Multi-factor portfolios: A new factor?
Limits of the static approach.

October, 25, 2017
Multi-factor portfolios: A new factor? Limits of the static approach.

Carmine De Franco, PhD and Bruno Monnier, CFA

October, 25, 2017

Abstract

We assess the value added of a multi-factor portfolio from a performance-agnostic point of view. First we introduce a broad general definition of factor, that encompasses usual factors like Size or Value, then we prove that static long-short multi-factor strategies (as the equal weighting of factors) are indeed factors according to our definition. This result is new in the literature and states that, by investing in a long-short static multi-factor strategy, one is indeed investing into a new (synthetic) factor. Finally we test the strength of such a synthetic factor compared to each single factor by looking at its predictive power. We empirically test the equal-weighting of Value, Size, Momentum and Low Volatility in the US and Europe. Our conclusion is very clear in both regions: the equal-weighting of these four standard factors is a synthetic factor that has no predictive power on stocks’ return, while each of the factors shows clear ability to distinguish among stocks. In other words, the measure that underlies this equal-weighting of factors has zero predictive power on cross-sectional differences in stocks’ returns.

Key words: Factor Investing, Multi-Factor Portfolio Construction, Predictive Power, Return Forecast.
1 Introduction

Factor investing has gained popularity among investors because it enables them to access well-known risk premia in the equity market, in a transparent and rule-based way. By using few robust and economically sound factors, investors can build modern portfolios by targeting their factor exposures, or by implementing their views on the behavior of each factor. Single-factor investment vehicles have known a large success, measured in terms of inflows and number of products. Worldwide listed Smart Beta products, the marketing labeling for factor investing, account for almost 700 USD billions and roughly 1,300 products according to ETFGI (2017)], with a growth rate of 30% per year since 2008. Eventually the logical move would be from the massive use of single-factor to multi-factor strategies. Literature, very little to our knowledge, unanimously agrees on the superior power of factor combinations, which usually bring value to the final portfolio (see for example Garff (2014) or Bender and Wang (2016)).

But we must recognize that multi-factor strategies, by definition, cannot be unique. Indeed, different investors should have different approaches as their final objectives can be significantly different, from both a risk and a horizon point of view (see Bender et al. (2013)). The three issues that are currently debated are the choice of factors, the allocation method and the use of a static versus dynamic approach.

Alighanbari and Chia (2016) compare static and dynamic multi-factor strategies. The static one is a simple equal weighting of six factors (Size, Value, Momentum, Volatility, Dividend and Quality) while the dynamic ones use fundamental data to weight each factor. They conclude that their fundamental based, dynamic strategies deliver higher returns, but come with higher turnover and greater complexity. We also distinguish between integrating multi-factor techniques (which allocate to stocks that have simultaneous exposures to the factors) and mixing techniques (that achieve the desired factor exposure by allocating across different single factors). Chow et al. (2017) for example show that the integrating technique is superior to the mixing technique, even if the former shows higher turnover and higher idiosyncratic risk. Superior properties of the integration technique are also shown in Bender and Wang (2015, 2016) and Fitzgibbons et al. (2017).

The static approach is nevertheless still favored in the industry, as it provides significant results without introducing complex frameworks. For example Blitz (2012, 2015) shows how a simple equal weighting of Value, Momentum and Low-Volatility delivers superior risk-adjusted returns when compared to each single factor. Similarly, Bender et al. (2010) find that using a combination of risk premia across different asset classes gives similar returns to a 60/40 equity/bond benchmark but with 65% less volatility. The most tempting dynamic approach would include a factor-timing mechanism. Proponents of such techniques usually adopt active-management decisions, based, among other considerations, on macro-economic and financial data, sentiments, political outcomes, etc. Systematic approaches, on the other hand, make use of valuations as the key driver for factor timing. For example Arnott et al. (2016a,b, 2017) link factor cheapness with positive performances. Opposed to such attempts, Asness and his coauthors
Asness, 2016; Asness et al., 2017) stress the difficulty of factor timing and how deceptively this could be, out-of-sample, when one accounts for fees and transaction costs. As Asness (2016) states ... “Factor timing has the potential of reintroducing a type of skill-based ‘active management’ (as timing is generally thought of this way) back into the equation.” Indeed, factor investing has been used as the framework within which active managers’ performances were decomposed into alpha and beta(s), so that alternative beta was not taken as alpha. The question whether one should seek factor timing or long-term static approaches remains unanswered. Our paper takes a different approach and focuses instead on the ability to forecast stock returns. First, we introduce a general definition of “factor” that encompasses usual factors like Size or Value. Our definition is very general and does not require the factor to be economically relevant, which is mandatory from a practical point of view.

To define a general factor, we need a measure (ex. stock valuation), a partition based on this measure (ex. three buckets) and an allocation scheme (ex. long on the highest bucket and short on the lowest, equal weighting within each bucket). Fama-French’s factors (Fama and French, 1993) fit our definition. With this definition, we are able to prove that static long-short multi-factor strategies (ex. equal weighting of factors) are indeed factors according to our definition and we are able to explicitly characterize the underlying measure, the partition and the scheme. This result is very new in the literature, and states that by investing in a static long-short multi-factor strategy, one is indeed investing in a new (synthetic) factor. Unfortunately, it is not easy to derive the economic logic behind the underlying measure (as it is for Size, Value, Low Volatility or Momentum). Finally, we test the strength of such a “synthetic factor” compared to each single factor by looking at its predictive power. The reason for this is that, from a performance/risk perspective, we already know that by mixing different factors, we end up with their average return (if we do not consider compounding effects) with lower volatility due to diversification. Our measure assesses the power of a factor when it comes to forecasting stock returns.

We empirically test the equal weighting of Value, Size, Momentum and Low Volatility in the US and Europe. Our conclusion is very clear in both realms: the equal weighting of these four standard factors is a synthetic factor that has no predictive power on stock returns, while each factor shows clear ability to discriminate among stocks, except for Size for which results are slightly puzzling. In other words, the measure that underlies this equal weighting of factors has zero predictive power on cross-sectional differences in stock returns. One of the reasons for this is that by equally weighing factors, exposures are mostly netted out so that the final stock weights are very noisy.

We truly believe that multi-factor portfolios are the logical evolution of modern portfolio construction. But, as of today, the majority of commercially available multi-factor strategies are mainly static, and their value added only come as a result of diversification, which is a good thing after all. Investors should be aware that combining different factors may lead to a final netted exposure similar to a random tilted portfolio.
Echoing the debate opposing proponents of factor timing to proponents of static allocations, we add that although factor timing is a difficult exercise, we remain cautious about the assumption that static allocation schemes are the best way of integrating factors in the portfolio.

2 Single factor

Since the very beginning of its academic foundation, a factor is a long-short portfolio that reflects the differences in performance between stocks with regard to a specific characteristic. The initial Size and Value factors in Fama and French (1993) are long-short portfolios that exploit the differences between small and large companies (the characteristic is the market capitalization) and between value and growth stocks (the characteristic is the book-to-market ratio). Together with the market portfolio, this so-called three-factor model extended the traditional CAPM model (Sharpe, 1964; Lintner, 1965). Later, Carhart (1997) extended the three factor model with Momentum, once again a long-short portfolio that exploits the empirical evidence of performance differences between stocks with respect to their most recent performances. Since then, both the financial and the academic communities have seen a surge of new factors. We do not want to overstate on their alleged validity and economic foundations as we remain skeptical. Actually, many of these factors turn out to be the result of data mining, errors or in-sample anomalies (Hsu and Kalenik, 2014; Harvey et al., 2016; McLean and Pontiff, 2016). The common traits of all these factors are summarized in the following:

**Measure:** The characteristic that is able to explain cross-sectional differences in stock returns.

**Factor Construction:** The transfer of the measure into portfolio weights.

As an example, the initial three-factor model by Fama and French (1993) used market capitalization (Size) and book-to-market (Value), and a factor construction technique that allows for reduction in correlation between them based on double sorting. In an attempt to generalize the modern factor construction, we introduce the following:

**Definition 2.1.** For a given investment universe of size \( n \), let \( (C_i)_{i=1,\ldots,n} \) be a generic characteristic measure associated with each stock in the investment universe and \( (L_s)_{s=1,\ldots,l} \) be a disjoint partition of the measure:

\[
\bigcup_{s=1}^{l} L_s = \left[ \min_i C_i, \max_i C_i \right]
\]

We denote \( l_s := \text{card} \{i = 1, \ldots, n \mid C_i \in L_s \} \), which essentially counts the number of stocks whose measures lie in the bucket \( L_s \). Let \( j \) be the map that identifies each \( C_i \) into the unique subset of the partition \( L \):

\[
\forall i = 1, \ldots, n, \; j(i) \text{ verifies } C_i \in L_{j(i)}
\]
and $\alpha = (\alpha_1, \ldots, \alpha_l)$ such that $\sum \alpha_s = 0$.

A Factor $F = F(C, L, \alpha)$ where $C$ is the measure and $L, \alpha$ the factor construction is the long-short portfolio where stock weights are defined as follows:

$$W_i := \frac{\alpha_j(i)}{l(j(i))} \quad (2.1)$$

We illustrate the Definition 2.1 on the Fama-French HML factor, defined as

$$HML = \frac{1}{2} (SmallValue + BigValue) - \frac{1}{2} (SmallGrowth + BigGrowth)$$

If $C_i := (BE_i, ME_i)$ denotes the couple book-to-market and market capitalization and

$L_1 := \{ C_i | ME_i \leq \text{median}(ME), BE_i \geq 70^{th} \text{percentile} \}$
$L_2 := \{ C_i | ME_i \leq \text{median}(ME), 30^{th} \text{percentile} \leq BE_i < 70^{th} \text{percentile} \}$
$L_3 := \{ C_i | ME_i \leq \text{median}(ME), BE_i < 30^{th} \text{percentile} \}$
$L_4 := \{ C_i | ME_i > \text{median}(ME), BE_i \geq 70^{th} \text{percentile} \}$
$L_5 := \{ C_i | ME_i > \text{median}(ME), 30^{th} \text{percentile} \leq BE_i < 70^{th} \text{percentile} \}$
$L_6 := \{ C_i | ME_i > \text{median}(ME), BE_i < 30^{th} \text{percentile} \}$

then HML is a factor as per Definition 2.1 if we consider $\alpha_{HML} = (1/2, 0, -1/2, 1/2, 0, -1/2)$. Exhibit 1 gives a graphic view of the factor construction.

Our definition, although very general, essentially formalizes the modern way of building factor as long-short portfolios: once one has a measure that represents the characteristics, it is very common to slice it into different blocks (buckets $L$) and within each block equally weighting stocks. Definition 2.1 does not require the factor to be economically relevant.
3 Multi-factor

Investors have nowadays accepted the existence of a (small) number of economically relevant factors backed by academic research. There are multiple ways to use these factors, among others:

**Performance analysis.** Investors measure their portfolios exposures to each factor and disentangle the sources of their performances into beta (coming from their exposures to various factors) and alpha (manager skills).

**Risk control.** Investors shape their portfolios to manage exposures to unwanted factors. Their financial objective remains unchanged, but the factor exposures are directly managed according to their guidelines.

**Tactical exposure.** Investors tilt their portfolios to various factors in line with their views of the market, the economy and their subjective beliefs.

**Multi-factor approach.** Investors understand the existence of a positive expected return (premium) attached to the factors over the long run. Therefore they build their portfolios exposed to all the factors while benefiting from potential diversification.

The multi-factor approach has gained interest among investors because it is an elegant solution and suits their needs: a well-diversified portfolio, with explicit and targeted risks and well-identified performance drivers. Furthermore, it can be efficiently implemented through passive and low-cost funds such as ETFs. However, the question surrounding the optimal combination of factors remains mostly unanswered. Because factors underperformance can last for several quarters, the most appealing solution would be factor timing. Unfortunately, factor timing is a very difficult exercise and could lead to higher risk. This is counter-intuitive given the fact that one of the initial goals of the multi-factor approach is to spread risk across uncorrelated drivers of performances in a controlled way. As a matter of fact, the majority of investors use very simple methods of factor allocation, and the equal weighting of factors is by far the most popular. Furthermore, many multi-factor products available in the market (especially in the ETF format) are indeed an equally weighted allocation of factors.

The main result of this paper is that equal weighting of factors synthetically creates a new factor. Although this is not necessarily a negative outcome, investors should be aware that their static equal weight multi-factor approach is essentially a new factor in the sense of Definition 2.1. This implies, among other things, that investors are expecting that the measure associated with this new factor should be able to discriminate stock performances.
Theorem 3.1. Let $W^h, h = 1, \ldots, K$ be the stock weightings within $K$ different factors as in (2.1). Then the equally-weighted multi-factor portfolio, whose weights are given by

$$W_{i}^{MF} := \frac{1}{K} \sum_{h=1}^{K} W_{i}^{h}$$

is also a factor in the sense of Definition 2.1. More precisely there exists a measure (i.e. a stock characteristic) $C^{MF}$, a disjoint partition $L^{MF}$ and a vector of coefficient $\alpha^{MF}$ such that

$$W_{i}^{MF} = \frac{\alpha_{j(i)}^{MF}}{\bar{\alpha}_{j(i)}}$$

The proof of this result is given in Appendix A.

The result can be easily extended to any other static combination of factors.

Corollary 3.2. Let $W^h, h = 1, \ldots, K$ be the stock weightings within $K$ different factors as in (2.1). If $\pi$ is an arbitrary static allocation scheme ($\pi_h \in \mathbb{R}$ and $\sum_h \pi_h = 1$) then the multi-factor portfolio, whose weights are given by

$$W_{i}^{MF} := \sum_{h=1}^{K} \pi_h W_{i}^{h}$$

is also a factor in the sense of Definition 2.1.

The proof of this result is given in Appendix B.

Corollary 3.2 sheds lights on the static combination of factors. Indeed, if the weighting scheme is static then it is possible to build a composite measure such that the multi-factor portfolio is a factor. This would not be true if the weighting scheme is dynamic and depends on a different source of external information. But how good is the (synthetic) measure that underlies the multi-factor portfolio in forecasting returns, especially when compared to the standard measures that underlie factors such as for Size, Value, Momentum or Volatility?

4 Predictive power of factors’ measures

Our definition of factor in Definition 2.1 is very general. However, in the real world, only very few stock characteristics (measures) are able to generate meaningful factors. Because we showed that static combinations of factors are indeed factors, it is appealing to check whether such a combination of factors brings benefits. Indeed, as long as single factors are meaningful, economically sound, robust and possibly with low correlation to each other, it would make sense, from a portfolio construction perspective, to add them.
to the portfolio. What is less certain is whether a static (and most of the time, naive) combination of factors is the right choice. The majority of academic and empirical work on factors usually addresses the question in two ways:

- By looking at financial performances and risks of the factor (ex: does the factor outperform its benchmark over a sufficiently long period?)
- By showing that the factor is able to explain the differences in stock return.

Unfortunately, these approaches do not tell us much about the significance of a static multi-factor portfolio. To find out more about it, we should analyze the synthetic measure that underlies the multi-factor and see if it is at least as good as the single measure that underlies each of the individual factors. In other words, we do know that stock valuation is able to explain variation in future performances (on average), but what about the measure that underlies the multi-factor?

4.1 The data

We consider two different investment universes: the US and Europe. The US universe is represented by the S&P 1500 Index (which is the union of the S&P 500 Index [Large Caps], S&P 400 Index [Mid Caps] and the S&P 600 Index [Small Caps]) over the period of March 1996 to October 2017. The European universe is represented by the Stoxx Europe TMI Index (which contains a variable number of stocks between 900 and 1100) over the period of December 2000 to October 2017. We consider four factors: Size (the relative measure is the inverse of the stock weight within the respective index), Value (the measure is the stock earning-to-price ratio), Momentum (the measure is the one-year stock return) and Volatility (the measure is the inverse of the one-year rolling stock volatility).

Single factors are constructed according to Definition 2.1, similar to Fama and French (1993) approach, except that we do not double-sort for the sake of simplicity. The factors go long stocks with higher measures and short stocks with lower measures. The factors are rebalanced on the last business day of the month and data is sampled four days before. We also consider versions with lower frequencies (quarterly, semi-annually and annually rebalancings). Stock prices, total returns, earnings and index weights are sourced from Thompson Reuters Datastream. The difference in the horizons between the US and European universes depends on the availability of the reference indices we use. Furthermore, for consistency, we use the EP ratio to build our Value factor instead of the EB ratio, since the latter does not exist in the database with a sufficiently long history across our 2500 stocks circa.

We consider several versions of factors depending on the number of buckets (the partitions in Definition 2.1). We let this number vary from 2 to 10 to assess the robustness of our results with respect to the factor construction. Stocks are equally weighted within each bucket and are proportional to the vector $\alpha$. We consider the following simple
choice of $\alpha$: if $k$ is the number of buckets then:

$$\alpha_j = \left\lceil \frac{|j - \frac{k+1}{2}|}{\frac{k}{2}} \right\rceil \text{sign} \left( j - \frac{k+1}{2} \right)$$

(4.1)

For example, the Size factor with two buckets uses $\alpha = [-1, 1]$, goes short on the stocks with lower measures of size (i.e. upper half by market capitalization), long in the other half, and equal weight stocks within each bucket. The Value factor with five buckets will use $\alpha = [-1, -1/2, 0, 1/2, 1]$. It will go short on the first quantile based on earning-to-price ratio and equally-weight stocks, it will go short on the second quantile, equal weight stocks but the weights within this bucket will be divided by 2, and so on.

4.2 The significance criterion

The perfect measure (which obviously does not exist) is able to give us the full information about future stock returns. A plot with stock measures against their forward returns would give a straight line. On the other side of the spectrum, useless measures (which obviously exist) do not give us any information. We call such a measure a random pick. In this case, the plot would simply be noise, as shown in Exhibit 2. The correlation of the perfect measure with forward returns is one, while the random pick has correlation equal to zero. However, correlation is not adapted to our analysis. The following example provides further information on this.

Exhibit 2: Hypothetical plots of a perfect measure and a random pick against stock future performances.

Assume we have a perfect measure and a random pick. We build a new measure by matching the perfect measure on the four extreme buckets and the random pick on the six central ones (Exhibit 3). It turns out that for the parameters we selected, the correlation of this measure with the forward measure is roughly 80%. This is very misleading since for 60% (6 buckets out of 10) of the stocks, this measure is just noise. We propose
Exhibit 3: Hypothetical plots of a perfect measure and a measure which is perfect on extreme buckets and random in central buckets.

a new way to assess the significance of a measure that is closely linked to the predictive power on the cross-sectional stock return differences.

Let $C$ be the vector containing stock measures (e.g., size) and $R$ be the forward return of these stocks at a given horizon (e.g., three months). Fix a partition $L$ for $C$ (as in Definition 2.1), let $l > 0$ be the number of buckets in the partition and consider the corresponding $l$-quantiles for $R$. We count the number of stocks that belong both to the $j$-th bucket of $C$ and $R$. Basically, our significance criterion increases when the measure is able to correctly map forward returns. It should be noted that we do not require the partition $L$ to be $l$-quantiles (i.e., they do not need to have the same number of elements) but we do require that the partition on $R$ be quantiles (i.e., they have the same number of elements).

An example is shown in Exhibit 4. The six ($l = 6$) elements of the partition $L_1, \ldots, L_6$ for $C$ are mapped with the six quantiles $Q_1, \ldots, Q_6$. The criterion will count the number of stocks $i$ for which the pair $(C_i, R_i)$ falls in the diagonal blocks over the total number of stocks:

$$H(C, R, L) := \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{l} 1_{C_i \in L_j \text{ and } R_i \in Q_j}$$

(4.2)

For the example presented in Exhibit 3, we find that the perfect measure unsurprisingly obtains $H=1$, while the matched-on-extremes measure obtains $H = 51\%$, which is more in line with the fact that only four out of ten buckets are matched with the perfect measure. This number compares far better than the 80\% correlation. It is important to note that the significance criterion $H$ depends on both the quality of the measure $C$ (its predictive power on forward returns) and the partition $L$. The score $H$ for a factor can

Multi-factor portfolios: A new factor? Limits of the static approach.
Exhibit 4: Blocks formed by the partition $L$ on the measure $C$ and the partition $Q$ in the returns. A measure $C$ and the partition $L$ will receive a high score $H$ if the diagonal blocks capture a significant proportion of stocks.

be calculated directly from the weights:

$$H(w, R, L) := \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{l} \frac{1}{w_i} = \frac{n}{l} \text{ and } R_i \in Q_j$$

where $\alpha$ is the shape in the definition of the factor and $l_j$ is the number of stocks whose measures fall in $L_j$. Equation (4.3) is more adapted in practice because we do not need to know the measure that underlies the factor, only the weights. This is particularly useful for the multi-factor portfolio since the calculation of the weights is straightforward, whereas the calculation of the measure (detailed in the Appendix in the proof of Theorem 3.1) is more complex. If the measure $C$ (and therefore weights $w$) is totally random without any predictive power on forward returns, then a probabilistic argument gives us (on average)

$$H(w, R, L) := \sum_{j=1}^{l} \mathbb{P}(C \in L_j) \ast \mathbb{P}(R \in Q_j) = \sum_{j=1}^{l} \frac{l_j}{n} \ast \frac{n}{l} = \frac{1}{l}$$

It makes then sense to center Equation (4.3) as follows:

$$H^{\text{cent}}(w, R, L) := \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{l} \frac{1}{w_i} = \frac{n}{l} \text{ and } R_i \in Q_j - \frac{1}{l}$$

$H^{\text{cent}}$ represents the ability of a measure $C$ and a partition $L$ to predict forward returns in excess of a random pick. We think of the centered measure $H^{\text{cent}}$ as a success rate over the random pick. For the example presented in Exhibit 3 we have $l = 10$ buckets so that $H^{\text{cent}} = 51\% - 1/10 = 41\%$. 

Multi-factor portfolios: A new factor? Limits of the static approach.
4.3 Empirical results: The US case

Over the period of March 1996 to October 2017, we consider the standard four factors Size, Value, Momentum and Volatility, as well as the equally weighted multi-factor portfolio (EW). Factors are built according to Definition 2.1. The rebalancing frequency and the shape coefficients $\alpha$ vary according to the different tests we implement, while partitions are, for the sake of simplicity, always k-quantiles. At each rebalancing, we calculate the $H^{\text{cent}}$ score of the factor against the vector of forward returns sampled at a horizon coherent with the factor rebalancing frequency. Of course, this vector is not observable and not known at the time the factor is built. Global $H^{\text{cent}}$ scores are then averaged over time. Exhibit 5 shows the average $H^{\text{cent}}$ scores for the factors and the EW when we let the horizon vary from 1 to 12 months. For this particular test, we built the factor using terciles (long on the upper tercile of the measure and short on the lower one) with $\alpha = (-1, 0, 1)$.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M</td>
<td>0.28%</td>
<td>2.18%***</td>
<td>2.55%***</td>
<td>0.84%**</td>
<td>0.19%**</td>
</tr>
<tr>
<td>3M</td>
<td>0.16%</td>
<td>2.63%***</td>
<td>2.32%***</td>
<td>1.10%*</td>
<td>0.21%</td>
</tr>
<tr>
<td>6M</td>
<td>-0.25%</td>
<td>3.14%***</td>
<td>2.12%***</td>
<td>1.32%</td>
<td>0.19%</td>
</tr>
<tr>
<td>12M</td>
<td>0.53%</td>
<td>4.41%***</td>
<td>1.64%</td>
<td>1.82%</td>
<td>0.46%</td>
</tr>
</tbody>
</table>

Exhibit 5: Averages of $H^{\text{cent}}$ for different factors and the EW portfolio with different horizons. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

Although these numbers are averaged over more than 20 years, over which factors have been through high and low performance periods according to market regimes (see for example Guidolin and Timmermann (2008); Zhang et al. (2009); De Franco et al. (2017)), it gives an important insight about the predictive power of the factors.

First of all, we find that stock valuation (Value) is indeed a characteristic with a significant predictive power, which also increases with the horizon. When we rebalance the Value factor monthly (1M) and we look at 1-month ahead performances, on average, Value is able to correctly predict the tercile in which the return will belong better than random pick. Indeed, random pick would do this job with 1/3 chances, while Value does it with a success rate of 2.18% over this reference, so 35.51%. These numbers increase with time, so that at a 12-month horizon, Value correctly estimates the tercile of 12-month returns with a success rate of 4.41%.

Momentum also does a fair job at low horizons (1 to 6 months), but loses significance at long horizons: at 12 months, Momentum is still able to beat the random pick by 1.64%, but the estimate is no longer statistically significant. Volatility also shows positive success rates, but they are significant only at lower horizons. Size is not able to achieve statistically significant success rates. To some extent this confirms the findings in Van Dijk (2011) and references therein, where the size premium is questioned (does it still exist?). Our test is not meant to prove whether the premium exists or not. It rather measures the goodness of the attached measure in predicting forward returns.
The lack of significant predictive power may be due to large estimation errors (Lo and MacKinlay, 1990) or strong dependence of the premium on market regime (Guidolin and Timmermann, 2008), which does not entail the existence of a real premium paid to the investors that hold small companies. On the other hand, the size premium is significantly lower now than when it was studied by Banz (1981). The EW portfolio has by far the lowest scores, if we consider that for size we can at least advocate its higher volatility in the estimation. EW is able to beat random pick only at a 1-month horizon, and even for this, the success rate is only 0.19%. At longer horizons, EW is not able to forecast forward returns better than random pick.

Exhibit 6 collects the results of a similar analysis where we now let the number of buckets in the factor construction vary from 2 to 10. The alphas are taken as in Equation (4.1) so that for two buckets we will use \( \alpha = [-1, 1] \), for three buckets \( \alpha = [-1, 0, 1] \), for four buckets \( \alpha = [-1, -1/2, 1/2, 1] \) and so on. The rebalancing frequency is fixed at 3 months.

<table>
<thead>
<tr>
<th>Buckets</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.51%</td>
<td>0.97%***</td>
<td>-0.39%</td>
<td>0.99%</td>
<td>-0.03%</td>
</tr>
<tr>
<td>3</td>
<td>0.16%</td>
<td>2.63%***</td>
<td>2.32%***</td>
<td>1.10%*</td>
<td>0.21%</td>
</tr>
<tr>
<td>4</td>
<td>0.14%</td>
<td>2.73%***</td>
<td>2.58%***</td>
<td>1.14%**</td>
<td>0.10%</td>
</tr>
<tr>
<td>5</td>
<td>0.33%</td>
<td>2.61%***</td>
<td>2.53%***</td>
<td>1.34%***</td>
<td>0.07%</td>
</tr>
<tr>
<td>10</td>
<td>0.49%***</td>
<td>1.82%***</td>
<td>1.92%***</td>
<td>1.11%***</td>
<td>0.04%</td>
</tr>
</tbody>
</table>

Exhibit 6: Averages of \( H^\text{cent} \) for different factors and the EW portfolio with different numbers of buckets in the factor construction. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

It is remarkable how all factors, except Value, do not outperform random pick with only two buckets. When we use only two buckets, the long and short legs are too diversified (each leg has roughly 750 stocks). With a higher number of buckets, Value and Momentum are effective. Volatility also sees its success rates increase with the number of buckets, showing that the volatility premium starts to appear when we are able to discriminate between high volatility stocks from low volatility ones (Haugen and Heins, 1975; Haugen and Baker, 1991; Ang et al., 2006; Haugen and Baker, 2010; De Franco et al., 2017). With 5 buckets, Volatility beats the random pick (which has a success rate of 1/5 = 20%) by 1.34%. We still do not find any superior predictive power for Size except for the 10 bucket version, where the success rate is 0.49%. This fact reflects the asymmetric nature of this premium, as shown in Pettengill et al. (2002). Once again, the EW portfolio is not able to show any significant predictive power. By increasing the number of buckets, the EW portfolio dilutes so extensively the predictive power of each single factor that, in the end, there is not enough useful information to differentiate it from random pick.

The final test we propose tries to measure the ability of the factors to correctly predict Multi-factor portfolios: A new factor? Limits of the static approach.
forward returns at least in the extreme buckets. The idea behind this is that, most often, spreads in stock performances are explained by specific stock characteristics when one compares the lowest versus the highest quantile in other words, where the risk premium is at its strongest. We modify $H_{\text{cent}}$ in (4.4) in order to count only extreme buckets (by restricting the summation over indexes $j$ that correspond to those extreme buckets) and changing the normalization term $1/l$ into $(\text{nb of extremes})/l^2$ to be coherent. Exhibit 7 collects the results when we consider factor construction based on 10 deciles, $\alpha = [-1, -4/5, -3/5, -2/5, -1/5, 1/5, 2/5, 3/5, 4/5, 1]$ and let the number of extremes vary from 1 to 5 (the case with 5 extremes corresponds to the initial $H_{\text{cent}}$). We observe that

<table>
<thead>
<tr>
<th>Extremes</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.34%***</td>
<td>1.27%***</td>
<td>1.41%***</td>
<td>0.82%***</td>
<td>0.01%***</td>
</tr>
<tr>
<td>2</td>
<td>0.31%**</td>
<td>1.40%***</td>
<td>1.44%***</td>
<td>0.92%***</td>
<td>0.01%**</td>
</tr>
<tr>
<td>3</td>
<td>0.33%**</td>
<td>1.48%***</td>
<td>1.46%***</td>
<td>1.02%***</td>
<td>0.02%**</td>
</tr>
<tr>
<td>4</td>
<td>0.36%**</td>
<td>1.59%***</td>
<td>1.62%***</td>
<td>1.10%***</td>
<td>0.02%*</td>
</tr>
<tr>
<td>All</td>
<td>0.49%***</td>
<td>1.82%***</td>
<td>1.92%***</td>
<td>1.11%***</td>
<td>0.04%</td>
</tr>
</tbody>
</table>

Exhibit 7: Averages of $H_{\text{cent}}$ for different factors and the EW portfolio with different numbers of extremes selected. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

for all factors the success rates are positive and significant if we restrict the analysis over the extremes only. This result is in line with the fact that the premium is very strong at the extreme (for example, very small caps versus very large caps, lowest volatility versus highest volatility and so on). When we consider, for example, only 1 extreme (the lowest deciles versus the highest decile), success rates go from $0.34\%$ for Size to $1.41\%$ for Momentum, given that the reference of the random pick is $2/10^2 = 2\%$. Momentum achieves then $2\% + 1.41\% = 3.41\%$ predictive power, an increase of $70.5\%$ over the random pick. Values success rate is $1.27\%$, an increase of $63.5\%$. Volatilities success rate is $0.82\%$, or $41\%$ in relative terms. The EW also achieves statistically significant results, but they are very minor compared to the other factors. This test partially explains why the EW ranks systematically lower than the other factor.

By equally weighting all the factors, most of the stock allocations will net out. In the end, we will find positive weights on the stocks that are in the highest quantiles of the majority factors, negative weights for those in the lowest quantiles of the majority of factors, and a very scattered allocation in between. Since the factors premia are stronger at the extremes, and the number of stocks that are at the same time in the highest (or the lowest) quantile of the majority of factors is low, this automatically results in a loss of predictive power.

4.4 Empirical results: the European case

We repeat the same analysis for the European case. The setup is similar except for the data time period (from December 2000 to October 2017) and the use of prices in
local currency for price-based measures (Volatility, Momentum and forward returns). Our choice is motivated by the fact that the predictive power of standard characteristics generally should exist at the local currency level while large swings in FX rates could in principle change the picture. However, results when all prices are converted in EUR remain in line with our findings.

Exhibit 8 compares with Exhibit 5 and shows the average $H^{cent}$ scores for the factors and the EW portfolio when the rebalancing frequency varies from 1 to 12 months and we use terciles with $\alpha = [-1, 0, 1]$. Like in the US case, we find in Europe a strong predictive power for Value and Momentum, with a success rate ranging from 2.44% to 5.13% for Value as the horizon increases and from 3.18% to 5.14% for Momentum. While Value was also strong in the US, we observe that in the European case Momentum is even stronger.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M</td>
<td>-0.28%</td>
<td>2.44%***</td>
<td>3.18%***</td>
<td>1.26%***</td>
<td>0.29%**</td>
</tr>
<tr>
<td>3M</td>
<td>-0.26%</td>
<td>2.98%***</td>
<td>3.75%***</td>
<td>1.47%</td>
<td>0.55%***</td>
</tr>
<tr>
<td>6M</td>
<td>0.20%</td>
<td>3.87%***</td>
<td>3.69%***</td>
<td>2.19%*</td>
<td>0.87%***</td>
</tr>
<tr>
<td>12M</td>
<td>-0.12%</td>
<td>5.13%***</td>
<td>5.14%***</td>
<td>2.86%</td>
<td>1.49%**</td>
</tr>
</tbody>
</table>

Exhibit 8: Averages of $H^{cent}$ for different factors and the EW portfolio at different frequencies. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

Volatility has positive success rates at all frequencies and these rates increase at lower frequencies, although the numbers are too noisy to get a statistically significant estimate. As for the US, we have a similar picture for Size, if not worse. The success rate is never significant and even negative. Finally, in contrast with the US, the EW portfolio shows small but statistically significant success rates that increase from 0.29% at a 1-month horizon to 1.49% at a 12-month horizon. Though positive, these numbers are significantly lower than three of the four factors used. It seems that by equally weighing the factors, we lose a significant proportion of the predictive power that each factor entails.

<table>
<thead>
<tr>
<th>Buckets</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.10%</td>
<td>2.07%***</td>
<td>2.13%**</td>
<td>1.04%</td>
<td>0.63%*</td>
</tr>
<tr>
<td>3</td>
<td>-0.26%</td>
<td>2.98%***</td>
<td>3.75%***</td>
<td>1.47%</td>
<td>0.55%***</td>
</tr>
<tr>
<td>4</td>
<td>0.03%</td>
<td>3.34%**</td>
<td>3.75%***</td>
<td>1.67%**</td>
<td>0.41%**</td>
</tr>
<tr>
<td>5</td>
<td>0.17%</td>
<td>3.07%***</td>
<td>3.60%***</td>
<td>1.65%**</td>
<td>0.26%***</td>
</tr>
<tr>
<td>10</td>
<td>0.16%</td>
<td>2.27%***</td>
<td>2.58%***</td>
<td>1.47%***</td>
<td>0.19%**</td>
</tr>
</tbody>
</table>

Exhibit 9: Averages of $H^{cent}$ for different factors and the EW portfolio with different numbers of buckets in the factor construction. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

Exhibit 9 is the equivalent of Exhibit 6 for the European universe. We fix the frequency at 3-month and let the number of buckets in the factor construction vary from 2 to 10,
while the alphas are given in Equation (4.1). The results are very similar for Value and Momentum, even if the success rates are higher for the European factors than for their US counterparts. Volatility starts to achieve positive and significant success rates when the number of buckets is sufficiently large, once again confirming that this specific premium strengthens when we discriminate between very low and very high volatility stocks. Size is not able to show positive success rates even with a very large number of buckets: for 10 buckets it has a not significant 0.16% success rate, while the US version shows a significant 0.49% (Exhibit 6). The EW portfolio has positive but low success rates that naturally decrease when we increase the number of buckets, signaling that equal weighting dilutes too extensively the power of each individual factor.

Finally, Exhibit Exhibit 10 reproduces the test in Exhibit 7 by removing the central buckets in the calculation of $H_{cent}$. Results for Value, Momentum and Volatility are in line with the US case: the success rate is significant and positive. Size has a positive and significant success rate only when we consider the lowest and the highest decile (one extreme), but it is not able to do better than the random pick starting from two extremes. Finally, the EW achieves significant scores, but they are very small: for example with two extremes considered (so four buckets in total, two on the lower side and two on the higher side of the measure), the random pick has a $H$ score of 4/100, Value obtains 5.68% (=4% + 1.68%), Momentum 5.96% (= 4% + 1.96%), Volatility 5.08% (= 4% + 1.08%) while EW is only at 4.03% (= 4% + 0.03%).

<table>
<thead>
<tr>
<th>Extremes</th>
<th>Size</th>
<th>Value</th>
<th>Momentum</th>
<th>Volatility</th>
<th>EW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.23%*</td>
<td>1.42%***</td>
<td>1.74%***</td>
<td>1.08%***</td>
<td>0.01%</td>
</tr>
<tr>
<td>2</td>
<td>0.19%</td>
<td>1.68%***</td>
<td>1.96%***</td>
<td>1.08%***</td>
<td>0.03%***</td>
</tr>
<tr>
<td>3</td>
<td>0.13%</td>
<td>1.81%***</td>
<td>2.04%***</td>
<td>1.25%***</td>
<td>0.04%***</td>
</tr>
<tr>
<td>4</td>
<td>0.19%</td>
<td>2.02%***</td>
<td>2.18%***</td>
<td>1.32%***</td>
<td>0.06%***</td>
</tr>
<tr>
<td>All</td>
<td>0.16%</td>
<td>2.27%***</td>
<td>2.58%***</td>
<td>1.47%***</td>
<td>0.19%***</td>
</tr>
</tbody>
</table>

Exhibit 10: Averages of $H_{cent}$ for different factors and the EW portfolio with different numbers of extremes selected. Significance: *** = 1% ** = 5% * = 10%. No stars = loading not significant.

4.5 Discussion

When we compare the US and European tests, we observe the following:

- The EW portfolio, a new factor as per Definition 2.1, is a factor derived from a measure that is not more powerful than a random pick when it comes to discriminating between low and high-performing stocks.

- In both US and Europe, Value and Momentum achieve the best success rates, with Volatility slightly lower. In Europe, these three factors have stronger predictive power than the US versions.
• Size does not have a strong predictive power according to our $H_{cent}$ measure. This does not question the validity of the Size premium, as confirmed by the literature which proves that there is a size/liquidity premium in US, Europe and international markets. However, the very volatile nature of this premium, the strong correlation it has with the business cycles and its lottery-like behavior make the estimations of success rates quite noisy.

• This logic does not apply to the EW portfolio because it should dampen these effects through diversification. But what we observe is that indeed the EW smooths out the predictive power of each single factor so that its success rate, on average, is not distinguishable from random pick.

Equal weighting (or more generally static combinations of factors) has its strengths and investors are already considering this approach to build their multi-factor allocation. Most of the time, the choice is driven by the ex-ante performance we can expect by smoothing out factors premia. Furthermore, investors are aware of the difficulty to time the factors (which are time varying, they depend on the market regime and are influenced by the business cycle). However, investors should keep in mind that simple static allocation, like equal weighting, is indeed a factor related to some synthetic measure and that this measure is very poor when it comes to discriminating between good and bad performers.

5 Conclusions

This work sheds new light on factor investing and multi-factor portfolio construction. We are able to show that in a very general framework, static multi-factor allocations are indeed new factors. It is then natural to assess to which extent these multi-factor portfolios are able to increase the predictive power that is usually linked to standard factors. Our empirical tests show that simple equal weighting usually dilutes too extensively the predictive power of each single factor and, on average, we are not able to distinguish it from a random pick. Equally weighing factors (as well as other static schemes) has many merits — it is simple to understand, it diversifies risks, it smooths the premium earned by each factor — and our research does not intend to diminish its interest from a practical point of view. But investors must be aware that static multi-factors are factors in the sense that they translate a synthetic characteristic into portfolios of stocks. Unfortunately, this new synthetic characteristic is not often better than a random pick.

References


Multi-factor portfolios: A new factor? Limits of the static approach.


Hsu, J. and V. Kalenik (2014). Finding Smart Beta in the factor zoo. *Research Affiliates (July).*


Multi-factor portfolios: A new factor? Limits of the static approach.


### A Proof of Theorem 3.1

*Proof.* We will prove the result by assuming the following on the partitions defining each factor:

**A1** All factor partitions are of fixed size \( l > 0 \), and all \( s \)-th elements of each partition have the same number of elements since it simplifies the calculations. At the end of the proof we will show that we can easily remove it, so that the proof is valid in its general form. From the definition of factors in 2.1 we can write:

\[
W_i^{MF} := \frac{1}{K} \sum_{h=1}^{K} W_i^{h} = \frac{1}{K} \sum_{h=1}^{K} \frac{\alpha_j^{h(i)}}{l_j^{h(i)}}
\]

and from the fact that cardinality of each element of the partition is the same across factor (Assumption **A1**) we can write:

\[
W_i^{MF} := \frac{1}{K} \sum_{h=1}^{K} \hat{\alpha}_j^{h(i)}
\]

where we simply defined \( \hat{\alpha}_s^{h} = \alpha_s^{h}/l_s \). We can imagine \( \hat{\alpha} \) as a \( l \times K \) matrix, where each column contains the alphas defining the factor. If:

\[
\mathcal{A} := \{ \sigma_i : \{1, \ldots, K\} \to \{1, \ldots, l\} \mid \sigma_i : (1, \ldots, K) \to (j^1(i), \ldots, j^K(i)) \}
\]

then:

\[
W_i^{MF} := \frac{1}{K} \sum_{h=1}^{K} \hat{\alpha}_j^{h(i)} = \frac{1}{K} \sum_{h=1}^{K} \hat{\alpha}_{\sigma_i(h)} := W^{MF}(\sigma_i)
\]

The map \( \sigma_i \) simply gives, for each factor \( (1, \ldots, K) \) the index of the relative measures \( (C_1^i, \ldots, C^K_i) \) into their respective partitions. The set of all possible values for \( W^{MF} \) is simply given by:

\[
\Omega := \left\{ W^{MF}(\sigma) \mid W^{MF}(\sigma) = \frac{1}{K} \sum_{h=1}^{K} \hat{\alpha}_{\sigma(h)} ; \sigma \in \mathcal{A} \right\}
\]
which can be written as \( \Omega := \{ \beta_1 \leq \beta_2 \leq \cdots \leq \beta_{M(K,l)} \} \). We denote the cardinality of \( \Omega \) by \( M(K,l) \) which is bounded from above by \( n \) (indeed \( \Omega \) can contain a lower number of elements if some betas coincide). Let:

\[
C_i^{MF} := \frac{1}{K} \sum_{h=1}^{K} \hat{\alpha}_{\sigma_i(h)}, \quad \sigma_i \in \mathcal{A}
\]

This composite measure is built in a very efficient computational way:

1. \((C^1_{i}, \ldots, C^K_{i}) \rightarrow (j^1(i), \ldots, j^K(i))\): from the vector of measures of each stock, compute the indexes of the respective buckets of their partitions.

2. \((j^1(i), \ldots, j^K(i)) \rightarrow \sigma_i\): build the corresponding map in \( \mathcal{A} \)

3. \(\sigma_i \rightarrow C_i^{MF}\): compute the composite measure.

The partition set \( L^{MF} \) is very straightforward: \( L^s_{MF} = \{ \beta_s \} \) for each \( \beta_s \in \Omega \) and \( l^s_{MF} = \text{card} \{ i \in \{1, \ldots, n\} | C_i^{MF} = \beta_s \} \). Finally

\[
\alpha^s_{MF} = l^s_{MF} \ast \beta_s, \ s = 1, \ldots, M(K,l)
\]

Putting all together, we finally obtain

\[
W^{MF}_i = W^{MF}(\sigma_i) = C_i^{MF} = \beta_{s_i} = \frac{\alpha^s_{Mi}}{l^s_{Mi}}
\]

where \( s_i \) is the index of \( C_i^{MF} \) in the partition \( L^{MF} \), i.e. \( C_i^{MF} \in L_{s_i} \), in other words \( s_i = j^i_{MF} \). We have then built a measure \((C^{MF})\), a partition \((L^{MF})\) and a vector of alphas \((\alpha^{MF})\) such that \( W^{MF} \) is the corresponding factor, which concludes the proof.

Let us now turn to the technical assumption \( A1 \). Each factor is characterized by its partition:

\[
\begin{align*}
\text{Factor}_1 & \rightarrow \{ L^1_1, \ldots, L^1_{l_1} \} \\
\text{Factor}_2 & \rightarrow \{ L^2_1, \ldots, L^2_{l_2} \} \\
\vdots & \\
\text{Factor}_K & \rightarrow \{ L^K_1, \ldots, L^K_{l_K} \}
\end{align*}
\]

\( l_1, \ldots, l_K \) being the size of each partition. Assume that the first partition has the maximum size \( l = l_1 = \max_h l_h \). We can simply extend each partition from 2 to \( K \) at size \( l \) by dividing any non-singleton element into a smaller subset until the size of the partition reaches \( l \). If the subset does not have enough elements inside it, we can repeat the operation with another one. We should then adjust the vector of respective alphas by adding new entries. As such, there is no loss of generality to assume that all partitions are made of \( l \) elements.
The second part of Assumption A1 assumes that $L_1^s, L_2^s, \ldots, L_K^s$ have the same number of elements. In other words, for each factor, the number of stocks whose measures fall into the $s$-th element of the partition is the same. If this is not the case, we can calculate the greatest common divisor of the different sizes:

$$l_s := \text{GCD} \left\{ l_h^s \mid h = 1, \ldots, K \right\}$$

and then for each $s$-th element whose size is not $l_s$ we can divide equally into subsets of size $l_s$. Once again, we should then adjust the vector of respective alphas by adding new entries. At the end of the two-step procedure, the new partitions are all of size $l' \geq l$ (because we may have created more subsets) but all corresponded buckets for each factor have the same size. This proves that Assumption A1 is not binding.

B Proof of Corollary 3.2

Proof. Each factor is characterized by a specific vector of alphas $\alpha^h$. Simple re-adjustments give us

$$W_{i}^{MF} := \sum_{h=1}^{K} \pi_h W^h_i = \frac{1}{K} \sum_{h=1}^{K} \tilde{W}_i^h$$

where $\tilde{W}_i^h$ are the stock weights of a new factor derived from $W^h_i$, for which we simply modified the relative $\alpha^h$ as follows:

$$\tilde{\alpha}^h = K \pi_h \alpha^h$$

Indeed the multi-factor $W^{MF}$ is the equal weighting of new adjusted factors, and then by Theorem 3.1 is itself a factor.

Multi-factor portfolios: A new factor? Limits of the static approach.
About Ossiam

Ossiam is a research-driven French asset management firm (authorized by the Autorité des Marchés Financiers) and specializes in delivering smart beta* solutions. Efficient indexing is at the core of Ossiam’s business model. The firm was founded in response to a post-subprime crisis demand from investors for simplicity, liquidity and transparency. Given the environment, there was a growing need among investors for enhanced beta exposure and risk hedging. Ossiam is focused on the development of innovative investment solutions for investors via a new generation of indices.

*‘Smart beta’ refers to systematically managed, non-market-cap-weighted strategies covering any asset class.

Multi-factor portfolios: A new factor? Limits of the static approach.