Risk optimizations on basis portfolios: The role of sorting*

Boris Fays[a], Nicolas Papageorgiou[b], Marie Lambert[a]

[a]*HEC Liege, Management School of the University of Liege*
[b]*HEC Montreal*

Abstract

This paper investigates the mean-variance and diversification properties of risk-based strategies performed on style or basis portfolios. We show that the performance of these risk strategies is improved when performed on portfolios sorted on characteristics correlated with returns and is highly sensitive to the sorting procedure used to form the basis assets. Whereas the extant literature provides mixed support for the outperformance of smart beta strategies based on scientific diversification, our designed strategies outperform both the market model and multifactor model. Our testing framework is based on bootstrapped mean-variance spanning tests and shows valid conclusions when controlling for multiple testing, transaction costs, and luck from random basis portfolio construction rules. Economically, our results are supported by diversification-based properties.

Keywords: Bootstrap, Mean-variance efficiency, Portfolio sorting, Risk-based optimization, Smart Beta, Style investing.

1. Introduction

At the core of the Modern Portfolio Theory, mean-variance portfolio optimization (MVO) poses serious practical issues. On the occasion of the 60^{th} anniversary of the Markowitz (1952) optimization, Kolm et al. (2014) debate on the common challenges induced by MVO, which make portfolio solutions often unimplementable in practice. Performing an optimization exercise *à la* Markowitz (1952) on a set of individual assets indeed induces large estimation errors which leads to poor stability of the estimators. Palczewski and Palczewski (2014) make a thorough review of the sources of the errors as well as their impact on the stability of the estimator. As a consequence, passive investors have considered for more than 60 years, capitalization-weighted (CW) indices as a proxy for the tangency portfolio, namely the Maximum Sharpe Ratio (MSR) portfolio. Although CW

*Boris Fays is a Research Associate at HEC Liège, ULiège. Rue Louvrex 14, Bldg. N1, 4000 Liège, Belgium. E-mail: boris.fays@uliege.be. Phone number: +32 474 48 92 20. Nicolas Papageorgiou is a Full Professor of Finance at HEC Montreal, and Vice-President and Head of R&D, Fiera Capital Corporation. Mailing address: 3000 Cote Ste Catherine, Montreal, Quebec. E-mail: Nicolas.Papageorgiou@hec.ca. Phone number: +1 514 340 7309. Marie Lambert is a Professor of Finance at HEC Liège, ULiège. She is also a Research Associate at EDHEC Risk Institute. Address: HEC Liège, Rue Louvrex 14, Bldg. N1, 4000 Liège, Belgium. E-mail: marie.lambert@uliege.be. Phone number: +32 4 232 74 32. All remaining errors are ours. Declarations of interest: none.

solutions provide a simple, cost-effective and intuitive manner to allocate stocks, they are exposed to certain inherent weaknesses, notably their embedded momentum bias and their concentration in large capitalization stocks. This situation has also led to the emergence of the so-called "low-risk portfolios" – such as minimum variance (MV), maximum diversification (MD), and risk parity (RP) – within the smart beta industry as the errors in estimating expected returns have been shown to pose a serious threat on the efficiency of the MVO output (Best & Grauer, 1991a; Best & Grauer, 1991b). Yet the extant literature has presented mixed evidence regarding the performance of these portfolios.

While low-risk portfolios typically outperform the CW benchmark over long horizons, Boudt et al. (2015) show that such strategies remain highly sensitive to market downturns. Taking into account that low-risk portfolios tend to have, by design, a low market beta, Anderson et al. (2012) note that risk parity portfolios require a certain level of leverage to achieve the significant gains that the academic literature associates with them. Scherer (2011) analytically show that minimum variance optimizations also imply higher weights to low beta assets and further evidence the underperformance of an (unleveraged) minimum variance portfolio against a combination of long-short (beta) portfolios and the CW market portfolio.

Besides, the inflation of multifactor models and risk factors has extended the investor's opportunity set to style portfolios.[1] This context raises some questions on the performance of smart beta solutions as they have mainly focused so far on low-risk strategies conducted on individual assets. This paper investigates the mean-variance properties of low-risk portfolios when applied on basis portfolios. We show that the performance of these portfolios is highly sensitive to its underlying assets or building blocks. Directly related to our research, Grinblatt and Saxena (2018) established a statistical technique to create a mean-variance efficient (MVE) portfolio starting from a set of characteristics- or style portfolios. This portfolio is shown to span the opportunity set formed from a 3-factor model (Fama & French, 1993). Ao et al. (2018) compare the properties of the MV and MVE portfolios for a large set of individual assets augmented with risk factors using both sample and robust estimates of the variance-covariance matrix. The authors design a new statistical approach to reduce estimation error and show that considering risk factors together with individual assets manages to deliver optimal risk-return properties. Both papers allow long and short positions into the extreme portfolios and might therefore constitute an unfeasible outcome for common investors. We differ from the previously cited works, by providing long-only investment solutions and by using the latest advances in the portfolio sorting literature to construct the basis portfolios (Chan et al., 2009; Chen & De Bondt, 2004; Hou et al., 2018; Kogan & Tian, 2015; Lambert et al., 2020). We show that the combination of low-risk optimizations with advanced portfolio sorts provide long-only

[1] Dimson et al. (2017) record over 6,000 ETFs/ETPs with 145 smart beta equity providers across 32 different countries in 2016. Quoted from Bloomberg, the number of ETFs reached roughly 5,000 in 2016 while outnumbering the number of listed securities, which was slightly above 4,000 over that year. This article is accessible at the following address: https://www.bloomberg.com/news/articles/2017-05-12/there-are-now-more-indexes-than-stocks.

solutions with attractive risk-return properties to the smart beta industry.

We compare the standard way to allocate stocks into these basis portfolios (using an independent scale with NYSE breakpoints) to the more recent dependent technique, which works in successive subportfolios. Whole-sample breakpoints are jointly used with a dependent sorting to obtain our opportunity set (Lambert & Hübner, 2013; Lambert et al., 2020). Hereafter, we refer to these two sets of basis portfolios as the dependent and independent basis portfolios.

Our empirical study proceeds as follows. First, we build on the "diversification return" from Booth and Fama (1992) and the extensions of Erb and Harvey (2006) and Willenbrock (2011) to infer the diversification properties of our basis portfolios. We show that risk optimizations on dependent basis portfolios outperform (in terms of Sharpe ratio and alpha) risk optimizations on independent basis portfolios. Second, we show that the risk optimizations on dependent basis portfolios span the traditional Fama and French (1993) three-factor model. Our empirical approach relies on the mean-variance spanning test of Kan and Zhou (2012), which evaluates the benefit of adding new investments to a baseline portfolio and discriminate these new investments according to their contribution to the baseline portfolio's mean-variance efficiency. We perform the mean-variance spanning test in a bootstrap setting, similar to that of Fama and French (2010) and Harvey and Liu (2019), to obtain spanning tests robust to the effect of multiple testing. Third, we apply the factor selection technique of Harvey and Liu (2019) to conduct a horse race between the different configurations of the sorting methods as well as the smart beta strategies. Our findings show that the MVE of the strategy is first subject to the definition of the sorting method, then to the choice of the risk optimization method. Our results are robust to the inclusion of transaction costs (see, e.g., Hasbrouck, 2009; Novy-Marx & Velikov, 2016). Finally, we demonstrate that our results do not hold for a random sort into portfolios, which suggests that efficient sorting procedures are important when characteristics are significantly correlated with returns.

Our approach is original and important as it relies on scientific diversification methods but is also driven by economic insights and market practice. For instance, our risk parity portfolio on size/book-to-market opportunity sets not only delivers interesting diversification properties but also makes sure each portfolio attribute contributes equally to the variance of the portfolio. Finally, even though our approach has a mean-variance focus, we check the higher-moment properties of our candidates and show that they manage to reduce left-asymmetry.[2]

The rest of the paper is organized as follows: Section 2 describes the data and methodology used to construct the basis portfolios. Section 3 presents smart beta strategies and their diversification properties. In Section 4, mean-variance spanning tests are used to compare smart strategies against

[2]For instance, maximum diversification optimizations on dependent portfolios offer, on average, Sharpe ratios (SR) adjusted for skewness and kurtosis that are 1.3 greater than on independent portfolios for the period ranging from July 1993 to December 2015. Adjusted Sharpe ratios are estimated by $SR(1 + (S/6)SR - (K - 3/24)SR^2)$ where S and K denote skewness and kurtosis respectively. For the sake of brevity, we do not report these descriptive statistics. However, results are available upon request.

single-index and multifactor models. In Section 5, we test the significance of our smart strategies to complement a multifactor model and explain the cross-section of characteristic-sorted portfolios. Section 6 investigates the performance of smart betas on portfolios constructed randomly. Section 7 concludes the paper.

2. Investment Opportunity Set

This section describes our opportunity set; i.e., the set of portfolios that constitute our basis assets. Our approach consists of stratifying the U.S. stocks universe in investment style portfolios under a classical angle; namely, size, book-to-market, and momentum portfolios.

Grouping stocks into portfolios offers several advantages. First, forming groups of stocks into style portfolios circumvents the burden of estimating a large covariance matrix of returns (Ao et al., 2018; Berk, 2000). Moreover, our framework is consistent with the stylized facts of Barberis and Shleifer (2003), who demonstrate the natural tendency of investors to allocate funds according to asset categories, and Froot and Teo (2008), who also observe that institutional investors tend to reallocate their funds across style groupings. Our objective to perform risk optimization techniques on investment style portfolios is, therefore, in line with the reallocation practice of institutional investors and avoids the implementation costs of working with a wide variety of individual securities.

Our stratification relies on two sorting methodologies. The first construction methodology is based on an independent sort of stocks into portfolios with NYSE-breakpoints and has become a standard in the asset-pricing literature for constructing characteristic-sorted portfolios (Fama & French, 1993, 1995, 2015). The second sorting methodology follows Lambert et al. (2020) and applies a dependent sort using whole-sample breakpoints; this strategy implies the sorting of stocks in successive subportfolios according to characteristics. We stratify the U.S. stock universe into six (2×3), nine (3×3) or twenty-seven (3×3×3) groups. The double sort is performed on size and book-to-market characteristics, while the 3×3×3 split is constructed on the momentum, firm size, book-to-market characteristics. More details of the two methodologies can be found below.

2.1. Data

The data are obtained by merging data from the Center for Research in Security Prices (CRSP) and Compustat. The CRSP database contains historical price information, whereas Compustat provides accounting information for all stocks listed on the major U.S. stock exchanges. The sample period ranges from July 1963 to December 2015 and covers all stocks listed on the NYSE, AMEX, and NASDAQ.[3] For stocks listed on the NASDAQ, the data collection starts in 1973. The analysis covers a total of 618 monthly observations. Following Fama and French (1993) to filter the database

[3] Data regarding Compustat and CRSP are available from January 1950 and January 1926, respectively. After correcting the databases for survival and backfill biases, the sample starts in July 1953. For comparison purpose, we start our empirical analyses from July 1963 onwards as in Fama and French (1993).

and construct cross-sectional portfolios, we keep stocks with a CRSP share[4] code (SHRCD) of 10 or 11 at the beginning of month t, an exchange code (EXCHCD) of 1, 2 or 3 available shares (SHROUT) and price (PRC) data at the beginning of month t, available return (RET) data for month t, at least 2 years of listing on Compustat to avoid survival bias and a positive book-equity value at the end of December of year $y-1$. We define the book value of equity as the Compustat book value of stockholders' equity (SEQ) plus the balance-sheet deferred taxes and investment tax credit (TXDITC). If available, we decrease this amount by the book value of the preferred stock (PSTK). If the book value of stockholders' equity (SEQ) plus the balance-sheet deferred taxes and investment tax credit (TXDITC) is not available, we use the firm's total assets (AT) minus its total liabilities (LT).

Book-to-market equity (B/M) is the ratio of the book value of equity for the fiscal year ending in the calendar year $y-1$ to market equity. Market equity is defined as the price (PRC) of the stock times the number of shares outstanding (SHROUT) at the end of June y to construct the size characteristic and at the end of December of year $y-1$ to construct the B/M ratio. Momentum is defined as in Carhart (1997); i.e., based on a $t-2$ until $t-12$ cumulative prior return.

2.2. Sorting Out Stocks

In the original Fama–French approach, portfolios are constructed using a 2×3 independent sorting procedure: two-way sorting (small and large) on market capitalization and three-way sorting (low, medium, high) on the book-to-market equity ratio. Six portfolios are constructed at the intersection of the 2×3 classifications and are rebalanced on a yearly basis at the end of June. These style classifications are defined according to the NYSE stock exchange only and then applied to the whole sample (AMEX, NASDAQ, and NYSE).[5] The authors motivate the use of NYSE breakpoints by the need to have approximately the same market capitalization across portfolios and the same number of NYSE firms in each portfolio.

The second sorting methodology is an extension of the Fama–French sorting methodology. Lambert et al. (2020) sort stocks in successive subportfolios according to various characteristics; moreover, they define sorting breakpoints based on the whole sample rather than considering only the NYSE. The authors indeed uncover that these NYSE breakpoints create an imbalance in the (total) number of stocks between small- and large-cap portfolios such that, an independent sorting leads to a higher number of stocks in small-value portfolios (Cremers et al., 2012). As from January 1963 to December 2015, the market equity and book-to-market equity of a firm were, on average, negatively correlated (-5%), using an independent sort on negatively correlated variables can induce, by design, a strong

[4]See Hasbrouck (2009, p. 1455):"*restricted to ordinary common shares (CRSP share code 10 or 11) that had a valid price for the last trading day of the year and had no changes of listing venue or large splits within the last 3 months of the year*".

[5]The NYSE is represented by stocks that account for the largest capitalization in the CRSP database. The exchange codes 1, 2 and 3 represent the NYSE, NASDAQ, and AMEX, respectively

tilt toward the extreme categories of inverse ranks, i.e., low-high and high-low.

Another practical consequence when sorting stocks into portfolios, as already stated by Chan et al. (2009), is that the original independent sorting with NYSE-breakpoints procedure could induce large value stocks to be categorized as growth stocks. Supportive evidence can be found in the recent work of Lettau et al. (2018) who characterize the holdings of value mutual funds using Daniel et al. (1997) methodology.[6] Lettau et al. (2018) show that value mutual funds tend to hold a large proportion of their investments in growth stocks. However, the ranking into quantiles relies on NYSE breakpoints. Lambert et al. (2020) document that the choices underlying the sorting methodology are important to draw robust inference on firm style characteristics. In particular, the standard procedure of NYSE breakpoints and the sequence of the dependent sort matter. If the sorting methodology is responsible for these empirical results, we claim that forming basis portfolios using this procedure will lead to a biased allocation of stocks into style portfolios and stratification of the U.S. equity universe, and therefore to a misleading optimization exercise.

To better understand the problem, we compare the Morningstar style classification of 8,739 mutual funds (focused on the U.S. equity market) to the ones implied by the dependent on all breakpoints and independent on NYSE breakpoints sorting procedures. For the dependent sort, the classification of stocks for growth and value characteristics is obtained by applying a first sort on the size characteristic of a firm and then performing a second sort on the book-to-equity market of a firm. We construct a matrix of 5×5 portfolios along the size and value characteristics of a firm. For the independent sort, the output is similar to the 5×5 size and value portfolios available on Kenneth French's website.

The sample of mutual funds is obtained from Morningstar and CRSP Mutual Fund databases over the period April 2002 to December 2015. Databases are merged according to two labels: funds' CUSIP and a phrase matching techniques applied on funds' name. Monthly performance and quarterly holdings are obtained from CRSP Mutual Fund Database. Style classifications are obtained from Morningstar. Next, we match the information of funds' holdings with the value-growth classification from the independent (with NYSE breakpoints) and dependent (with whole sample breakpoints) sorting methodologies. The classification is applied according to accounting information obtained from Compustat at the end of June for each stock. The stock universe is then split according to a 1–5 scale: 1 represents a growth tilt, 3 represents a blend/neutral style, and 5 represents a value tilt as in the work of Lettau et al. (2018).

Figure 1 illustrates the distribution of funds along the dependent-name breakpoints (hereafter referred simply to dependent sort) and independent-NYSE (hereafter referred simply to independent sort) frameworks for the following Morningstar categories: growth (left), blend (middle), value

[6] Daniel et al. (1997) sort stocks at the end of June of each year to form 125 portfolios along a triple dependent sort with the first sort on firms size, the second sort on firms' industry adjusted book-to-market and the final sort on firms' momentum (cumulative return from t-2 to t-12).

(right). The distributions across the growth and value styles demonstrate that the BM-score of mutual funds computed using a dependent scale instead of an independent scale is better aligned with the style classification of the fund. Indeed, the distribution for growth funds is more skewed to the left for the dependent sort as shown by the 21.32% (dependent sort) vs. 9.17% (independent sort) of the observations falling under the first quintile of the distribution. Similarly, for value funds, the distribution is more skewed to the right for the dependent sort given that 7.30% (dependent sort) vs. 4.81% (independent sort) of the observations falling under the last quintile of the distribution. Lastly, the mode of the distribution of blend mutual funds under a dependent scale falls around the third quintile as 49.42% of the observations are found in the quintile 2 and 4. Using an independent scale, the mode is shifted to values below 3, which are representative of growth stocks. This would wrongly indicate that these funds hold more growth than value stocks and suggest that the sorting definition could mislead the allocation.

In summary, value (growth) mutual funds have a higher probability of being categorized as value (growth) funds under a dependent sorting procedure than an independent sorting procedure. Blend mutual funds also show better neutrality to the value-growth categorization using a dependent sort.

[Figure 1 about here.]

2.3. Pair-Wise Correlation of Style Portfolios

Figure 2 illustrates the stock distribution when the number of portfolios is increased either by a larger split of the sample (from a 2×3 to a 3×3 split) or by adding a new characteristic (3×3×3). The 3×3×3 splits are constructed based on the size, value, and momentum characteristics of a firm. We observe that using an independent sort results in an imbalance of stocks across the portfolios, and this effect becomes larger when more groups are constructed.

[Figure 2 about here.]

We expect the higher level of diversification induced by the dependent sort and by the higher dimensional space representation of the U.S. equity market to deliver additional diversification benefits for risk-based optimizations with regard to independent basis portfolios.

To verify this hypothesis, in Table I, we compute the average correlation between the investment style portfolios. It can be shown that the correlation is lower when stocks are sorted dependently and are split into a larger number of groups (i.e., 3×3×3). Here, the basis portfolios are cap-weighted portfolios to mitigate the impact of small cap stocks and rebalanced annually at the beginning of July consistent with the approach of Fama and French (1993).[7]

[Table I about here.]

[7] Our work could be further extend as in the work of Brandt et al. (2009) who allocated the weights of stocks in portfolios according to the level of their characteristics as to maximize an CRRA investor's utility. However, we leave this option to more interested readers as we are more interested to review the consequences of stock classification methods for risk-based optimizations rather than the allocation scheme inside the basis portfolios.

3. Smart Investment Strategies

Table II recalls the analytic forms of the risk-based allocations that serve as a practical base in our empirical analysis; namely, *minimum variance* (MV), *maximum diversification* (MD), and *risk parity* (RP). Following Ao et al. (2018), Ardia et al. (2018), Grinblatt and Saxena (2018), and Roncalli and Weisang (2016) among others, these risk-based allocations are rebalanced on a monthly basis.

[Table II about here.]

To feed these low-risk investment strategies (MV, MD, RP), we form 6, 9, and 27 cap-weighted portfolios and use 60 daily returns to estimate the covariance matrix.[8] In the most extreme case (27 portfolios), we are left with 0.17 data points per parameter. Even this simplified situation might create large sampling errors if we only consider the sample covariance matrix in our optimizations. In our applications, we use a traditional shrinkage methodology developed by Ledoit and Wolf (2004) to estimate the covariance matrix with lower sampling errors. Further details on the shrinkage method used can be found in the Appendix A.

3.1. Diversification Properties

This section compares the diversification returns achieved through implementing risk-based optimization based on dependent and independent basis portfolios and further decomposes the diversification return into its two components and performs a paired difference test.

The diversification return, according to Booth and Fama (1992), is defined as the difference between the compound return of a portfolio and the weighted average of the compound return of its constituent assets. This relationship assumes that the portfolios are rebalanced so that the weights are held constant and moments higher than the second are very small. In this situation, the diversification return increased with the spread between the individual asset variance and its covariance with the portfolio.

Denoting the geometric average return as g, the volatility as σ, and the arithmetic average return as μ, the geometric return of a portfolio p can be expressed as follows:

$$g_p = \mu_p - \frac{\sigma_p^2}{2} \tag{1}$$

[8]We use a range of 60-day to estimate variance-covariance matrices for two reasons; first, Fama and French (2018) use 60 days of lagged returns to estimate the monthly variance of stocks, and second, real-life applications on tradable assets would also impose practical constraints over the length available for time-series (Idzorek & Kowara, 2013).

The diversification return (DR) can be written as follows (Booth and Fama (1992) and Willenbrock (2011)):

$$\mathrm{DR} = g_p - \sum_i^N w_i g_i \tag{2}$$

where i stands for the i^{th} security in the portfolio p, and g refers to the geometric return. Weights (w_i) are assumed to be constant over the estimation period. We refer to fixed-weight diversification return using the superscript (FW).

Substituting (1) in (2), we obtain

$$\mathrm{DR}^{FW} = \mu_p - \frac{\sigma_p^2}{2} - \sum_i^N w_i \left(\mu_i - \frac{\sigma_i^2}{2}\right) \tag{3}$$

Rearranging the terms,

$$\mathrm{DR}^{FW} = \underbrace{\mu_p - \sum_i^N w_i \mu_i}_{\mathrm{DR}_1^{FW} = 0 \text{ if weights are constant}} + \underbrace{\frac{1}{2}\left(\sum_i^N w_i \sigma_i^2 - \sigma_p^2\right)}_{\mathrm{DR}_2^{FW} = \text{variance reduction benefit}} \tag{4}$$

In the last part of the equation, we retrieve the variance reduction benefit (DR_2^{FW}) of Booth and Fama (1992) and Willenbrock (2011). Note that in theory, w_i should be determined at inception and remain constant over the life of the strategy. To implement equation (4) for rebalancing strategies (non fixed weight), Erb and Harvey (2006) use the average of the weights over the sample period ($\overline{w_i} = \frac{1}{T}\sum_1^T w_i^t$). As the computation of the diversification return induces a comparison with a static portfolio endogenous to each strategy, it is difficult to compare the total diversification gains across a pair of smart beta strategies, which shift systematically assets weights. We, therefore, extend equation (4) to consider a rebalanced portfolio p and its diversification return with regard to an EW benchmark. We chose the equal-weighted strategy because this is the only allocation for which we know ex-ante the value of w_i, that is (1/N), as long as the amount of securities (N) remains constant in the portfolio.[9] In this alternative framework, we impose that two smart beta strategies constructed on an equivalent number of basis portfolios (N) share the same benchmark ($1/N$). We denoted the principle that the diversification return is compared to an EW strategy

[9]Due to the simplicity and the out-of-sample performance of the strategy, DeMiguel et al. (2009, p. 1948) also recommend the "$1/N$" portfolio as "the first obvious benchmark" for evaluating other weighting schemes.

using the superscript (EW) as follow,

$$\mathrm{DR}^{EW} = \underbrace{\mu_p - \frac{1}{N}\sum_i^N \mu_i}_{\mathrm{DR}_1^{EW}} + \underbrace{\frac{1}{2}\left(\frac{1}{N}\sum_i^N \sigma_i^2 - \sigma_p^2\right)}_{\mathrm{DR}_2^{EW}} \tag{5}$$

With this benchmark, it may not be immediately clear whether the measure departures from the essence of diversification, which concerns the interaction among the constituents of a single portfolio. We may see this interaction as DR^{EW} simply adds another strategic return, i.e., $\sum_i^N \left(\overline{w_i} - \frac{1}{N}\right) g_i$, to DR^{FW}, that is,

$$\begin{aligned}
\mathrm{DR}^{EW} =& \mathrm{DR}^{FW} + \underbrace{\sum_i^N \left(\overline{w_i} - \frac{1}{N}\right) g_i}_{\mathrm{DR}_3^{FW} \;=\; \text{gain in strategic return over EW}} \\
=& \mathrm{DR}^{FW} + \sum_i^N \left(\overline{w_i} - \frac{1}{N}\right)\left(\mu_i - \frac{\sigma_i^2}{2}\right) \\
=& \mathrm{DR}^{FW} + \sum_i^N \left(\overline{w_i} - \frac{1}{N}\right)\mu_i - \frac{1}{2}\sum_i^N \left(\overline{w_i} - \frac{1}{N}\right)\sigma_i^2 \\
=& \mathrm{DR}^{FW} + \sum_i^N \overline{w_i}\mu_i - \frac{1}{N}\sum_i^N \mu_i - \frac{1}{2}\sum_i^N \overline{w_i}\sigma_i^2 - \frac{1}{2}\frac{1}{N}\sum_i^N \sigma_i^2 \\
=& \mu_p - \frac{1}{N}\sum_i^N \mu_i + \frac{1}{2}\left(\frac{1}{N}\sum_i^N \sigma_i^2 - \sigma_p^2\right) = \mathrm{DR}_1^{EW} + \mathrm{DR}_2^{EW}
\end{aligned} \tag{6}$$

where the term DR_3^{FW} is the spread between the weighted average and the simple average of the geometric return of the portfolio's assets. It can be interpreted as the hypothetical gain in geometric return obtained by selecting a rebalancing strategy different from an equal-weighted allocation. The term might thus be useful to measure the return attributed to the strategic decision which allocates the opportunity set. For example, it could tell us whether a *minimum variance* offer a higher strategic return than an *equal-weight* for allocating a set of independent-sorted portfolios (or dependent-sorted portfolios).

To test the statistical difference in diversification return brought by a pair of strategies performed on two opportunity sets, we follow the indirect bootstrap framework of Ledoit and Wolf (2008), which is initially constructed to compare if a pair of strategies have statistically equivalent Sharpe ratios. In their conclusion, Ledoit and Wolf suggest extending their model to other mean-variance performance measures. We thus revisit their framework to a spread in diversification return between a pair of strategies. We provide more details on our extension in the Appendix B. In short, we aim to compare the spread in diversification return (Δ Dep-Ind) estimated in the original sample to an empirical distribution of spreads constructed from bootstrapped samples, and then infer the level of significance of this spread.

We report, in Table III, the results of the diversification return for the low-risk investment strategies based on 2×3, 3×3, and 3×3×3 basis portfolios. We observe that the spread in variance reduction benefit (DR_2^{FW} or DR_2^{EW}) is in 8 out of 9 times statistically greater for the set of dependent basis portfolios at the usual significance level. According to DR_3^{FW}, it also seems more interesting to perform any smart beta strategy on dependent portfolios rather than independent portfolios.

For 7 out of 9 risk-based optimizations, the dependent opportunity set offers significantly higher total diversification returns (DR^{EW}) than the independent sort. Consistent with Grinblatt and Saxena (2018), risk-return improvement can be achieved by allocating basis portfolios with optimization techniques, which departure from the traditional equal-weight allocation. However, our results uncover that it is only valid when advanced sorting methods are used to construct the basis portfolios. In the Table, the dependent opportunity sets systematically outperform the independent opportunity sets.

[Table III about here.]

The next section is dedicated to providing a methodological analysis on the mean-variance performance of the smart beta strategies.

4. Mean-Variance Spanning Test

Mean-variance spanning *à la* Huberman and Kandel (1987) means that a set of K risky assets spans a larger set of $K + N$ assets if the efficient frontier made of the K assets is identical to the efficient frontier comprising the $K + N$ assets. We initially set R_1 to a K-vector of the returns on K benchmark assets, R_2 to a N-vector of the returns on N test assets, and R to the raw returns on $K + N$ assets. Huberman and Kandel (1987) define the following regression test:

$$R_2^t = \alpha + \beta R_1^t + e^t \tag{7}$$

The null hypothesis H_0 sets $\alpha = 0$ and $\delta = 1 - \beta = 0$ and implies mean-variance spanning as the benchmark assets dominate the test assets; both assets have the same mean, but the K benchmarks have a lower variance than the test assets.

Considering an efficient frontier comprising $K + N$ assets, the following two formulas express the optimal weights of the N assets into the tangent MSR (Qw_1) and GMV (Qw_2) portfolios:[10]

$$\begin{aligned}Qw_1 &= \frac{QV^{-1}\mu}{1'_{N+K}V^{-1}\mu} = \frac{\Sigma^{-1}\alpha}{1'_{N+K}V^{-1}\mu}\\ Qw_2 &= \frac{QV^{-1}1_{N+K}}{1'_{N+K}V^{-1}1_{N+K}} = \frac{\Sigma^{-1}\delta}{1'_{N+K}V^{-1}1_{N+K}}\end{aligned} \tag{8}$$

where $Q = [0_{N\times K}, I_N]$ with I_N, an $N \times N$ identity matrix, $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$ which comes from V the variance-covariance matrix of the K benchmark assets (R_1) plus the N test assets (R_2) that is,

$$V = Var[R_1^t, R_2^t] = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{9}$$

The value of alpha will determine whether the tangency portfolio is improved by the introduction of the N assets, while testing beta will determine whether a significant change is induced in the GMV portfolio by the addition of the N assets. Huberman and Kandel (1987) jointly test these two conditions. The rejection of mean-variance spanning could thus find two sources: an improvement in the slope of the tangency portfolio or an improvement in the risk-return properties of the GMV portfolio. However, beta can be estimated more accurately than alpha, as it does not depend on the expected returns of the assets (see equation 8). Therefore, the statistical significance of the change in the composition of the GMV portfolio can be reached without implying economic significance. To circumvent this problem, Kan and Zhou (2012, hereafter KZ) propose to test the two conditions separately and to adjust the significance threshold of the two tests to economic significance. If the GMV condition is rejected more easily, the significance threshold should be reduced.

The KZ step-down test proceeds as follows: The first test defines the null hypothesis for the tangent portfolio such that $\alpha = 0_N$ using the OLS regression. The tangency portfolio is improved when the null hypothesis is rejected.

$$H_0^1 : \alpha = 0_N \tag{10}$$

Kan and Zhou (2012) perform a test for the statistical significance of the hypothesis similar to a GRS F-test. The F-test for the first hypothesis (H_0^1) is

$$F_1 = \frac{T-K-N}{N}\left[\frac{\hat{a}-\hat{a}_1}{1+\hat{a}_1}\right] \tag{11}$$

[10] To make a clear distinction between the risk-optimization that minimizes the portfolio variance and the ex-post global minimum variance portfolio, we denote the former MV and the latter GMV in the rest of the paper. MSR will denote the tangent portfolio, i.e., maximum Sharpe ratio.

where T is the number of observations; K is the number of benchmark assets; N is the number of test assets; $\hat{a}_1 = \hat{\mu}_1' \hat{V}_{11}^{-1} \hat{\mu}_1$ represents the squared Sharpe ratio of the K benchmark assets (R_1), with $\hat{V}_{11}$ denoting the variance and $\hat{\mu}_1$ the vector of mean return of the benchmark assets; and $\hat{a}$ takes the same notation as $\hat{a}$ but refers to the benchmark assets plus the new test asset (R).

The second test of the step-down procedure defines the null hypothesis for the GMV portfolio. This second test is conditional on the first test, $\alpha = 0_N$, and verifies whether $\delta = 1_N - \beta 1_K = 0_N$. Only when both conditions are rejected does the test suggest that the GMV portfolio is improved by adding N assets to the K benchmark assets.

$$H_0^2 : \delta = 1_N - \beta 1_K = 0_N | \alpha = 0_N \tag{12}$$

The F-test for the second hypothesis (H_0^2) is

$$F_2 = \frac{T-K-N+1}{N} \left[\frac{\hat{c}+\hat{d}}{\hat{c}_1+\hat{d}_1} \frac{1+\hat{a}_1}{1+\hat{a}} - 1 \right] \tag{13}$$

where $\hat{c}_1 = 1_K' \hat{V}_{11}^{-1} 1_K$ and $\hat{d}_1 = \hat{a}_1 \hat{c}_1 - \hat{b}_1^2$ are the efficient set (hyperbola) constants with $\hat{a}_1 = \hat{\mu}_1' \hat{V}_{11}^{-1} \hat{\mu}_1$ and $\hat{b}_1 = \hat{\mu}_1' \hat{V}_{11}^{-1} 1_K$ for the benchmark assets (R_1). $\hat{\mu}_1$ and $\hat{V}_{11}$ denote the vector of mean return and the variance of the benchmark assets. $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{d}$ are the equivalent notations for the benchmark assets plus the new test assets (R).

In Figure 3a, we graphically illustrate a significant improvement in the tangency portfolio when a test asset (R_2) is added to the benchmark assets (R_1). In Figure 3b indicates a significant improvement in the GMV portfolio when a test asset (R_2) is added to the benchmark assets (R_1).

[Figure 3 about here.]

Mean-variance spanning implies that both null hypotheses hold (H_0^1 and H_0^2). The benchmark assets R_1 are said to span the test assets R_2 if the weight attributed to the N test assets within the efficient frontier comprising $K+N$ assets is trivial. Put differently, discarding the N test assets does not significantly change the efficient frontier of the K benchmark assets from a statistical standpoint. By testing the two hypotheses separately, we gain understanding of the reason for mean-spanning rejection. If the mean-variance test is rejected, the test assets improve either the slope of the tangency portfolio or the risk-return properties of the GMV portfolio. Assuming the existence of a risk-free rate, investors are mostly concerned by the difference in the tangency portfolios.

Our application of mean-variance spanning tests whether smart investment strategies span existing benchmarks, such as the single-factor model or the multi-factor model of Fama and French (1993) (Section 4.1). Spanning tests between the different configurations of low-risk portfolios are also performed to investigate the consequences of the use of different opportunity sets (Section 4.2).[11]

[11]The MATLAB code is available on Prof. Guofu Zhou's website.

Notice that all the following tables report the p-values of the F-tests while controlling for multiple testing, which are more restrictive than the p-values from the original test of Kan and Zhou (2012) (Appendix C explains in detail the steps of the multiple testing method).

4.1. Mean-Variance Spanning Test of the Traditional Multi-Factor Models

We assume the market model of Sharpe (1964) and the Fama and French (1993) 3-factor model as our initial choices for the benchmark portfolio R_1 and construct multiple scenarios to test the superiority of portfolio sort configurations for risk-optimization strategies.

In Scenario 1, R_1 comprises two assets: an investment in a 30-year U.S. treasury bond (B30) and the market portfolio, which are both given in excess of the risk-free rate (one-month T-bill from Ibbotson). The market portfolio and the risk-free rate are obtained from Kenneth French's website while the 30-year U.S. treasury bond (B30) is obtained from CRSP U.S. Treasury and Inflation Indexes.

In Scenario 2, R_1 comprises four assets: the 30-year U.S. treasury bond (B30) and the market portfolio (Mkt), the size (SMB) and value (HML) factors obtained from Kenneth French's data library.

In Scenario 3, R_1 comprises two assets: an investment in a 30-year U.S. treasury bond (B30) and the gross return of one smart beta strategy while R_2 is now the market portfolio (Mkt).

In all scenarios, R_2 is the gross return of one smart beta strategy (taken in excess of the risk-free rate but without taking transaction costs into account). We consider the strategies gross of transaction costs because the risk factors used as explanatory variables are also gross of transaction costs. We provide further evidence on the performance of the strategies net of transactions latter in the next sub-section of the paper.

The step-down spanning test proceeds as follows: We first test the null hypothesis H_0^1 that the α is equal to 0, meaning that no improvement is obtained in the efficient frontier by adding the smart beta strategy to the initial benchmark portfolio (R_1). We consider the usual significance thresholds; i.e., 1%, 5%, and 10%. Consistent with post-publication concerns claimed by Mclean and Pontiff (2016), our results will be further split into two sub-periods: the period for the full sample and the period after the publication date of the seminal Fama and French (1993) paper.

We report in Table IV the results for the dependent and independent opportunity sets. Only bootstrapped p-values are reported as these ones control for multiple testing and are consequently more conservative than the standard p-values found in the MVE test from Kan and Zhou (2012). Model (1) shows that the traditional CAPM model does not span an expanded set augmented with risk-optimization strategies. The tangent portfolio level (F_1) is significantly improved when adding the smart beta strategies using both the dependent and independent basis portfolios (all p-values are significant, with a 99% confidence level). However, the results of Model (2) indicate that a three-factor model spans the larger set comprising the original assets supplemented by a smart beta factor defined using *independent* basis portfolios (Panel A). Yet, several smart betas performed on a *dependent* opportunity set (Panel B) improve the tangency portfolio implied by the three-factor model: 4 strategies out of 9 improve the initial 3-factor portfolio at the 90% confidence level. Finally, Model (3) shows that smart beta strategies performed on dependent basis portfolios span (7 out of 9 cases) the tangent portfolio made of the traditional cap-weighted market portfolio. These results do not hold for independent basis portfolios. This evidence makes the latter sub-optimal with regard to low-risk strategies implemented on *dependent* basis portfolios (i.e., 3×3 and 3×3×3).

[Table IV about here.]

Next, we present in Table V the results on the post-publication period of the Fama and French (1993) 3-factor model. Findings suggest that all low-risk strategies performed on a *dependent* opportunity set reject the mean-variance spanning hypothesis of the CAPM and 3-factor model as both sub-hypotheses (on alpha and delta) are statistically different from 0. This means that two portfolios of the mean-variance frontier (the MSR and the GMV) are improved under a *dependent* framework. However, the three-factor model continues to span four low-risk portfolios that are formed on the *independent* opportunity, especially the strategies aiming at maximizing the portfolio diversification, i.e., maximum diversification (MD). This last evidence is particularly important as it confirms that the traditional *independent* sorting can not compete with a dependent sort when forming basis portfolios that offer sufficient cross-sectional variation.

[Table V about here.]

In summary, our results on the post-publication period highlight the improvement brought by considering low-risk portfolios constructed on style basis portfolios against the related multi-factor model. These results might be explained by the increasing market diversity offering a higher potential for diversification and the increase in volumes traded on the U.S. stock exchanges; this necessitates performing the optimization exercise on basis portfolios or factors rather than individual stocks. Our findings also support the outperformance of low-risk strategies performed on the dependent-sorted opportunity set. A horse race between the two sorting approaches for constructing basis portfolios will be performed in the next subsection.

4.2. Horse Race Between Dependent and Independent Basis Portfolios

The previous subsection suggests that dependent basis portfolios offer better properties to perform risk-based optimization. We, therefore, carry out a horse race between the opportunity sets made of basis portfolios formed after dependent and independent sorting. The spanning test considers whether a portfolio (R_1) composed of the U.S. government 30-year bonds and a smart beta formed on the sorting configuration A spans this set of portfolios (R_1) plus the same smart beta but performed on the sorting configuration B. In our applications, this leads to two different scenarios. In Scenario 1, R_2 is a smart beta formed on the independent-sorted portfolios (SB_{ind}) while in

Scenario 2, R_2 is the same smart beta but performed on the dependent-sorted portfolios (SB_{dep}). Here, smart betas are *net* of transaction costs. Details on the estimation of transaction costs can be found in Appendix D. Both H_1^0 (the test on the tangency portfolio) and H_2^0 (the test on the GMV) are tested and only bootstrapped p-values are reported as they control for testing both candidate simultaneously and consequently, are more conservative than standard p-values.

Table VI presents the results for cap-weighted basis portfolios. Both scenarios demonstrate that the dependent opportunity sets outperform the independent set. In Model (1), we test whether the low-risk strategies formed on dependent opportunity sets span a larger universe augmented with independent sets. For all low-risk strategies, we cannot reject mean-variance spanning at the 10% confidence level. This means that the efficient frontier comprising a low-risk optimization of dependent portfolios and an investment in a long-term U.S. government bond cannot be improved using an independent opportunity set. However, Model (2) indicates that the MD (2×3, 3×3, and 3×3×3) and MV (2×3, 3×3) strategies performed on a dependent opportunity set improve both the tangency and the GMV portfolios formed on an independent opportunity set. This is evidenced by the levels of p-values attached to F-tests on H_1^0 and H_2^0 when the dependent portfolio is used as R_2. Empirically, the best improvement is found for the MD_{3x3x3}^{dep} with a monthly abnormal net return of 0.23% (2.75% annually) over a combination of the long-term U.S. bond and MD_{3x3x3}^{ind}.

[Table VI about here.]

5. MVE Benchmark Selection

We follow the method of Harvey and Liu (2019) to select the most appropriate (without luck) MVE benchmark among the low-risk portfolios and the original CW portfolio for explaining the cross-section of expected returns. Our test assets are the 2×3 and 3×3 portfolios sorted on size and book-to-market or the 3×3×3 when the sorting procedure first pre-condition on a firm's momentum. The MVE benchmark should best complement a basis multi-factor model comprising a long-term U.S. Government rate as a proxy for the risk-free rate (B30) and the size (SMB) and the value (HML) factors of Fama and French (1993).

The method is an alternative to the test developed by Gibbons et al. (1989). It departs, however, from the GRS test as it allows the initial model to be sub-optimal and tests the incremental contribution of the additional factor.

To measure the incremental contribution of the selected candidate, Harvey and Liu (2019) define a scaled intercept (SI) measure and look at the spread between the scaled (by the standard error of the estimated intercept) intercept of the augmented and initial model. Using equivalent notations as the authors, the measure is defined as follow,

$$SI_{ew}^{med} = \frac{median(\{|a_i^g|/s_i^b\}_{i=1}^J) - median(\{|a_i^b|/s_i^b\}_{i=1}^J)}{median(\{|a_i^b|/s_i^b\}_{i=1}^J)} \tag{14}$$

where $median(.)$ is the median value of the ratio $|a_i^g|/s_i^b$ or $|a_i^b|/s_i^b$. Here the superscript b is for the baseline model and g is for the augmented model, the subscript i refers to the i-th portfolio among the J test assets, and s denotes the standard errors for the regression intercept a.

A negative value of the SI means that the augmented model outperforms the baseline model to explain the variations of the J test assets returns. To define a statistical level of confidence to the measure, Harvey and Liu (2019) use the bootstrapping method presented in Step 2 of Section C.

To orthogonalize the MVE candidates, the authors regress the returns of R_2^i, where i denotes the i-th candidate among the list of K candidates, against the baseline benchmark R_1 and then subtract the intercept from the time-series R_2^i, as follows:

$$\begin{aligned} R_2^i &= \alpha^i + \beta^i R_1 + e^i \\ R_2^{\alpha,i} &= R_2^i - \alpha^i = \beta^i R_1 + e^i \end{aligned} \tag{15}$$

In our applications, R_1 is composed of the 30-Year U.S. Bond (B30), the size (SMB), and the value (HML) factors. $R_2^{\alpha,i}$ is defined as a linear combination of the benchmark assets (R_1), i.e., the 30-Year U.S. Bond (B30), the size (SMB) and the value (HML) factors such that it does not bring any additional information to the baseline model.

Then in each sample of the B bootstrap, a score for the scaled intercept SI_{ew}^{med} can be obtained for the K number of orthogonalized candidates (i.e, $R_2^{\alpha,i}$ with the $i = \{1, 2, ..., K\}$ candidates). Hence, the single test p-value for the i-th candidate is given by,

$$\text{p-val} = \frac{\#\{SI^o > SI^b\}}{B} \tag{16}$$

To control for multiple testing, the authors suggest taking the minimum value among K estimates of SI in the b-th bootstrap as follow,

$$SI^{b,*} = \underbrace{min}_{i \in \{1,2,...,K\}} \{SI^{b,i}\} \tag{17}$$

Hence, the multiple test p-value for the i-th candidate is written as,[12]

$$\text{p-val} = \frac{\#\{SI^o > SI^{b,*}\}}{B} \tag{18}$$

Next, our objective is to apply the method to a multiple set of MVE candidates and filter them to find the best candidate. For example, the first natural candidate to consider is the traditional

[12] Note that the sign of the indicator function is important. Here, we want to count the number of bootstrapped scaled intercepts (SI^b) that have lower values (improvement of the model) than the scaled intercept from the original sample(SI^o). In other words, when the test is performed on the time-series of R_2 from the original data.

cap-weighted market portfolio. But smart beta strategies on the Fama-French's independent 2×3 size and value portfolios or on the dependent 2×3 size and value portfolios can also constitute MVE candidates to augment the baseline model. We also extend the 2×3 size and value grid, to a 3×3 or 3×3×3 splits with an additional sort on firms' momentum and end up with a set of 7 candidates as MVE portfolio. We run the test sequentially, as in Harvey and Liu (2019), until the single test p-value of each candidate is greater than a pre-specified threshold. In our application, we set the threshold to 10%. In each run, the selected candidate has a single test p-value but will also be attributed with a multiple test p-value to control for data snooping. The candidate is only accepted if the multiple test p-value is significant at a 90% confidence level.

Table VII presents the results for the different types of basis assets and smart beta portfolios. The table shows the single-test p-value for each MVE candidate as well as the final joint p-value for the selected candidates considering the multiple testing framework. Note that except for the cap-weighted market portfolios, all smart beta portfolio candidates are net of transaction costs as computed in Appendix E. For 2×3 portfolios, the optimal MVE candidate comes from the same family as the set of basis portfolios to be explained, i.e., independent for Panel A and dependent for Panel B. However, as soon as the dimension of the sort increases, and therefore the dispersion between portfolios, the dependent candidates win the horse race (Panel D to F).

In summary, our results can be explained by two elements documented by recent academic research. First, the spaces of 2×3 and 3×3 test assets are "rank deficient" as coined by Grinblatt and Saxena (2018), which means that the dimensions are too low to provide a robust statistical framework (Lewellen et al. (2010)). Second, the imbalance of the distribution of stocks in portfolios under an independent sort is too sensitive to a number of macro-economic factors (Daniel and Titman (2012)), and lead to the construction of sub-optimal basis portfolios. However, smart beta strategies can benefit from a 3-dimensional dependent sort, which overcomes the issues of portfolio diversification relatable to macro-economic factors (Lambert et al. (2020)).

[Table VII about here.]

6. Alphabet Portfolios

This section investigates the role of the underlying characteristics to build basis portfolios. Our objective is to show the impact of the sorting method when the underlying characteristics command a significant relationship with future stock returns. To that end, we construct smart betas on "Alphabet" portfolios instead of portfolios sorted across size, value and momentum dimensions. These "Alphabet" portfolios are formed on random characteristics, which are defined by the letters found in the ticker of stocks. More precisely, we assign each year at the end of June a random value obtained from a standard normal to each letter of the alphabet and allocate stocks at the beginning of July in basis portfolios according to these random values. The first characteristic is based upon the first letter of the ticker while the second characteristic is related to the second letter, and so on.[13] That way, the characteristics should be time-varying in a random manner.

The underlying construction methodology for the basis portfolios follows the independent sort with NYSE breakpoints of Fama and French (1993) and the dependent sort with whole sample breakpoints of Lambert et al. (2020). We form cap-weighted portfolios using 2x3, 3x3 and 3x3x3 splits. Next, we allocate these basis portfolios into one final smart beta strategy, i.e., risk parity, minimum variance, and maximum diversification, which rebalances the basis portfolios every month. Finally, we run the bootstrap mean-variance spanning test described in the previous section and verify whether these random allocations have significantly better Sharpe ratio than the CW market portfolio and the 3-factor model. In short, we want to test whether the definition of the underlying characteristics matters when constructing smart betas on basis portfolios.

By definition, the purpose of a sort into portfolios resides in factoring characteristics into returns for which the characteristics command a linear relationship with $E[R]$. For instance, a *positive* relationship between a characteristics and expected return would imply that $E[R_i] < E[R_j]$ where $i < j$ and correspond to the i-th and j-th portfolios. However, if the characteristic does not command any relationship with expected return, we should get $E[R_i] \approx E[R_j] \approx \beta E[R_m]$ and $\sigma[R_i] \approx \sigma[R_j] \approx \sigma[R_m]$ when the number of basis portfolios gets larger. Consequently, any smart betas (linear combination) of the N basis portfolios should not systematically outperform the market portfolio.

However, results displayed in Model (1) of Table VIII show that smart betas outperforming the market portfolio can still be found when basis portfolios are formed on random characteristics, i.e., when alphas are positive, and F_1 are significant. And that even when controlling for the Fama-French 3-factor model as evidenced by results in Model (2). Nonetheless, contrasting these results with the Model (2) from Table IV, we can formulate three remarks for smart betas on random portfolios: alphas are (i) not dependent of the sorting methods, (ii) at least 2x lower than for sorts on determinant characteristics, and (iii) greater for Risk Parity optimizations. When $\sigma[R_i] \approx \sigma[R_j]$, Risk Parity optimizations attribute a weight to each portfolio which is close to $1/N$ and consequently, the source of the abnormal return earned over the cap-weighted market portfolio might be attributed to the monthly rebalancing of the random portfolios rather than to the choice of the allocation itself as shown in Plyakha et al. (2015).

In summary, the test demonstrates that sorting methods are important when stock characteristics are correlated to each other and related to expected return while they are not when characteristics are uncorrelated to each other. Also, smart betas constructed on random signals can only outperform the cap-weighted market portfolio by maintaining constant weights and frequent rebalancing. These evidence might thus be helpful to filter the large universe of ETFs among which some smart betas

[13] If the ticker is missing in the database; instead, we use the first letters of the company names from CRSP (COMNAM).

are potentially designed on "fake" signals and only rely on frequent rebalancing schemes.[14]

[Table VIII about here.]

7. Concluding Remarks

New tendencies have emerged in passive investing toward smart beta strategies and style investing as a channel to obtain mean-variance efficient portfolios through low-risk objectives. Our paper reconciles the trends by applying long-only risk-based strategies to characteristic-sorted equity portfolios and addresses the mean-variance efficiency of these products. The exercise is economically important for two reasons. First, the recent inflation of discovered risk factors questions the capitalization-weighted market portfolio as a mean-variance efficient candidate. Second, there is a common practice among institutional investors to reallocate funds across style groupings (e.g., Froot & Teo, 2008).

We show that the methodology for grouping stocks in different style buckets has substantial implications for the performance of the selected smart beta strategy. To categorize stocks in investment style portfolios, we stratify the universe along the academic standard dimensions of size, value, and momentum characteristics. We implement two sorting methodologies: (a) a dependent sort on all-breakpoints and (b) an independent sort on NYSE-breakpoints. We demonstrate through a set of state-of-the-art mean-variance tests that risk optimizations on dependent portfolios obtain superior performance, and that after controlling for transaction costs and multiple testing. Contrary to conventional wisdom, our results show that the trade-off between allocating more weight to smaller capitalization stocks and transaction costs can be offset by the benefit of diversification achieved by a sophisticated sorting method, i.e., a dependent sort. Because a dependent sort controls for correlated variables and stratifies the stock universe in well-diversified portfolios (Lambert et al., 2020), this sorting methodology delivers significant diversification benefits for smart beta strategies. We substantiate this point by extending the approach of Booth and Fama (1992) on diversification return. Economically, we infer from our findings that our method, which reconciles style investing with smart beta strategies, not only reduces the curse of dimensionality in portfolio optimizations but is also well-aligned with current practices in the mutual fund industry.

This study thus contributes to the development of novel sorting methods to obtain efficient basis portfolios, which can improve the mean-variance performance of smart beta strategies.

Bibliography

Anderson, R. M., Bianchi, S. W., & Goldberg, L. R. (2012). Will My Risk Parity Strategy Outperform? *Financial Analysts Journal*, *68*(6), 75–93. https://doi.org/10.2469/faj.v68.n6.7

[14]Here, fake signals refer to characteristics which might correlate with expected return in-sample, but fail to do so out-of-sample or post-publication as it has been documented in the work of Mclean and Pontiff (2016) among others.

Ao, M., Li, Y., & Zheng, X. (2018). Approaching Mean-Variance Efficiency for Large Portfolios. *Review of Financial Studies*, *00*(0), 1–30. https://doi.org/10.1093/rfs/hhy105

Ardia, D., Boudt, K., & Nguyen, G. (2018). Beyond Risk-Based Portfolios: Balancing Performance and Risk Contributions in Asset Allocation. *Quantitative Finance*, *18*(8), 1249–1259. https://doi.org/10.1080/14697688.2018.1424349

Barberis, N., & Shleifer, A. (2003). Style Investing. *Journal of Financial Economics*, *68*(2), 161–199. https://doi.org/10.1016/S0304-405X(03)00064-3

Berk, J. B. (2000). Sorting Out Sorts. *Journal of Finance*, *55*(1), 407–427. https://doi.org/10.1111/0022-1082.00210

Best, M. J., & Grauer, R. R. (1991a). Sensitivity Analysis for Mean-Variance Portfolio Problems. *Management Science*, *37*(8), 980–989. https://doi.org/10.1287/mnsc.37.8.980

Best, M. J., & Grauer, R. R. (1991b). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Review of Financial Studies*, *4*(2), 315–342. https://doi.org/10.1093/rfs/4.2.315

Booth, D. G., & Fama, E. F. (1992). Diversification Returns and Asset Contributions. *Financial Analysts Journal*, *48*(3), 26–32. https://doi.org/10.2469/faj.v48.n3.26

Boudt, K., Darras, J., Nguyen, G. H., & Peeters, B. (2015). Smart Beta Equity Investing Through Calm and Storm. *Risk-based and factor investing* (pp. 195–225).

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *Review of Financial Studies*, *22*(9), 3411–3447. https://doi.org/10.1093/rfs/hhp003

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, *52*(1), 57–82. https://doi.org/10.2307/2329556

Chan, L. K. C., Dimmock, S. G., & Lakonishok, J. (2009). Benchmarking Money Manager Performance: Issues and Evidence. *Review of Financial Studies*, *22*(11), 4554–4599. https://doi.org/10.1093/rfs/hhp016

Chen, H. L., & De Bondt, W. (2004). Style Momentum within the S&P-500 Index. *Journal of Empirical Finance*, *11*(4), 483–507. https://doi.org/10.1016/j.jempfin.2004.04.005

Choueifaty, Y., & Coignard, Y. (2008). Toward Maximum Diversification. *The Journal of Portfolio Management*, *35*(1), 40–51. https://doi.org/10.3905/JPM.2008.35.1.40

Clarke, R., de Silva, H., & Thorley, S. (2013). Risk Parity, Maximum Diversification, and Minimum Variance: An Analytic Perspective. *The Journal of Portfolio Management*, *39*(3), 39–53. https://doi.org/10.3905/jpm.2013.39.3.039

Cremers, M., Petajisto, A., & Zitzewitz, E. (2012). Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation. *Critical Finance Review*, *2*, 1–48. https://doi.org/10.1561/104.00000007

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance*, *52*(3), 1035. https://doi.org/10.2307/2329515

Daniel, K. D., & Titman, S. (2012). Testing Factor-Model Explanations of Market Anomalies. *Critical Finance Review*, *1*, 103–139. https://doi.org/10.1111/j.1540-6261.1997.tb02724.x

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Dimson, E., Marsh, P., & Staunton, M. (2017). Factor-Based Investing: The Long-Term Evidence. *Journal of Portfolio Management*, *43*(5), 15–37. https://doi.org/https://doi.org/10.3905/jpm.2017.43.5.015

Elton, E. J., & Gruber, M. J. (1973). Estimating the Dependence Structure of Share Prices. *Journal of Finance*, *28*(5), 1203–1232.

Erb, C. B., & Harvey, C. R. (2006). The Strategic and Tactical Value of Commodity Futures. *Financial Analysts Journal*, *62*(2), 69–97. https://doi.org/10.2469/faj.v62.n2.4084

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns Stocks and Bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (1995). Size and Book-to-Market Factors in Earnings and Returns. *Journal of Finance*, *50*(1), 131–155. https://doi.org/10.1111/j.1540-6261.1995.tb05169.x

Fama, E. F., & French, K. R. (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns. *Journal of Finance*, *65*(5), 1915–1947. https://doi.org/10.1111/j.1540-6261.2010.01598.x

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

Froot, K., & Teo, M. (2008). Style Investing and Institutional Investors. *The Journal of Financial and Quantitative Analysis*, *43*(4), 883–906. https://doi.org/10.1017/S0022109000014381

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, *57*(5), 1121. https://doi.org/10.2307/1913625

Grinblatt, M., & Saxena, K. (2018). When Factors Don't Span Their Basis Portfolios. *Journal of Financial and Quantitative Analysis*. https://doi.org/10.2139/ssrn.2788387

Harvey, C. R., & Liu, Y. (2019). Lucky Factors. *Working Paper*. https://doi.org/10.2139/ssrn.2528780

Hasbrouck, J. (2009). Trading Costs and Returns for U. S. Equities: Estimating Effective Costs from Daily Data. *Journal of Finance*, *64*(3), 1445–1477. https://doi.org/10.1111/j.1540-6261.2009.01469.x

Hou, K., Xue, C., & Zhang, L. (2018). Replicating Anomalies. *Review of Financial Studies*, *hhy131*, 1–115. https://doi.org/10.2139/ssrn.2961979

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

Idzorek, T. M., & Kowara, M. (2013). Factor-Based Asset Allocation vs. Asset-Class-Based Asset Allocation. *Financial Analysts Journal*, *69*(3), 19–29. https://doi.org/10.2469/faj.v69.n3.7

Kan, R., & Zhou, G. (2012). Tests of Mean-Variance Spanning. *Annals of Economics and Finance*, *13*(1), 139–187. https://doi.org/10.2139/ssrn.231522

Kogan, L., & Tian, M. (2015). Firm Characteristics and Empirical Factor Models: A Model-Mining Experiment. *Working Paper*.

Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 Years of Portfolio Optimization: Practical Challenges and Current Trends. *European Journal of Operational Research*, *234*(2), 356–371. https://doi.org/10.1016/j.ejor.2013.10.060

Lambert, M., & Hübner, G. (2013). Comoment Risk and Stock Returns. *Journal of Empirical Finance*, *23*, 191–205. https://doi.org/10.1016/j.jempfin.2013.07.001

Lambert, M., Fays, B., & Hübner, G. (2020). Factoring Characteristics into Returns: A Clinical Study on the SMB and HML Portfolio Construction Methods. *Journal of Banking and Finance*, *114*(105811). https://doi.org/10.1016/j.jbankfin.2020.105811

Ledoit, O., & Wolf, M. (2004). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Ledoit, O., & Wolf, M. (2008). Robust Performance Hypothesis Testing with the Sharpe Ratio. *Journal of Empirical Finance*, *15*, 850–859. https://doi.org/10.1016/j.jempfin.2008.03.002

Lettau, M., Ludvigson, S., & Mandel, P. (2018). Characteristics of Mutual Fund Portfolios: Where Are the Value Funds? *Working paper*. https://doi.org/10.3386/w25381

Lewellen, J., Nagel, S., & Shanken, J. (2010). A Skeptical Appraisal of Asset Pricing Tests. *Journal of Financial Economics*, *96*(2), 175–194. https://doi.org/10.1016/j.jfineco.2009.09.001

Maillard, S., Roncalli, T., & Teïletche, J. (2010). The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, *36*(4), 60–70. https://doi.org/10.3905/jpm.2010.36.4.060

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2011). Commodity Liquidity Measurement and Transaction Costs. *Review of Financial Studies*, *25*(2), 599–638. https://doi.org/10.1093/rfs/hhr075

Mclean, R. D., & Pontiff, J. (2016). Does Academic Research Destroy Stock Return Predictability? *Journal of Finance*, *71*(1), 5–32. https://doi.org/10.1111/jofi.12365

Novy-Marx, R., & Velikov, M. (2016). A Taxonomy of Anomalies and Their Trading Costs. *The Review of Financial Studies*, *29*(1), 104–147. https://doi.org/10.1093/rfs/hhv063

Palczewski, A., & Palczewski, J. (2014). Theoretical and Empirical Estimates of Mean-Variance Portfolio Sensitivity. *European Journal of Operational Research*, *234*(2), 402–410. https://doi.org/10.1016/j.ejor.2013.04.018

Plyakha, Y., Uppal, R., & Vilkov, G. (2015). Why Do Equal-Weighted Portfolios Outperform Value-Weighted Portfolios? *Working Paper*, 1–25. https://doi.org/10.2139/ssrn.2724535

Politis, D., & Romano, J. (1992). A Circular Block-Resampling Procedure for Stationary Data. *John wiley* (pp. 263–270).

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, *39*(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Roncalli, T., & Weisang, G. (2016). Risk Parity Portfolios with Risk Factors. *Quantitative Finance*, *16*(3), 377–388. https://doi.org/10.2139/ssrn.2155159

Scherer, B. (2011). A Note on the Returns from Minimum Variance Investing. *Journal of Empirical Finance*, *18*(4), 652–660. https://doi.org/10.1016/j.jempfin.2011.06.001

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, *19*(3), 425–442. https://doi.org/10.2307/2977928

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, *68*(5), 1097–1126. https://doi.org/10.1111/1468-0262.00152

Willenbrock, S. (2011). Diversification Return, Portfolio Rebalancing, and the Commodity Return Puzzle. *Financial Analysts Journal*, *67*(4), 42–49. https://doi.org/10.2469/faj.v67.n4.1

Figures

Figure 1. Distribution of BM-scores of Mutual Funds: Independent vs Dependent Sorts

The figure shows the kernel distribution in BM-score of mutual funds with a focused on the U.S. equity market for which Morningstar attributes a value-growth classification. The value-growth classification applied to the mutual funds present in the CRSP mutual funds database. For each point in time where a fund reports its holdings, we associate a BM-score from a 1–5 scale according the Fama–French's 5x5 size and value *independent* sorting methodology or a 5x5 size and value *dependent* sorting methodology. The fund's BM-score is then calculated as the percentage of Total Net Assets (TNA) weighted average of the 1–5 scale of the securities the fund holds. Distributions are displayed for 3 Morningstar classifications of funds: growth (left), blend (middle), value (right). The sample period ranges from April 2002 to December 2015.

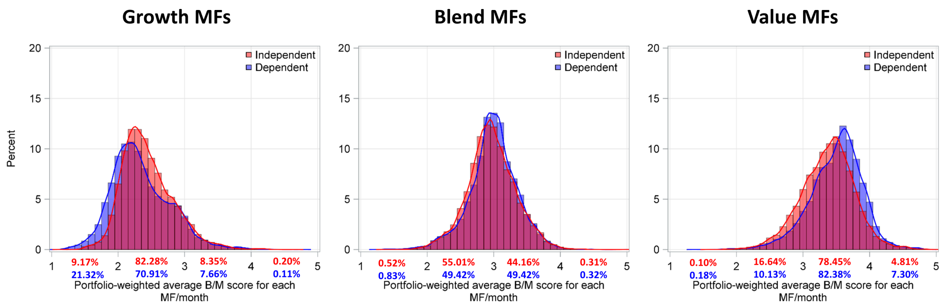


Figure 2. Average Stock Distribution with Independent vs Dependent Sorting

These plots show the stock distribution among the 2×3 and 3×3 characteristic-sorted portfolios based on size (low, medium and high) and the book-to-market equity ratio (low, medium and high) for the independent- and dependent-sorting methodologies. We also report the average percentage of stock distribution among the 3×3×3 characteristic-sorted portfolios when momentum is added as a third variable. For clarity, we group the 27 portfolios according to their size classifications (small, medium, and large). The period is the interval from July 1963 to December 2015.

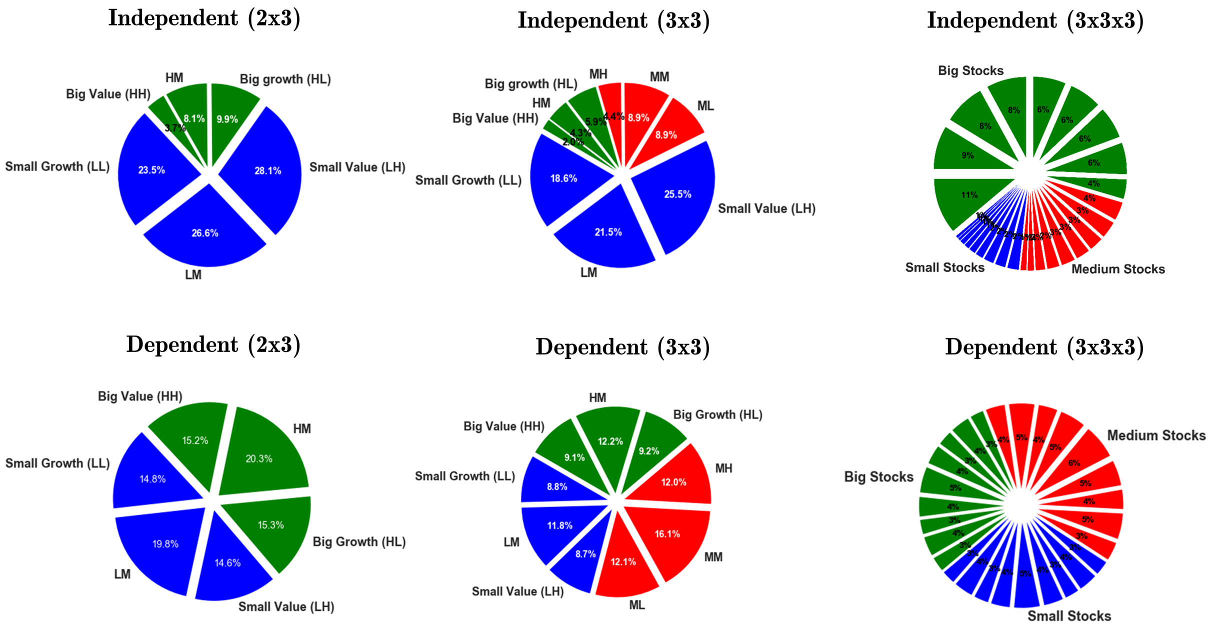
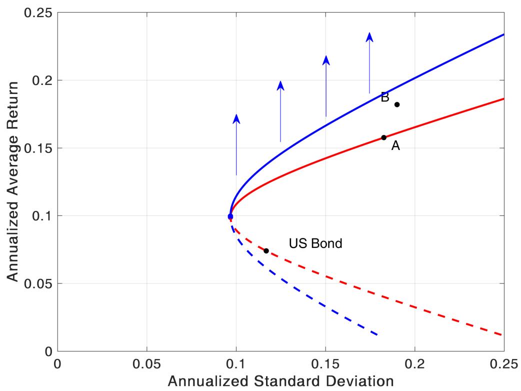


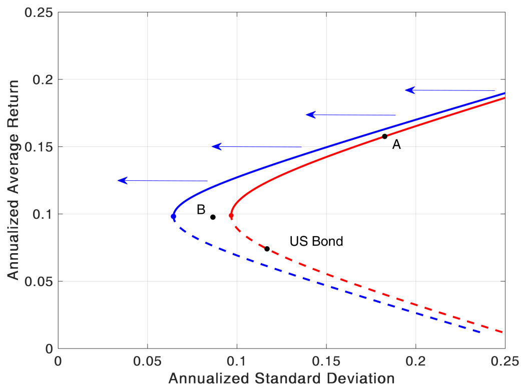
Figure 3. Improving the MSR (**a**) and GMV (**b**) Portfolios

The figure displays the spanning illustration for opportunity sets comprising the benchmark assets (R_1), i.e., the 30-Year U.S. Treasury Bond and Portfolio A, in the color red. The benchmark assets plus a test asset (R_2), i.e., Portfolio B, are displayed in the color blue. The x-axis reports the annualized standard deviation (in %), and the y-axis reports the annualized average return (in %). This example is fictitious but illustrates in Figure A (Figure B) an improvement of the MSR (GMV) portfolio after Portfolio B is added to the benchmark assets.

(a) MSR Portfolio



(b) GMV Portfolio



Tables

Table I
Correlation Between Characteristic-Sorted Portfolios

The table reports the average correlation (in %) for the characteristic-sorted portfolios constructed using independent and dependent sorting methodologies. The third column specifies the difference in the average correlation between the independent and dependent sorting results. Correlations are estimated based on daily returns, and the sample period extends from 01/07/1963 to 31/12/2015.

#Number of portfolios	Independent Sort (1)	Dependent Sort (2)	Difference (1)-(2)
	Panel A: Cap-weighted Portfolios		
2×3	84.99	78.00	6.99
3×3	84.99	75.81	9.18
3×3×3	78.38	66.8	11.58

Table II
List of the Smart Beta Strategies' Objective Functions

The table decomposes the smart beta strategies' objective function applied on the constituents' weights. The first column refers to the common name of the strategy. The second column specifies the main authors who have analyzed the strategy. The third column reports the objective function for minimization or maximization. In the objective function, w refers to the weights, N is the total amount of assets introduced in the optimization, i and j denote the i-th asset and the j-th asset, σ_{ij} is the covariance between the i-th asset and j-th asset, p refers to portfolio, and $(\Sigma w)_i$ is the risk contribution of the i-th asset. All objective functions are submitted to long-only budget constraints, i.e., $w_i \in [0,1]$ and $\sum_{i=1}^{N} w_i = 1$.

Strategy	Referenced Authors	Objective Function
Minimum Variance (MV)	Clarke et al. (2013)	$\min\ f(w) = \sum_i^N \sum_j^N w_i \sigma_{ij} w_j$
Maximum Diversification (MD)	Choueifaty and Coignard (2008)	$\max f(w) = \frac{\sum_i^N w_i \sigma_i}{\sqrt{\sum_i^N \sum_j^N w_i \sigma_{ij} w_j}}$
Risk parity (RP)	Maillard et al. (2010)	$\min\ f(w) = \sum_i^N \sum_j^N (w_i \times (\Sigma w)_i - w_j \times (\Sigma w)_j)^2$

Table III
Diversification Returns: Equal-Weight Benchmark

The table reports the spread of diversification return obtained from equation (5) for the three different strategies: MD, MV, and RP. These strategies are applied to portfolios that are sorted independently (ind) or dependently (dep). These portfolios are rebalanced on a monthly basis and the number of portfolios is either six (2×3), nine (3×3) or twenty-seven (3×3×3). The components of diversification return are reported in percentage and on a monthly basis. The sample period extends from July 1963 to December 2015. We then provide the p-value of the hypothesis that the spread in the component of diversification are equivalent for a pair of strategies applied on independent or dependent portfolios. To extract estimate a p-value for this static measures, we use the framework on hypothesis testing with the Sharpe ratio from Ledoit and Wolf (2008) and substitute the Sharpe ratio by the measures of diversification return. The p-values identified by *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively.

	Δ Dep-Ind					
	DR_1^{FW}	DR_2^{FW}	DR_3^{FW}	DR_1^{EW}	DR_2^{EW}	DR^{EW}
MD_{2x3}	0.002	0.012***	0.079***	0.083***	0.011***	0.094***
MD_{3x3}	-0.035	0.019***	0.091***	0.060	0.015***	0.075*
MD_{3x3x3}	-0.034	0.034***	0.112**	0.097	0.014***	0.112*
MV_{2x3}	-0.038	0.007	0.162***	0.136*	-0.004	0.131*
MV_{3x3}	-0.035	0.020***	0.119***	0.093	0.012**	0.105*
MV_{3x3x3}	-0.097*	0.023***	0.089***	-0.003	0.018***	0.015
RP_{2x3}	0.000	0.012***	0.035***	0.036**	0.010***	0.047***
RP_{3x3}	-0.005	0.014***	0.032***	0.029*	0.013***	0.041**
RP_{3x3x3}	-0.009	0.023***	0.031***	0.024	0.022***	0.046***

Table IV
Spanning Tests with Multiple Factor: Full Sample

The table reports the results for the bootstrap mean-variance spanning test from Kan and Zhou (2012). The mean-variance test goes as follow: we test whether a benchmark portfolio R_1 have a significant improvement at the tangent (F_1), or at the GMV (F_2) portfolio level when a test asset (R_2) is added to the benchmark assets (R_1). The test is performed twice, given that we have two proxies for R_1, that is a smart beta strategy on independent-sorted or dependent-sorted portfolios. The outcomes of the test are the following, (i) the abnormal return of the candidates (α), (ii) the F-tests and, (iii) the bootstrap p-values that control for multiple testing for which *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively. The regression models are as follow:

$$
\begin{aligned}
&(1) \quad R_1 = MKT + B30, \quad R_2 = SB \\
&(2) \quad R_1 = MKT + B30 + SMB + HML, \quad R_2 = SB \\
&(3) \quad R_1 = SB + B30, \quad R_2 = Mkt
\end{aligned}
$$

Results presented below are composed of 1,000 simulations for each smart beta (SB) strategy, i.e. maximum diversification (MD), minimum variance (MV), risk parity (RP). All strategies are taken in excess of the risk-free rate except the long-short size (SMB) and value (HML) factors. B30 refers to the 30-Year U.S. Treasury Bonds in excess of the risk-free rate. The sample period is composed of monthly returns from July 1963 to December 2015.

Models	(1)			(2)			(3)		
	α	F_1	F_2	α	F_1	F_2	α	F_1	F_2
Panel A: Independent									
MD_{2x3}	0.20%	11.69***	2.37	0.01%	0.21	1419.96***	-0.13%	5.92**	7.93**
MD_{3x3}	0.22%	11.89***	2.96	0.02%	0.66	1392.77***	-0.14%	5.51**	9.40**
MD_{3x3x3}	0.19%	7.21**	1.42	-0.03%	0.65	824.06***	-0.10%	2.22	17.25***
MV_{2x3}	0.32%	13.63***	11.83**	0.09%	2.22	317.78***	-0.17%	3.97*	12.51***
MV_{3x3}	0.29%	10.60***	6.68*	0.06%	1.02	438.93***	-0.14%	2.63	18.31***
MV_{3x3x3}	0.29%	13.60***	10.24**	0.07%	2.22	564.35***	-0.17%	4.60**	10.31**
RP_{2x3}	0.21%	13.03***	1.62	0.02%	0.92	1786.24***	-0.14%	6.82**	9.52**
RP_{3x3}	0.22%	11.57***	0.49	0.02%	0.88	2015.93***	-0.14%	5.35**	16.33***
RP_{3x3x3}	0.23%	12.09***	0.25	0.03%	1.16	1895.22***	-0.14%	5.50**	19.14***
Panel B: Dependent									
MD_{2x3}	0.31%	13.00***	13.52***	0.11%	4.89**	578.95***	-0.16%	4.20*	8.11**
MD_{3x3}	0.34%	10.66***	21.30***	0.11%	3.13	365.03***	-0.13%	1.99	10.40**
MD_{3x3x3}	0.38%	9.45***	19.39***	0.12%	2.22	285.69***	-0.10%	0.92	19.97***
MV_{2x3}	0.49%	15.55***	23.12***	0.22%	6.21**	249.22***	-0.16%	2.31	21.01***
MV_{3x3}	0.44%	13.39***	28.21***	0.18%	4.59*	254.99***	-0.14%	1.87	15.30***
MV_{3x3x3}	0.34%	10.02***	19.87***	0.10%	1.92	310.91***	-0.12%	1.45	14.54***
RP_{2x3}	0.27%	10.60***	7.48**	0.07%	3.21	882.21***	-0.14%	3.44*	10.80***
RP_{3x3}	0.29%	8.70***	8.49**	0.07%	2.36	801.67***	-0.11%	1.81	16.07***
RP_{3x3x3}	0.31%	10.28***	7.34**	0.10%	4.27*	811.01***	-0.13%	2.46	17.64***

Table V
Spanning Tests with Mutliple Factor: Sub Sample

The table reports the results for the bootstrap mean-variance spanning test from Kan and Zhou (2012). The mean-variance test goes as follow: we test whether a benchmark portfolio R_1 have a significant improvement at the tangent (F_1), or at the GMV (F_2) portfolio level when a test asset (R_2) is added to the benchmark assets (R_1). The test is performed twice, given that we have two proxies for R_1, that is a smart beta strategy on independent-sorted or dependent-sorted portfolios. The outcomes of the test are the following, (i) the abnormal return of the candidates (α), (ii) the F-tests and, (iii) the bootstrap p-values that control for multiple testing for which *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively. The regression models are as follow:

$$
\begin{aligned}
&(1) \quad R_1 = MKT + B30, \quad R_2 = SB \\
&(2) \quad R_1 = MKT + B30 + SMB + HML, \quad R_2 = SB \\
&(3) \quad R_1 = SB + B30, \quad R_2 = Mkt
\end{aligned}
$$

Results presented below are composed of 1,000 simulations for each smart beta (SB) strategy, i.e. maximum diversification (MD), minimum variance (MV), risk parity (RP). All strategies are taken in excess of the risk-free rate except the long-short size (SMB) and value (HML) factors. B30 refers to the 30-Year U.S. Treasury Bonds in excess of the risk-free rate. The sample period is composed of monthly returns from July 1993 to December 2015.

Models		(1)			(2)			(3)	
	α	F_1	F_2	α	F_1	F_2	α	F_1	F_2
				Panel A: Independent					
MD_{2x3}	0.20%	3.77	6.48*	0.01%	1.52	384.37***	-0.13%	0.74	3.25
MD_{3x3}	0.22%	5.08*	7.26**	0.02%	3.71*	382.82***	-0.14%	1.10	4.15
MD_{3x3x3}	0.19%	1.86	5.87	-0.03%	0.02	249.23***	-0.10%	0.02	7.24*
MV_{2x3}	0.32%	9.03***	19.93***	0.09%	6.52**	42.73***	-0.17%	1.35	5.11*
MV_{3x3}	0.29%	7.55**	11.39***	0.06%	5.89**	92.31***	-0.14%	0.88	10.74**
MV_{3x3x3}	0.29%	11.36***	18.15***	0.07%	11.43***	124.93***	-0.17%	2.54	4.27*
RP_{2x3}	0.21%	4.73*	4.94	0.02%	3.50	563.63***	-0.14%	1.15	4.53*
RP_{3x3}	0.22%	5.02**	3.44	0.02%	5.07*	652.92***	-0.14%	1.10	7.88**
RP_{3x3x3}	0.23%	6.19**	4.37	0.03%	7.72**	618.11***	-0.14%	1.53	7.3**
				Panel B: Dependent					
MD_{2x3}	0.31%	12.7***	16.12***	0.11%	16.78***	118.87***	-0.16%	3.02	5.50*
MD_{3x3}	0.34%	9.14**	16.55***	0.11%	9.31***	63.08***	-0.13%	0.89	11.18***
MD_{3x3x3}	0.38%	9.38**	16.66**	0.12%	8.88***	39.76***	-0.10%	0.65	15.00***
MV_{2x3}	0.49%	14.13***	17.8***	0.22%	14.39***	44.38***	-0.16%	1.08	21.32***
MV_{3x3}	0.44%	13.10***	21.37***	0.18%	13.19***	43.67***	-0.14%	1.04	16.16***
MV_{3x3x3}	0.34%	9.92***	19.77***	0.10%	9.86***	51.57***	-0.12%	0.89	11.11***
RP_{2x3}	0.27%	9.39***	9.07**	0.07%	14.80***	217.80***	-0.14%	2.27	7.03**
RP_{3x3}	0.29%	6.64**	7.98**	0.07%	9.23***	174.22***	-0.11%	0.70	13.32***
RP_{3x3x3}	0.31%	8.69***	9.42**	0.10%	13.91***	158.94***	-0.13%	1.34	11.52***

Table VI
Horse Race Boostrap Test

The table reports the results for the bootstrap mean-variance spanning test from Kan and Zhou (2012). The mean-variance test goes as follow: we test whether a benchmark portfolio R_1 have a significant improvement at the tangent (F_1), or at the GMV (F_2) portfolio level when a test asset (R_2) is added to the benchmark assets (R_1). The test is performed twice, given that we have two proxies for R_1, that is a smart beta strategy on independent-sorted or dependent-sorted portfolios. The outcomes of the test are the following, (i) the abnormal return of the candidates (α), (ii) the F-tests and, (iii) the bootstrap p-values that control for multiple testing for which *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively. The regression models are as follow:

$$\begin{aligned} &(1) \quad R_1 = B30 + SB^{net}_{dep}, \quad R_2 = SB^{net}_{ind} \\ &(2) \quad R_1 = B30 + SB^{net}_{ind}, \quad R_2 = SB^{net}_{dep} \end{aligned}$$

Results presented below are composed of 1,000 simulations for each smart beta (SB) strategy, i.e. maximum diversification (MD), minimum variance (MV), risk parity (RP). All smart beta strategies are net of transaction costs, which are estimated as in Hasbrouck (2009). B30 refers to the 30-Year U.S. Treasury Bonds. The sample period is composed of monthly returns from July 1963 to December 2015.

	Model (1)			Model (2)			MSR	GMV
	α	F_1	F_2	α	F_1	F_2	Candidate	Candidate
MD_{2x3}	-0.05%	0.80	0.06	0.13%	5.82**	11.17**	Dependent	Dependent
MD_{3x3}	-0.01%	0.02	0.12	0.14%	4.16*	17.58***	Dependent	Dependent
MD_{3x3x3}	0.00%	0.00	2.47	0.23%	5.96**	18.9***	Dependent	Dependent
MV_{2x3}	0.03%	0.12	6.52*	0.20%	5.25**	10.17**	Dependent	Dep $\approx$ Ind
MV_{3x3}	-0.01%	0.04	0.09	0.19%	6.10**	24.21***	Dependent	Dependent
MV_{3x3x3}	0.08%	1.49	3.21	0.06%	0.79	9.72**	Dep $\approx$ Ind	Dependent
RP_{2x3}	0.00%	0.01	0.30	0.06%	1.74	5.81*	Dep $\approx$ Ind	Dependent
RP_{3x3}	0.01%	0.04	0.12	0.07%	1.50	9.12**	Dep $\approx$ Ind	Dependent
RP_{3x3x3}	0.00%	0.01	0.09	0.08%	2.32	9.24**	Dep $\approx$ Ind	Dependent

Table VII
Lucky MVE Candidates

The table reports sequential test developed by Harvey and Liu (2019) to select robust MVE candidates. As MVE candidates, we use the traditional cap-weighted market portfolio obtained from Kenneth French's website, and the smart beta strategies constructed on the two sets of basis portfolios: one from an (ind)ependent sort and the other from a (dep)endent sort. All strategies are taken in excess of the risk-free rate and net of transactions costs. The test assets used in the MVE test are the basis portfolios used to construct the smart beta strategies. These ones are indicated next to the referenced Panel. The baseline model used in our test is the 30-Year U.S. Treasury Bonds in excess of the risk-free rate (B30), and the size (SMB) and value (HML) factors obtained from Kenneth French's website. We report in the first row of each panels the Gibbons et al. (1989) test and in the second row its respective p-value; the third row presents the value of the scaled intercept (SI); the fourth row displays the single test p-value for SI; the fifth row shows the numbering sequence for which SI and its p-values are referring; the penultimate row reports the selected MVE candidates among the smart beta strategies; and the final row displays the p-value of the selected candidate when controlling for multiple testing. The sample period ranges from July 1963 to December 2015.

MVE Candidates →	Mkt	MV$_{dep}$	MD$_{dep}$	RP$_{dep}$	MV$_{ind}$	MD$_{ind}$	RP$_{ind}$	Mkt	MV$_{dep}$	MD$_{dep}$	RP$_{dep}$	MV$_{ind}$	MD$_{ind}$	RP$_{ind}$
	Baseline = B30 + SMB + HML							Baseline = B30 + SMB + HML						
	Panel A: 2x3 cap-weighted independent portfolios as test assets							Panel B: 2x3 cap-weighted dependent portfolios as test assets						
GRS	4.836	4.189	4.391	4.538	4.155	4.445	4.341	12.058	10.947	11.344	11.589	11.552	11.888	11.962
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scaled intercept (SI)	0.042	0.070	0.041	0.036	0.066	-0.012	-0.842	0.049	0.061	-0.838	0.023	0.009	-0.001	-0.008
Single test p-value	0.701	0.984	0.929	0.894	0.893	0.325	0.000	0.880	0.865	0.000	0.782	0.569	0.547	0.468
SI sequence	2	2	2	2	2	2	1	2	2	1	2	2	2	2
Selected candidate(s)				**RP**$_{ind}$							**MD**$_{dep}$			
Multiple test p-value				[0.000]							[0.000]			
	Panel C: 3x3 cap-weighted independent portfolios as test assets							Panel D: 3x3 cap-weighted dependent portfolios as test assets						
GRS	3.403	2.792	2.892	2.975	2.909	2.918	2.929	8.228	7.757	7.959	7.950	8.116	8.108	8.108
p-value	0.000	0.003	0.002	0.002	0.002	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scaled intercept (SI)	-0.006	0.041	0.255	0.373	-0.876	0.100	0.169	0.210	0.249	0.221	-0.806	0.001	0.150	0.107
Single test p-value	0.542	0.755	0.883	0.895	0.000	0.697	0.755	0.937	0.979	0.916	0.000	0.528	0.861	0.849
SI sequence	2	2	2	2	1	2	2	2	2	2	1	2	2	2
Selected candidate(s)				**MV**$_{ind}$							**RP**$_{dep}$			
Multiple test p-value				[0.000]							[0.000]			
	Panel E: 3x3x3 cap-weighted independent portfolios as test assets							Panel F: 3x3x3 cap-weighted dependent portfolios as test assets						
GRS	2.262	2.135	2.164	2.213	2.085	2.332	2.110	4.591	4.366	4.335	4.283	4.335	4.608	4.379
p-value	0.000	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scaled intercept (SI)	0.245	0.030	-0.790	0.301	0.283	0.111	0.225	-0.015	0.007	-0.668	-0.039	0.006	0.013	0.020
Single test p-value	0.943	0.661	0.000	0.970	0.962	0.816	0.952	0.566	0.571	0.000	0.431	0.610	0.686	0.739
SI sequence	2	2	1	2	2	2	2	2	2	1	2	2	2	2
Selected candidate(s)				**MD**$_{dep}$							**MD**$_{dep}$			
Multiple test p-value				[0.000]							[0.000]			

Table VIII
Spanning Tests with Multiple Factor: Alphabet Portfolios

The table reports the results for the bootstrap mean-variance spanning test from Kan and Zhou (2012) when basis portfolios are based on the "alphabet" characteristics. The mean-variance test goes as follow: we test whether a benchmark portfolio R_1 have a significant improvement at the tangent (F_1), or at the GMV (F_2) portfolio level when a test asset (R_2) is added to the benchmark assets (R_1). The test is performed twice, given that we have two proxies for R_1, that is a smart beta strategy on independent-sorted or dependent-sorted portfolios. The outcomes of the test are the following, (i) the abnormal return of the candidates (α), (ii) the F-tests and, (iii) the bootstrap p-values that control for multiple testing for which *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively. The regression models are as follow:

$$
\begin{aligned}
(1) \quad & R_1 = MKT + B30, \quad R_2 = SB \\
(2) \quad & R_1 = MKT + B30 + SMB + HML, \quad R_2 = SB
\end{aligned}
$$

Results presented below are composed of 1,000 simulations for each smart beta (SB) strategy, i.e. maximum diversification (MD), minimum variance (MV), risk parity (RP). All strategies are taken in excess of the risk-free rate except the long-short size (SMB) and value (HML) factors. B30 refers to the 30-Year U.S. Treasury Bonds in excess of the risk-free rate. The sample period is composed of monthly returns from July 1963 to December 2015.

Models		(1)			(2)	
	α	F_1	F_2	α	F_1	F_2
			Panel A: Independent			
MD_{2x3}	0.04%	6.84**	0.57	0.03%	4.70*	0.26
MD_{3x3}	0.04%	5.66**	0.06	0.03%	3.09*	4.66*
MD_{3x3x3}	0.05%	3.05	3.68	0.02%	0.47	23.15***
MV_{2x3}	0.01%	0.08	1.66	-0.01%	0.03	0.04
MV_{3x3}	0.07%	4.70*	0.07	0.07%	4.23*	0.14
MV_{3x3x3}	0.04%	0.65	0.18	0.02%	0.18	3.00
RP_{2x3}	0.04%	8.87***	0.78	0.04%	6.51**	0.20
RP_{3x3}	0.05%	9.34***	0.11	0.04%	5.68**	7.05**
RP_{3x3x3}	0.07%	9.11***	2.17	0.04%	3.69*	34.27***
			Panel B: Dependent			
MD_{2x3}	0.03%	2.52	0.08	0.02%	1.38	1.28
MD_{3x3}	0.05%	10.53***	0.01	0.04%	7.51**	3.39
MD_{3x3x3}	0.07%	8.71***	6.21**	0.04%	3.46*	59.95***
MV_{2x3}	0.03%	0.82	0.71	0.02%	0.25	1.91
MV_{3x3}	0.05%	2.44	0.51	0.05%	2.25	0.21
MV_{3x3x3}	0.02%	0.26	3.43	0.01%	0.02	5.63
RP_{2x3}	0.03%	2.48	0.11	0.02%	1.37	1.22
RP_{3x3}	0.05%	11.9***	0.19	0.04%	8.12***	3.96*
RP_{3x3x3}	0.07%	11.03***	7.96***	0.04%	4.46*	80.98***

Appendices

A. Estimation of the Covariance Matrix

In this section, we briefly describe a shrinkage methodology used in our applications to estimate the covariance with lower sampling errors following Ledoit and Wolf (2004). In their model, the authors build on Elton and Gruber (1973), who use a constant correlation coefficient to shrink the assets' covariance toward a global average correlation estimator.

The constant correlation coefficient is determined using

$$\hat{\rho} = \frac{1}{N(N-1)} \left(\sum_{i}^{N} \sum_{j}^{N} \hat{\rho}_{ij} - N \right) \tag{A.1}$$

where N is the number of portfolios – in our applications, either 6, 9 or 27. The term $\hat{\rho}_{ij}$ is the historical correlation estimate between the i^{th} portfolio and the j^{th} portfolio. Ledoit and Wolf (2004) then obtain an optimal structure for the covariance matrix and reduce the sampling error of a traditional sample covariance matrix (S) as follows:

$$\Sigma = \delta F + (1-\delta)S \tag{A.2}$$

where Σ is the output covariance matrix obtained from the shrinkage estimation, and δ is the optimal shrinkage intensity.[15] S is the sample covariance matrix from our 60 daily returns, and F is the structured covariance matrix with the assets' covariance estimated via the constant correlation estimator in equation (A.1).[16] In our empirical study, the estimations of the sample and the structured covariance matrices are based on 60-day rolling windows to accommodate for gradual changes in the return distribution and short-term variations. A real-life application with tradable assets (Idzorek & Kowara, 2013) would impose constraints on the historical information available to replicate our results. For this reason - to stay as close as possible to what real-world applications may offer - we limit our optimizations on 60-day windows. This choice is also consistent with Fama and French (2018) who estimate the monthly variance of stocks using 60 days of lagged returns.

B. Testing the Incremental Diversification Return

Ledoit and Wolf (2008) propose an "indirect" bootstrap methodology to construct an empirical distribution of the spread in a function of the underlying first and second moments of two time-series. They test the significance of the spread by considering whether a 1-α confidence interval (e.g., 90%) contains zero.

[15] Matlab code is available at Prof. Wolf's website.

[16] The covariance of the matrix F is given by $\sigma_{ij} = \hat{\rho}\sigma_i\sigma_j$.

The authors first consider that the difference between the true first and second moments of the two series converge towards their sample estimate such that,

$$\sqrt{T}(\hat{u} - u) \xrightarrow{d} N(0, \Omega) \tag{B.1}$$

where $\hat{u} = (\hat{\mu_i}, \hat{\mu_j}, \hat{\sigma_i}^2, \hat{\sigma_j}^2)$ are the sample estimates of $u = (\mu_i, \mu_j, \sigma_i^2, \sigma_j^2)$, $\xrightarrow{d}$ refers to the convergence in distribution of the parameters, T is the length of the time-series, and Ω refers to variance of the estimator distribution.

Considering the sample uncentered second moments instead of the sample estimated variances, i.e., $\hat{\gamma_i} = E(r_i^2)$ and $\hat{\gamma_i} = E(r_j^2)$, and taking into account non-normality and auto-correlation in returns, the relationship (B.1) becomes

$$\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0, \Psi) \tag{B.2}$$

where $\hat{v} = (\hat{\mu_i}, \hat{\mu_j}, \hat{\gamma_i}, \hat{\gamma_j})$ is the sample estimates of $v = (\mu_i, \mu_j, \gamma_i, \gamma_j)$.

The estimator Ψ is estimated through a heteroskedasticity and autocorrelation (HAC) robust kernel method. We refer to the paper of Ledoit and Wolf (2008) for a more detailed discussion on the computation of this estimator.

The standard error of the spread $\hat{\Delta}$ in a function $f(\hat{v})$ can be defined as,

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}{T}} \tag{B.3}$$

where $\nabla' f(\hat{v})$ is the gradient function of $f(\hat{v})$ and T is the length of the time-series.

To obtain a confidence interval attached to $\hat{\Delta}$, we resample the original time-series using the block-bootstrap method of Politis and Romano (1992) and construct an empirical (bootstrap) distribution of a studentized test statistic (d^b) defined as

$$d^b = \frac{|\hat{\Delta}^b - \hat{\Delta}|}{s(\hat{\Delta}^b)} \tag{B.4}$$

where the superscript b denotes the b-th bootstrap sample and where $s(\hat{\Delta}^b)$, for the b-th bootstrap is obtained by using both the gradient of $f(\hat{v}^b)$ and the HAC kernel estimator $\hat{\Psi}^b$ and defined as follows,

$$s(\hat{\Delta}^b) = \sqrt{\frac{\nabla' f(\hat{v}^b)\hat{\Psi}^b\nabla f(\hat{v}^b)}{T}} \tag{B.5}$$

The boostrap 1-α confidence interval is defined as:

$$\left[\hat{\Delta} - z^b_{|\cdot|,1-\alpha/2} s(\hat{\Delta}), \hat{\Delta} + z^b_{|\cdot|,1-\alpha/2} s(\hat{\Delta})\right] \tag{B.6}$$

with $z^b_{|.|,1-\alpha}$ the quantile of the distribution function of the studentized statistic estimated from the bootstrap and denoted $\mathcal{L}(d^b)$.

In our applications, we use a block-bootstrap of 10 observations and runs 4999 simulations.[17] The bootstrap process works as follow: First, we set a length for the block of observations (e.g., 10) that we want to resample in order to capture serial autocorrelation. Second, we match the length of the original time-series in the bootstrap samples to preserve the uncertainty and the degree of freedom from the original data. Third, we randomly resample (with replacement) the sequence of time-series for the b-th bootstrap and keep the same sequence for resampling the time-series of the strategies and their underlying opportunity sets. This way, we make sure to preserve the cross-sectional correlation across the assets (see, e.g. Fama & French, 2010; Harvey & Liu, 2019). Lastly, we repeat the operation B times, e.g., 4999, to construct an empirical distribution of centered studentized test statistics in which the standard error,

Defining a studentized test statistic (d) on the original time-series as follows,

$$d = \frac{|\hat{\Delta}|}{s(\hat{\Delta})} \tag{B.7}$$

The p-value attached to the test of the spread $\hat{\Delta}$ in a function $f(\hat{v})$ is computed as,

$$\text{p-val} = \frac{\#\{d^b \geq d\} + 1}{B + 1} \tag{B.8}$$

Ledoit and Wolf (2008) apply this framework to a test of Sharpe ratio for a pair of strategies. They consider the following function:

$$f(a, b, c, d) = \frac{a}{\sqrt{c - a^2}} - \frac{b}{\sqrt{d - b^2}} \tag{B.9}$$

Where $a = \hat{\mu_i}$, $b = \hat{\mu_j}$, $c = \hat{\gamma_i}$, and $d = \hat{\gamma_j}$.

The gradient of the function is defined as $\nabla' f(\hat{v}) = \left(\frac{c}{(c-a^2)^{1.5}}, -\frac{d}{(d-b^2)^{1.5}}, -\frac{1}{2}\frac{a}{(c-a^2)^{1.5}}, \frac{1}{2}\frac{b}{(d-b^2)^{1.5}} \right)$ is the gradient function of $f(\hat{v})$.

Our estimates of $\hat{\Delta}$ are defined as follows,

$$\begin{aligned} \hat{\Delta}(DR) &= DR^{Dep} - DR^{Ind} \\ \hat{\Delta}(DR_1) &= DR_1^{Dep} - DR_1^{Ind} \\ \hat{\Delta}(DR_2) &= DR_2^{Dep} - DR_2^{Ind} \end{aligned} \tag{B.10}$$

[17]Our results are not sensitive to the choice of the block length. As a matter of fact, we run the test with blocks of length $\{2, 4, 6, 8, 10\}$ and found very similar results (available upon request). Also, we run 4999 simulations to stay aligned with the recommendations of Ledoit and Wolf (2008, p. 858).

However obtaining the gradient for these functions is cumbersome because additionally to the pair of strategies, we are left with a number N of dependent-sorted and independent-sorted portfolios. Given that in our applications, this amount N can take the value of 6, 9 or 27, finding the gradient for this large amount of parameters is difficult. Consequently, we make the assumption that if there are large deviations in the spread of diversification return for a pair of strategies, then there should also be large deviations in their spread of Sharpe ratio. This assumption is helpful as we can now only substitute the numerator in equations (B.4) and (B.7) by the spread in diversification return while keeping the standard error from the spread in Sharpe ratio derived in the initial framework of Ledoit and Wolf (2008). To test the sensitivity of this assumption, we also substitute the standard error from the spread in the Sharpe ratio by the standard error from the spread in geometric return, and also by the standard error of the spread in geometric return scaled by the standard deviation. Their gradient function $(\nabla' f(\hat{v}))$ are respectively given by $(a+1, -b-1, -0.5, 0.5)$ and $\left(\frac{0.5a^3-0.5ac-c}{a^2-c\sqrt{c-a^2}}, -\frac{0.5b^3+0.5bd+d}{b^2-d\sqrt{d-b^2}}, \frac{-0.25a^2+0.5a+0.25c}{a^2-c\sqrt{c-a^2}}, \frac{0.25b^2-0.5b-0.25d}{b^2-d\sqrt{d-b^2}}\right)$. We obtained qualitatively similar results under all robustness tests. Results are available upon request.

C. Multiple Test: A Bootstrap Approach

To test the robustness of our results, we extend the mean-variance spanning tests to address the multiple testing concern. We implement the bootstrap method used in Harvey and Liu (2019). The method proceeds in 4 steps:

Step 1: Orthogonalization Under the Null

The goal of this step is to modify the original times series of R_2 such that the null hypothesis appears to be true in-sample (Harvey & Liu, 2019; White, 2000). To do this, we perform the following regression,

$$R_2^t = \alpha + \beta R_1^t + R_{2,e}^t \tag{C.1}$$

Then we can work on last equation to obtain an orthogonal time-series, denoted by the subscript $\perp$, that satisfies $Qw_1 = 0$ and $Qw_2 = 0$ while preserving the dependence between R_2 and R_1 as follow,

$$R_{2,\perp}^t = \beta_\perp R_1^t + R_{2,e}^t \tag{C.2}$$

where the term $\beta_\perp = \frac{\beta}{\beta 1_K}$ is a simple re-scale of the original vector of slopes (β) that satisfies $\delta = 1 - \beta_\perp 1_K = 0$. Moreover, one can easily identify that last equation also satisfies $\alpha = 0$. We use this new time-series $R_{2,\perp}$ in our bootstrap to estimate the statistical validity of the F-values from both hypothesis tests (H_0^1 and H_0^2).Note that only H_0^2 is a joint test that $\delta = 1_N - \beta 1_K = 0_N$ conditional on $\alpha = 0_N$.

Step 2: Bootstrap

The bootstrap procedure is a random selection of monthly observations of the strategies with replacement (i.e., R_1 and $R_{2,\perp}$). We jointly resample the monthly observations to preserve the cross-sectional correlations across strategies, as in Fama and French (2010). Also, we make sure that the new time-series have the same size as the original time frame (630 months) to ensure that the degrees of freedom in the measurements of the bootstrap F-tests remain equal to the F-test from the original sample (Harvey & Liu, 2019).

Step 3: MVE spanning test

We apply the mean-variance spanning from Kan and Zhou (2012) on the bootstrapped samples according to the benchmark assets (R_1) and the test asset ($R_{2,\perp}$). We repeat the operation B times (1,000) to construct an empirical distribution of the performance measures. In these bootstrap samples, the null is valid in-sample, and a significant value for the F-tests simply arises from the resampling (or luck). The empirical distribution serves as a threshold for the critical value of the F-tests. Each bootstrap contains four F-tests: two for the tangent ($F^b_{1,ind}$ and $F^b_{1,dep}$) and two for the GMV ($F^b_{2,ind}$ and $F^b_{2,dep}$), with the subscript b denoting the b-th bootstrapped sample while *ind* and *dep* denote independent and dependent-sorted portfolios.

Step 4: Controlling for multiple testing

To control for multiple testing, we follow the framework of Harvey and Liu (2019) and adjust the confidence intervals of the original F-tests by keeping for each bootstrap the maximal measure of each hypothesis (tangent and GMV). For instance for the tangent hypothesis, the reference point for the b-th bootstrap is $F_1^b = max(F^b_{1,ind}, F^b_{1,dep})$. Hence, we take care of the multiple testing issue by comparing the distribution of the B maximal statistic measures to the ones found in the original sample $F^o_{1,ind}$ and $F^o_{1,dep}$, where o denotes the original sample.

The frequency of observations in the bootstrap sample that are greater than the F-test under the original sample defines the bootstrap p-value. Thus, the p-value is the sum of indicator value $I\{F_1^o < F_1^b\}$ divided by the total number of bootstraps B.

D. Transaction Costs

To consider transaction costs, we follow an approach similar to that of Novy-Marx and Velikov (2016) and use the individual stock estimates from the Gibbs sampling developed in Hasbrouck (2009). This approach is practically useful as we trade stocks on NYSE-NASDAQ-AMEX exchanges and consequently have to differentiate between transaction costs for small and large-cap stocks.

Novy-Marx and Velikov (2016) uncover a minor drawback to Hasbrouck's estimation technique, which requires relatively long series of daily prices to perform the estimation (250 days), resulting in a number of missing observations (mostly for non-NYSE stocks), for which the authors perform a non-parametric matching method and attribute equivalent transaction costs to the stock with a missing value according to its closest match to a stock with non-missing value according to their size and idiosyncratic volatility. However, according to the authors, these missing observations represent only 4% of the total market capitalization universe. Instead, we replace the missing values with a transaction cost of 0.50%. We employ this value because (1) we see from Figure E.4 that only a very little amount of estimates from Hasbrouck's algorithm have breached a trading cost of 50 bps since 1963, (2) this choice will more strongly impact illiquid stocks with a small number of daily observations (small-capitalization stocks), and (3) Plyakha et al. (2015) also choose to set this threshold for transaction costs from 1993 onwards.

In Figure E.4, we show the annual box-and-whisker plot for the CRSP/Gibbs estimates of transaction costs (variable c from equation (E.4)) from 1963 to 2015.

[Figure 4 about here.]

The next subsection describes the estimations process of the effective costs as in Hasbrouck (2009).

E. Transaction Costs: Gibbs Estimates

A traditional model to estimate the trading costs of a security is documented by Roll (1984) and simply use the autocovariance of the change in trade price (Δp_t) to find an effective estimate of spread such that,

$$c^{Roll} = \begin{cases} \sqrt{-Cov(\Delta p_t, \Delta p_{t-1})} & \text{if } Cov(\Delta p_t, \Delta p_{t-1}) < 0 \\ 0 & \text{if } Cov(\Delta p_t, \Delta p_{t-1}) \geq 0 \end{cases} \tag{E.1}$$

In the last equation, we see that when the autocovariance is positive, the model fails to provide a fair estimate of effective costs. For this reason, Hasbrouck (2009) extends the measure under Roll (1984)'s framework on the price dynamics in a market with transaction costs. In this framework, the model only requires information about the daily trade price, the prior midpoint of the bid-ask prices, and the sign of trade to perform the estimation. Formally, the price dynamic is written as follows:

$$\begin{aligned} m_t &= m_{(t-1)} + u_t \\ p_t &= m_t + cq_t \end{aligned} \tag{E.2}$$

where m_t is the log midpoint of the prior bid-ask price (the efficient price), p_t is the log trade price (the real price), q_t is the sign of the last trade of the day (+1 for a buy and −1 for a sale), c is the effective cost, and u_t is assumed to be unrelated to the sign of the trade (q_t).

Since we use the logarithm for the price variables in equation (E.2), the daily change in price is given by

$$\begin{aligned}\Delta p_t &= p_t - p_{t-1} \\ &= m_t + cq_t - m_{t-1} - cq_{t-1} \\ &= c\Delta q_t + u_t\end{aligned} \tag{E.3}$$

Hasbrouck (2009) extends Roll (1984)'s model with a market factor to capture a larger part of the changes in prices not due to transaction costs. They estimate the effective trading costs using Bayesian Gibbs sampling applied to the daily prices of U.S. equities retrieved from CRSP data.[18] The market-factor model is presented as follows:

$$\Delta p_t = c\Delta q_t + \beta_{rm} rm_t + u_t \tag{E.4}$$

where rm_t is the market return on day t and β_{rm} is the parameter estimate obtained from a Bayesian regression on the market return.

The Bayesian methodology estimates the effective costs (c) based on a sequence of iterations where the initial prior for c is strictly positive and follows a normal distribution with a mean of 0.01 and variance equal to 0.01^2, denoted N^+ ($\mu = 0.01$, $\sigma^2 = 0.01^2$). This initial prior of β_{rm} follows a normal distribution with mean and variance of 1, i.e. N ($\mu = 1$, $\sigma^2 = 1$) and the prior of σ_u^2 follows an inverted Gamma distribution initiated at IG ($\alpha = 10^{-12}$, $\beta = 10^{-12}$).[19] The objective of the Gibbs sampling is to estimate the value of the parameters c and β_{rm} conditional on the values drawn for q_t, which is based on the sign of trade (Δp_t), and the error term (u_t). Initially, q_1 is set to +1 and σ_u^2 is set to 0.001. Next, the sampler runs as follow,

for 1 to 1,000 sweeps

1. Perform a Bayesian OLS regression on a 250-day of lagged observations to estimate the new values of c and β_{rm}, update the posterior distribution of the parameters and make a new draw of the coefficients.

[18]The SAS code is available on Prof. Hasbrouck's website.

[19]These initial values of the priors are the ones found in the SAS code made available by Prof. Hasbrouck. According to Hasbrouck (2009), the initial values of the prior should not impact the final estimate of the effective cost of a stock because the first 200 iterations (of 1,000) are disregarded to compute the average of the estimated values for the trading cost (c).

2. Back out u_t from the values of c, β_{rm}, Δp_t, rm_t, Δq_t as follow,

$$u_t = \Delta p_t - \beta_{rm} rm_t - c\Delta q_t \tag{E.5}$$

3. Update the posterior σ_u^2 according to the series of u_t,

4. Draw new series of q_t knowing the new value of σ_u^2. Given that $u_t = \Delta p_t - \beta_{rm} rm_t - cq_t + cq_{t-1}$, estimate u_t if $q_t = +1$ or $q_t = -1$. Find the probability of $u_t(q_t = +1)$ and $u_t(q_t = -1)$ given that $u_t \sim N(0, \sigma_u^2)$ and compute the odds ratio for a buy order as follow,

$$\text{Odds} = \frac{f(u_t(q_t = +1))}{f(u_t(q_t = -1))} \begin{cases} q_t = +1 & \text{if Odds} > 1 \\ q_t = -1 & \text{if Odds} < 1 \end{cases} \tag{E.6}$$

end

The process is repeated 1,000 times and the final value for c is the average of the last 800 estimations of the procedure ("burn in" the 200 first observations). For more information on simulating the probability distributions of q_t and u_t as well as on the iterative process, interested readers should refer to Hasbrouck (2009, p. 1449-1951).[20]

[20]Further details regarding the application of the estimation technique can also be found in Marshall et al. (2011) and Novy-Marx and Velikov (2016).

Figures

Figure E.4. Variation of Transaction Cost Estimates

The figure presents a boxplot of the distribution of individual stock transaction costs estimated as in Hasbrouck (2009). The sample period is the interval from 1963 to 2015. The whiskers represent the distribution of the 5^{th} to 95^{th} percentile, and the upper and lower edges of the boxes correspond to the 25^{th} and 75^{th} percentiles. The gray dots represent outliers.