Vanguard®

Can active equity managers be cloned using factor funds? A practical testing method

Inna Zorina, CFA; Reimar Scholz; and Douglas M. Grim, CFA

- Given developments in academic research, technology, and access to product strategies, more investors are assessing whether traditional active managers are generating excess returns that cannot be explained by well-known style factors, which can now be purchased directly and cheaply by investors worldwide. Some may be content with what they find. Others, however, may be surprised to learn that what they have been getting from a manager is just a fairly consistent tilt toward one or more of these factors, when their objective for hiring the manager may have been solely to generate positive alpha.

- In cases where traditional active managers are not adding unique value or are charging too much, replacing them with factor strategies may provide the end investor with a way to generate excess return with greater transparency, more risk control, and lower implementation costs.

- We develop an accessible method to test to what extent the returns of a traditional active manager can be replicated using factor-based strategies. This framework has a number of practical applications, including replacing a manager who is not delivering net alpha above investable factor exposure and performing due diligence on prospective managers.

Introduction

Performance attribution, including attribution done through a factor lens, has been a standard part of top-tier active manager due diligence for decades. **Figure 1** illustrates the evolution of factor-based evaluation in assessing active equity managers. In some cases, investors may learn that the returns they have been getting from their active equity manager come from a consistent tilt toward one or more well-known factors (such as value or quality)—although their sole objective in hiring the manager may have been to generate positive, manager-specific alpha.[1,2] The continued development of investable strategies to harvest factor returns in a low-cost manner has effectively raised the bar for many managers, who need to produce attractive returns to justify their higher fees. We aim to empower the average professional investor by providing a simple but powerful testing framework that can reveal the extent to which a manager's returns can be cloned using widely available, lower-cost, and more transparent factor products.

Evaluating the factor profile of an active equity manager

Methods of assessing the factor profile of an active manager date back at least to the late 1960s (Jensen 1968; Fama, 1972). Since that time, a plethora of techniques have been proposed in the academic literature. At the highest level, there are two general approaches: returns-based style analysis (RBSA) and holdings-based style analysis (HBSA).[3] Both methods are useful (Israel and Ross, 2017) but have well-documented limitations, and neither stands out as the clearly superior approach. In this paper, we focus on a specific RBSA approach.[4]

RBSA has a number of key practical benefits. It requires fewer inputs and is relatively easy to perform and customize (Christopherson and Sabin, 1999). RBSA aims to match the month-to-month historical performance of a manager with that of a buy-and-hold set of factors. The approach attempts to reflect the typical behavior of the manager, and corresponding factor weights and differences in performance can be easily measured. Lastly, with RBSA the investor can decide which factors to include in the analysis and which factor strategy to use to represent the performance of each factor.

Figure 1. Attribution models have evolved to distinguish factor exposure from true alpha contribution

Evolution of active manager attribution models
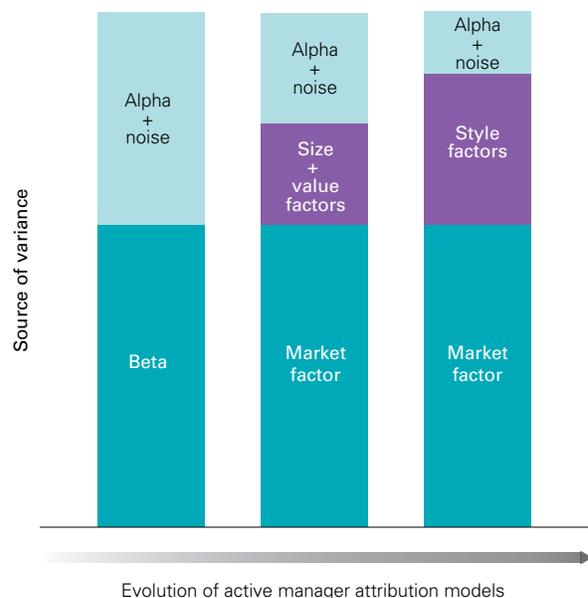
Notes: This is a hypothetical scenario for illustrative purposes only. *Noise* refers to the fact that through any period, some degree of statistical randomness affects results.

Source: Vanguard, 2020.

---

1 For the purposes of this paper, our definition of *alpha* is value added or subtracted by an active manager through difficult-to-replicate security selection or sector/factor timing.

2 For general information on equity factors, see Grim et al. (2017).

3 HBSA is also referred to in the literature as portfolio-based style analysis or characteristics-based style analysis.

4 A detailed comparison is beyond the scope of this paper. Bender, Hammond, and Mok (2014) note that RBSA is commonly used by industry practitioners. More sophisticated investors with access to the proper tools, data, and technical expertise can complement our method with an HBSA approach and a direct manager discussion for a more comprehensive performance assessment.

Selecting a suitable RBSA approach for the job

Choosing an appropriate RBSA approach requires clarifying the objective for the analysis.[5] In fact, there are many practical applications for factor models, including market efficiency testing, manager style consistency evaluation, peer group classification, benchmarking, performance driver identification, and portfolio fit assessment.

Our goal is to determine to what extent a manager's performance can be replicated simply by systematically tilting toward well-known style factors. If the manager's performance is almost entirely explained by such factor tilts, the investor should determine whether the same results could be achieved in a more transparent, risk-controlled, and lower-cost way using available off-the-shelf equity factor products.

For this purpose, we require an RBSA method that employs long-only, investable factor proxies to create a fair alternative investment option to a long-only manager or a multimanager portfolio that is subject to real-world fees and transaction costs.[6] These added criteria narrow our search to methods that, directly or indirectly, build off of the pioneering work by Sharpe (1988, 1992), such as those set forth in Davis et al. (2008), Bender, Hammond, and Mok (2014), and Berk and van Binsbergen (2015). To make the factor portfolio investable and the RBSA output more intuitive, Sharpe (1992) suggested including two constraints: (1) that the mimicking portfolio be fully invested and (2) that it use no shorting or explicit leverage.

Employing these constraints ensures that the factor coefficients are straightforward to interpret as the portfolio weight to each factor that would have led to the best fit between the returns of the factor-mimicking portfolio and those of the manager for a given period.[7] If the net alpha for the manager is negative or not statistically significant, it's likely that the manager did not deliver unique value to end investors for the period studied.

Constrained RBSA

The approach we chose to use could be described as constrained optimization to find the portfolio with the best historical fit:

$$R_t^M = \alpha + \sum_{j=1}^{J} w_j^F R_{j,t}^F + \varepsilon_t$$

where

R_t^M represents the manager's return for month t;

α represents alpha;

F represents selected factors;

w_j^F represents factor coefficients (i.e., factor weights);

$R_{j,t}^F$ represents the factor proxy return for month t;

j represents factor proxy number between 1 and J, where J is the number of proxies; and

ε_t represents residual returns (i.e., error term).

We define *alpha* as the manager's excess return relative to the custom factor portfolio. It could be driven by the manager's security selection and timing skill, or by good or bad luck.[8] Similar to Sharpe (1992), we impose two constraints:

$$\sum_{j=1}^{J} w_j^F = 1$$ Factor weights must sum to 1. This assumes the manager is fully invested in equities.

$$0 \le w_j^F \le 1$$ All factor weights must be equal to or greater than 0 and less than or equal to 1. This assumption limits all factor exposures to long-only, while not allowing any leverage or short exposures.

---

5 See, for example, the methods described in: Jensen (1968), Sharpe (1992), Fama and French (1993), Carhart (1997), Davis et al. (2008), Bender, Hammond, and Mok (2014), Fama and French (2015), Berk and van Binsbergen (2015), Kahn and Lemmon (2016), Stambaugh and Yuan (2017), Fama and French (2018), and Hou et al. (2018). Technically, each of these approaches is performed with returns-based analysis using styles; however, we acknowledge that some researchers assume that RBSA only represents approaches that use long-only styles.

6 For instance, Berk and van Binsbergen (2015) and Madhavan, Sobczyk, and Ang (2018) highlight the importance of investability with proxies chosen for RBSA with certain objectives.

7 In technical terms, the method uses quadratic programming to determine the buy-and-hold weights for each factor strategy that together will minimize the squared error between the active manager and the long-only factor-mimicking portfolio's month-to-month returns, subject to the binding constraints of the mimicking portfolio being fully invested without any shorting or explicit leverage.

8 Alpha could also result from model misspecification (e.g., suboptimal choice of factor proxies or omitted factors).

Is there a hindsight bias with RBSA?

Berk and van Binsbergen (2015) and Frazzini, Kabiller, and Pedersen (2018) rightly argue that investors must be fair in how they judge managers with a long track record of tilting toward certain factors that were not recognized by most market participants. A manager who identifies these stock characteristics as drivers of outperformance before they are widely publicized and has the discipline to stick with the strategy, particularly during out-of-favor periods, should get credit for that as a form of alpha. The purpose of our framework, however, is not to determine whether a manager had alpha in the past. Instead, the objective is to identify managers who have exhibited a consistent style profile and could be cloned using low-cost factor strategies that are investable today.

RBSA: Part art, part science

RBSA done correctly can lead to valuable insights. However, the output can be misleading if a robust system of checks and balances is not in place. Although RBSA is a quantitative approach, professional judgment is still essential at different steps in the process. Improper model specifications can lead to incorrect interpretations of results and to imprudent investment decisions (Israel and Ross, 2017). In this section, we address the most important RBSA methodology considerations to mitigate statistical biases and errors that can occur.

Which factors should be included?

There is no industry-wide consensus on which factors should be considered when attempting to mimic the style performance of a manager. In theory, only the factors the manager is likely to be exhibiting should be included. However, that may not always be obvious, and at times the manager's style profile may not align with expectations.[9] A prudent starting point would be to consider factors that have a sound economic rationale and extensive empirical evidence for driving equity return premiums.[10] Some managers may exhibit modest factor tilts versus the broad market. In such cases, including a total market index fund in the RBSA could make sense.[11]

What strategy should serve as the proxy for each factor?

Once the factor candidates have been selected, the investor must decide which investable strategy should represent the historical performance for each factor. Judgment can help determine what would be most appropriate given the manager's strategy, as there is no universally agreed-upon way to define a factor (Buetow, Johnson, and Runkle, 2000).[12] For example, Hsu, Kalesnik, and Kose (2019) describe how some quality factor strategies target very different stock characteristics than others do.

Factor proxy selection is critical, because in any given year, strategies that tilt toward the same factor can generate very different performance results (Brown and Mott, 1997). Material cross-sectional performance dispersion is typically driven by factor proxy construction differences such as eligible stock universes, stock characteristic choices, screening cutoffs (for example, top 50%), weighting scheme (for example, capitalization, factor strength), and rebalancing and reconstitution frequency.

---

9 For example, Swinkels and van der Sluis (2006) argue that a portion of funds have misleading names, have vague investment objectives, or pursue a different style than advertised. They also cite a few empirical studies that provide evidence of fund misclassification.

10 Beck et al. (2016) found that characteristics targeting value, momentum, quality, minimum volatility, and liquidity have the largest number of supporting academic research articles.

11 Another option is to include multifactor products, as they can exhibit a different month-to-month return pattern than a combination of single factor funds provided they are constructed in a bottom-up manner (i.e., by assessing stocks across multiple factors before deciding whether and how to include them in an equity portfolio). An investor could also include a cash-equivalent proxy such as a money market fund to capture any return a manager generates if they do not typically keep their cash position fully equitized.

12 For instance, stock characteristics of price-to-book and price-to-earnings are different fundamental metrics, but they tend to proxy for the same factor theme (namely, value).

Figure 2 illustrates how a set of transparent, rules-based strategies targeting the same factor yielded strikingly different performance for 2019, with 12-month returns ranging from 18% to 32%. For example, the Russell 2000 Value Index and Russell 2000 Pure Value Index generated returns of 22% and 18%, respectively, even though they were constructed by the same index provider. Although both indexes select from the Russell 2000 stock universe, their weighting schemes differ greatly. The value index, for example, weights stocks based on market capitalization, while the pure value index weights stocks based on the quantitative attractiveness of their value characteristics as defined by FTSE Russell.

This is not a bug, but rather represents an intentional feature, as index providers typically produce indexes with different construction methods in order to meet the preferences of different sets of investors.

Because different factor proxies can generate different RBSA results—which in turn affects the calculated alpha, the tracking error, and even the stability of the factor exposures—we recommend conducting the analysis in multiple rounds, varying the factor proxy chosen for a given factor.[13]

Figure 2. Strategies that target the same factor can deliver materially different results
Gross returns for 2019
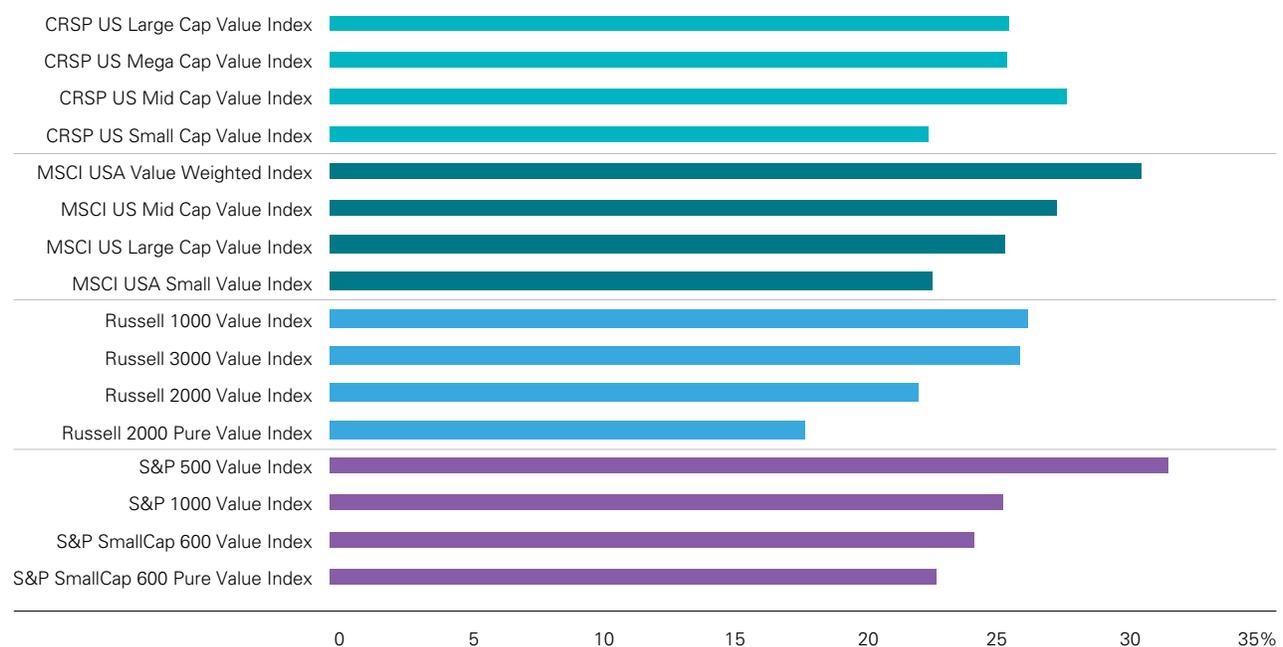
Notes: This represents a sample, not an exhaustive list, of U.S. equity value indexes. Past performance is no guarantee of future returns. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index. Benchmark comparative indexes represent unmanaged or average returns for various financial assets, which can be compared with funds' total returns for the purpose of measuring relative performance.
Sources: Vanguard calculations, using data from Morningstar from January 1, 2019, to December 31, 2019.

---

13 Tracking error indicates how much a manager's return differs from a specified benchmark (for the purposes of this analysis, the "benchmark" would be the factor-mimicking portfolio). The most common industry measure is the annualized standard deviation of excess return between the manager and benchmark.

Manage multicollinearity

Using a set of long-only equity factor strategies to conduct RBSA comes with a potential statistical issue: multicollinearity.[14] This is because their returns can be highly correlated. If two factors move in unison, it becomes difficult to distinguish if one or both are influencing the manager's returns.[15] Since each factor strategy is fully invested in stocks, the primary driver of month-to-month variability is the market factor. As a result and unsurprisingly, there has been a strong positive directional relationship with the performance of factors. Figure 3 shows this relationship—but it also shows that the magnitude of yearly performance differences between factors can be quite large.

Ignoring the potential impact of multicollinearity can lead to a false sense of accuracy in factor weights.

Standard errors, which are widely used as a statistical measure to assess the accuracy of regression coefficients, can generate misleading estimates if they are not adjusted for multicollinearity.[16] Lobosco and DiBartolomeo (1997) developed an approach that helps take multicollinearity into account, adjusting standard errors for the amount of information that each factor adds to the RBSA.[17] When a factor's returns are very close to a linear combination of other factors used in the analysis, the added (or, "unexplained") return and added (unexplained) volatility of that factor will be relatively small and the standard error for the factor will be relatively large.

Figure 3. Long-only factors have exhibited significant differences in performance

Legend: Market, Minimum volatility, Momentum, Quality, Size, Value
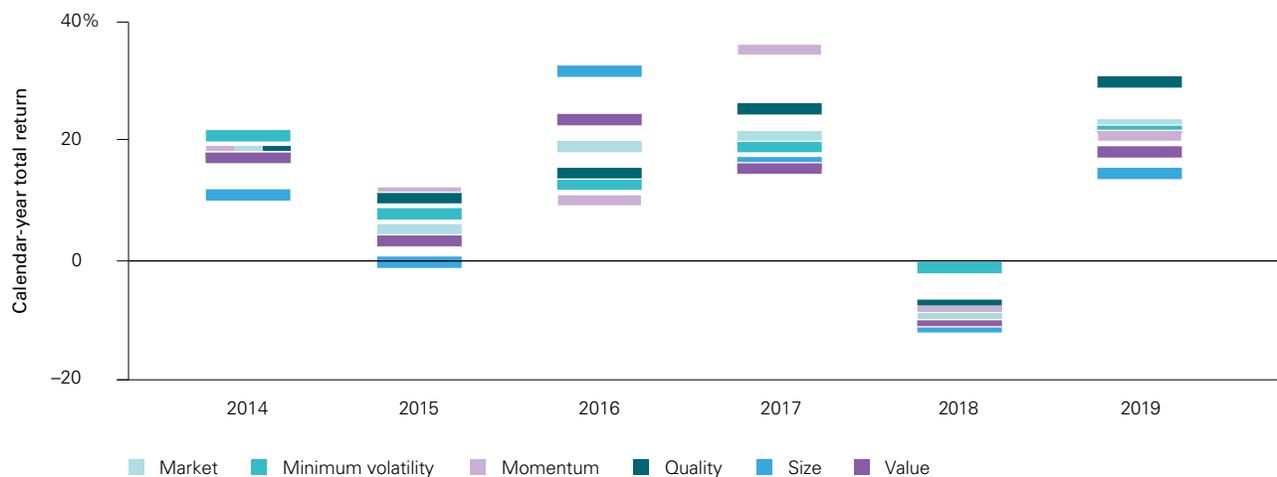
Notes: Factor benchmark returns are calculated using MSCI USA gross return indexes: MSCI USA Index (USD), MSCI USA Minimum Volatility Index (USD), MSCI USA Momentum Index (USD), MSCI USA Quality Index (USD), MSCI USA Small Cap Index (USD), and MSCI USA Value Index (USD). The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index. Benchmark comparative indexes represent unmanaged or average returns on various financial assets, which can be compared with funds' total returns for the purpose of measuring relative performance.

Sources: Vanguard calculations, using data from Morningstar from January 1, 2014, through December 31, 2019.

---

14 Multicollinearity occurs when two or more explanatory variables in a multiple regression model are highly linearly related. Its presence results in large standard errors of the affected coefficients. Another issue with multicollinearity is that small changes to the input data can lead to large changes in the model.

15 A key goal of regression analysis is to isolate the relationship between each independent variable (factor return) and the dependent variable (manager return). A regression coefficient represents the average change in the dependent variable for one unit change in an independent variable holding all of the other independent variables constant. However, when independent variables are correlated, changes in one variable may be associated with shifts in another variable. The stronger the correlation, the more difficult it is to estimate the actual effect of each variable.

16 In this context, a larger standard error indicates a lower degree of confidence in the factor weight estimate from the RBSA output.

17 Adjusting standard errors in this manner is not typically part of a standard statistical tool kit, but it can be done using standard programming software.

In addition to making this adjustment, it's also considered a best practice to err on the side of including fewer factors in the final factor-mimicking portfolio. This helps reduce the potential impact of multicollinearity. A factor that only marginally improves the "clonability" of a manager could thus be removed, lowering the number of factors.

It should be noted that when investors are building the factor portfolio in an attempt to replicate a manager, adding long-only equity factor strategies will reduce the weight of the broad market index. In some cases, as we discuss later, the broad market weight may be completely replaced by the long-only equity factor strategies, as market exposure is often a key driver of these strategies' month-to-month performance.

Costs count

A final important consideration with the factor-mimicking portfolio is that the comparison must take into account implementation frictions, such as expenses that are ongoing (like management fees) and ones that are transaction-related (commissions, bid/ask spreads, for example). If investors use indexes or a back-tested paper portfolio to represent all or part of the historical factor returns, they need to estimate how real-world costs would affect those results. Doing so allows for a fair comparison of results achievable in practice, since managers cannot avoid such costs.[18]

Assessing the clonability of a manager

To judge whether the manager could be substituted with one or more factor strategies using RBSA, the investor must evaluate the goodness of fit, confidence intervals for factor weights, and factor stability.

Assessing goodness of fit and confidence intervals

There are two key metrics that researchers use to evaluate goodness of fit, or how similarly a custom factor-mimicking portfolio performed versus a manager: adjusted R-squared and tracking error. Although there is no consensus on the appropriate thresholds for these metrics, our analysis shows that an adjusted R-squared above 95% and a tracking error below 3% are generally indicative of a high goodness of fit.[19]

Adjusted standard errors help the investor evaluate confidence intervals for factor weights. This is critical for assessing the statistical significance of the factor weights to improve the robustness of the results.

Evaluating factor stability

An implicit assumption when conducting standard RBSA is that there is persistence in a manager's factor exposures (Christopherson, 1995). Sharpe (1992) and Lucas and Riepe (1996), among others, rightly point out that tilts a manager exhibited over a prior period may have varied and may not represent the style profile of the manager in the future.[20] Since our RBSA method calculates the average weight of each factor over a selected time frame, it can mask time series variation.

---

18 These costs vary by strategy, asset size, and manager. Using Morningstar data, Vanguard estimates that as of December 31, 2019, the asset-weighted U.S. equity strategic beta ETF expense ratio was 15 basis points (bps) and the equal-weighted expense ratio was 33 bps. For more information on real-world investment related costs with factors, see Grim et al. (2017) and the citations therein.

19 These thresholds are also similar to those used by other industry participants (see ESMA, 2016). This is also consistent with Tidmore, Walker, and Kinniry (2019), who found that for the period from December 31, 2003, to December 31, 2018, the lowest-quintile tracking error for U.S. active equity managers versus a 9-box style category benchmark averaged less than 2.9%.

20 Some investors wonder whether managers who exhibit time-varying factor exposure can be cloned. As Hasanhodzic and Lo (2007) point out in a study on replicating hedge funds, frequent allocation changes seem to be counter to the passive spirit of the cloning endeavor.

What options does an investor have for determining whether the manager's style has been consistent in the past and whether it is likely to persist in the future?

Ask the manager
Discuss why the manager has exhibited that style profile in the past and whether or not it can be expected to persist.

Perform an "eyeball test"
Since there is no scientifically proven optimal period of time to assess a manager's style, calculating the factor weights over different periods is a prudent practice. Sharpe (1992) provides examples of this by showing area graphs with factor weights using 36-month rolling regression results.[21] Investors can then assess whether the weights appear fairly stable across different time periods. They can also decide whether there should be tolerance limits on factor weight variation over the tested period (for example +/-5%, +/10%). If the variation of a factor weight exceeds the threshold, it may not be a good candidate for cloning.
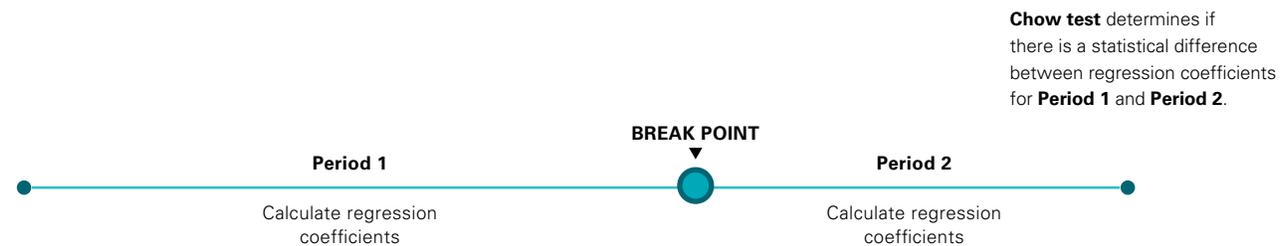
Test for constant variance
A more robust, albeit complex, way to check the stability of the manager's factor exposure is to test whether the variance of the error term remains constant over the entire sample (that is, there is no heteroskedasticity).[22] If the manager's factor exposure changes substantially over time, the regression residuals (errors) relative to the factor-mimicking portfolio should also experience substantial variability. There are several ways to test for the presence of heteroskedasticity. For our case studies, we use the White test, which establishes to what extent the regression residuals could be considered independent and identically distributed random noise.[23]

Test for a structural break
The investor can divide the return history into subsamples and test whether the subsample regression results have statistically significant differences in weights. This can be done using a Chow test or Quandt test.[24] The choice of date for splitting the sample for the Chow test is subject to professional judgment. If there was a manager change during the period, the date of that change could serve as a break. It could also make sense to choose a date when the biggest change in rolling regression output is observed.[25] For example, using the illustration in **Figure 4**, let's assume there was a change at the end of Period 1. The investor could conduct a test that compared the factor strategy weights generated from RBSA analysis of Period 1 with those generated from Period 2.[26]

Figure 4. A structural break test can help determine if a manager's factor profile has changed

Chow test determines if there is a statistical difference between regression coefficients for **Period 1** and **Period 2**.

BREAK POINT

Period 1
Calculate regression coefficients

Period 2
Calculate regression coefficients

Source: Vanguard, 2020.

---

21 There is no consensus in the academic literature on what constitutes the appropriate length of time for evaluation. Swinkels and van der Sluis (2006) note that window length in empirical work tends to range from 24 months to 60 months. We chose 36 months as it is a common length used in similar studies.

22 Heteroskedasticity occurs when the standard deviations of a predicted variable, monitored over time periods (or over different values of an independent variable), are nonconstant. Heteroskedasticity is a violation of the assumptions for the ordinary least squares linear regression analysis, and it can affect standard errors and invalidate inferences obtained from this analysis.

23 See White (1980) for more information.

24 With a Chow test, the investor must specify the historical point to split the data into subsamples before conducting the test. An alternative approach is the Quandt test, which builds on the Chow test by computing the Chow test at all possible breakpoints within the sample. The largest Chow test statistic is then chosen as the Quandt statistic to determine whether the weights (coefficients) are statistically different for the subsamples used for that particular test. The Quandt test requires advanced software and strong econometric knowledge.

25 In order for the test to be effective, both subsamples need to have a sufficient amount of return history (e.g., at least 36 months).

26 We demonstrate outputs of a modified Chow test for selected case studies in the Appendix.

Manager cloning framework: A step-by-step guide

Figure 5 summarizes the ways we have discussed for assessing the factor profile of a manager as well as a number of considerations that are important to keep in mind when implementing RBSA. Next we showcase examples using practical case studies.

Figure 5. A practical testing framework for assessing the extent to which a manager can be cloned

Step 1: Select an active manager to test
Choose an active equity fund in your portfolio or one being considered for inclusion, with a sufficient return history (preferably more than 36 months).

Step 2: Select factor candidates
Look for unique and investable factors with the most supporting evidence that they could mimic the manager's return profile.

Step 3: Select investable proxies for each factor
Choose low-cost, long-only investable proxies from an investment universe similar to the manager's. Make sure they have a sufficient return history.

Step 4: Conduct returns-based style analysis (RBSA)
Run a constrained regression of the manager's returns on the factor proxy returns, iterating through different subsets of factor proxies and limiting analysis to only a few factors.

Step 5: Assess goodness of fit, factor stability, and alpha
Test which factors and factor proxies can be used to substitute for the fund. Assess goodness of fit measures, standard errors and confidence intervals of the factor weights, factor stability, and the net alpha of the fund to determine the extent to which the fund's returns can be cloned.

Software for conducting constrained RBSA

Morningstar Direct Presentation Studio has constrained regression functionality. The program can examine the basic criteria of both tracking error and adjusted R-squared, and it allows for eyeballing factor stability. This is a reasonable option for garnering basic insight into the manager's behavior.

Python is another powerful tool for RBSA. It not only allows the user to run constrained regressions but can also provide more sophisticated statistical tests such as adjusted standard errors and structural break tests. While some code has to be written, open-source packages such as scipy.optimize.minimize make implementation feasible with some basic coding background. We used Python for the analysis in this white paper. Other software packages that could undertake this analysis include RStudio and MATLAB.

Alternatively, some consulting firms and asset managers also provide the capability to conduct constrained RBSA.

Implementation: Two hypothetical case studies

In this section, we showcase how investors can apply our practical testing framework to determine to what extent a manager can be cloned. Our hypothetical case studies attempt to explain active U.S. equity manager returns using the U.S. broad market and five factors: minimum volatility, momentum, quality, size, and value.[27] Our investable factor proxies consist of low-cost ETFs that provide exposure to the U.S. broad market and those five well-known U.S. style factors.[28] As of 2019, the net expense ratio for each of these factor strategies was 20 basis points (bps) or less per year.[29] Funds in each example are white-labeled (that is, we do not disclose the names of the active managers or factor ETFs used for the case studies). We use the gross total return time series for the managers and factor proxies.[30] Each case study uses all available data; different data availability resulted in different time frames.

Although these case studies have been undertaken for U.S. equity managers and factors, the framework can be applied globally, depending on the availability of factor products.

Case study A: Cloning a manager

In the first case study, we discuss a fund that can be cloned using our framework. Fund A is an actively managed, large-cap growth fund with an expense ratio of 69 bps. We start our analysis with the U.S. broad market, then add U.S. style factors, and then modify RBSA inputs until we get the final custom factor portfolio.

Fund A has outperformed a broad market index on a gross basis by 146 bps annually over the last 10 years. By adding factor funds to the RBSA (see **Figure 6**, Round 1 on page 11), we find that much of the historical outperformance relative to the market can be explained by the factor portfolio, with a tracking error of 3.1% and adjusted R-squared of 0.95.

Standard errors help estimate the confidence interval for the factor weights. After removing factors that do not generate statistically significant weights, we rerun the analysis with the momentum, quality, and size factors (Round 2 in Figure 6).[31] Importantly, removing the insignificant factors does not have an adverse effect on either the tracking error or adjusted R-squared of the regression. This gives us confidence that they are not important drivers of the manager's performance.

This case study demonstrates how adding long-only equity factor strategies to the portfolio can reduce the weight of the broad market index: Because Fund A's broad market exposure has been "absorbed" by other long-only factor strategies, its broad market weight becomes not statistically significant.

We can also see the decrease in the standard error of the quality factor weight that comes from removing the market, minimum volatility, and value factors. This example illustrates the importance of limiting the number of factors used for RBSA.

---

[27] There is an ongoing debate in the industry as to whether size is a true stand-alone factor or merely serves as a rough proxy for a liquidity factor. As a result, some investors may consider using liquidity factor funds instead of size factor funds.

[28] Using investable factor proxies from the same index provider or asset manager is a useful starting point but not a requirement, as the framework is able to accommodate factor proxies from various providers. See page 15 for further discussion of the impact of factor proxies on the analysis.

[29] For the purposes of our case studies, we assume that the factor-mimicking portfolios are rebalanced monthly with no transaction-related costs. Depending on the ETFs or mutual funds used and the size of the trades, the effect of the added costs could range from immaterial to significant.

[30] A benchmark or simulated factor return history adjusted for costs may be a good option when the factor fund's live return time series is not long enough for the analysis and the alternative return history is derived from a strategy that is likely to be substantially similar to the factor fund's unique style profile.

[31] Unless otherwise stated, the statistical significance threshold used in this paper is 5%. If adjusted standard errors are not available, the evaluator can experiment by removing factors with very low weights and observing how this affects the tracking error. Because multicollinearity can be an issue, we recommend trying to find a balance of parsimony (using as few factors as possible) and goodness of fit.

We assess the factor exposure stability of Fund A using these three factors and two approaches: full sample regression (test for constant variance) and rolling window regressions (eyeball test). The best way to assess factor stability using the full sample regression is to test if the variance of the error term remains constant over the entire sample. If the manager's factor exposure changes significantly over time, the regression residuals (errors)

relative to the factor-mimicking portfolio should also experience substantial variability. We run the White test for the entire sample, which results in a p-value of 0.03; the null hypothesis of constant variance is rejected at the 5% significance level, and we conclude that the error variance is not constant (that is, heteroskedasticity is present).

Figure 6. Constrained RBSA results for Fund A, 2010–2019

| | Alpha (annualized) | Factor weights | | | | | | Tracking error | Adjusted R-squared |
		Market	Minimum volatility	Momentum	Quality	Size	Value		
Round 1 (all factors)	27 bps	5%	0%	**19%**	**49%**	**27%**	0%	**3.1%**	**0.95**
Standard error	(32 bps)	(16%)	(6%)	(5%)	(10%)	(3%)	(10%)		
Round 2 (significant factors only)	20 bps			**20%**	**52%**	**28%**		**3.1%**	**0.95**
Standard error	(30 bps)			(5%)	(6%)	(3%)			
			After Round 2, assess factor stability using momentum, quality, and size factors (see Figure 7).						
Round 3 (2014–2019)	–3 bps			**29%**	**55%**	**16%**		**2.6%**	**0.95**
Standard error	(34 bps)			(6%)	(7%)	(4%)			
			After Round 3, reexamine factor stability with momentum, quality, and size factors for the period 2014–2019 only (see Figure 8).						

■ Significant at 1%   ■ Significant at 5% (none indicated)   ☐ Insignificant

Notes: Statistical significance is assessed using adjusted standard errors (Lobosco and DiBartolomeo, 1997). Standard error for the alpha term is approximate. Factor weights may not add to 100% due to rounding. For the dynamic regression results performed after Rounds 2 and 3, see Figures 7 and 8.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2010, to December 31, 2019.

We also assess factor stability by reviewing a plot of dynamic rolling window regressions, estimating the 36-month rolling factor weights that most closely mimic the manager's returns over each subperiod (eyeball test). **Figure 7** shows that while the exposures were relatively stable over time, there seemed to be a material style change where the tilt toward size was gradually replaced by momentum from 2014. We can see the impact of this on the entire sample factor weights, as they are biased by those initial few years.

Figure 7. Dynamic regression results (36-month rolling window) for Fund A, 2010–2019
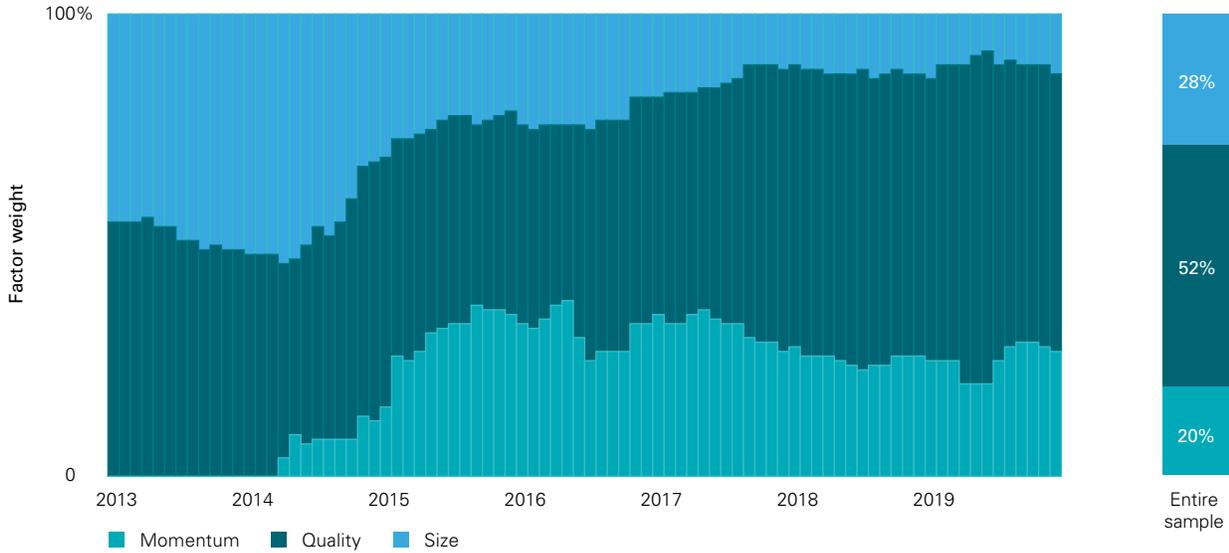
Notes: The dynamic regression is run from 2010; the chart starts in 2013 because the regression window length is 36 months. An examination of a more sophisticated stability test using Chow tests is provided in the Appendix.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2010, to December 31, 2019.

We use the most recent time frame to calculate our final weights as it provides the most accurate representation of the manager's current factor exposure. To do this, we limit the sample to returns for the years 2014 through 2019 in Round 3 of Figure 6 and once again examine stability, as shown in **Figure 8**. With exposures more stable, we end up with final factor weights of 29% momentum, 55% quality, and 16% size.

Figure 8. Dynamic regression results (36-month rolling window) for Fund A, 2014–2019
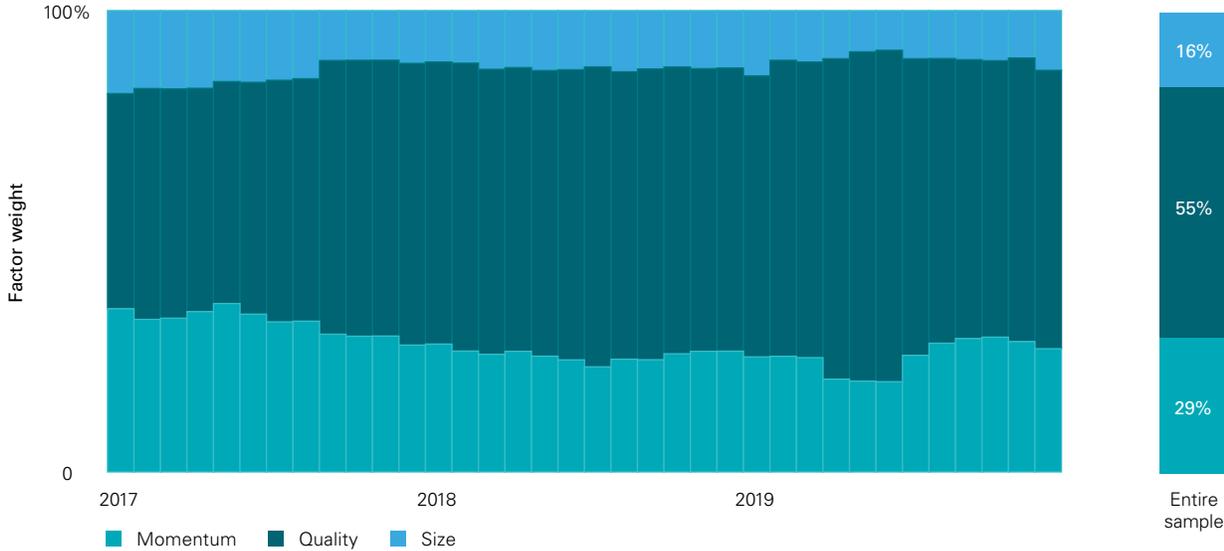
Notes: The dynamic regression is run from 2014; the chart starts at 2017 because the regression window length is 36 months. An examination of a more sophisticated stability test using Chow tests is provided in the Appendix.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2014, to December 31, 2019.

This portfolio tracks the actual manager's returns very closely at 2.7% tracking error and 0.95 adjusted R-squared. The performance difference (gross alpha) between this portfolio and the manager is not statistically significant at 10%, confirming that the majority of excess returns have been delivered through stable factor tilts. Reducing the sample to 2014–2019 resulted in more stable factor exposures as confirmed by both an eyeball test and a formal test for heteroskedasticity.[32]

We conclude that Fund A can be cloned using cheaper (approximately 16 bps versus 69 bps) and more transparent off-the-shelf factor products.[33] **Figure 9** confirms the close tracking of the returns of this final portfolio and the returns of the manager.

Figure 9. Fund A's historical returns and those of the final factor-mimicking portfolio show very similar behavior
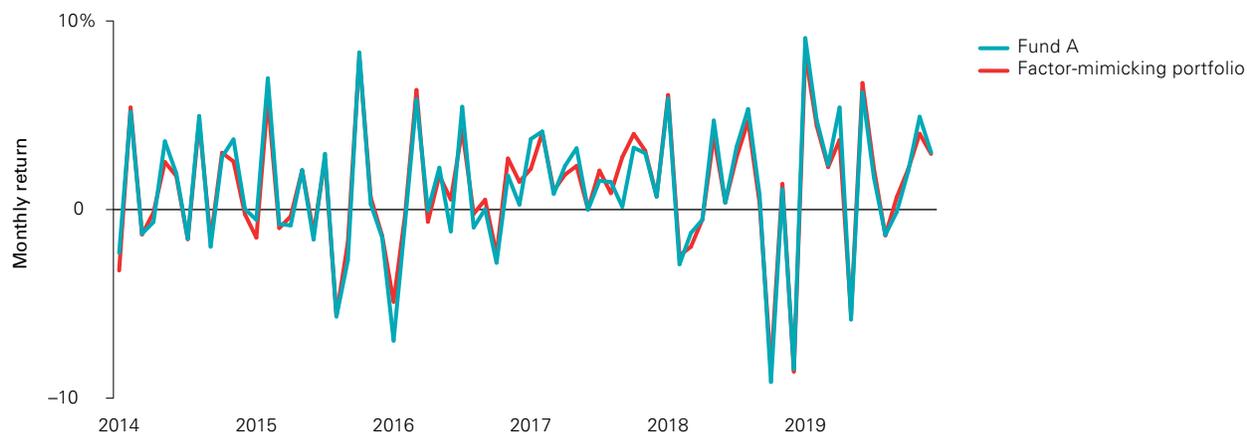
Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2014, to December, 31, 2019.

---

32 The White test for the 2014–2019 sample results in a p-value of 0.856; the null hypothesis (homoscedasticity) is not rejected. We therefore conclude that the error variance is constant (i.e., heteroskedasticity is not present).

33 The first value, 16 bps, represents the weighted expense ratio of the factor proxy products used in the final factor-mimicking portfolio.

What if we had selected different factor proxies for the analysis?

To illustrate the impact of different factor proxies on the cloning results, we apply the same approach but with a different investable strategy representing each factor. As stated previously, since factor proxies for the same factor produce a wide range of cross-sectional returns, it may be beneficial to test a few different options.

Figure 10 demonstrates the difference that proxies can make. Although there is some overlap in the significant factors, the market factor remains significant and retains its position in the portfolio, but the size factor has "disappeared." Tracking error and adjusted R-squared are similar at 3.1% and 0.95. Interestingly, alpha is not fully explained away with the new set of factor proxies, which means the manager cannot be represented as well using these alternative factor proxies for a custom factor-mimicking portfolio.

Figure 10. Regression results for Fund A using a different set of proxies, 2010–2019

| | Alpha (annualized) | Factor weights | | | | | | Tracking error | Adjusted R-squared |
		Market	Minimum volatility	Momentum	Quality	Size	Value		
Round 1 (all factors)	76 bps	28%	0%	18%	51%	4%	0%	3.1%	0.95
Standard error	*(33 bps)*	*(13%)*	*(4%)*	*(4%)*	*(10%)*	*(4%)*	*(3%)*		
Round 2 (significant factors only)	70 bps	33%		20%	48%			3.1%	0.95
Standard error	*(31 bps)*	*(11%)*		*(4%)*	*(10%)*				

After Round 2, assess factor stability with market, momentum, and quality factors (see Figure 11).

■ Significant at 1% ■ Significant at 5% □ Insignificant

Notes: Statistical significance is assessed using adjusted standard errors (Lobosco and DiBartolomeo, 1997). Standard error for the alpha term is approximate. Factor weights may not add to 100% due to rounding. For the dynamic regression results performed after Round 2, see Figure 11.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2010, to December, 31, 2019.

Figure 11 shows that the factor exposures are much more unstable over time relative to the initial set of factor proxies, resulting in a decreased confidence in the final factor weights chosen. Compared with the other set of factor proxies, factor exposures have been stable for a much shorter period of time.

The clonability of a manager as assessed by its alpha, goodness of fit, and factor stability measures can be very dependent on the factor proxies chosen. We encourage investors to try different combinations of factors and proxies for those factors until they are comfortable with the result.

Case Study B: Can every manager be cloned?

Not every manager generates returns solely by stable tilts toward factors. Managers are often trying to add value for end investors through stock selection; this is particularly true for managers who maintain relatively few holdings and tend to have high levels of idiosyncratic risk that factor strategies cannot mimic. Another source of potential manager alpha is factor timing, which results in factor exposures varying over time and can be challenging to clone.

Figure 11. Dynamic regression results (36-month rolling window) for Fund A with a different set of proxies, 2010–2019
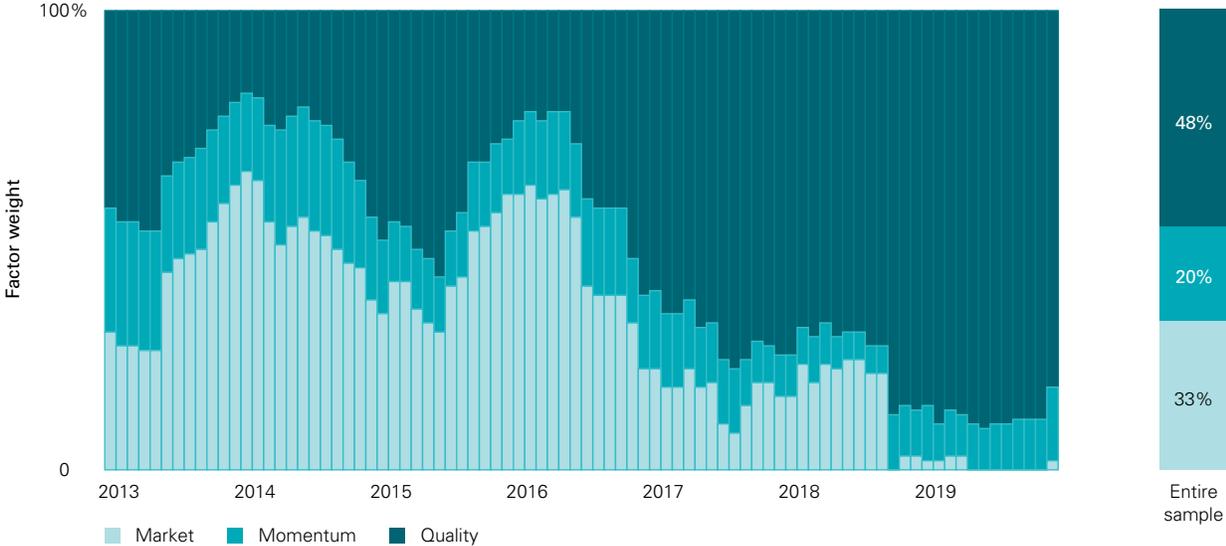
Notes: The White test for the entire sample results in a p-value of 0.75; the null hypothesis (homoscedasticity) is not rejected. We conclude that the error variance is constant (i.e., heteroskedasticity is not present). The dynamic regression is run from 2010; the chart starts in 2013 because the regression window length is 36 months.
Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2010, to December 31, 2019.

In the following case study, we discuss a fund that ends up not being clonable after applying our framework. Fund B represents a large-cap growth manager with an expense ratio of 74 bps. We start with the broad U.S. market and then add U.S. style factors, assessing alpha, goodness of fit, and the statistical significance of the coefficients (see **Figure 12**).

Fund B has outperformed the broad market on a gross basis by a staggering 462 bps per year. Its tracking error to the market is also very high at 10%, and it has a fairly low adjusted R-squared of 0.58. This means not only that

the fund has outperformed the market by a wide margin, but that its return profile is also very different.

Experimenting with adding style factor funds and assessing their significance, we remove market, momentum, quality, and size. The final custom factor-mimicking portfolio comprising minimum volatility and value has a high tracking error of 9.2% and a low adjusted R-squared of 0.65, suggesting that Fund B has not generated historical excess return through a strategic tilt to these factors. Further, the manager still outperforms this factor portfolio by a statistically significant 370 bps.

Figure 12. Regression results for Fund B, 2000–2019

| | Alpha (annualized) | Factor weights | | | | | | Tracking error (%) | Adjusted R-squared |
		Market	Minimum volatility	Momentum	Quality	Size	Value		
Round 1 (all factors)	**360 bps**	0%	**38%**	14%	1%	0%	**47%**	**9.1%**	**0.65**
Standard errors	*(64 bps)*	*(34%)*	*(13%)*	*(9%)*	*(20%)*	*(6%)*	*(21%)*		
Round 2 (significant factors only)	**370 bps**		**50%**				**50%**	**9.2%**	**0.65**
Standard errors	*(62 bps)*		*(10%)*				*(10%)*		

Assess factor stability with minimum volatility and value factors (see Figure 13).

■ Significant at 1% ■ Significant at 5% (none indicated) □ Insignificant

Notes: Statistical significance is assessed using adjusted standard errors (Lobosco and DiBartolomeo, 1997). Standard error for the alpha term is approximate. Factor weights may not add to 100% due to rounding. For the dynamic regression results performed after Round 2, see Figure 13.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2000, to December 31, 2019.

Figure 13 shows that Fund B's factor exposures were very unstable over time. It is possible that Fund B has employed factor timing and thus delivered excess returns that could not be imitated using static tilts to off-the-shelf factor products.

While it may be possible to find some other, more suitable off-the-shelf investable factor proxies to replicate the return profile of Fund B, we conclude that this fund is not clonable using the style factors and factor proxies we selected. Figure 14 also confirms that the final factor-mimicking portfolio does not clone Fund B's historical returns very well.

Figure 13. Dynamic regression results (36-month rolling window) for Fund B, 2000–2019
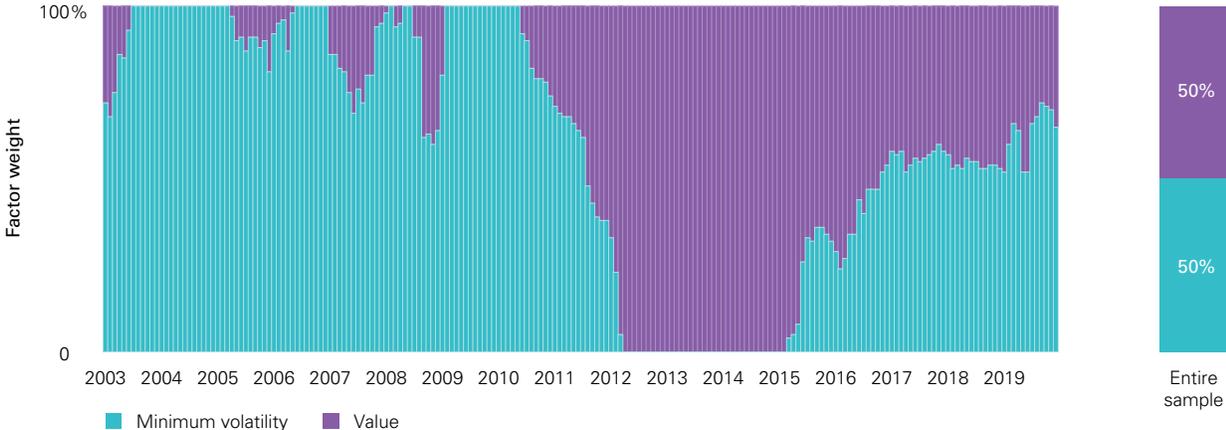
Notes: The White test for the entire sample results in a p-value of 0.00; the null hypothesis (homoscedasticity) is rejected at the 1% significance level. We conclude that the error variance is not constant (i.e., heteroskedasticity is present). The dynamic regression is run from 2000; the chart starts at 2003 because the regression window length is 36 months.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2000, to December 31, 2019.

Figure 14. Fund B's historical returns and those of the final factor-mimicking portfolio do not move in lockstep
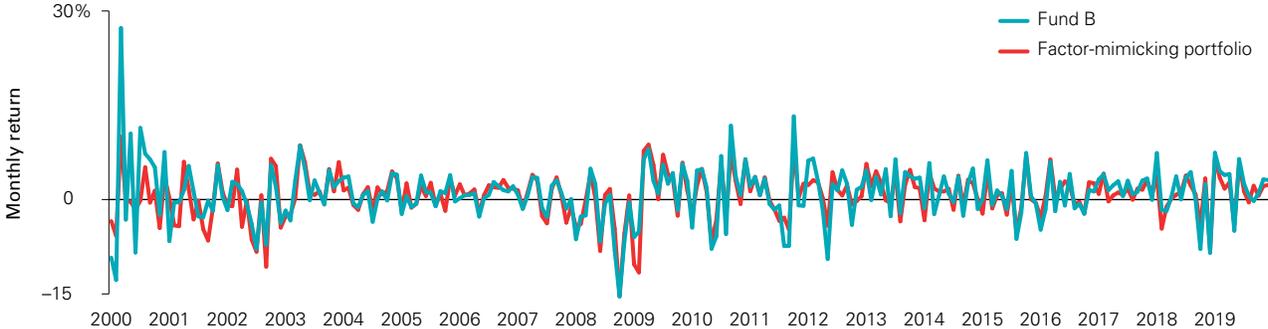
Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2000, to December 31, 2019.

Conclusion

We develop an accessible method to test to what extent a traditional active manager has been generating a return profile that could be imitated using stable tilts toward well-known factors.

This approach augments previous returns-based style analysis research. It describes both the full process needed for a robust test and the judgment needed for addressing important statistical features, to ensure that investors do not draw incorrect inferences from the results. It employs long-only investable factor strategies to create a fair alternative investment option to a long-only active equity manager or a multimanager portfolio that is subject to real-world costs. The framework builds a bridge between pure academic research on both factor investing and econometric testing and the practical implementation of the insights for investors through increasingly available off-the-shelf equity factor products.

Our approach has a number of practical applications, including enhancing the factor-oriented due diligence conducted on current and prospective managers. If the active manager is found to have generated returns that could have been more or less mimicked by factor strategies, investors may be able to produce a similar style profile with greater transparency, more risk control, and lower costs, thereby raising the bar for some active managers. This type of analysis helps investors both evaluate active managers and employ the disruptive force that factor investing is having on the active management industry.

References

Beck, Noah, Jason Hsu, Vitali Kalesnik, and Helge Kostka, 2016. Will Your Factor Deliver? An Examination of Factor Robustness and Implementation Costs. *Financial Analysts Journal* 72(5): 58–82.

Bender, Jennifer, P. Brett Hammond, and William Mok, 2014. Can Alpha Be Captured by Risk Premia? *The Journal of Portfolio Management* 40(2): 18–29.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015. Measuring Skill in the Mutual Fund Industry. *Journal of Financial Economics* 118(1): 1–20.

Brown, Melissa R., and Claudia E. Mott, 1997. Understanding the Differences and Similarities of Equity Style Indexes. In *The Handbook of Equity Style Management,* 2nd edition, pages 21–53. T. Daniel. Coggin, Frank J. Fabozzi, and Robert D. Arnott, editors. New Hope, Pa.: Frank J. Fabozzi Associates.

Buetow, Gerald W., Robert R. Johnson, and David E. Runkle, 2000. The Inconsistency of Return-Based Style Analysis. *The Journal of Portfolio Management* 26(3): 61–77.

Carhart, Mark M., 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance* 52(1): 57–82.

Christopherson, Jon A., 1995. Equity Style Classifications. *The Journal of Portfolio Management* 21(3): 32–43.

Christopherson, Jon A., and Frank C. Sabin, 1999. How Effective Is Effective Mix? *Russell Research Commentary:* 1–18.

Davis, Joseph H., Yesim Tokat, Glenn Sheay, and Nelson W. Wicas, 2008. Evaluating Small-Cap Active. *The Journal of Investing* 17(3): 64–74.

ESMA, 2016. *Supervisory Work on Potential Closet Index Tracking.* February 2; available at www.esma.europa.eu/document/public-statement-supervisory-work-potential-closet-index-tracking.

Fama, Eugene F., 1972. Components of Investment Performance. *The Journal of Finance* 27(3): 551–567.

Fama, Eugene F., and Kenneth R. French, 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33(1): 3–56.

Fama, Eugene F., and Kenneth R. French, 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116(1): 1–22.

Fama, Eugene F., and Kenneth R. French, 2018. Choosing Factors. *Journal of Financial Economics* 128(2): 234–252.

Frazzini, Andrea, David Kabiller, and Lasse Heje Pedersen, 2018. Buffett's Alpha. *Financial Analysts Journal* 74(4): 35–55.

Grim, Douglas M., Scott N. Pappas, Ravi G. Tolani, and Savas Kesidis, 2017. *Equity Factor-Based Investing: A Practitioner's Guide.* Valley Forge, Pa.: The Vanguard Group.

Hasanhodzic, Jasmina, and Andrew W. Lo, 2007. Can Hedge-Fund Returns Be Replicated?: The Linear Case. *Journal of Investment Management* 5(2): 5–45.

Hou, Kewei, Haitao Mo, Chen Xue, Lu Zhang, 2018. Q5. NBER Working Paper No. 24709. Cambridge, Mass.: National Bureau of Economic Research.

Hsu, Jason, Vitali Kalesnik, and Engin Kose, 2019. What Is Quality? *Financial Analysts Journal* 75(2): 44–61.

Israel, Ronen, and Adrienne Ross, 2017. Measuring Factor Exposures: Uses and Abuses. *The Journal of Alternative Investments* 20(1): 10–25.

Jensen, Michael C., 1968. The Performance of Mutual Funds in the Period 1945–1964. *The Journal of Finance* 23(2): 389–416.

Kahn, Ronald N., and Michael Lemmon, 2016. The Asset Manager's Dilemma: How Smart Beta Is Disrupting the Investment Management Industry. *Financial Analysts Journal* 72(1): 15–20.

Lobosco, Angelo, and Dan DiBartolomeo, 1997. Approximating the Confidence Intervals for Sharpe Style Weights. *Financial Analysts Journal* 53(4): 80–85.

Lucas, Lori, and Mark W. Riepe, 1996. *The Role of Returns-Based Style Analysis: Understanding, Implementing, and Interpreting the Technique.* Chicago: Ibbotson Associates, Inc.

Madhavan, Ananth, Aleksander Sobczyk, and Andrew Ang, 2018. What's in Your Benchmark? A Factor Analysis of Major Market Indexes. *The Journal of Portfolio Management* 44(4): 46–59.

Sharpe, William F., 1988. Determining a Fund's Effective Asset Mix. *Investment Management Review* 2(6): 59–69.

Sharpe, William F., 1992. Asset Allocation: Management Style and Performance Measurement. *The Journal of Portfolio Management* 18(2): 7–19.

Stambaugh, Robert F., and Yu Yuan, 2017. Mispricing Factors. *The Review of Financial Studies* 30(4): 1270–1315.

Swinkels, Laurens, and Pieter J. van der Sluis, 2006. Return-Based Style Analysis With Time-Varying Exposures. *European Journal of Finance* 12(6–7): 529–552.

Tidmore, Chris, David J. Walker, and Francis M. Kinniry Jr., 2019. *The Urban Legends of Active Share.* Valley Forge, Pa.: The Vanguard Group.

White, Halbert, 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica: Journal of the Econometric Society* 48(4): 817–838.

Appendix

Statistical structural break testing for Fund A

To supplement the visual analysis presented in Figures 7 and 8, we employ the Chow test, a statistical method of detecting structural breaks. We do this in two ways (see **Figure A-1**):

1. We run traditional Chow tests (henceforth, static break tests) on all possible break dates where the minimum length of a period is 36 months. Period 1 extends from the beginning of the sample until the break date and Period 2 extends from the break date until the end of the sample.

2. We also run Chow tests on all possible break dates, where the period length is always 36 months (henceforth, rolling break tests). This could also be described as a "rolling Chow test," where Period 1 always starts 36 months before the break date and

Period 2 always ends 36 months after the break date. Performing the test helps account for multiple breaks in the sample not being detected in a traditional Chow test.

While the static break test does not detect any breaks, the rolling break test detects some at multiple points—but not the seemingly obvious one around 2014. These break tests can be useful to tease out the data; however, any obvious breaks will also show up in the eyeball test of the dynamic rolling regression.

Statistical structural break testing for Fund B

Both the static and rolling break tests detect breaks. The rolling test (**Figure A-2** on page 23) seems to work better at detecting short-term shifts for unstable factor exposures, while the static test (**Figure A-3** on page 23) indicates that a big structural break took place between 2012 and 2017.
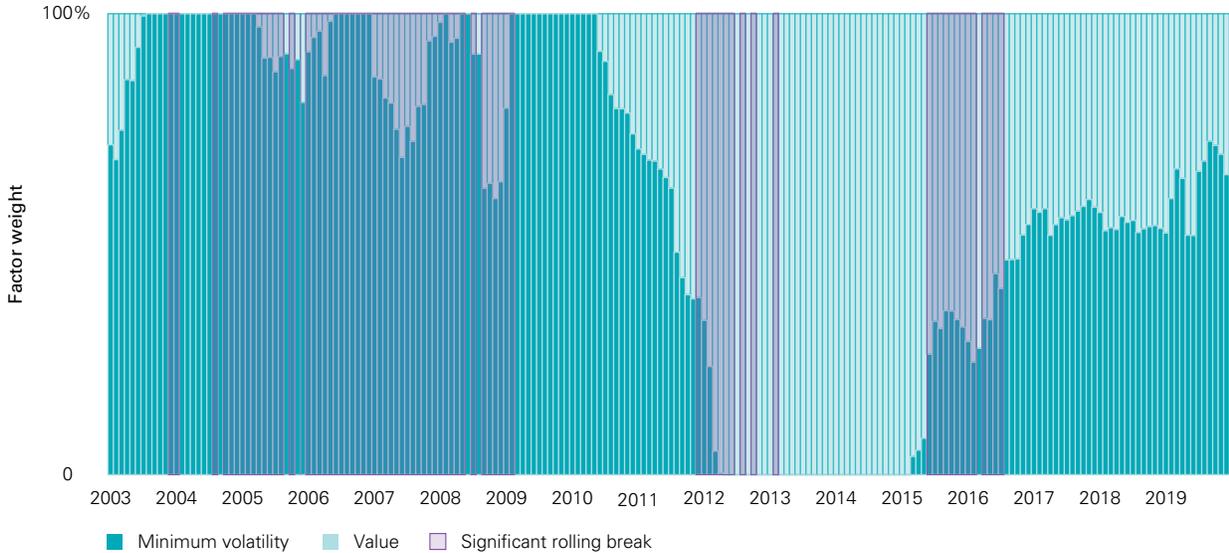
Figure A-1. Rolling and static break tests on Fund A with dynamic rolling regression as backdrop, 2010–2019
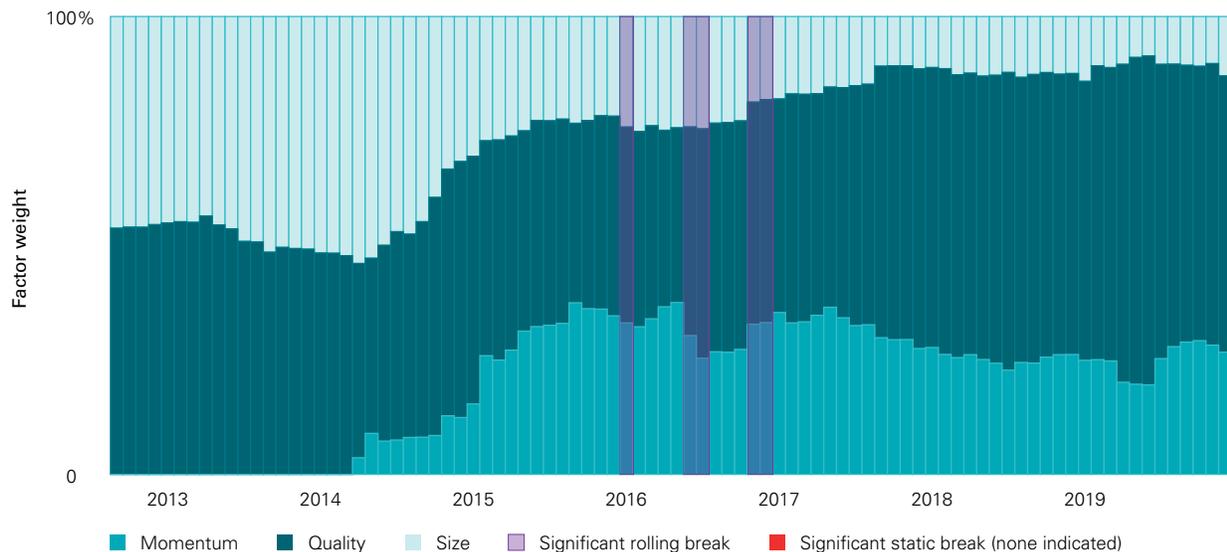
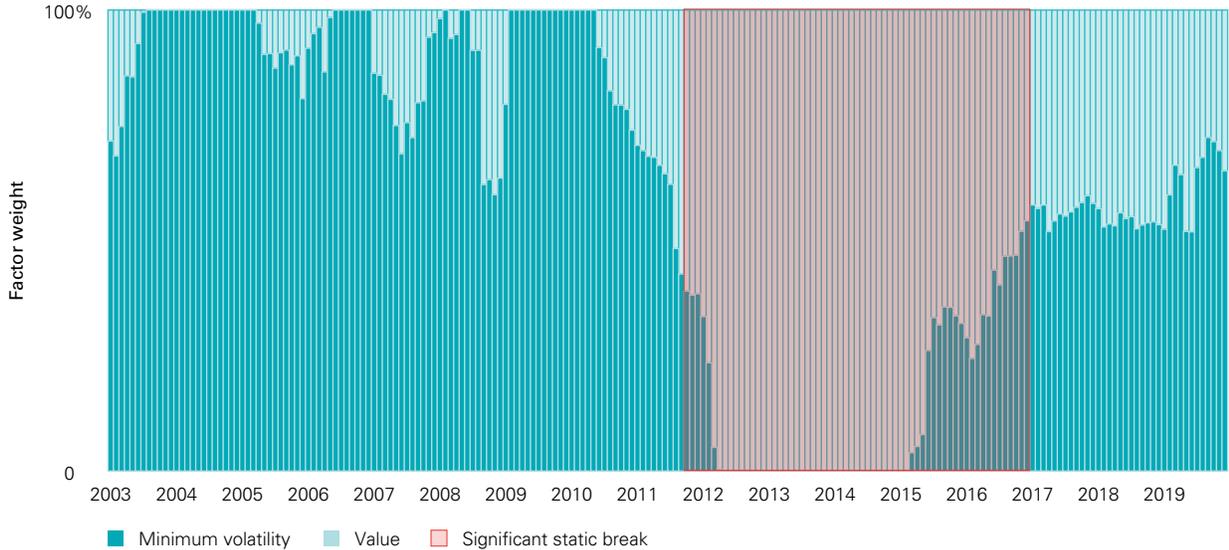Note: Vertical lines indicate structural breaks detected at a 1% significance level.

Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2010, to December 31, 2019.

Figure A-2. Rolling break tests on Fund B with dynamic rolling regression as backdrop, 2010–2019

Note: Vertical lines indicate structural breaks detected at a 1% significance level.
Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2000, to December 31, 2019.

Figure A-3. Static break tests on Fund B with dynamic rolling regression as backdrop, 2010–2019

Note: Vertical lines indicate structural breaks detected at a 1% significance level.
Sources: Vanguard calculations, using monthly gross return data from Morningstar from January 1, 2000, to December 31, 2019.

Connect with Vanguard® > vanguard.com

All investing is subject to risk, including the possible loss of the money you invest.
Past performance is not a guarantee of future returns. Be aware that fluctuations in
the financial markets and other factors may cause declines in the value of your account.
There is no guarantee that any particular asset allocation or mix of funds will meet
your investment objectives or provide you with a given level of income.

CFA® is a registered trademark owned by CFA Institute.

Vanguard®