An ERI Scientific Beta Publication

Scientific Beta
An EDHEC-Risk Institute Venture

Methodological Differences across Multi-Factor Index Offerings

August 2016

The authors can be contacted at contact@scientific beta.com.

Table of Contents

About the Authors

Frédéric Ducoulombier is Associate Professor of Finance at EDHEC-Risk Institute and Corporate Director at ERI Scientific Beta. He has most recently co-authored EDHEC-Risk Institute's contributions to international regulatory consultations on ETFs, indices and benchmarks as well as publications on the purported risks of ETFs, non-financial risks in fund management, index transparency and governance and factor investing. Prior to joining ERI Scientific Beta, he established the executive education arm and Asian operations of EDHEC-Risk Institute and EDHEC Business School's PhD in Finance. He serves on the Consultative Working Group of the Financial Innovation Standing Committee established by the European Securities and Markets Authority.

Felix Goltz is Research Director, ERI Scientific Beta, and Head of Applied Research at EDHEC-Risk Institute. He carries out research in empirical finance and asset allocation, with a focus on alternative investments and indexing strategies. His work has appeared in various international academic and practitioner journals and handbooks. He obtained a PhD in finance from the University of Nice Sophia-Antipolis after studying economics and business administration at the University of Bayreuth and EDHEC Business School.

Jakub Ulahel is a Quantitative Equity Analyst with ERI Scientific Beta. He holds an MSc in Financial Markets from EDHEC Business School in France, where he was included in the Dean’s list. His work focuses on the analysis of alternative equity portfolios. Previously, he worked as Business Analyst for Microsoft after graduating from the Charles University in Prague with a bachelor´s degree in Economics.

Introduction

Introduction

Multi-factor indices are one of the newest additions to the family of rules-based equity strategies, and have attracted substantial inflows over the last couple of years. Factor investing advocates focusing on exposure to the underlying drivers of security returns identified in academic research, such as value, momentum, and other factors. With the realisation that the returns from active management are to a very large extent attributable to exposure to well-documented systematic factors, factor indices are increasingly regarded as a cost-efficient, straightforward and transparent way of implementing desired factor tilts.

In particular, while discretionary active managers may generate performance from tilting their portfolios towards such factors, they may destroy the performance benefits of these factor tilts if they charge the high fees that are typical of active management and/or make wrong idiosyncratic calls. For example, Fama and French (2010) find that active managers on average deliver -1% return per year after adjusting for their market, value, size and momentum exposures. If investors could access index-based strategies that simply delivered the returns of their factor-tilted benchmark at low fees, this would be an interesting alternative to investing with discretionary managers, who the evidence shows underperform their factor-tilted benchmark.

Multi-factor indices are a natural extension of indices based on single factors. While any single-factor index typically targets improving risk-adjusted returns over cap-weighted reference indices, there is strong intuition suggesting that multi-factor allocations will tend to result in improved risk-adjusted performance relative to single factor indices. Intuitively, since different factors work at different times, allocating across multiple rewarded factors will increase the probability of over-performance in the short- and medium-term. Moreover, investors who allocate across factors using single-factor indices subsumed into a multi-factor solution or held within the same mandate will enjoy implementation benefits. Indeed, some of the rebalancing trades necessary to maintain exposure to different factors may actually cancel each other out. Consider the classic example of an investor who pursues an allocation across the value and momentum tilts. If a stock included in the value strategy experiences a rapid price increase then, other things being equal, its weight in the value-tilted portfolio will tend to decrease at the next rebalancing, while its weight in the momentum-tilted portfolio will tend to increase. If both strategies are rebalanced at the same time and held within the same multi-factor index or mandate, then rebalancing trades can be crossed and turnover reduced.

While sharing the same objective, indices aiming to provide multiple factor exposures may opt for very different implementation methods, which reflect differences in underlying beliefs on multi-factor investing. This article reviews the current offerings in the world of multi-factor indices and looks at the conceptual considerations involved in designing the different approaches. The key issues that we discuss involve the robustness and consistency of the multi-factor indices as well as the (lack of) diversification among the various products.

Introduction

We first provide a brief overview of several multi-factor indices published by five different index providers. Afterwards, we discuss the design choices some of the indices in this group have made, as well as the conceptual underpinnings of these choices. In particular, we look at the difference between proprietary and consensual definitions of factors and the issue related to the use of composite factor scores. Subsequently we turn our attention to the importance of consistency in index design, before focusing on the important issue of diversification, which is too frequently ignored. We show the difference between top-down and bottom-up approaches to constructing multi-factor indices and discuss the issues relating to both.

Introduction

2. Overview of Multi-Factor Indices

2. Overview of Multi-Factor Indices

In this section, we provide a brief overview of the multi-factor indices that we are analysing. We look at the current multi-factor offerings from Scientific Beta, FTSE Russell, MSCI, Goldman Sachs and S&P.

In the next exhibit, we provide a short summary of the different index methodologies to provide the reader with a high-level overview of the situation. There are pronounced differences in methodology across the indices. For example, among other differences, some of the multi-factor indices that we look at use single-factor indices as building blocks while others apply a multi-factor methodology directly at the stock level. The choice of targeted factors also changes from index to index.

Exhibit 1: Overview of Multi-Beta indices

Provider	Index	Targeted factors	Short description of the methodology
Scientific Beta	Multi-Beta Multi-Strategy EW	Value, Size, Momentum, Volatility (4-factor version)	Equal-weighted combination of single factor indices.
	Multi-Beta Multi-Strategy ERC	Value, Size, Momentum, Volatility, Profitability, Investment (6-factor version)	The weight of single factor indices is determined so that the allocation equates the contribution of individual smart-factor portfolios to relative risk as represented by tracking error relative to the capitalisation-weighted reference.
FTSE Russell	FTSE Global Diversified Factor	Value, Volatility, Momentum, Size	Universe split into 40 regional industry universes (4 regions x 10 industries). Regional industry universes weighted in proportion to their inverse volatility. Factor ranking for individual security is a weighted average composite of individual factor rankings with weights proportional to inverse volatility of the factors.
	FTSE Russell Comprehensive	Quality, Value, Volatility, Momentum, Size	Z-score for individual factors translated into s-score using cumulative standard normal distribution. Composite factor score is a product of the 5 individual s-scores. Final weights in the index are proportionate to the product of the composite factor score and the market cap.
MSCI	MSCI Quality Mix	Value, Volatility, Quality	Equal-weighted combination of single factor indices.
	MSCI Diversified Multiple-Factor	Value, Momentum, Size, Quality	Optimisation focused on maximising an equal-weighted combination of factor exposure to the four rewarded factors. Optimisation controls exposure to non-rewarded factors and country and sector exposures.
Goldman Sachs	Goldman Sachs Equity Factor	Value, Momentum, Size, Quality, Low Beta	Optimisation-based index that aims to maximise aggregate basket score. Basket score is weighted average of individual factor scores weighted in proportion to the inverse volatility of the factors.
S&P	S&P GIVI	Value, Low Beta	Low beta stock selection (70%) and weighting by intrinsic value calculated from a discounted cash flow valuation model based on Residual Income.

It should be noted that additional differences exist among these indices that are not captured by the schematic overview above. These differences are notable when it comes to factor definitions. We will discuss the issues relating to factor definitions in the next section. Apart from this, indices naturally differ in terms of other implementation details including universe definition, implementation rules etc.

2. Overview of Multi-Factor Indices

In the next sections, we will review some conceptual issues related to the construction of multi-factor indices and provide examples of how the different indices included in our analysis approach these issues.

2. Overview of Multi-Factor Indices

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

Providers across the board put strong emphasis on the academic grounding of their factor indices[1]. At the same time, product providers try to differentiate their products using proprietary elements in their strategy, often leading to the creation of products using new factors or novel strategy construction approaches which may or may not be consistent with the broad consensus in the academic literature on empirical asset pricing. As for factor definitions, many factor indices show considerable divergence from academic definitions.

A key result from analysing the literature is that well-established factor premia are not simply based on "back-tests" similar to those used by product providers, but instead have been subjected to extensive empirical analyses, including assessments over very long-term data and post-publication data as well as cross-sample validation, notably when factors uncovered in the cross-section of U.S. stock returns have been confirmed in international data or in other asset classes. Moreover, some common factor premia have been explained using formal economic models providing a rationale for the persistence of such premia. Recalling these results is useful in clarifying the fact that novel strategies or "enhancements" with respect to such factors should be subjected to similar levels of scrutiny before conclusions on their relative merits are drawn.

For example, the Fama and French (2012, 2014) factor definitions, which are widely used in academic research, are based on straightforward stock selection criteria such as price-to-book for value for example, even though professionals often like to make the method more complex to pick the "best" value stocks. Extensive empirical evidence is available for the academic factor definitions, but not for the ad-hoc approaches, which are often justified by fairly limited in-sample performances.

More generally, for most factor or multi-factor offerings, product providers typically favour more complex factor definitions which may indeed reflect a stark disagreement with how academic research defines these factors. For example, some factor scores are calculated relative to the industry or regional groups a stock belongs to. Interestingly, some providers use such industry or regional adjustments for certain variables within a given factor score while not using it for other variables making up the same factor score. Moreover, providers often use variables which are quite far removed from the original factor definition, such as, for example, change in asset turnover in quality scores, as compared to the more straightforward profitability measures used in academic research. In fact, most of the Quality indices on offer have more to do with the precepts of stock-picking gurus than with the academic literature that has identified profitability and investment as asset pricing factors.

The next exhibit contains examples of selected indices among the ones included in this analysis, where the objective is to show how commonly-used definitions of factor indices deviate from the standard definitions used in the literature. Three different sets of definitions used by index providers are contrasted with the standard definitions widely employed in the literature. Rather than providing an exhaustive overview across all factors and index providers, we focus on selected examples where a deviation from academic consensus is easily apparent. However, the issue is not limited to these examples.

1 - For example, consider the following quotes from marketing material of index providers: "MSCI currently identifies six equity risk premia factors…. They are grounded in academic research…"; "In developing the Russell High Efficiency Factor Index.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

Exhibit 2: Mismatch with academic factor definitions – examples

Source	Value	Momentum	Profitability
Academic reference	Price-to-Book (Fama and French, 1993, 2012, 2015, Carhart, 1997)	Past 12 months returns omitting last month (Carhart, 1997, Fama and French, 2012)	Return On Equity[2] (Fama and French, 2015 and Hou, Xue, Zhang, 2015) or Gross Profitability (Novy-Marx, 2013)
Scientific Beta Multi-Beta Multi-Strategy indices	Follows Fama and French and Carhart	Follows Fama and French and Carhart	Follows Novy-Marx
Goldman Sachs Equity Factor Index World	Value score from proprietary risk model (Axioma), relative to stock's regional industry group	Residuals from cross-sectional regression of twelve-month return (omitting last month) on stock volatility	Composite based on Asset Turnover, Liquidity, Return On Assets, Operating Cash Flow-to-Assets, Accruals, Gross Margin, Leverage
MSCI Diversified Multiple-Factor	Sector-relative composite based on Enterprise Value/Operating Cash Flow, Forward P/E, Price-to-Book	Exposure from the Barra Equity Model based on 12-month relative strength (25% weight), 6-month relative strength (37.5% weight), historical alpha (37.5% weight)	Sector-relative composite based on Return on Equity, Earnings Variability, Debt-to-Equity
FTSE Global Factor Index Series	Composite of Cash Flow Yield, Earnings Yield and Sales-to-Price	Residual Momentum - Mean/Std. dev. of "avg. residual" from 11 rolling window regressions of past 36 months returns on country and industry index	Composite of Return on Assets, Change in Asset Turnover, Accruals and Leverage Ratio

When considering the descriptions in Exhibit 2, the mismatch of the provider definitions with the standard academic definitions is indeed striking. While the definitions found in the reference academic research rely on straightforward variables and make a choice of transparently and simply selecting one key metric to come up with a factor score for each stock, the proprietary definitions from most providers use different sets of variables, as well as various adjustments, and often consist of complex combinations of several variables.

The implications of the mismatch with academic factor definitions might not be immediately obvious. Nevertheless, any mismatch creates two problems. The first, which we have already mentioned, is that it is difficult to refer to academic evidence to justify one's factor offering and at the same time distance oneself from the empirical framework used for that same research with factor definitions that are different from those used by the researchers cited. The second is that this complexification and/or creation of ad-hoc proprietary factors is a source of potential data-mining problems. We discuss this second issue in the following two sub-sections. First, we look at selection bias, which relates to testing different variations in factor definitions. Afterwards, we discuss the issue of over-fitting bias, which may arise when using composite scoring to capture single or multiple factors.

3.1 Data-mining risk: selection bias with proprietary variables

Selecting proprietary combinations or making proprietary tweaks to variable definitions offers the possibility of improving the performance of a factor index in a back-test. In general, proprietary factor definitions increase the amount of flexibility providers have in testing many variations of factors and thus pose a risk of data-mining. In fact, it appears that providers

2 - Operating profits minus interest expense divided by book equity in Fama and French (2015) and income before extraordinary items divided by book equity in Hou, Xue and Zhang (2015).

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

sometimes explicitly aim at selecting ad-hoc factor definitions which have performed well over short-term back-tests.

The question is whether the improvement of the "enhanced" factor definition will also hold going forward, especially if there is no solid economic foundation for it. There is clearly a risk that one ends up with what academics have termed "lucky factors." Harvey and Liu (2015) show that by snooping through data on a large number of candidate factors and retaining those with the highest t-stat, one takes the risk of uncovering flukes, which will not repeat out of sample. Perhaps even more importantly, it is unclear what – if anything – factors with extensive proprietary tweaks still have in common with the factors from academic research. Therefore, the empirical evidence in favour of the academic factors and their economic grounding cannot be transposed to such new proprietary factors.

As an example of a proprietary factor found among the multi-factor indices that we analyse, consider the highly original factor definition of the value factor used in the S&P GIVI index. This index is marketed as a multi-factor index, but really only qualifies as a two-factor combination whose main goal is to create exposure to low beta stocks weighted by their "true" value – the intrinsic value. The intrinsic value used by S&P is based on valuation methods that use discounted cash flows, such as discounted residual income. This valuation is based on many assumptions and simplifications and is mechanically used across the entire spectrum of stocks without accounting for the subtleties that any valuation based on discounted cash flows entails. Even though the basic premise of value investing is to buy stocks whose market price understates their true value, proper intrinsic value estimation requires a lot of company-specific information, significant due diligence and intelligent estimates of several key variables. A unified, mechanical approach to hundreds of stocks using simplified assumptions may not truly reflect the intrinsic value of a diverse portfolio of stocks. The value factor definition used by S&P in this index is thus truly proprietary.

In the absence of a clear relationship to standard academic factors, such proprietary factor strategies can be regarded as ad-hoc constructs resulting from product back-tests. In fact, to find out whether any of these new proprietary factors are indeed related to the well-documented academic factors one would first need to assess how they align empirically with standard factors.

3.2 Illustration of selection bias: fishing for an enhanced value factor

We illustrate the problem with selection bias, otherwise known as "factor fishing", in the following section, where we consider alternatives to the standard value definition. The academic literature works with parsimonious and time-proven definitions of the value factor, the most popular being the book-to-market ratio, due largely to Fama and French (1993). When deviating from the established definitions, one has to keep in mind that these definitions have been confirmed by out-of-sample results ever since the seminal papers were published. This out-of-sample stability gives researchers and practitioners greater confidence that the uncovered value premium is not simply a product of a data-mining exercise.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

On the other hand, if we allow ourselves the flexibility of looking for the "best" or "improved" definition of value, such an exercise can easily lead to relying on promising in-sample results that do not hold out-of-sample.

Commercial back-tests are typically performed over a very short time-frame, in which around ten years of data is frequently used. Since different factor definitions will ultimately lead to different past performance, using a short time period to decide which one to pick might lead to unstable solutions.

Below, we illustrate the problems with variable selection and excessive reliance on back-tests with stylised examples. First, we study the change in back-tested performance over time, to see how (un)stable the results of variable selection are. Afterwards, we focus on the out-of-sample decay of performance benefits of in-sample variable selection.

We study the rolling spreads between the annualised performances of portfolios constructed based on different value proxies. This will allow us to gain a perspective on how back-tests may have looked like over the years and, more importantly, how the change over time impacted these results.

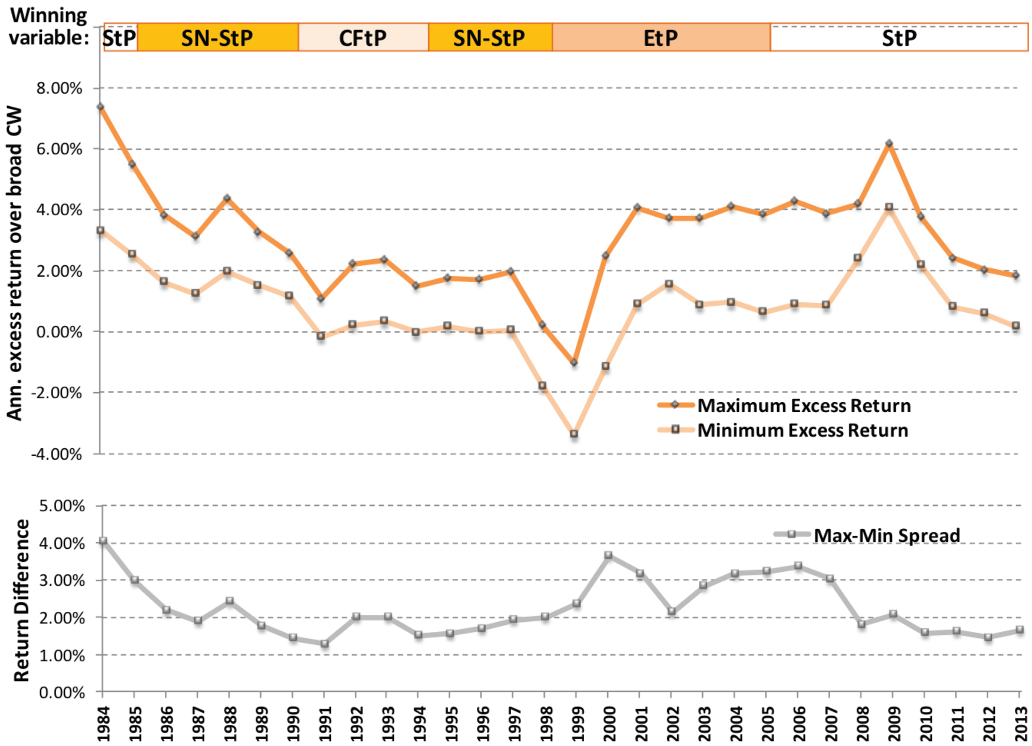
Our question is whether we can do better than using the book-to-market measure. We select among ten alternative value metrics – earnings-to-price, cash-flow-to-price, sales-to-price, dividend-to-price and payout-to-price, using both an unadjusted and a sector-neutral version for each. These metrics serve as the basis for forming a value-tilted portfolio where the portfolios simply select the 50 percent of stocks with the highest value score on an annual basis and cap-weight the selected stocks. The time series of these portfolios will serve as the basis for the empirical analysis below.

Consider the following exhibit. Every year, we look back ten years and plot the maximum and minimum annualised relative returns of ten value strategies over the broad cap-weighted index in that particular period. This is done on a rolling basis between 1984 and 2013. Every year thus represents a different potential starting point for a ten-year back-test. Naturally, the excess returns change over time, but one should pay closer attention to the changing spread between the maximum and the minimum.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

Exhibit 3: Extremes of annualised excess returns of ten cap-weighted Value strategies for ten year look-back periods
This chart plots the maximum and minimum of annualised excess returns with respect to the broad US market cap-weighted benchmark of 500 stocks (from CRSP) of annually-rebalanced cap weighted value tilted strategies with 50% stock selection out of the universe of 500 US stocks based on 10 Value variables - Earnings-to-Price, Cash-flow-to-Price, Sales-to-Price, Dividend-to-Price and Payout-to-Price, both plain-vanilla and sector-neutral versions for each. The analysis is based on daily total returns from 31/12/1973 to 31/12/2003. Ten year trailing returns are obtained with annual step size. Every year thus represents a ten-year back-test

PANEL A



Over the years, the difference in annualised returns of the possible back-tests ranges from slightly over 4% p.a. to a little under 1% p.a. This large spread between the different definitions suggests that considerable value can be added, at least within a back-test, when improving the variable selection. However, it is also worth noting that the best-performing variable changes over time, as the next table shows.

PANEL B – Best-performing variables based on 10- year rolling window back-tests

Period	Best-performing variable based on 10-year back-test
1984	Sales-to-Price
1985-1989	Sales-to-Price, sector-neutral
1990-1993	Cash-Flow-to-Price
1994-1997	Sales-to-Price, sector-neutral
1998-2004	Earnings-to-Price
2005-2013	Sales-to-Price

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

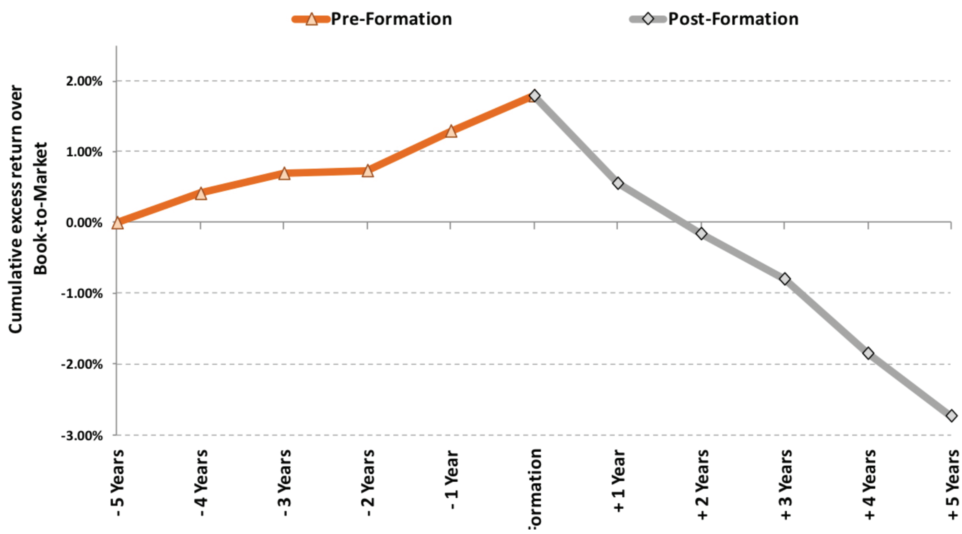
This clearly illustrates that back-tests that search for the best past performer over a short time period might be very unstable and caution should be exercised in evaluating strategies purely constructed on the basis of in-sample performance.

Below, we focus on the performance of strategies based on alternative value definitions relative to the performance of a portfolio based on book-to-market. We know that we can enhance back-tests, but should we? Our illustrations will reveal the out-of-sample decay of the data-mined solutions that rely on picking the best in-sample winners.

In the following exercise, we use a five-year formation window at the end of which we select the best-performing strategy based on its in-sample performance. Then, we hold the strategy for five years and compare the cumulative returns of this alternative strategy with respect to the portfolio based on the book-to-market measure. We do this every year between 1984 and 2009 to obtain 26 different event studies and we study the average performance.

Exhibit 4 shows the average cumulative relative returns of the best-performing alternative value definition with respect to book-to-market, both pre- and post-formation. As the chart below clearly shows, the average alternative variable definition ultimately underperforms book-to-market and drives the cumulative relative returns way below zero. Picking the past winner yields cumulative outperformance over book-to-market of +1.79% in sample. However, over the following five years, picking the in-sample winner leads to cumulative underperformance of -2.72% out of sample. This is evidence that searching for a better value definition in-sample does not beat book-to-market.

Exhibit 4: Comparison of cumulative relative returns of the average best in-sample alternative Value strategy with respect to a portfolio based on Book-to-Market
This chart plots the cumulative excess returns of ten annually-rebalanced cap-weighted value-tilted strategies with 50% stock selection out of the universe of 500 US stocks based on ten alternative value strategies, with respect to a similarly constructed portfolio based on Book-to-Market. Between 1984 and 2009, the five-year formation period is used to pick the best portfolio based on alternative Value definitions and this portfolio is held for another five years. This is done every year for a total of 26 event studies. The chart plots the average outperformance pre- and post-formation with respect to the Book-to-Market portfolio. The alternative value definitions are Earnings-to-Price, Cash-Flow-to-Price, Sales-to-Price, Dividend-to-Price and Payout-to-Price, both plain-vanilla and sector-neutral versions for each. The graph is smoothed by using yearly values.



3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

The previous exercise demonstrates that alternative value definitions hardly present a suitable replacement for book-to-market overall, based on the event study approach. To illustrate this point clearly and more convincingly, we now turn to simulating the experience of an actual investor in alternative value strategies.

Starting in 1984, we allow the investor to select the best-performing value variable, again using the ten alternative variable definitions specified above. After formation, the portfolio is held for a certain period and re-evaluated again at the end of it. We thus create active strategies that the investor sticks to for the duration of the holding period. We use two lengths of the calibration period (10 and 5 years) as well as four lengths of the holding period (2 to 5 years) for a total of eight active strategies, to capture the variability of the performance. We compare the performance of the active strategies based on alternative value definitions with a simple portfolio based on book-to-market in Exhibit 5. The results clearly show that none of the active strategies beats book-to-market, with the average active strategy lagging 61 basis points behind. Relative to their in-sample performance, the variable picking strategies on average create an out-of-sample degradation in performance of 128 basis points.

Exhibit 5: Performance of variable picking strategies for value tilted portfolios
This table shows the performance of 8 active strategies formed on the basis of a calibration period and held for a holding period from ten alternative Value definitions. The alternative Value definitions are Earnings-to-Price, Cash-Flow-to-Price, Sales-to-Price, Dividend-to-Price and Payout-to-Price, with both plain-vanilla and sector-neutral versions for each. At formation, the best-performing strategy based on the calibration period of 10 or 5 years is selected and held for a holding period of 2 to 5 years. The Book-to-Market portfolio is formed annually by cap-weighting the 50% selection of the stocks with the highest Book-to-Market ratio. The in-sample results select the best returns ex-post of the alternative Value strategies. All portfolios are based on the top 500 US stocks between 01/01/1984 and 31/12/2013.

	Book-to-Market	In-sample results			Out-of-sample results of variable picking strategies								
		10 Years	5 Years	Average	Calibration Period = 10 years				Calibration Period = 5 years				Average
					HP=5	HP=4	HP=3	HP=2	HP=5	HP=4	HP=3	HP=2	
Ann. Returns	13.1%	13.7%	13.9%	13.8%	12.7%	12.5%	12.7%	12.6%	12.4%	12.3%	12.3%	12.4%	12.5%
Ann. Volatility	19.0%	18.4%	17.9%	18.1%	18.6%	18.6%	18.6%	18.5%	18.1%	18.6%	18.4%	18.3%	18.5%
Sharpe Ratio	0.47	0.52	0.55	0.53	0.46	0.45	0.46	0.46	0.46	0.44	0.45	0.45	0.45
Return Difference with B-t-M	-	0.56%	0.78%	0.67%	-0.37%	-0.60%	-0.43%	-0.46%	-0.69%	-0.80%	-0.77%	-0.75%	-0.61%
Ret. Diff with "in-sample"	-0.67%	-	-	-	-0.93%	-1.17%	-1.00%	-1.03%	-1.47%	-1.58%	-1.55%	-1.53%	-1.28%

Overall, our empirical illustrations suggest that it is quite possible to enhance back-tests by selecting variables that "work" in sample. However, the strong out-of-sample degradation of performance suggests that such an approach leads to a risk of overstated back-test performance. We emphasise also that we consider our illustration to correspond to a mostly harmless data-mining experiment, which is likely to understate the actual bias that could result in more flexible data-mining exercises. In particular, we use a relatively small number of variables that remain economically sensible proxies for value, and which are by construction highly correlated among one another. Data-mining biases would obviously be much higher if we used a much larger number of variables, economically less sensible proxies or variables that are less correlated with one another.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

Ultimately and more generally, many value-tilted indices include other large sets of ad-hoc methodological choices, opening the door to data-mining. It can be argued that the use of straightforward single variables in factor definitions can be an effective safeguard against data-snooping or factor-fishing biases.

Our comparison of different approaches to factor definitions thus shows two different underlying philosophies. One is to avoid creating a mismatch with academic factors. This can be achieved easily by referring to indicators for which academic research has provided thorough tests and economic explanations, and by refraining from proprietary "tweaks." The benefit is that this approach creates indices that are directly based on the academic groundings of factor investing. It also allows investors to understand the return drivers of the indices, and make sure that the rationale and empirical evidence of such drivers has been confirmed in vastly scrutinised and independent academic research. That is the choice that EDHEC-Risk Institute made for the Scientific Beta smart factor indices.

Another approach is to try to improve upon the academic consensus through tweaked proprietary definitions. When using novel or proprietary factors, one needs to make sure that they are thoroughly tested (i.e. tested with very long-term data, across asset classes, for robustness to data-mining and to transaction costs) as well as linked to economic mechanisms. In contrast to academic factor definitions which have survived such analyses, the same amount of scrutiny has not necessarily been applied to proprietary tweaks, which carry the risk of being driven by marketing innovation rather than by genuine research advances. Therefore, investors should hold providers of proprietary factors to higher standards and conduct thorough due diligence on the soundness of their particular definitions.

3.3 More data-mining risk: over-fitting bias with composite scores

While the selection bias potentially exists for any strategy, there is an additional bias that is specific to so-called composite scoring approaches. These are factor definitions which draw on combinations of multiple variables. A recent paper by Novy-Marx (2015) analyses the bias inherent in back-tests of composite scoring approaches[3]. Novy-Marx argues that the use of composite variables in the design and testing of smart beta strategies yields a "particular pernicious form of data-snooping bias." He shows that creating a composite variable based on the in-sample performance of single variable strategies generates an over-fitting bias. To make matters worse, this over-fitting bias interacts with the selection bias. The presence of both biases in composite variable smart beta strategies increases the data-mining problems exponentially.

Novy-Marx analyses the bias occurring in an analysis of back-tested performance by considering strategies that combine random signals. The results show that combining signals that happened to perform well in the past leads to even better *past* performance. Given that signals are uninformative by construction, the past performance of these composite strategies of course does not imply any capacity to generate performance *out of sample*. The author concludes that "combining signals

3 - Novy-Marx cites the MSCI Quality Index and Research Affiliates Fundamental Indices as industry examples of such multi-signal approaches.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

that back-test positively can yield impressive back-tested results, even when none of the signals employed to construct the composite signal has real power."

The analysis also underlines the severity of the overall bias in composite scoring approaches where the selection bias and over-fitting bias interact. Novy-Marx finds that a back-test based on composite scoring using the "best k of n" variables, is almost as biased as a back-test of a strategy where one selects the single variable that had the best performance of n to the power of k candidate variables. For example, using a composite score where one selects three variables out of six candidate variables is as biased as selecting with hindsight a single variable from 216 (6 to the power of 3) candidate variables. Likewise, selecting a composite of five variables out of ten based on back-tested performance is almost as bad as selecting a single variable among 100,000 (10 to the power of 5) candidate variables. This result underlines that the use of composite scores may lead to severe data-snooping bias. As the author concludes, by "combining spurious, marginal signals, it is easy to generate back-tested performance that looks impressive."

A simple reason for why composite scores may be more prone to generating biased results is that a composite variable requires more inputs and thus increases the number of possible choices. There seems to be wide ranging awareness that composite strategies, by having more inputs, will lead to increased data-mining risk. Pedersen (2015) makes a case against excessive back-testing, arguing that "we should discount back-tests more if they have more inputs and have been tweaked or optimised more." Likewise, Ilmanen (2013) states that analysis involving "tweaks in indicator specification" is "even more vulnerable to data-mining than is identification of the basic regularities."

In Exhibit 6 we present examples of composite scores used in the construction of multi-factor indices. It is immediately apparent that it is typical for factor index providers to use composite scores. For example, and focusing on the definition of Quality, MSCI combines three variables, FTSE Russell ups the ante with four metrics and Goldman Sachs mixes no fewer than seven metrics.

Exhibit 6: Examples of the use of composite scores in the construction of a multi-factor index

Index	Composite score
FTSE Russell 1000 Comprehensive	The Quality component is a composite of Profitability and Leverage and three individual measures make up Profitability
Goldman Sachs Equity Factor Index World	Quality metric uses a composite based on Asset Turnover, Liquidity, Return On Assets, Operating Cash Flow-to-Assets, Accruals, Gross Margin, Leverage
MSCI Diversified Multiple-Factor	The Quality component uses a sector-relative composite based on Return on Equity, Earnings Variability, Debt-to-Equity

For investors conducting due diligence on commonly-offered smart beta strategies, it thus appears important to investigate not just the back-tested performance but also the underlying data-snooping risk, given that both selection bias and over-fitting bias may be present when proprietary

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

composite scores are being used. Moreover, one can argue that back-tests of strategies that do not employ complex proprietary scores are naturally more robust and the back-tested performance of such strategies needs to be discounted less than that of complex proprietary factor definitions. In the next section, we further investigate the biases stemming from methodological choices, in particular what happens when a consistent approach may be lacking in the index design.

3. Factor Definitions: Proprietary Variables Versus Consensual Variables, and Problems with Composite Scores

4. Inconsistencies in Methodologies

4. Inconsistencies in Methodologies

As we discussed in the previous sections, a major source of potential data-mining bias that may result in overstated back-tested performance is the flexibility offered by the testing of many variations in search of the winning one. Such flexibility is obviously increased when a provider allows index methodologies to be inconsistent, across indices and/or time.

On the contrary, a very effective mechanism to avoid data-mining is to establish a consistent framework for smart beta index creation. Such a framework can limit ad-hoc choices while providing the necessary flexibility needed for smart beta index construction. A consistent framework is the best safeguard against post-hoc index design, or model mining, i.e. the testing of a large number of smart beta strategy variations, and selection of the ones that have good in-sample results. Perhaps surprisingly, while most major index providers argue that cap-weighted indices should have a consistent set of rules across regions to avoid unintended investment outcomes, consistency is often forgotten for factor indices. Below, we draw on several examples of inconsistencies for the indices discussed in this article.

4.1 Inconsistency or Consistency Across Factors

An important aspect of a robust methodology in the case of smart beta indices is consistency in the design across the indices for different factors. Indeed, it is surprising to see the same provider rely on radically different approaches to index construction to establish exposure to different factors. It is arguably even more surprising to see indices that were built using widely different methodologies being combined into a multi-factor index.

The next table outlines the design framework of the factor-based strategy indices that constitute the components of the MSCI Quality Mix index, and compares them to the factor indices used by ERI Scientific Beta as components of its multi-factor indices. This exhibit, in essence, compares two strikingly different approaches to constructing indices for different factors. While the Scientific Beta single factor indices apply the same index construction methodology and thus serve as fairly uniform building blocks for the multi-factor index, the underlying indices of the MSCI Quality Mix index all follow a different design path.

MSCI employs different stock selection schemes, weighting schemes and risk control options for the three different component indices in its “Quality Mix” multi factor index. For example, the Value factor index includes all stocks in the universe and reweights them by their value-related scores. The Quality factor component takes a different approach and first selects a fixed number of stocks with the highest Quality score. A relevant question is why the Value index does not do the same and first selects stocks or instead, why the Quality index does not use all stocks and simply reweights them by Quality score, to be consistent with the Value component. The third component capturing the Low Volatility factor follows yet another methodology. It obtains its Low Volatility factor tilt implicitly through a weighting scheme (Minimum Volatility). Again, one wonders why the same objective cannot be obtained with a methodology that would be consistent with e.g. the

4. Inconsistencies in Methodologies

Quality component and simply select and weight stocks by a risk measure for example. It is also worth noting that the three indices use different types of constraints on individual stock weights or sector weights. Such lack of uniformity in index design across factor indices leads to the question of what justifies the differences across factors and how back-tests of indices following such design approaches are impacted by data-mining risks.

Exhibit 7: Comparison of consistency in index construction framework between component factor indices used in MSCI multi factor indices and ERI Scientific Beta multi factor indices.

Factor	Index	Stock selection	Weighting scheme	Risk controls
MSCI Index Methodologies (for Components of Quality Mix Index)				
Value	MSCI USA Value Weighted Index	All stocks in CW parent index universe	Value score derived from four fundamental metrics, adjusted by investability factor	None
Low Vol.	MSCI USA Minimum Volatility Index	All stocks in CW parent index universe	Optimisation to minimise portfolio risk	Sector weight constraints. Cap on multiple of market cap of individual security
Quality	MSCI USA Quality Index	Fixed number of stocks by three-metric factor score	Selected stocks weighted by product of market cap and three-metric quality score	Issuer weights capped at 5%
Scientific Beta Index Methodologies (for Components of SciBeta Multi-Beta Multi-Strategy Index)				
Size	SciBeta USA Mid Cap Div. Multi-Strategy Index	Half the stocks in the universe by relevant single-metric factor score	Same weighting scheme for selected stocks (Diversified Multi-Strategy by default)	Cap on multiple of market cap and weight of individual securities
Value	SciBeta USA Value Div. Multi-Strategy Index			
Mom.	SciBeta USA High Momentum Div. Multi-Strategy Index			
Low Vol.	SciBeta USA Low Volatility Div. Multi-Strategy Index			

We can also see from the previous exhibit that Scientific Beta follows a dramatically different approach by using a consistent methodology across each of the four component indices of its multi factor index, containing four main rewarded factors. The implication of such consistency is that the number of potential variations that may have been tested is limited by construction.

Such an approach aligns with common sense recommendations to avoid the pitfalls of data-snooping. For example, Lo (1994) argued that we need "some kind of framework to limit the number of possibilities that we search over."

4.2 Inconsistency Across Time

Data-mining risks are further exacerbated by inconsistencies among index offerings across time. If providers change their mind frequently on what a good proxy for a given factor is, this inevitably increases the flexibility of index design and increases the potential to show inflated back-test performance. If providers launch new and enhanced versions of indices for the same

4. Inconsistencies in Methodologies

factor to replace old indices capturing the same factor, one may ask whether that new version was engineered to produce a simulated track record that would distract from the poor live performance of the erstwhile flagship product, which if correct would not bode well for the robustness of the new product's performance.

To show an example of changes in methodology over time, Exhibit 8 contrasts the factor definitions used for implementing the MSCI multi-factor approach as described in 2013 and 2015[4]. These single factor definitions are relevant in the design of two of the multi factor indices we analyse in this document.

The MSCI Quality Mix index is an equal-weighted index of three single factor indices targeting the Value, Low Volatility and Quality factors respectively. The component indices of the MSCI Quality Mix index for the Value and Quality factors follow the 2013 definitions, outlined in the table below. On the other hand, the optimisation-based MSCI Diversified Multi Factor index, targeting the Value, Quality, Size and Momentum factors, currently uses the four factor definitions from 2015.

It appears indeed that the 2015 factor definitions are at odds with earlier single- and multi-factor offerings. For example, in the MSCI Quality Mix index, the Value component is a fundamentally-weighted index aggregating book value, sales, earnings and cash earnings, and the Quality component is not sector relative and winsorised. In MSCI's Diversified Multi Factor index, the Value score is a sector-neutral composite score based on earnings-to-price, book-to-market and cash flow to enterprise value.

Exhibit 8: Inconsistency in factor definitions among MSCI Multi-Factor indices over time
This exhibit shows the difference between the definitions of factors used by MSCI in their "Deploying Multi Factor Allocations" White Paper (2013) and the definitions for the same factors used in creating the MSCI Diversified Multiple-Factor index (2015).

	Scoring		Adjustments	
	2013	2015	2013	2015
Value	Sales, book value, earnings and cash earnings	Price-to-book value, price-to-forward earnings and enterprise value-to-cash flow from operations	No sector control	Sector-relative scoring
	Past 3 year average values	Current values		
	Simple average across variables	Average of z-score for each variable		
Quality	Return on equity, debt-to-equity and earnings variability	Return on Equity, Debt-to-Equity and Earnings Variability	No sector control	Sector-relative scoring
Size	None	Negative of the exposure from the Barra Equity Model: Barra uses a z-score based on the logarithm of the market cap of the relevant firm		Country control (the Barra descriptor is on a country-relative basis)
Momentum	12-month and 6-month local price performance	Exposure from the Barra Equity Model based on 12-month relative strength (25% weight), 6-month relative strength (37.5% weight), historical alpha (37.5% weight).	Momentum score is risk-adjusted	No explicit risk adjustment (use of Barra exposure)

4 - Please refer to "Deploying Multi-Factor Index Allocations in Institutional Portfolios," Research Insight, MSCI, December 2013 and "The MSCI Diversified Multi-Factor Indexes - Maximizing Factor Exposure While Controlling Volatility," Research Insight, MSCI, May 2015.

4. Inconsistencies in Methodologies

Therefore, it appears that two multi-factor indices launched at different points in time by the same provider use different definitions of the Value factor[5]. This may be surprising, especially for the Value component, as Value seems to be among the most standard factors. Just like inconsistencies across factors open the room for a large number of variations in index design, it is clear that inconsistencies over time further increase such flexibility.

Such inconsistency across time is however widely present among index offerings. Amenc et al. (2015) emphasise that "Russell launched new factor indices to create a new brand known as 'High Efficiency' (HE) indices when it already had the following factor indices in the market – Russell 1000 High Momentum, Russell 1000 Low Volatility and Russell 1000 Value. The new indices have the same objective as the old ones but different construction principles." Interestingly, these "high efficiency" factor indices have since been abandoned by the provider for most of the factors and replaced by yet another suite of factor indices for the same factors using yet another methodology. It thus appears that inconsistency over time is all but day-to-day business for index providers.

5 - None of which is consistent with the definition used in MSCI's oldest Value index, aptly named Value index, which is based on a composite of book value-to-price, 12-month forward earnings to price and dividend yield.

4. Inconsistencies in Methodologies

5. Concentrated Indices and Stock-Level Optimisation

5. Concentrated Indices and Stock-Level Optimisation

An important issue that can be easily neglected when constructing a multi-factor index is diversification. Positive exposure to rewarded factors is obviously a strong and useful contributor to expected returns. However, products that aim to capture explicit risk-factor tilts often neglect adequate diversification. This is a serious issue because diversification has been described as the only "free lunch" in finance. Diversification allows a given exposure to be captured with the lowest level of risk required. In contrast, gaining factor exposures exposes investors to additional types of risk, and therefore, such exposures do not constitute a "free lunch." They instead constitute compensation for risk in the form of systematic factor exposures. Such capturing of risk premia associated with systematic factors is attractive for investors who can accept the systematic risk exposure in return for commensurate compensation.

However, factor-tilted strategies, when they are very concentrated, may also take on other, non-rewarded, risks. Non-rewarded risks come in the form of idiosyncratic or firm-level risk, as well as potential risk for sector concentration, currency, sovereign or commodities risk exposure, etc. Financial theory does not provide any reason why such risk should be rewarded. Therefore, a sensible approach to factor investing should not only look at obtaining a factor tilt, but also at achieving proper diversification within that factor tilt. To illustrate this point, we focus on the value factor as an example below, but the discussion carries over to other factors as well.

In fact, if the objective was to obtain the most pronounced value tilt, for example, the only unleveraged long-only strategy that corresponds to this objective is to hold 100% in a single stock, the one with the largest value tilt, as measured for example by its estimated sensitivity to the value factor or its book-to-market ratio. This thought experiment clearly shows that the objective of maximising the strength of a factor tilt is not reasonable. Moreover, this extreme case of a strong factor tilt indicates what the potential issues with highly concentrated factor indices are. Even if the appropriateness of such an approach had been established, any Value premium so captured would necessarily come with a large amount of idiosyncratic risk. This risk is not rewarded and therefore we should not expect the strategy to lead to an attractive risk-adjusted return. Additionally, it is unlikely that the same stock will persistently have the highest value exposure within a given investment universe. Therefore, a periodically-rebalanced factor index with such an extreme level of concentration is likely to generate 100% one-way turnover at each rebalancing date, as the stock held previously in the strategy is replaced with a new stock that displays the highest value exposure at the rebalancing date. While practical implementations of concentrated factor-tilted indices will be less extreme than this example, we can expect problems with high levels of idiosyncratic risk and high levels of turnover whenever index construction focuses too much on concentration and pays too little attention to diversification. This goes for both single-tilt and multi-factor indices. In the next sub-sections, we discuss concentration in the context of different approaches to designing multi factor indices.

5. Concentrated Indices and Stock-Level Optimisation

5.1 Top-down approaches

One of the possible ways to construct a multi-factor index is to combine different single factor indices. Among the indices we analyse in this article, this is the approach chosen by Scientific Beta and is also the chosen methodology of the MSCI Quality Mix index.

For such combinations of single factor indices, there will of course be a certain level of deconcentration resulting from the fact that different indices are combined. However, a relevant question is whether such multi-factor indices are constructed using well-diversified building blocks.

Amenc et al. (2016) show that well-diversified factor indices which pursue a diversification objective through an alternative weighting scheme based on a relatively broad stock selection provide considerable benefits over more concentrated single factor indices. Their results suggest that well-diversified factor portfolios or indices outperform their highly-concentrated counterparts in terms of risk-adjusted performance, because concentrated factors may be highly exposed to unrewarded factors. In addition, they show that factor-tilted portfolios on narrow stock selections present implementation drawbacks such as higher turnover.

The Scientific Beta multi factor indices that are part of our analysis use well-diversified factor indices as building blocks. These single factor indices are also termed "smart factor indices" (see Amenc et al. (2014)). In this approach, each single factor index is well-diversified and multi-factor allocation across several such indices additionally smoothes returns over factor cycles.

On the other hand, looking at the MSCI Quality mix index from a conceptual perspective, we can observe that the index does not have an explicit diversification objective which may lead to concentration, depending on the specific parameters used for weighting and stock selection. The index involves simple market-cap weighting adjusted by quality scores (in its quality component), weighting by firm fundamentals (in its value component, which uses their value-weighted approach) and an approach that is notorious for producing high concentration (to establish its low-volatility exposure) and therefore may end up being at least as concentrated as cap-weighted approaches.

5.2 Bottom-up approaches

Concentration may also arise in multi-factor indexing methodologies which, rather than combining single factor indices, actually build multi-factor indices from the stock level up. If the methodology targets exposure to stocks with the highest composite multi-factor score for example, concentration may be expected quite naturally as a result of the objective of strong multi-factor exposure. If the stock-level information on factor scores is integrated in an optimisation approach, concentration issues may be exacerbated.

5. Concentrated Indices and Stock-Level Optimisation

The optimisation approach is for example followed by the MSCI Diversified Multiple-Factor index. This index maximises the ratio of a weighted average composite multi-factor score to portfolio volatility, which corresponds to mean-variance optimisation when stock-level expected returns are proxied by the composite factor score.

As a result of such an optimisation, one may observe high levels of concentration. Indeed, it is interesting to note that MSCI report that their MSCI World Diversified Multi-Factor index (where the top 10 stocks account for 14.5% of the index capitalisation at the end of June 2016) is more concentrated than the broad capitalisation-weighted MSCI World index (where the Top 10 stocks account for 10.08% of the total index capitalisation).

More generally, methodologies that optimise on the basis of stock-level factor scores as proxies for expected returns may result in high levels of concentration if suitable deconcentration mechanisms are not included in the methodology. One may end up with a portfolio with high weighted-average factor scores, but also high idiosyncratic risk.

As an example of high idiosyncratic risk, consider the following case study analysing the exposure of the European version of the MSCI index, the MSCI Europe Diversified Multiple-Factor index, to Volkswagen AG. Exhibit 9, taken from Amenc, Sivasubramanian and Ulahel (2015), shows that, at the time when the emissions scandal erupted, the MSCI Europe Diversified Multiple-Factor index was very strongly exposed to the risk of Volkswagen AG stock. This poor diversification of the specific risks led to the MSCI index considerably underperforming the SciBeta Extended Europe Multi-Beta Multi-Strategy EW index.

The table shows that the MSCI Europe Diversified Multiple-Factor index contained roughly 16 times more Volkswagen AG stock than the Scientific Beta Extended Europe Multi-Beta Multi-Strategy EW index, with respective weights of 0.05% and 0.80% as of 31 August, 2015. Similarly, the MSCI multi-factor index overweighted Volkswagen stock more than twice with respect to the reference cap-weighted Stoxx Europe 600 index, with a 0.80% weight compared to 0.35% in the reference index.

Exhibit 9: Impact of Volkswagen Scandal on Stoxx Europe 600 vs. SciBeta Extended Europe Multi-Beta Multi-Strategy EW Index and the MSCI Europe Diversified Multiple-Factor Index, taken from Amenc, Sivasubramanian and Ulahel (2015)
Analysis is based on weekly total returns in USD from 31-Aug-2015 to 30-Sep-2015 on the Extended Europe Universe. The SciBeta Extended Europe Multi-Beta Multi-Strategy EW index is the equal-weighted combination of the four factor-tilted multi-strategy indices – Mid Cap, High Momentum, Low Volatility and Value. Each factor-tilted multi-strategy index selects 50% of stocks from the universe based on the factor score. The multi-strategy weighting scheme is the equal-weighted combination of 5 weighting schemes: Maximum Deconcentration, Maximum Decorrelation, Efficient Minimum Volatility, Efficient Maximum Sharpe Ratio and Diversified Risk Weighted. Each factor-tilted score-weighted index selects 50% of stocks from the universe based on the factor score.

31-Aug-2015 - 30-Sep-2015	Performance and Weight Analysis		
	Stoxx Europe 600	SciBeta Extended Europe Multi-Beta Multi-Strategy EW	MSCI Europe Diversified Multiple-Factor Index
Volkswagen AG Weights as of 31-Aug-2015	0.35%	0.05%	0.80%
Active Weights		-0.30%	0.45%
Cumulative Returns	-4.41%	-2.94%	-3.82%
Attributable to Volkswagen AG	-0.15%	-0.02%	-0.34%
Cumulative Excess Returns	-	1.47%	0.59%
Attributable to Volkswagen AG		0.13%	-0.19%

5. Concentrated Indices and Stock-Level Optimisation

Ultimately, the MSCI Europe Diversified Multiple-Factor Index underperformed the SciBeta Extended Europe Multi-Beta Multi-Strategy EW index by 88 basis points, with excess returns relative to the Stoxx Europe 600 index for the month of September 2015 of 0.59% compared to 1.47% for the SciBeta Extended Europe Multi-Beta Multi-Strategy EW index. The analysis of the Volkswagen case also provides a good understanding of how the search for strong factor exposure can lead to overconcentration in a particular stock.

Another example of possible concentration issues that may arise when working from the bottom up as well as using the previously discussed composite scores is the so-called "tilt-tilt" methodology employed by the FTSE Russell Comprehensive indices. These indices multiply several factor scores for each stock and combine them with market cap weights to arrive at a final stock weight in the index. The multiplicative scoring across factors means that stocks will be overweight not necessarily when they have high average exposure to the different factors but rather when they have positive exposure to each and every factor. This approach thus incorporates the idea of looking for champion stocks which rank well according to all factor attributes for several factors at the same time.

In the next exhibit we compare the Scientific Beta Multi-Beta Multi-Strategy Six-Factor EW index (a top-down index based on well-diversified single-tilt building blocks) with a stylised bottom-up test portfolio using a multiplicative scoring methodology to adjust market cap weights by a multi-factor score. This strategy, which we call the Multiplicative Scoring Strategy, uses a composite factor score for six factors – lower Size, positive Momentum, Low Volatility, Value, Low Investment and High Profitability. The composite is based on multiplying the s-scores for each factor and adjusting the market cap-weights by this multiplicative score. The table, analysing a period of 40 years in the USA, shows the risk-adjusted performance of the two indices as well as interesting weight-related measures such as turnover and effective number of stocks. The effective number of stocks, calculated as the inverse of the sum of squared stock weights, is a good measure of diversification and allows the concentration levels of the portfolios to be compared.

Exhibit 10: Example of concentration of composite factor index
The time period of analysis is 31-Dec-1974 to 31-Dec-2014 (40 years). The composite factor index is constructed by multiplying the s-score for the six factors – size, momentum, low volatility, value, low investment, and high profitability – for each individual stock and then combining this composite factor score with the market cap to arrive at final weights. This composite factor index is rebalanced annually and is constructed using a US stock universe that contains the 500 largest stocks by total market capitalisation. The analysis is done using total returns in USD. The yield on secondary market US Treasury Bills (3M) is the risk-free rate. Scientific Beta US Long-Term Track records are used to obtain the Scientific Beta LTTR Multi-Beta Multi-Strategy Six-Factor EW index as well as the cap-weighted benchmark. Effective number of stocks is given by the inverse of the sum of squared constituent weights. The reported turnover and effective number of stocks is an average across the time period of analysis.

31-Dec-1974 to 31-Dec-2014	Multiplicative Scoring Strategy	SciBeta Long-Term United States Multi-Beta Multi-Strategy Six-Factor EW
Annualised Return	15.64%	16.01%
Volatility	15.22%	15.52%
Sharpe Ratio	0.69	0.70
Relative Returns	3.48%	3.85%
Tracking Error	5.22%	4.73%
Information Ratio	0.67	0.81
Annualised One-Way Turnover	34.82%	25.03%
Weighted Average Market Cap (M$)	11,653	11,607
Effective Number of Stocks	170	345

5. Concentrated Indices and Stock-Level Optimisation

Comparing the effective number of stocks between the Multiplicative Scoring Strategy and the Scientific Beta multi-factor index, we learn that the levels of diversification are quite different. Indeed, Scientific Beta's multi-factor index has more than twice the effective number of stocks compared to the Multiplicative Scoring Strategy, with 345 versus 170. Apart from the concentration in fewer stocks, the Multiplicative Scoring Strategy also experienced higher turnover.

As the performance metrics reveal, the composite multiplicative factor scoring approach did not produce better risk-adjusted performance, with an information ratio of 0.67 lagging behind 0.81 for the Scientific Beta index.

In addition to concentration, stock level approaches contain further issues which we turn to now.

When using multi-factor scores in portfolio optimisation, it should not be forgotten that the score is ultimately used as a proxy for expected returns. It is well known for example that mean-variance optimisation that integrates expected returns can result in an "error maximisation exercise" since expected return are hard to estimate at the individual stock level, and since mean-variance optimisers are very sensitive to estimation error for expected returns (Best and Grauer, 1991).

Achieving high absolute factor scores at the portfolio level by concentrating on picking champion stocks that score highly on all targeted factor dimensions is probably intuitively attractive but it is predicated on a high-precision relationship between factor scores and returns at the stock-level. There is no question that factor investing is motivated by an attempt to capture higher long-term returns through the right risk exposures. However, return estimation at the stock level is notoriously difficult. Black (1993) distinguishes between explaining returns, which is easy because it is really explaining variance, and predicting returns, which is hard. He contends that the accurate estimation of average expected return requires decades of data. For variance, he notes, "We can use daily (or more frequent) data to estimate covariances. Our estimates are accurate enough that we can see the covariances change through time." To estimate expected returns, on the other hand he writes "Daily data hardly help at all." and "We need such a long period to estimate the average that we have little hope of seeing changes in expected return." These observations are consistent with the unavoidable statistical fact that estimators of risks are convergent/consistent (the more data points the more precise the estimation) while estimators of returns are non-convergent/consistent (the frequency of observation does not help, only the length of time does), as underlined in the first appendix to Merton (1980).

The search for champion stocks as measured by their factor scores is a stock-picking exercise that relies implicitly but heavily on the accuracy of expected return predictions. As alpha envy appears to contaminate smart beta and factor investing, it is important to pause and remember that it is precisely the lack of persistent success in stock picking that has led an increasing number of institutional investors to shift towards passive strategies and that it is the realisation that the bulk of the performance of active management programmes comes from exposure to well-documented systematic factors that has reignited the interest in factor-based investing.

5. Concentrated Indices and Stock-Level Optimisation

Attempting to improve stock-level return forecasts, even when this is done with the support of a factor model, is a largely futile exercise. This should probably remain the preserve of professional stock pickers. If efforts are to be made to improve the adjusted returns of factor investing, it is more on the risk dimension side, where we can rely on sixty years of progress in financial econometrics to estimate convergent estimators of volatilities and covariances.

When academics have tested standard factors, they have done so by running portfolio sorts, and assessing return differences at the portfolio level, not by assessing returns at the stock level. For example, they have observed that, on average, value stocks tend to have higher returns than growth stocks over the long-term. If one now tries to design strategies based on very fine distinctions at the stock level, such relations may be drowned in noise. More generally, making very fine distinctions at the stock level is prone to capturing estimation error.

Indeed, an implication of the "error maximisation" issue is that stock-level optimisation, which considers expected returns (or, equivalently, composite factor exposures), will not only be highly concentrated in a few stocks, but will actually tend to assign the highest weights to the stocks with the highest estimation errors. We should underline that optimisation-based smart beta strategies (such as minimum volatility, equal-risk contribution etc.) had avoided using direct estimates of expected returns in optimisation precisely because it is well known that this leads to the error maximisation problem.

Thus any stock-level approach needs to be handled with care and one needs to assess whether suitable mechanisms have been built in to achieve robustness.

Furthermore, optimisation-based approaches frequently come with stringent constraints attached. These constraints are intended to avoid extreme solutions and produce more acceptable portfolios. This not only reveals a guarded faith in optimisation, but also creates model mining risks as the solution may end up being primarily driven by the constraints and not the objective. Furthermore, it is not clear how sensitive the strategy is to such constraints. Last but not least, there may be provisions in the ground rules that allow the optimiser to be changed, thus potentially introducing discretion into a supposedly rules-based design.

5. Concentrated Indices and Stock-Level Optimisation

6. Conclusions

6. Conclusions

The offerings in the area of multi-factor indices are multiplying rapidly and investors have to assess how such indices match their investment needs. Given that most products have been launched recently, analysis of risk and performance is mostly limited to back-tested data. Therefore, the methodological principles behind index construction should become a key area of attention in the assessment of these indices. Analysing robustness requires an assessment of index design principles and the conceptual considerations underlying index design. Our brief review of offerings aims to shed light on several issues such as complex proprietary factor definitions, potential inconsistencies in methodologies, and concentration issues.

We have discussed the all-important issues of data-mining, which can present real problems in many cases. The proprietary factor definitions and the use of composite scores in index construction may lead to overstated back-tested performance. This is of major interest to investors as the new index products are mostly being sold on the strength of good back-tested performance. However, flexibility in design choices and the ability to test many variations of factor definitions and portfolio construction models can severely bias any historical simulation. In addition, we have argued about the importance of consistency in index design. The lack of a well-defined methodological framework, or frequent changes to it, increases the amount of flexibility that providers have and thus potentially biases the historical tests further. Similarly, uniformity and consistency across the various index offerings is a surprisingly overlooked aspect of index design.

We also compared the top-down and bottom-up approaches to multi-factor index construction and found that concentration could arise in many scenarios, thus exposing investors to undesired idiosyncratic risks. Diversification has been described as the only "free lunch" in finance and unwanted concentration does not do much more than erode it.

In principle, multi-factor indices aim at a common goal – outperforming cap-weighted benchmarks by providing exposure to multiple rewarded factors. As discussed here, the ways to do this are nonetheless quite diverse. A key consideration for investors is how robust the performance presented in back-tests is expected to be. Highly parameterised approaches naturally contain higher risks of overstated back-test performance than more parsimonious index design methods. In particular, since the bottom-up approach is more flexible, it can more easily fall prey to data-mining. It is always possible to find a combination of factor definitions, multi-factor scoring and weighting schemes that will select the right stocks in sample. In-sample over-fitting, however, would lead to disappointing out-of-sample performance. In terms of due diligence, the bar on innovative bottom-up methods should be set higher than for classic top-down approaches, and investors would be well advised to ask for live track records of a significant length when a provider shows a lot of creativity.

There is no doubt that more elaboration on factor definitions and the use of more granular stock-level information allow the data to be fitted better and help to produce back-tests that suggest superior performance, but the ultimate question investors should ask is that of the robustness of the advertised index performance in live conditions.

References

References

- Amenc, N., F. Ducoulombier, F. Goltz, A. Lodh, and S. Sivasubramanian. 2016. Diversified or Concentrated Factor Tilts? *Journal of Portfolio Management* 42(2): 64-76.
- Amenc, N., F. Goltz., A. Lodh, and S. Sivasubramanian. 2015. Robustness of Smart Beta Strategies. *Journal of Index Investing* 6(1): 17-38.
- Amenc, N., S. Sivasubramanian and J. Ulahel. 2015. The Limitations of Factor Investing: Impact of the Volkswagen Scandal on Concentrated versus Diversified Factor Indices. EDHEC Working Paper.
- Amenc, N., F. Goltz, A. Lodh, and L. Martellini. 2014. Towards Smart Equity Factor Indices: Harvesting Risk Premia without Taking Unrewarded Risks. *Journal of Portfolio Management* 40(4): 106-122.
- Bender, J., R. Briand, R., D. Melas, R. Subramanian and M. Subramanian, 2013. Deploying multi-factor index allocations in institutional portfolios. MSCI Research Insights, December.
- Best, M. J., and Grauer, R. R. 1991. Sensitivity analysis for mean-variance portfolio problems. *Management Science* 37(8): 980-989.
- Doole, S., C.. Chia, P. Kularni and D. Melas. 2015. The MSCI Diversified Multi-Factor Indexes - Maximizing Factor Exposure While Controlling Volatility, MSCI Research Insights, May.
- Fama, E. F. and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bond. *Journal of Financial Economics* 33: 3-56.
- Fama, E. F. and K. R. French. 2010. Luck versus Skill in the Cross Section of Mutual Fund Returns. *Journal of Finance* 65(5): 1915-1947.
- Fama, E. F., and K. R. French. Size, Value, and Momentum in International Stock Returns. 2012. *Journal of Financial Economics* 105(3): 457-472.
- Fama, E. F., and Kenneth. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116(1): 1-22.
- Harvey, C. R. and Yan Liu. 2015. Lucky Factors, Working Paper, available at SSRN.
- Hou, K., Xue; C., Zhang, L. 2015. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28 (3): 650-705.
- Ilmanen, A., 2011. Expected Returns: An Investor's Guide to Harvesting Market Rewards, John Wiley & Sons.
- Lo, A. W. 1994. Data-Snooping Biases in Financial Analysis. In AIMR Conference Proceedings. 1994(9): 59-66. Association for Investment Management and Research.
- Merton, R. 1980. On Estimating the Expected Return on the Market - An Exploratory Investigation. *Journal of Financial Economics* 8, 323 -361.
- Novy-Marx, R. 2013. The Other Side of Value: The Gross Profitability Premium. *Journal of Financial Economics* 108: 1-28.
- Novy-Marx, R. 2015. Back-testing Strategies Based on Multiple Signals. Working Paper. National Bureau of Economic Research.
- Pedersen, L. H. 2015. Efficiently Inefficient: How Smart Money Invests & Market Prices Are Determined, Princeton University Press.

About ERI Scientific Beta

About ERI Scientific Beta

EDHEC-Risk Institute set up ERI Scientific Beta in December 2012 as part of its policy of transferring know-how to the industry. ERI Scientific Beta is an original initiative which aims to favour the adoption of the latest advances in "smart beta" design and implementation by the whole investment industry. Its academic origin provides the foundation for its strategy: offer, in the best economic conditions possible, the smart beta solutions that are most proven scientifically with full transparency of both the methods and the associated risks. Smart beta is an approach that deviates from the default solution for indexing or benchmarking of using market capitalisation as the sole criterion for weighting and constituent selection.

EDHEC-Risk Institute considers that new forms of indices represent a major opportunity to put into practice the results of the considerable research efforts conducted over the last 30 years on portfolio construction. Although these new benchmarks may constitute better investment references than poorly-diversified cap-weighted indices, they nevertheless expose investors to new systematic and specific risk factors related to the portfolio construction model selected.

Consistent with a full control of the risks of investment in smart beta benchmarks, ERI Scientific Beta not only provides exhaustive information on the construction methods of these new benchmarks but also enables investors to conduct the most advanced analyses of the risks of the indices in the best possible economic conditions.

Lastly, within the context of a Smart Beta 2.0 approach, ERI Scientific Beta provides the opportunity for investors not only to measure the risks of smart beta indices, but also to choose and manage them. This new aspect in the construction of smart beta indices has led ERI Scientific Beta to build the most extensive smart beta benchmarks platform available which currently provides access to 3,817 smart beta indices.

ERI Scientific Beta Publications

ERI Scientific Beta Publications

2016 Publications

- Amenc, N., F. Goltz, A. Lodh and S. Sivasubramanian. Robustness of Smart Beta Strategies (July).
- Amenc, N., E. Christiansen, F. Goltz and K. Gautam. Scientific Beta Low Carbon Multi-Beta Multi-Strategy Indices (June).
- Amenc, N., F. Goltz, K. Gautam and S. Sivasubramanian. The Dimensions of Quality Investing: High Profitability and Low Investment Smart Factor Indices (June).
- Amenc, N., F. Ducoulombier, F. Goltz and S. Sivasubramanian. Scientific Beta Multi Smart Factor Indices: A Double Diversification Approach to Factor Investing (June).
- Amenc, N. and F. Goltz. Long-Term Rewarded Equity Factors: What Can Investors Learn from Academic Research? (May).
- Gautam, K. and E. Shirbini. Scientific Beta Global Universe. (May).
- Amenc, N., F. Goltz and S. Sivasubramanian. Investability of Scientific Beta Indices (May).
- Amenc, N., F. Goltz and S. Sivasubramanian. Scientific Beta Multi Smart Factor Indices: An Introduction (April).
- Amenc, N., F. Ducoulombier and A. Lodh. ERI Scientific Beta Defensive Strategies: Bringing Diversification to, and Going Beyond, Traditional Approaches (March).
- Amenc, N., F. Goltz, A. Lodh and S. Sivasubramanian. Scientific Beta Multi-Strategy Factor Indices: Combining Factor Tilts and Improved Diversification (March).

2016 Factsheet

- Scientific Beta Developed Multi-Smart Factor Indices (February).

2015 Publications

- Gautam, K., A. Lodh and S. Sivasubramanian. Scientific Beta Efficient Maximum Sharpe Ratio Indices. (December).
- Lodh, A. and S. Sivasubramanian. Scientific Beta Diversified Risk Weighted Indices. (December).
- Goltz, F. and S. Sivasubramanian. Scientific Beta Maximum Decorrelation Indices. (November).
- Goltz, F. and A. Lodh. Scientific Beta Efficient Minimum Volatility Indices. (October).
- Gonzalez, N., S. Sivasubramanian and S. Ye. Scientific Beta Maximum Deconcentration Indices. (September).
- Amenc, N., F. Goltz, A. Lodh and S. Sivasubramanian. Robustness of Smart Beta Strategies (September).
- Goltz, F. and N. Gonzalez. Overview of Diversification Strategy Indices. (July).
- Lodh, A. and S. Sivasubramanian. Scientific Beta Diversified Multi-Strategy Index. (July).
- Le Sourd, V. and A. Lodh. Scientific Beta Analytics: Examining the Performance and Risks of Smart Beta Strategies. (January).

2014 Publications

- Amenc, N., F. Goltz and A. Lodh. Scientific Beta USA Long-Term Track Records (November).
- Amenc, N., F. Goltz and L. Martellini. Smart Beta 2.0. (June).
- Goltz, F. and N. Gonzalez. Risk Managing Smart Beta Strategies. (June).
- Martellini, L., V. Milhau and A. Tarelli. Estimation Risk versus Optimality Risk : An Ex-Ante Efficiency Analysis of Heuristic and Scientific Equity Portfolio Diversification Strategies. (March).

ERI Scientific Beta Publications

- Coqueret, G. and V. Milhau. Estimating Covariance Matrices for Portfolio Optimisation (January).
- Gautam, K. and N. Gonzalez. A Comparison of Construction Methodologies and Performances of Cap-Weighted Indices (January).
- Gautam, K. and N. Gonzalez. Scientific Beta Cap-weighted Indices: Tools for Relative Performance and Risk Analysis of Scientific Beta's Smart Beta Strategies (January).

2013 Publications

- Amenc, N., F. Goltz and A. Lodh. Alternative Equity Benchmarks. (February).

Disclaimer

For more information, please contact:
Séverine Cibelly on: +33 493 187 863 or by e-mail to: severine.cibelly@scientificbeta.com

ERI Scientific Beta HQ & Asia
1 George Street
#07-02
Singapore 049145
Tel: +65 6438 0030

ERI Scientific Beta R&D
393 promenade des Anglais
BP 3116 - 06202 Nice Cedex 3
France
Tel: +33 493 187 863

ERI Scientific Beta—Europe
10 Fleet Place, Ludgate
London EC4M 7RB
United Kingdom
Tel: +44 207 871 6742

ERI Scientific Beta—North America
One Boston Place, 201 Washington Street
Suite 2608/2640, Boston, MA 02108
United States of America
Tel: +1 857 239 8891

ERI Scientific Beta—Japan
East Tower 4th Floor, Otemachi First Square,
1-5-1 Otemachi, Chiyoda-ku, Tokyo 100-0004
Japan
Tel: +81 352 191 418

www.scientificbeta.com

Scientific Beta
An EDHEC-Risk Institute Venture