, Department of Finance, Saint Louis University, 3674 Lindell Boulevard, DS 307, St. Louis, MO 63108, USA; email: thomas.doellman@slu.edu

[3]Department of Finance, Kansas State University, 2097 Business Building, 1301 Lovers Lane, Manhattan, KS 66506, USA; email: ssardarli@ksu.edu

(Received March 20, 2023; revised August 4, 2023; accepted October 2, 2023)

ABSTRACT

A set of assets is said to span the mean–variance space if the efficient frontier it generates cannot be improved upon with additional assets. Mean–variance spanning is used to determine empirically whether or not particular assets should be included in a given portfolio. Because of typical issues relating to parameter estimation in mean–variance optimization, the results of this empirical approach may differ from those of optimization, which assumes known parameters. In this paper, we show that the Wald tests used to account for short sales are prone to numerical instability. To address this, we exploit the uniqueness of the stochastic discount factor in

Corresponding author: F. AitSahlia

Print ISSN 2047-1238 | Online ISSN 2047-1246

the presence of a risk-free rate, leading to more robust tests. We also show that the purported Wald tests that have appeared in the literature on retirement plans in the United States do not correspond to mean–variance optimality and that their proper implementation leads to significantly different results.

Keywords: mean–variance spanning; constrained multivariate regression; Wald test; mutual funds; short sales; defined contribution plans.

1 INTRODUCTION

A set of assets is said to span the mean–variance space if the efficient frontier it generates cannot be improved upon with additional assets (Huberman and Kandel 1987). Mean–variance spanning has been used to, among other things, assess the benefits of international diversification, to evaluate mutual fund performance, to test linear-factor asset pricing models and to consider the relevance of cryptocurrencies as an alternative asset class (see, for example, Errunza *et al* 1999; De Roon *et al* 2001; Fama and French 2015, 2018; Petukhina *et al* 2021).

Empirical testing for mean–variance spanning has been conducted via multivariate regression since the seminal paper of Huberman and Kandel (1987). Mean–variance spanning was studied extensively by Kan and Zhou (2012), who, incidentally, point out a misprint by Huberman and Kandel (1987) relating to a crucial F-statistic, which has been repeated unquestioningly in a number of subsequent papers.

The mean–variance literature mostly assumed an absence of short sales and transaction costs, until these were addressed by De Roon *et al* (2001). In this paper we show that the implementation of the associated Wald tests is subject to numerical instability, and hence they yield unreliable results. Further, we also show that the Wald tests implemented by Elton *et al* (2006) and Tang *et al* (2010) in their evaluation of the efficiency of defined contribution retirement plans, such as 401(k) and 403(b) plans in the United States, do not correspond to the optimality conditions for mean–variance spanning and that they lead to excessive values.

The remainder of the paper is organized as follows. In Section 2 we review the methodology of mean–variance spanning with short-sales constraints and address the potential numerical instability of the related Wald tests via the uniqueness of the stochastic discount factor in the presence of a risk-free asset. In Section 3 we show how these tests are incorrectly implemented in the extant literature regarding the efficiency of defined contribution retirement plans in the United States. In Section 4 we provide an empirical illustration of the resulting discrepancy relative to a proper implementation. In Section 5 we state our conclusions.

2 A REEXAMINATION OF MEAN–VARIANCE SPANNING TESTS UNDER SHORT-SALES CONSTRAINTS

We review the regression-based mean–variance spanning test methodology when short sales are prohibited and highlight some issues related to its implementation as presented by De Roon *et al* (2001). Due to the multitude of stochastic discount factors, De Roon *et al* (2001) suggest using only the smallest and the largest mean discount factors. However, these are not observable and must be inferred. In addition, we show that following this approach leads to numerical instability. As a remedy, we instead appeal to the uniqueness of the mean discount factor in the presence of a risk-free rate, which is observable, with the former being the inverse of 1 plus the latter.

We start by defining our terminology. Between dates t and $t+1$, "return" refers to the simple net (raw) return, defined as $(P_{t+1} + I_t)/P_t - 1$, where P_t, P_{t+1} and I_t are, respectively, the asset value at times t and $t+1$ and the related income in that interval.

To assess the efficiency of K assets relative to a benchmark of N other assets (eg, index funds), we determine whether the mean–variance efficient frontier associated with K assets coincides with that generated with an augmented set of $K+N$ assets. In other words, we determine whether the K assets are "sufficient" to span the efficient frontier of the $K+N$ assets.

Let R and r be the $K \times 1$ and $N \times 1$ return vectors, respectively, of the K assets and the N benchmark indexes. Denote by μ_R and μ_r their corresponding expected return vectors. The related covariance matrixes are defined as follows.

- $\Sigma_{R,R}$ denotes the $K \times K$ covariance matrix between the K assets with returns vector R.
- $\Sigma_{r,r}$ denotes the $N \times N$ covariance matrix between the returns of the N benchmark indexes captured by vector r.
- $\Sigma_{R,r}$ denotes the $K \times N$ covariance matrix of the K asset returns with the N benchmark returns.
- $\Sigma_{r,R}$ denotes the $N \times K$ transpose of $\Sigma_{R,r}$ (ie, $\Sigma_{r,R} = \Sigma'_{R,r}$).[1]

[1] We use a prime symbol to denote matrix transposition.

These covariance matrixes are then concatenated across all $K+N$ assets into the $(K+N)\times(K+N)$ covariance matrix Σ, defined as

$$\Sigma = \begin{pmatrix} \Sigma_{R,R} & \Sigma_{R,r} \\ \Sigma_{r,R} & \Sigma_{r,r} \end{pmatrix}. \tag{2.1}$$

Similarly, we denote by $\mu \equiv \binom{\mu_R}{\mu_r}$ the $(K+N)\times 1$ (concatenated) vector of expected returns across the $K+N$ assets.

Given short-sales constraints, the mean–variance optimization problem across the $K+N$ assets consists in determining the K-dimensional vector $\omega_R \geqslant \mathbf{0}_R$ and the N-dimensional vector $\omega_r \geqslant \mathbf{0}_r$ that maximize

$$(\omega_R', \omega_r') \begin{pmatrix} \mu_R \\ \mu_r \end{pmatrix} - \tfrac{1}{2}\gamma(\omega_R', \omega_r') \begin{pmatrix} \Sigma_{R,R} & \Sigma_{R,r} \\ \Sigma_{r,R} & \Sigma_{r,r} \end{pmatrix} \begin{pmatrix} \omega_R \\ \omega_r \end{pmatrix}, \tag{2.2}$$

subject to $\omega_R' \cdot i_R + \omega_r' \cdot i_r = 1$.[2] If the assets in the plan span, then the optimal mean–variance allocation (ω_R^*, ω_r^*) is such that $\omega_r^* = 0_N$.

Since the work of Huberman and Kandel (1987), the formal empirical analysis of (2.2) has been tied to the multivariate regression specification

$$r = \alpha + \beta R + \varepsilon. \tag{2.3}$$

Accounting for short-sales restrictions, the necessary and sufficient conditions for mean–variance spanning can be derived as (De Roon *et al* 2001, (15), p. 727)

$$\upsilon\alpha + \beta i_R - i_r \leqslant 0, \tag{2.4}$$

where υ is the mean of any stochastic discount factor that prices the assets. Given the range of values for υ, De Roon *et al* (2001, pp. 729–730) suggest that, for spanning, it is equivalent to jointly test

$$\left.\begin{aligned} 1\alpha + \beta i_K - i_N &\leqslant 0, \\ \upsilon_{\min}\alpha + \beta i_K - i_N &\leqslant 0, \end{aligned}\right\} \tag{2.5}$$

where $\upsilon = 1$ is the upper bound, $\upsilon_{\min} = 1/E[R^{\mathrm{GMV}}]$ is the lower bound and $E[R^{\mathrm{GMV}}]$ is the mean gross return (ie, 1+net return) of the global minimum-variance portfolio.

To test the inequalities in (2.5), De Roon *et al* (2001) follow Kodde and Palm (1986) and use the Wald statistic

$$\xi = \min_{\gamma \geqslant 0} (\tilde{\gamma} - \gamma)' \tilde{\Sigma}^{-1} (\tilde{\gamma} - \gamma), \tag{2.6}$$

[2] We denote by $\mathbf{0}_R$ a vector of zeros with the same dimension (K) as the vector R. In the remainder of the paper we do not use boldface and we omit the subscript when referring to dimension if the context is evident.

where

$$\tilde{\gamma} = \begin{pmatrix} -\hat{\alpha} - \hat{\beta} \times i_K + i_N \\ -\dfrac{1}{1+\mu}\hat{\alpha} - \hat{\beta} \times i_K + i_N \end{pmatrix}, \tag{2.7}$$

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

$$\tilde{\Sigma} = \begin{pmatrix} -I_N & -A \\ -\dfrac{1}{1+\mu} I_N & -A \end{pmatrix} \Omega \begin{pmatrix} -I_N & -A \\ -\dfrac{1}{1+\mu} I_N & -A \end{pmatrix}', \tag{2.8}$$

with I_N defined as the $N \times N$ identity matrix, A the Kronecker product $I_N \otimes i_K'$, and Ω the $(N + NK) \times (N + NK)$ covariance matrix between the multivariate intercept α and the loading matrix β in the regression (2.3), and where $\hat{\alpha}$ and $\hat{\beta}$ refer to estimates of α and β, respectively. Note that the estimates for μ are typically two orders of magnitude smaller than unity. Therefore, when they are indeed very small for global minimum-variance portfolios, as often occurs, the first N rows (columns) in the matrix before (after) multiplying by Ω are almost identical to the last N rows (columns), making $\tilde{\Sigma}$ nearly singular and frequently resulting in incomputable inverses, as we experienced in our empirical implementation on 401(k) plans. As a result, we instead appeal to the fact that in the presence of a risk-free rate, say r_{f}, there is only one stochastic discount factor, with mean $1/(1 + r_{\text{f}})$. Consequently, instead of the two sets of inequalities in (2.5), we need only to deal with the one in (2.4), where $\nu = 1/(1 + r_{\text{f}})$, and for (2.7) and (2.8) we now have

$$\tilde{\gamma} = -\frac{1}{1+r_{\text{f}}}\hat{\alpha} - \hat{\beta} \times i_K + i_N,$$
$$\tilde{\Sigma} = \left(\frac{1}{1+r_{\text{f}}} I_N - A\right) \Omega \left(\frac{1}{1+r_{\text{f}}} I_N - A\right)'.$$

3 A CRITIQUE OF THE CURRENT LITERATURE ON 401(K) PLAN SPANNING TESTS

In the presence of short-sales constraints, both Elton *et al* (2006) and Tang *et al* (2010) use excess returns relative to the risk-free rate to test the null hypothesis

$$\alpha^* \leqslant 0, \tag{3.1}$$

where the N-dimensional α^* is a Jensen-type alpha (ie, $\alpha^* = (\mu_r - r_{\text{f}}) - \beta(\mu_R - r_{\text{f}})$). While they do not use any formal optimization model, Elton *et al* (2006, p. 1304) justify their choice by arguing that “if short sales are forbidden, then only the addition of an asset with positive alpha can improve the efficient frontier". Similarly, Tang *et al* (2010, p. 1078) state: "As short-sales are not allowed for [the] market benchmark index, if none of the α_i 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."

Next, we argue that (3.1) does not in fact reflect an optimality condition with short-sales constraints. In other words, an asset may have a negative alpha relative to a given set of other assets and still improve the efficient frontier of that set when short sales are precluded. Further, we also show that (3.1), which is expressed for excess returns, does not correspond to the necessary and sufficient optimality condition (2.4) for raw returns.

First, recall that the covariance matrixes of the excess returns and the raw returns are the same, and that we retrieve the same variance-minimizing portfolios subject to the expected returns' constraints regardless of whether the optimization problem is stated for raw or excess returns.[3] Now consider assets 1, 2 and 3 with the 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 excess returns $\mu_1 - r_{\mathrm{f}} = 0.043$, $\mu_2 - r_{\mathrm{f}} = 0.001$ and $\mu_3 - r_{\mathrm{f}} = 0.028$, respectively. We then have

$$\mu_2 - r_{\mathrm{f}} = -0.012 + 0.1689(\mu_1 - r_{\mathrm{f}}) + 0.1416(\mu_3 - r_{\mathrm{f}}).$$

For a given expected excess return of 3%, the variance-minimizing allocation when only assets 1 and 3 are considered is 0.1333 and 0.8667, respectively, with the resulting standard deviation of the portfolio return equal to 0.1243. On the other hand, with the addition of asset 2 for the same 3% level of expected excess return, the variance-minimizing strategy yields 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 (ie, 0.0773), despite the negative alpha of asset 2.

We now show that, for excess returns, (3.1) does not correspond to condition (2.4) when reexpressed in terms of raw returns. Based on raw returns, regression (2.3) yields

$$\beta = \Sigma_{r,R}\Sigma_{R,R}^{-1}. \tag{3.2}$$

[3] Since $\omega'(\mu - r_{\mathrm{f}}) = \bar{\mu} - r_{\mathrm{f}}$ is equivalent to $\omega'\mu = \bar{\mu}$ when $\omega'\mathbf{1} = 1$, where μ is a vector of expected raw returns, $\bar{\mu}$ a target expected raw return, r_{f} a risk-free rate of return and $\mathbf{1}$ a vector of elements all equal to 1.

Letting $r^* = r - r_f i_r$ and $R^* = R - r_f i_R$ denote the excess returns associated with r and R, respectively, the corresponding regression

$$r^* = \alpha^* + \beta^* R^* + \epsilon^* \tag{3.3}$$

leads to

$$\beta^* = \Sigma_{r^*,R^*} \Sigma^{-1}_{R^*,R^*}, \tag{3.4}$$

where the covariance matrixes above are associated with r^* and R^*. Clearly, since r_f is deterministic, $\beta = \beta^*$. From (2.3) and (3.3) we can infer that

$$\alpha^* = \alpha + r_f(\beta i_R - i_r).$$

With $\nu = 1/(1 + r_f) > 0$, condition (3.1) is equivalent to

$$\nu\alpha + \frac{r_f}{1 + r_f}(\beta i_R - i_r) \leqslant 0, \tag{3.5}$$

which is not exactly (2.4), even though, given that r_f is significantly less than unity, the two conditions could be approximately similar. Next, we show that, in fact, the discrepancy in the test results can be significant in practice.

4 EMPIRICAL ILLUSTRATION

Our sample consists of 7975 defined contribution plans with data provided by BrightScope, Inc., an information provider of retirement plan ratings and investment analytics, with returns covering the period from 2004 to 2008, which overlaps with the periods covered by Elton *et al* (2006) and Tang *et al* (2010). This sample is significantly larger than the 417 plans of Elton *et al* (2006) and the 1003 plans of Tang *et al* (2010). Return data for benchmark funds are from Datastream. Our benchmark set is identical to those used by Elton *et al* (2006) and Tang *et al* (2010), consisting of the Barclays Capital Aggregate Bond Index, the Credit Suisse High Yield Bond Fund and the Citigroup Non-US Dollar World Government Bond Index for the returns of fixed income securities; the Russell 1000 Growth Index, the Russell 1000 Value Index, the Russell 2000 Growth Index and the Russell 2000 Value Index for the returns of large-, mid and small-cap equities; and the MSCI EAFE Index for international exposure. Descriptive statistics for our data are summarized in Tables 1 and 2 and Figure 1.

Table 1 shows that our final sample has an average of 18 funds per plan, a median of 17 funds per plan, a standard deviation of 7 funds (approximately) and an interquartile range of 13 to 23 funds across the 7975 plans, the sizes of which have bottom and top deciles of, respectively, 10 and 28 funds. This distribution of funds illustrates the variety of plans tested in our analysis. We individually test (via a Wald

TABLE 1 Plan-level descriptive statistics.

(a) Original sample, $N = 17\,386$

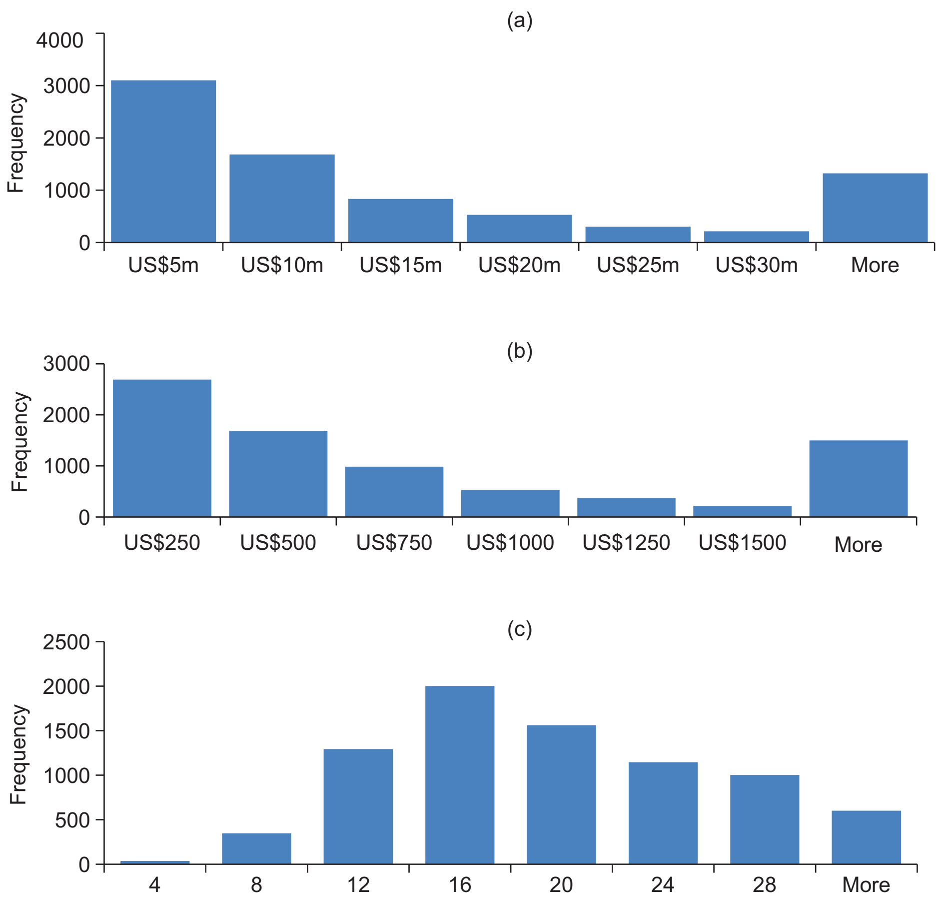
	Mean	SD	10th percentile	25th percentile	Median	75th percentile	90th percentile
Plan size (thousands)	22 100	111 000	1157	2534	5894	13 700	34 200
Average fund balance (thousands)	1 228	8 456	63	133	313	734	1 865
Number of fund options	22.13	15.00	11	15	19	26	32

(b) Final sample, $N = 7975$

	Mean	SD	10th percentile	25th percentile	Median	75th percentile	90th percentile
Plan size (thousands)	31 827	136 058	1141	2684	6697	17 172	52 676
Average fund balance (thousands)	1 891	8 564	77	180	426	1 059	3 110
Number of fund options	18.21	6.78	10	13	17	23	28

This table presents the summary data used for our empirical illustration. It includes descriptive statistics for plan size and fund balances (both in US dollars) for 401(k) plan data provided by BrightScope, Inc. Based on data filters such as full return availability for the full sample period 2004–8, the original sample of 17 386 plans is reduced to 7975. “Plan size” is the sum of all fund balances held within a given plan, while “average fund balance” is the balance within each mutual fund. “Number of fund options” refers to the number of funds in the menu offered through a given plan. SD, standard deviation.

FIGURE 1 Detailed histograms associated with the summary statistics for the final sample in Table 1.



(a) Fund size. (b) Average fund balance (thousands). (c) Number of fund options.

test) each plan to see if it satisfies the null hypothesis (2.4) with $\nu = 1/(1 + r_f)$ for spanning (see the appendix of AitSahlia *et al* (2023) for further details of this test). Our empirical analysis focuses on the proportion of funds across all plans for which the null hypothesis is not rejected (ie, which “span”; see below).

For a given plan, a rejection of the null hypothesis suggests, with a 5% level of confidence, that its mean–variance spanning frontier can be improved with additional funds from the benchmark set. When the null hypothesis is not rejected, for the sake of expository simplification, especially when comparing our results with

TABLE 2 Benchmark index set.

	Monthly returns	
	Mean (%)	**Standard deviation (%)**
MSCI EAFE Index	0.28	4.76
Barclays Capital Aggregate Bond Index	0.37	1.34
Credit Suisse High Yield Bond Fund	–0.04	2.96
Citigroup Non-US Dollar World Government Bond Index	0.50	2.36
Russell 1000 Growth Index	–0.22	4.09
Russell 1000 Value Index	0.01	3.78
Russell 2000 Growth Index	–0.05	5.64
Russell 2000 Value Index	0.14	4.99

This table reports descriptive statistics on the performance of the benchmark index set used for the mean–variance spanning tests. The performance covers the period 2004–8 and is measured as monthly net-of-fees returns. This benchmark index set is consistent with those of Elton *et al* (2006) and Tang *et al* (2010).

TABLE 3 Spanning test results.

Span	3630	45.52%
No span	4345	54.48%
Total	7975	100.00%

This table reports a summary of the spanning test results at a 5% significance level relative to the benchmark index set in Table 2. Note that, for ease of cross-referencing with Elton *et al* (2006) and Tang *et al* (2010), we use the label “span” to refer to the failure of the rejection of spanning (the null hypothesis) and “no span” to the rejection of the null hypothesis at the 5% level.

those of Elton *et al* (2006) and Tang *et al* (2010), we retain the “spanning” label they adopted, instead of the strictly correct “spanning not rejected”. We find that 46% of plans span (Table 3), which is much lower than the 53% and 97% found by Elton *et al* (2006) and Tang *et al* (2010), respectively. This difference could be attributed to the disparity between our samples. Elton *et al* (2006) have a median of 8 funds per plan, with 11% of the plans offering 13 or more fund choices, for a total of 417 plans. On the other hand, Tang *et al* (2010) have 1007 plans with a median of 13 funds per plan, an interquartile range of 10 to 16 funds and bottom and top deciles of 8 and 19 funds, respectively. However, in view of Tang *et al* (2010) having a closer distribution of funds across plans than we do, the difference between our spanning results and those of these two sets of authors is also due to the fact that they both test the null hypothesis (3.1), $\alpha \leqslant 0$, which is a necessary condition, but not a sufficient one. As a result, when they fail to reject the null hypothesis, they may incorrectly conclude in favor of spanning. In contrast, we test the necessary and sufficient condition (2.4).

5 CONCLUSION

Regression-based mean–variance spanning tests are ubiquitous in empirical finance. This paper centered on challenges that arise when these tests are implemented in the context of short-sales constraints, as in defined contribution retirement plans such as the 401(k) plans of US employees. While the standard Wald testing methodology for mean–variance spanning relies on the implied means of unobservable discount factors, we exploited the fact that the mean discount factor is uniquely determined in the presence of a risk-free asset to help with its efficient implementation, thus avoiding potential issues of numerical instability. We also showed how the incorrect implementation of mean–variance spanning in the retirement literature has led to vastly overstated results.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

AitSahlia, F., Doellman, T., and Sardarli, S. (2023). Menu simplification for portfolio selection under short-sales constraints. *European Financial Management* **29**(1), 3–21 (https://doi.org/10.1111/eufm.12390).

De Roon, F. A., Nijman, T. E., and Werker, B. J. M. (2001). Testing for the mean–variance spanning with short sales constraints and transaction costs: the case of emerging markets. *Journal of Finance* **56**(2), 721–742 (https://doi.org/10.1111/0022-1082.00343).

Elton, E. J., Gruber, M. J., and Blake, C. R. (2006). The adequacy of investment choices offered by 401(k) plans. *Journal of Public Economics* **90**(6–7), 1299–1314 (https://doi.org/10.1016/j.jpubeco.2005.08.001).

Errunza, V., Hogan, K., and Hung, M.-W. (1999). Can the gains from international diversification be achieved without trading abroad? *Journal of Finance* **54**(6), 2075–2107 (https://doi.org/10.1111/0022-1082.00182).

Fama, E. F., and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**(1), 1–22 (https://doi.org/10.1016/j.jfineco.2014.10.010).

Fama, E. F., and French, K. R. (2018). Choosing factors. *Journal of Financial Economics* **128**(2), 234–252 (https://doi.org/10.1016/j.jfineco.2018.02.012).

Huberman, G., and Kandel, S. (1987). Mean–variance spanning. *Journal of Finance* **42**(4), 873–888 (https://doi.org/10.1111/j.1540-6261.1987.tb03917.x).

Kan, R., and Zhou, G. (2012). Tests of mean–variance spanning. *Annals of Economics and Finance* **13**(1), 139–187.

Kodde, D. A., and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* **56**(5), 1243–1248 (https://doi.org/10.2307/1912331).

Petukhina, A., Trimborn, S., Härdle, W. K., and Elendner, H. (2021). Investing with cryptocurrencies: evaluating their potential for portfolio allocation strategies, *Quantitative Finance* **21**(11), 1825–1853 (https://doi.org/10.1080/14697688.2021.1880023).

Tang, N., Mitchell, O. S., Mottola, G. R., and Utkus, S. P. (2010). The efficiency of sponsor and participant portfolio choices in 401(k) plans. *Journal of Public Economics* **94**(11–12), 1073–1085 (https://doi.org/10.1016/j.jpubeco.2010.06.016).