THE JOURNAL OF
INVESTMENT CONSULTING

A reprinted article from Volume 20, Number 1, 2020

Nonlinear Factor Attribution

By Sanne de Boer, PhD, CFA®

Nonlinear Factor Attribution

Sanne de Boer, PhD, CFA®

ABSTRACT

Factor attribution based on linear regression often fails to satisfactorily explain the performance of systematic investment strategies. Sizeable attribution residuals that do not average out to zero over time suggest latent exposures to nonlinearities in factor returns. Our proposed adjustment takes a portfolio manager's perspective in attributing the impact thereof, identifying which factor tilts were most responsible for the unexplained performance. The resulting nonlinear attribution better reconciles realized returns with the investment process and is testable for statistical significance, with R code provided for evaluation purposes. We illustrate how this deeper understanding of factor interactions may guide client discussions and point to strategy enhancements.

INTRODUCTION

Factor attribution helps investment managers and their clients understand the drivers of a portfolio's total or benchmark-relative return. The portfolio's realized performance is decomposed into contributions from its factor, geographic, and industry exposures. Any residual return left unexplained is commonly attributed to stock-specific risk. Fischer and Wermers (2013) include a comprehensive review of the standard methodology.

For quantitative investment managers, finding the residual dominating performance attribution poses a problem in strategy reviews with clients; meaningful levels of stock-specific return are inconsistent with a diversified systematic strategy. Anecdotally, this appears to be a common issue.[1] Although the positioning of sizeable residuals in short-term factor attribution constitutes a communication challenge, their average not converging to zero over time signifies an issue with more serious investment ramifications: misspecification of the return model underlying the strategy in a way that impacts long-term performance. To shed light on the impact thereof, we develop an extension of standard regression-based factor attribution that better reconciles realized performance with the investment process. Taking a portfolio manager's perspective rather than an econometrician's, our proposal classifies stocks based on their squared standardized factor exposures, identifies which factor tilts were most responsible for the unexplained portfolio return,

Our proposed adjustment takes a portfolio manager's perspective in attributing the impact thereof, identifying which factor tilts were most responsible for the unexplained performance.

and adjusts the attribution accordingly. The nonlinear adjustments of the estimated factor contributions can be corroborated through statistical testing as unlikely to have resulted from truly stock-specific risk. The resultant deeper understanding of performance drivers may help guide client discussions and suggest enhancements to the investment process.

Factor attribution issues are part of the larger class of "factor alignment problems," which arise due to the interaction between alpha forecasts, the factor risk model, and constraints in mean-variance optimization. A recap of the broader literature is outside our scope, but relevant work includes Lee and Stefek (2008), Ceria et al. (2012), and Saxena et al. (2013). As a simple step toward mitigating the problem of attribution residuals, Vandenbussche (2016) recommends aligning the estimation universe of factor returns with the strategy's investment opportunity set. Menchero and Poduri (2008) and Sivaramakrishnan and Stubbs (2013) stress the importance of using consistent sets of factors in portfolio construction and performance attribution. Grinold (2006, 2011) maps strategy holdings to a set of explanatory factor portfolios. Recent ideas to capture nonlinear interaction effects include asymmetric factor return estimates (Vandenbussche 2016), additional attribution factors (De Boer and Jeet 2016), and a second-stage time series regression of the attribution residuals on the estimated factor contributions to adjust the portfolio's factor loadings (Stubbs and Jeet 2016). All of these proposals have merits, but each might fail to mitigate the attribution residual and potentially could worsen it. Our proposal aims to be a robust alternative in the practitioner's toolkit to better reconcile realized returns with the investment process.

A different body of literature focuses on the impact of investment constraints. Clarke et al. (2002, 2005) expand Grinold's (1989) fundamental law on the value-add of active management with a transfer coefficient that approximates the resulting performance degradation. It distinguishes between two causes thereof: the muting of factor tilts and the increase in stock-specific risk. Stubbs and Vandenbussche (2010) isolate the impact of individual constraints on expected and realized performance. In contrast, our proposal aims to identify factor contributions that deviated meaningfully from their unconstrained estimates, and to revise the attribution accordingly.

A REVIEW OF FACTOR ATTRIBUTION

We first briefly review the mathematical foundation of factor attribution. We assume stock returns follow a linear factor model (e.g., Grinold and Kahn 2000). Let $\boldsymbol{r}$ denote a vector of all n stock returns over the universe and period of interest. Let $\boldsymbol{B}$ be an n by m matrix of n stock exposures to m factors. We focus on fundamental factor models for which the stocks' factor exposures are pre-specified and known, typically based on valuation ratios, price momentum, and other characteristics that have been found to explain return dispersion. The factor returns that drive co-movements in share prices are random variables denoted by the m-vector $\boldsymbol{r_b}$. Lastly, let $\boldsymbol{\varepsilon}$ be an n-vector of zero-mean stock-specific returns, assumed Gaussian with diagonal covariance matrix $\boldsymbol{\Omega}$. Then the uncertainty around stock returns is described by:

$$\boldsymbol{r} = \boldsymbol{B}\boldsymbol{r_b} + \boldsymbol{\varepsilon} \tag{1}$$

For portfolios, we use $\boldsymbol{w}$ to denote the vector of beginning-of-period weights for all n stocks in the investable universe. We use $\boldsymbol{b}$ to indicate the beginning-of-period factor exposures of the portfolio ($\boldsymbol{b} = \boldsymbol{B'w}$). For benchmarked long-only portfolios with a tracking error constraint, $\boldsymbol{w}$ and $\boldsymbol{b}$ are understood to represent active portfolio weights and factor exposures, respectively.

Factor attribution aims to explain a strategy's performance for a given period in the context of the return model specified by equation (1). A portfolio's realized return is decomposed into the contributions of the individual model factors as measured by factor exposure times estimated factor return, and the residual left unexplained.[2] Realized factor returns are unobservable and thus estimated with error, typically using weighted least squares of the realized stock returns over the period on the factor exposures.[3] Such estimates are known to have an interpretation as the returns to efficient "factor-mimicking portfolios," dollar-neutral sleeves $\boldsymbol{w}_k^{wls}$ with unit exposure to one factor, zero exposure to all others, and minimum specific risk (see the appendix for details). It follows that regression-based factor attribution reflects the following decomposition of the portfolio weights:

$$\boldsymbol{w} = \sum\nolimits_{k=1}^{m} b_k \boldsymbol{w}_k^{wls} + \boldsymbol{w}_{resid} \tag{2}$$

The estimated factor-driven return component equals the return to a reference portfolio that emulates its factor exposures with minimum specific risk. The residual portfolio $\boldsymbol{w}_{resid}$ captures deviations from this optimal reference portfolio necessitated by investment constraints. Intuitively, position limits and transaction cost penalties in portfolio construction cause the residual portfolio to be long the specific risk of unattractive stocks that the investor would have liked to further underweight and short attractive stocks the investor would have liked to hold more of.

The return of the residual portfolio is uncorrelated with the return to the factor-mimicking portfolios. Because their combination decomposes the strategy's total risk, the volatility of the attribution residual reflects the impact of constraint-induced implementation inefficiency.[4] Clarke et al. (2002) refer to this residual portfolio as "weights not taken" and to its return as "constrained-induced noise." As a simple step toward mitigating the noisiness of the attribution residual, the investor may thus want to consider relaxing any self-imposed position limits and other constraints, to the extent possible. To identify the contributions of individual stocks to the attribution residual, we note it can be minimally decomposed into the residual returns $\hat{\boldsymbol{\varepsilon}}$ multiplied by the residual portfolio weights $\boldsymbol{w}_{resid}$. In what follows, we use this information to explore any nonlinear relation between the attribution residual and factor exposures.

PUTTING THE RESIDUAL TO THE TEST

Some short-term residual volatility is to be expected for a constrained systematic investor, but the long-term impact of stock-specific risk on the attribution should be relatively inconsequential under our assumed return model. When the attribution residuals are well outside the expected range or do not average out to zero over time, they can no longer be dismissed as noise. As pointed out by Clarke et al. (2002), this is an indication that the attribution model is misspecified and the residual is a proxy for a missing return factor.

To assess whether even a sizeable residual conceivably reflects a confluence of stock-specific returns, we propose to include the slightly transformed residual weights as an added attribution factor:

$$\boldsymbol{B}_{resid} = \frac{\boldsymbol{\Omega}\boldsymbol{w}_{resid}}{(\boldsymbol{w}_{resid}{}'\boldsymbol{\Omega}\boldsymbol{w}_{resid})} \tag{3}$$

It is easy to show that the portfolio has unit active exposure to $\boldsymbol{B}_{resid}$ while its estimated coefficient from weighted least squares equals the attribution residual. To our knowledge, the ability to statistically test the attribution residual through casting it as a regression variable is a new contribution to the literature.

The t-statistic of the attribution residual puts it in perspective relative to the degree of stock-specific risk during the performance window. If the contribution of the added factor is statistically significant, it likely reflects some systemic return effect that the attribution should account for, rather than idiosyncratic noise. We will propose a solution for examining this in the next section. If the threshold for statistical significance is not met yet the stock-specific return contribution was nevertheless meaningful, this suggests portfolio construction would benefit from better diversification. Both are actionable insights for factor-based investment strategies.

In practice, factor attribution will span different time windows, linking bursts of trading activity that changed the portfolio's holdings and factor exposures. The periodical factor returns can be estimated through panel regression, with residual-generating factors $\boldsymbol{B}^{t}_{resid}$ defined for each time window.[5] Rejection of the joint null hypothesis of all periodical attribution residuals equaling zero indicates a missing risk factor affecting strategy returns, and statistical significance of their aggregated estimate suggests its impact was priced.[6] This provides easy-to-implement specification tests of the linear factor model of equation (1) and its suitability for attributing the returns of a given strategy.

APPLYING A NONLINEAR FACTOR LENS

The expanded attribution that includes the additional explanatory factor specified in equation (3) allows evaluating the magnitude of the original residual, but it sheds no light on the causes thereof. De Boer and Jeet (2016) illustrate how investment constraints may result in latent exposures to possible nonlinearity in factor returns. For instance, a long-only trend-follower may not benefit from the sustained collapse of a narrow segment of an otherwise uneventful market, while a turnover-constrained bargain-hunter might load heavily on perennially cheap value traps. However, meaningfully reducing the attribution residual by identifying and including these missing factors is challenging, as illustrated in the empirical results below.

In this paper, we develop a "nonlinear factor lens" that provides additional insight without directly addressing the misspecification of the return model. De Boer and Jeet (2016) show that restricting factor return estimates from linear regression to exactly explain a portfolio's performance redistributes the original attribution residual roughly in proportion to its squared standardized factor exposures. Building on their top-down portfolio-level approach, we propose to apply this allocation rule to the residual return contributions of individual stocks so as to identify from the bottom-up which segments of the investable universe contributed most to the original attribution residual. We first select a set of $p \geq 2$ factors to explore attributing unexplained performance to. When a stock has a high absolute exposure to any of these factors, we allocate most of its residual return contribution thereto. To implement this systematically, we define an n by p classification matrix as:

$$\boldsymbol{B}^{i,k}_{classif} = \frac{\boldsymbol{B}^2_{i,k}}{\Sigma^{p}_{k=1}\boldsymbol{B}^2_{i,k}} \ (i = 1, \dots, n;\ k = 1, \dots, p) \qquad (4)$$

This matrix has rows summing to one, reflecting the relative importance of each included factor k for each stock i.[7] Recall from our review of factor attribution that the attribution residual decomposes at the individual stock level into weight deviations $\boldsymbol{w}_{resid}$ from the optimal unconstrained portfolio times the estimated stock-specific returns $\hat{\boldsymbol{\varepsilon}}$. Classifying all stocks' contributions $\hat{\varepsilon}_i w^i_{resid}$ to the attribution residual by equation (4) then implies the following nonlinear adjustment $\boldsymbol{\Delta}_{nl}$ to the estimated linear contributions of the factors included in $\boldsymbol{B}_{classif}$:

$$\boldsymbol{\Delta}_{nl} = \hat{\boldsymbol{\varepsilon}}' diag(\boldsymbol{w}_{resid})\boldsymbol{B}_{classif} \qquad (5)$$

We caution that under the assumptions of the return model specified in equation (1), $\boldsymbol{\Delta}_{nl}$ merely reshuffles the idiosyncratic "noise" that constitutes the attribution residual roughly equally among all factors in the classification scheme. The adjustment only evidences return nonlinearities when one or more factors explain a disproportionate share of the original attribution residual to an extent unlikely to have resulted from truly stock-specific risk. The appendix shows how to statistically test the significance of the nonlinear reallocation based on this logic, both jointly (F-test) and individually (t-values). R code that implements this logic is available for research and evaluation purposes at https://github.com/sdb-research/nonlinear-factor-attribution.

It is important to note that this attribution approach aims to identify patterns in the cross-section of stock returns and is nonlinear in that it may identify contributions from factors to which the portfolio has no linear exposure, unlike the top-down implementation by De Boer and Jeet (2016) that inspired it. We also note that when two selected interaction factors in equation (4) are strongly correlated, either positively or negatively, the proposed nonlinear approach will allocate stocks' residual return contributions about equally between them because the impacts cannot be properly differentiated. In contrast, any such multi-collinearity for the linear attribution may result in highly unintuitive and noisy factor return estimates in an attempt to precisely isolate the impact of each of them.

We now provide an example to illustrate the investment intuition behind the dry symbology of equation (5). Consider performance attribution for a long-only multi-factor portfolio that tilts to value, quality, and momentum. Suppose that the value and quality exposures contributed negligibly to performance, and the returns to the momentum factor were particularly strong in its tails; that is, the very strongest past winners performed better than a linear return to the momentum factor would estimate (positive residuals), and the very worst past losers did even worse (negative residuals). In addition, suppose that liquidity-based position limits prevented the portfolio from fully emulating the large active bets on the strongest past winners in the momentum-mimicking portfolio (negative constraint-induced deviations). Similarly, the long-only constraint prevented the portfolio from fully emulating the large active bets against the worst past losers therein (positive deviations). In this case, standard regression-based factor attribution would show a negative residual because the portfolio's return fell short of the unconstrained potential of its momentum tilt. Extreme winners and losers both contributed to the negative attribution residual and all would be predominantly identified as momentum bets under the proposed classification scheme in equation (4). As a result, the proposed adjustment, equation (5), re-applies most of the negative original attribution residual to give a haircut to the over-extrapolated momentum contribution. The nonlinear attribution better matches the portfolio's realized return while still identifying the primary driver thereof.

To illustrate the calculations involved, suppose the strongest momentum stock in the example above was held at 1 percent less than optimal due to a liquidity limit. This would result in a weight of −1 percent in the residual portfolio $\boldsymbol{w}_{resid}$ from holdings decomposition, equation (2). If its return over the attribution window was 1,000 basis points (bps) higher than its estimated factor-driven return component from the return model of equation (1), this would be reflected by a residual return $\hat{\varepsilon}$ of +10 percent from the regression estimation. Its contribution to the attribution residual would then equal −1% × 10% = −10 bps. Assuming standardized exposures of 3 to momentum, 1 to value, and 0 to quality, equations (4) and (5) would result in a 9-bps haircut to the linear estimate of momentum's contribution to the portfolio's performance, and 1-bps haircut to the estimated contribution from value. With similarly negative adjustments from the other extreme momentum bets, both long and short, the brunt of the original attribution residual would be reallocated to bring down the estimated contribution from the momentum exposure. In contrast, had the regression residuals reflected true idiosyncratic risk rather than return nonlinearity at the tails of the momentum factor, the attribution residual would have been reallocated about equally across the three factors.

We note that the philosophy behind this approach is different from the adjustment regression framework proposed by Stubbs and Jeet (2016). Both approaches build on standard attribution and make data-driven adjustments to the estimated factor contributions without directly addressing the misspecification of the underlying return model. However, Stubbs and Jeet (2016) propose to adjust a portfolio's factor exposures when a time-series regression identifies a statistically significant relation between the periodical attribution residuals and estimated factor contributions. In contrast, we aim to identify segments of the investable universe that contributed disproportionately to the attribution residual even for a single period. The attribution is adjusted accordingly if evidence suggests it is unlikely to have resulted from true stock-specific noise. This may identify factors that failed to deliver for the portfolio and on average fell short from their estimated contribution, which may not always show up in any time-series correlation.

IMPLEMENTATION CONSIDERATIONS

The proposed nonlinear attribution is intended to facilitate periodical performance reviews for internal and client reporting purposes. The resulting adjustment can be reported separately rather than replace the existing linear attribution template to maintain full visibility on the analysis results. This may help craft a more nuanced narrative around performance, as illustrated in the next section. We caution that in our experience it is rare to reach statistical significance for shorter-term attributions, but confidence levels and t-values can be an integral part of this supplementary report.

Although theoretically the nonlinear part of the attribution might include all attribution factors ($p=m$), we can choose a subset thereof to focus on those deemed most relevant for the underlying strategy. We also recommend including an intercept in equation (4). This implies that stocks with relatively small factor exposures will be predominantly left unclassified. The residual return allocated to this category can then be interpreted as the new constraint-induced attribution residual that cannot be reasonably explained by return nonlinearities in any of the included factors. Although the resultant factor attribution will no longer fully add up, the residual should more closely resemble stock-specific risk. In our experience, the default intercept level of one is a reasonable choice relative to standardized factor exposures, but it does not necessarily result in the nonlinear attribution that best fits the data. In general, the higher the intercept, less of the residual gets reallocated but more is driven by the tails of the factor exposures range. Evaluating the statistical significance of the adjusted attribution for different levels of the intercept adds further insight into the nature of any nonlinearity in factor returns.

The proposed approach also can be used for research purposes to explore possible enhancements to the investment process. Including only the main alpha factors such as value, momentum, and quality would result in systematically adjusting the reported contribution of any of the portfolio's targeted factor exposures with performance that deviated from its unconstrained estimate. If investment constraints consistently impede the ability to earn a certain targeted factor premium, return forecasts used by the strategy should be adjusted accordingly. Alternatively, including portfolio-construction related factors such as size, liquidity, and risk might identify which thereof are most responsible for the portfolio's shortfall or pick-up in factor returns. For instance, De Boer et al. (2018), in a nontechnical illustration of the proposed methodology, highlight a strategy that targets the most volatile stocks in portfolio construction, for which the nonlinear attribution confirms that factor performance was more pronounced. Lastly, including industry dummies in the classification scheme specified in equation (4) could identify parts of the investable universe where factor performance was differentiated, potentially meriting the development of sector-specific stock selection

Table 1

AVERAGE FACTOR EXPOSURES FOR HYPOTHETICAL SIMULATED VALUE/MOMENTUM PORTFOLIO RELATIVE TO CAPITALIZATION-WEIGHTED BENCHMARK

Monthly average active exposure over attribution window; all factor exposures cross-sectionally z-scored

	Linear			Targeted Interactions		
	Value	Momentum	Size	Abs (Value)	Abs (Momentum)	Avg {Size × M, Size × V}
U.S.: Jan '15–Dec '17	0.76	0.56	-0.07	0.07	0.09	0.12
EUR: Jan '15 Dec '17	0.77	0.36	-0.07	0.04	0.11	0.25
U.S.: Jan '14–Dec '15	0.94	0.50	-0.10	0.04	0.03	0.04
EUR: Jan '14–Dec '15	0.82	0.47	-0.08	0.10	0.12	0.33

Table 2

STANDARD FACTOR ATTRIBUTION FOR HYPOTHETICAL SIMULATED VALUE/MOMENTUM PORTFOLIO

In percent; annualized from monthly averages ignoring compounding; attribution per Axioma US4 and WW4 models; "Other Style" includes earnings and dividend yield, growth, volatility, leverage, profitability, foreign-exchange sensitivity, market sensitivity, and a mid-cap indicator (US4 only)

		Attribution						
	Active Return	Value	Momentum	Size	Other Style	Industry	Country	FX
U.S.: Jan '15–Dec '17	0.53	-1.05	1.66	-0.37	0.49	-1.45		
EUR: Jan '15–Dec '17	1.20	-2.12	0.98	0.25	-0.31	0.70	0.33	-0.16
U.S.: Jan '14–Dec '15	-1.80	-0.18	1.27	-0.16	-0.96	-3.20		
EUR: Jan '14–Dec '15	-2.79	0.11	-0.06	0.74	-4.62	0.87	0.23	-1.15

models. Each of these "nonlinear lenses" may lead to actionable recommendations for the strategy under examination.

EMPIRICAL ILLUSTRATION

We now provide a simple empirical example of the attribution method developed above. We emphasize the goal is not to assess the efficacy of factor models in explaining stock returns or advocate any particular investment strategy but solely to illustrate the proposed toolkit that allows such exploration. We calculate a composite score of each stock's standardized "value" (book to market) and "momentum" (trailing twelve months return) exposure as of the beginning of each month between January 2015 and December 2017. We then create a hypothetical portfolio that invests, on a capitalization-weighted basis, in the highest-scoring quintile among the largest 1,000 U.S. stocks at that point in time. This is a very basic example of factor investing that ignores liquidity and trading cost considerations. We consider this same strategy on the point-in-time constituents of the MSCI Europe Index, with returns denominated in euros (EUR). For both geographies, we include a shorter-term attribution for just the preceding year.

We included value, momentum, and size in the nonlinear residual reallocation approach, equation (4), as well as a unit intercept. For comparison with the proposal by De Boer and Jeet (2016) to directly target potential nonlinear interaction effects, we also report results with the z-scored absolute value of the value and momentum exposures as well as their respective cross-product with a standardized "size" (log of market capitalization) exposure included as additional attribution factors. The former aims to capture the impact of the long-only nature of the portfolio in case of asymmetric factor returns. The latter reflects any differentiated factor performance among large-cap stocks where its active positions are concentrated.

All data are per Axioma's US4 risk model (version WW4 for the European strategy), which also is used for the factor attribution.[8] Because any model of stock returns is at best a useful tool, rather than a fully accurate characterization thereof, all reported p-values should be taken as approximations used to corroborate attributions and guide discussions.[9]

We now illustrate how standard attribution supplemented with the proposed nonlinear factor lens may highlight the portfolio's positioning and identify drivers of its performance. Table 1 shows that the hypothetical simulated portfolios were exposed to both of their intended return drivers, albeit most so to value, with a slight small-cap bias. These imbalances reflect the mega-cap growth tilt inherent to capitalization-weighted benchmarks. The slightly positive exposure to the absolute value of value and momentum reflect that it is indeed more difficult to position against the short side of these factors for a benchmark-relative long-only portfolio. Lastly, the moderately positive average exposure to the size interactions confirms that the value and momentum tilts are concentrated among large-cap stocks.

Table 2 shows active performance and the standard regression-based attribution thereof. The hypothetical simulated value/momentum portfolio eked out a modest gain in both the United States and Europe over the three-year window, with the positive return to momentum offsetting all or at least part of the disappointing return to value. Asynchronous factor performance is an important rationale for multi-factor strategies, and the success thereof in this case is worth highlighting to clients.

Table 3

DISTRIBUTION OF UNEXPLAINED RESIDUAL FOR HYPOTHETICAL SIMULATED VALUE/MOMENTUM PORTFOLIO BY ATTRIBUTION METHOD

In percent; annualized mean and standard deviation from monthly attributions ignoring compounding

	Mean			Standard Deviation		
	Standard	With targeted interactions	With residual reallocation	Standard	With targeted interactions	With residual reallocation
U.S.: Jan '15–Dec '17	1.26	0.85	0.19	1.10	1.02	0.28
EUR: Jan '15–Dec '17	1.53	1.29	0.30	1.18	0.93	0.30
U.S.: Jan '14–Dec '15	1.43	-0.63	0.22	1.29	1.01	0.32
EUR: Jan '14–Dec '15	1.08	1.05	0.45	0.89	0.97	0.30

Table 4

TESTING FOR NONLINEARITIES IN FACTOR RETURNS, BY UNDERLYING ATTRIBUTION METHOD

F-test results (*p*-value) of stated null hypothesis for attribution-based test metric; cases for which evidence for nonlinearity is statistically significant at 90-percent confidence are marked with '*'

Test metric from attribution	Aggregate residual of standard attribution	Targeted interaction factors	Aggregate residual reallocation
Null hypothesis ("H0")	Stock-specific noise	Zero aggregate return contributions	Proportional to average classification
# estimates jointly tested	1	4	3
U.S.: Jan '15–Dec '17	6.10% *	10.85%	24.96%
EUR: Jan '15–Dec '17	2.10% *	7.99% *	9.92% *
U.S.: Jan '14–Dec '15	15.68%	0.26% *	7.89% *
EUR: Jan '14–Dec '15	32.37%	44.31%	90.79%

We note that for the one-year window, "Industry" and "Other Style" effects were the predominant drivers of the hypothetical portfolio's underperformance. We have deliberately not explored this in more detail because it is not relevant for purposes of our illustration. For actual client discussions, this would raise questions about unintended risk exposures of the strategy that would need to be addressed.[10]

Table 3 compares the distribution of the unexplained residuals. For the three-year attributions, substantially all of the outperformance reported in table 2 remains unexplained, while the one-year attributions overextrapolated how much the strategy should have underperformed. Adding targeted interaction factors generally reduces this residual, albeit modestly. However, for the one-year U.S. attribution, it overextrapolates and flips the sign of the average residual, while for the one-year EUR attribution it raises its volatility. By design, the residual reallocation approach reliably leads to residuals of much smaller magnitude, more comprehensively linking performance to the factor-based investment strategy. The stocks that contributed most to the reduced attribution residual can be identified and included in client discussions, because true idiosyncratic risk is difficult to fully diversify away in active strategies. We note that lowering the intercept in the classification scheme specified in equation (4) would further reduce the remaining residual, with the statistical significance of the resulting reallocation providing a gauge of this being empirically justified.

Table 4 shows the *p*-values of testing for nonlinearities in factor returns through each attribution framework. In our experience, the nonlinear adjustments are rarely statistically significant for shorter-term attributions due to the impact of stock-specific risk. For discussion purposes, we therefore put the bar low and flag potential systematic interactions that appear to have affected performance with 90-percent confidence, rather than the more commonly used 95-percent threshold.

First, we note that the sizeable aggregate residual from standard attribution for both three-year attributions is unlikely to have resulted from stock-specific risk, suggesting it captured the impact of nonlinear factor pay-offs during this period. We note that the *p*-value for the aggregate residual from the one-year U.S. attribution is low as well.

Second, for the linear attribution including targeted interaction factors, we test for the estimated aggregated return contributions thereof being jointly zero. For the three-year EUR and the one-year U.S. attributions, and very nearly so for the three-year U.S. case, the statistical significance thereof provides evidence of nonlinearities in factor returns that are at least partially captured by the added interaction terms.

Third, we note that for the three-year EUR and one-year U.S. attributions the nonlinear reallocation of the unexplained performance under the standard approach is meaningfully different from what might be expected had stocks' residual return contributions truly reflected stock-specific events. For the three-year U.S. case, although the sizeable aggregate residual of standard attribution suggests latent exposure to some omitted risk factor, the residual reallocation rule appears to shed

VOLUME 20
NUMBER 1
2020

Table 5

NONLINEAR RESIDUAL REALLOCATION

Percent contribution (t-value); statistically significant at 90-percent confidence are marked with *

	Jan '15–Dec '17		Jan '14–Dec '15	
	U.S.	EUR	U.S.	EUR
Nonlinear Value	0.25 (1.19)	0.27 (1.31)	0.94 (2.53) *	0.27 (0.83)
Nonlinear Momentum	0.32 (1.78) *	0.52 (2.85) *	0.06 (0.21)	0.08 (0.24)
Nonlinear Size	0.5 (1.66) *	0.44 (1.75) *	0.21 (0.53)	0.28 (0.69)

little light on the impact thereof. However, we found that the reallocation did become statistically significant for an intercept of 2 and above, suggesting that interactions are stronger at the tails of the factor exposures range. The remaining unexplained performance after making this change, albeit larger than for a unit intercept, is still reduced by 66 percent from the standard regression approach (not shown).

Looking at all three tests combined, only the one-year EUR performance was fully consistent with a linear factor model. We also note that the lower p-values of the targeted interactions approach suggest higher power than the residual reallocation framework in uncovering possible nonlinearity in factor payoffs. The trade-off between the two nonlinear approaches is more specific insight into possible enhancements of the attribution model, versus a more encompassing understanding of the performance drivers of the particular strategy under review. This paper prioritizes the latter.

Table 5 reports how nonlinear residual reallocation would attribute most of the performance originally left unexplained. Both three-year attributions suggest that an interaction between momentum and size has benefitted performance. Drilling deeper into individual stocks' contributions to the residual, as much as 65 percent of the positive U.S. attribution residual indeed was caused by just three mega-cap momentum positions (two overweight, one underweight; not shown).[11] The one-year U.S. attribution suggests the portfolio's value-driven positions performed much better than linear regression had assessed, dominated by the tails thereof per our earlier observation (overweight deep value, or underweight those most overpriced). For these three strategies, the supplementary nonlinear attribution analysis provided a more nuanced narrative about performance. In contrast, for the one-year EUR attribution, we again find no indication of nonlinearity. There is no reason to believe its attribution residual, meaningful but smallest out of our four test cases on both an absolute basis and relative to active return, reflects anything but the performance impact of true stock-specific events.

CONCLUSIONS

Standard factor attribution assumes a linear relation between factor exposures and returns, yet investment constraints create a nonlinear relationship between factor exposures and portfolio weights. As a result, the approach generally fails to fully reconcile the performance of constrained factor-driven strategies. Sizeable attribution residuals that do not average out to zero over time suggest latent exposures to nonlinearities in factor returns. We have proposed a nonlinear adjustment that ascribes the impact thereof in an intuitive way that can be statistically tested. The resulting attribution better links realized performance to the investment process by mitigating both the long-term average as well as the volatility of the residuals.

We emphasize that our nonlinear residual reallocation proposal reflects a portfolio manager's perspective rather than an econometrician's. The approach is easy to implement, supplementing existing regression-based linear attribution, and it may help practitioners build a comprehensive, factor-driven narrative around performance. However, although the resultant nonlinear adjustment can be tested for statistical significance, it lacks the theoretical underpinning of an alternative asset pricing model.

This is certainly not the final word in the quest to find more comprehensive factor attributions. Areas of future research include testing alternative specifications of the kernel-like classification rule, equation (4). A much different implementation of the same idea is to apply machine learning techniques such as tree-based methods to stocks' residual return contributions. As factor investing increasingly moves outside the realm of linear models (e.g., Gu et al. 2020), matching this with nonlinear attribution techniques becomes all the more relevant.

Sanne de Boer, PhD, CFA®, is director of Quantitative Equity Research at Voya Investment Management. Contact him at sanne.deboer@voya.com.

The views expressed in this paper are the author's own and do not necessarily reflect those of the author's employer. The author thanks the anonymous referees for helpful suggestions for presenting the attribution framework, as well as his former colleagues at Invesco Quantitative Strategies for constructive comments when developing the proposed approach.

APPENDIX

De Boer (2012) shows that factor attribution exactly explains the return to fully invested unconstrained mean-variance optimal (MVO) portfolios when the same factor model is used for return and risk prediction as for attribution, and factor returns are estimated through generalized least squares (GLS). Generalizing a result by Jagannathan and Ma (2003), De Boer (2012) proceeds to show that constraints in MVO can be absorbed into a shrunk covariance matrix, and that attribution with GLS estimates of factor returns using this adjusted covariance matrix fully explains the constrained MVO portfolio's return. Mathematically, this approach turns out to redistribute the original attribution residual proportionately to factors' expected return contributions. De Boer and Jeet (2016) estimate factor returns subject to the restriction that attribution explains the strategy's performance. They show this has a closed-form solution that redistributes the attribution residual roughly proportional to the portfolio's squared factor exposures. Assuming the portfolio has approximately the highest exposures to factors with the highest expected return, both heuristics will give similar reallocations of the residual.

Although the resultant linear attributions provide a more comprehensive perspective on realized performance, albeit using inefficient factor return estimates, it is uninformative with regard to any model misspecification: The redistribution weights are based on the portfolio's active factor exposures rather than driven by return data. To address this shortcoming, we can apply the adjustment at the individual stock level rather than the portfolio level, assigning each stock's residual return contribution to the factors by which it is most characterized.[12] This idea is the basis of the nonlinear residual reallocation approach explored in this paper.

To facilitate statistical testing of the nonlinear residual reallocation, equation (5), we first define this adjustment in terms of realized rather than residual returns. Let $\boldsymbol{w}_k^{wls}$ denote the standard factor-mimicking portfolios from weighted least squares, which equal the columns of the projection matrix $\boldsymbol{P} \triangleq \boldsymbol{\Omega}^{-1}\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{\Omega}^{-1}\boldsymbol{B})$ that maps stock returns to linear factor returns (e.g., Grinold and Kahn 2000). Letting $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{BP}')$ denote the residuals-generating matrix from the weighted regression, we note that $\boldsymbol{w}_{resid} = \boldsymbol{M}'\boldsymbol{w}$ in decomposition, equation (2), of active portfolio weights underlying factor attribution. It follows that:

$$\boldsymbol{\Delta}_{nl} = \boldsymbol{r}'\boldsymbol{P}_{realloc}\,;\ \boldsymbol{P}_{realloc} \triangleq \boldsymbol{M}'diag(\boldsymbol{w}_{resid})\boldsymbol{B}_{classif}$$

Much like for the linear part of the attribution, the columns of the n by p matrix $\boldsymbol{P}_{realloc}$ can be interpreted as portfolios with returns equal to the proposed nonlinear adjustments. Because these portfolios have no linear exposure to any of the factors, they are subject only to stock-specific risk. To facilitate notation in this section, we decompose the covariance matrix of the stock-specific returns ε in equation (1) as $\sigma^2\boldsymbol{\Omega}$. Here σ^2 is a scalar that reflects the overall level of specific risk, and the diagonal matrix $\boldsymbol{\Omega}$ captures the dispersion thereof across stocks.[13] Under the common assumption that the dispersion has been prespecified, this leaves only its realized level ($\hat{\sigma}^2$) to be estimated. For technical reasons, we propose to do so through a weighted return regression where the factors of equation (1) are supplemented with the following nonlinear interaction terms:

$$\boldsymbol{B}_{inter} = \frac{1}{(\boldsymbol{w}_{resid}{}'\,\boldsymbol{\Omega}\boldsymbol{w}_{resid})}\boldsymbol{\Omega}\boldsymbol{P}_{realloc}$$

It can be shown that the resultant estimate $\hat{\sigma}^2$ is independent of the nonlinear adjustments $\boldsymbol{\Delta}_{nl}$, allowing reporting proper t-statistics to gauge their statistical significance.[14] If stocks' contributions to the attribution residual have no correlation to any of the included interaction terms, the reattribution is proportional to the nonlinear classification of the average stock (roughly equally, in our experience):

$$\boldsymbol{\Delta}_{nl}^{0} = \boldsymbol{r}'\boldsymbol{P}_{realloc}^{0}\,;\ \boldsymbol{P}_{realloc}^{0} \triangleq \tfrac{1}{n}\,\boldsymbol{w}_{resid}\,\boldsymbol{e}'\boldsymbol{B}_{classif}$$

To validate whether the residual reallocation, equation (5), truly captures nonlinearities in factor returns, we therefore consider the difference thereof to this "proportional reallocation":

$$\boldsymbol{\delta}_{nl} = \boldsymbol{\Delta}_{nl} - \boldsymbol{\Delta}_{nl}^{0}$$

The distribution of $\boldsymbol{\delta}_{nl}$ is degenerate as the differences net to zero, but we can run an F-test on the joint statistical significance of any subset of $(p-1)$ elements thereof. In a multiperiod context, testing can be done again jointly for all periods or on an aggregate basis, the latter being most relevant to the reported attribution. If $\boldsymbol{\delta}_{nl}$ is statistically significantly different from zero, this suggests the nonlinear classification by factor meaningfully differentiated residual return contributions.

For completeness, we provide the exact definition of the test statistics we propose to report as part of the attribution. First, we define the aggregate residual reallocation $\boldsymbol{\Delta}_{nl}^{aggr}$ and the difference thereof to the aggregated proportional reallocation $\boldsymbol{\delta}_{nl}^{aggr}$:

$$\boldsymbol{\Delta}_{nl}^{aggr} = \textstyle\sum_{t=1}^{T}\boldsymbol{\Delta}_{nl}^{t}\,;\ \boldsymbol{\delta}_{nl}^{aggr} = \textstyle\sum_{t=1}^{T}(\boldsymbol{\Delta}_{nl}^{t} - \boldsymbol{\Delta}_{nl}^{0,t});$$

The pooled estimate of the specific risk level per the residuals $\hat{\boldsymbol{\varepsilon}}_{inter}^{t}$ of the expanded periodical regressions equals:

$$\hat{\sigma}_{aggr}^{2} = \frac{1}{n_{aggr}}\textstyle\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_{inter}^{t}{}'\boldsymbol{\Omega}_t^{-1}\hat{\boldsymbol{\varepsilon}}_{inter}^{t}\,;\ n_{aggr} = \textstyle\sum_{t=1}^{T}(n_t - m_t - p)\,;$$

The joint covariance of aggregated residual reallocation $\boldsymbol{\Delta}_{nl}^{aggr}$ equals:

$$\sigma^2\textstyle\sum_{t=1}^{T}\boldsymbol{P}_{realloc}^{t}{}'\boldsymbol{\Omega}_t\boldsymbol{P}_{realloc}^{t} = \sigma^2\boldsymbol{\psi}_{aggr}^{nl}$$

Similarly, the joint covariance of the cumulative difference $\boldsymbol{\delta}_{nl}^{aggr}$ thereof with the uninformative adjustments equals:

$$\sigma^2\sum_{t=1}^{T}(\boldsymbol{P}_{realloc}^{t} - \boldsymbol{P}_{realloc}^{0,t})'\boldsymbol{\Omega}_t(\boldsymbol{P}_{realloc}^{t} - \boldsymbol{P}_{realloc}^{0,t}) = \sigma^2\boldsymbol{\phi}_{aggr}^{nl}$$

Under the null hypothesis of a linear factor model, the aggregated residual reallocation to interaction factor k normalized by its estimated volatility then follows the t-distribution:

$$[\boldsymbol{\Delta}_{nl}^{aggr}]_k/\sqrt{\hat{\sigma}_{aggr}^{2}[\boldsymbol{\psi}_{aggr}^{nl}]_{kk}} \sim t(n_{aggr})$$

Lastly, our test for nonlinearity in the aggregate factor payoffs considers the difference between the residual reallocations and the uninformative adjustments jointly being zero. The calculation below drops the *pth* interaction from consideration, but the exact choice thereof does not impact the test statistic.

$$\frac{[\boldsymbol{\delta}_{nl}^{aggr}]'_{-p}\,[\boldsymbol{\Phi}_{aggr}^{nl}]^{-1}_{-p,-p}\,[\boldsymbol{\delta}_{nl}^{aggr}]_{-p}/(p-1)}{\hat{\sigma}^2_{aggr}} \sim F(p-1,\, n_{aggr})$$

Much like for regular linear regression, the accuracy of the reported test statistics relies on the validity of the return model specified in equation (1), and the assumption of Gaussian distributed specific risk is less critical for sufficiently large estimation universes.

ENDNOTES

1. From conversations with other industry practitioners during a panel discussion at the Axioma Quant Forum, New York, NY (fall 2014).
2. For simplicity, the analysis has focused on arithmetic attribution and aggregation thereof over time. See Carino (1999), among others, for proposals on how to link single-period factor attributions to explain properly compounded multi-period active performance.
3. The textbook approach that minimizes estimation error (e.g., Greene, 2003) is to use regression weights inversely proportional to stocks' specific risk. In practice, the regression weights are taken to be the square root of each stock's market capitalization.
4. Additional but avoidable frictions affecting the residual include any disconnect between regression weights and the specific risk predictions used for portfolio construction, as well as any misalignment between the factors used for return prediction, risk modeling, and performance attribution.
5. The distribution of idiosyncratic risk for each subperiod should continue to reflect cross-sectional dispersion in stock-specific volatilities but also be proportional to its duration and possibly reflect any variation in the overall risk level over time.
6. Testing linear constraints on regression coefficients is standard econometrics (e.g., Greene 2003).
7. It's assumed here that factor exposures have been demeaned with roughly symmetric positive and negative exposures, excepting segment indicators such as country and sector flags. This implementation is based on quadratic factor exposures, which applies the restricted least squares attribution framework by De Boer and Jeet (2016) at the individual stock level, and is found to be empirically robust and flexible; but other classification schemes of stocks' residual return contributions remain to be explored.
8. The factors of equation (1) typically include an intercept reflecting the market as well as indicator variables for each relevant industry. To deal with the resultant perfect multicollinearity, we fit the model with the constraint that the weighted sum of the estimated regression coefficients across this segmentation equals zero, rendering an interpretation as active industry returns. Each additional set of segmentation variables such as a country mapping is incorporated using an additional such constraint. For attributions spanning multiple currencies, we run the regression using local-currency stock returns, calculating currency contributions directly and reporting them separately.
9. To simplify implementation, all statistical tests assume that the cross-sectional dispersion of stock-specific variances in each month is inversely proportional to the square root of market capitalizations, the choice of regression weights. The appendix touches upon how the dispersion in Axioma's specific risk might have been used, both cross-sectionally among stocks and over time, in pursuit of more accurate test statistics.
10. Mitigating unintended industry and country tilts as well as other style exposures in portfolio construction also might have reduced the hypothetical strategy's realized tracking error of about 5 percent per annum, potentially boosting its information ratio. Any such enhancements are outside the scope of this paper.
11. The sign of this adjustment is likely period-specific. Hong et al. (2000) in fact report a higher historical momentum premium for small-cap stocks, an interaction we could consider including in equation (1).
12. We thank Anna Gulko for this suggestion.
13. The multiplicative indeterminacy inherent to this decomposition is inconsequential for these purposes, but it may be specified that the average of the diagonal elements of $\boldsymbol{\Omega}$ equals 1 to resolve it.
14. The proof hinges on the proposed adjustment $\boldsymbol{\Delta}_{nl}$ being a linear transformation of the estimated coefficients of the added regressors, which are known (e.g., Greene 2003) to be independent of the corresponding estimate of $\hat{\sigma}^2$. The exposures to the added interaction factors can be shown to be non-negative and sum to 100 percent, thus decomposing the residual volatility, while the resulting regression-based attribution fully explains the portfolio's active return. However, (unreported) empirical tests suggest the multi-collinearity of $\boldsymbol{B}_{inter}$ renders the resultant attribution too unstable to be directly useable. Note that the addition of the interaction effects does not change the linear part of the attribution. Only if the regression weights are not consistent with the dispersion in stock-specific risk assumed for statistically testing the nonlinear adjustment does least squares have to be run twice: once to generate the linear attribution and residual portfolio weights and separately to estimate the level of specific risk.

REFERENCES

Carino, D. R. 1999. Combining Attribution Over Time. *Journal of Performance Measurement* 4, no. 3 (summer): 5–15.
Ceria, S., A. Saxena, and R. A. Stubbs. 2012. Factor Alignment Problems and Quantitative Portfolio Management. *Journal of Portfolio Management* 38, no. 2 (winter): 29–43.
Clarke, R., H. de Silva, and S. Thorley. 2002. Portfolio Constraints and the Fundamental Law of Active Management. *Financial Analysts Journal* 58, no. 5 (September/October): 48–66.
———. 2005. Performance Attribution and the Fundamental Law. *Financial Analysts Journal* 61, no. 5 (September/October): 70–83.
De Boer, S. 2012. Factor Attribution That Adds Up. *Journal of Asset Management* 13, no. 6 (December): 373–383.
De Boer, S., and V. Jeet. 2016. Aligning Factor Attribution with Latent Exposures, *Journal of Asset Management* 17, no. 7: 502–525.
De Boer, S., J. Keuerleber, and C. Rother. 2018. Performance Attribution Through a Factor Lens. *Invesco Risk & Reward* 2: 32–38.
Fischer, B. R., and R. Wermers. 2013. *Performance Evaluation and Attribution of Security Portfolios*, 1st ed. Waltham, Massachusetts: Academic Press.
Greene, W. H. 2003. *Econometric Analysis*, 5th ed. Englewood Cliffs, New Jersey: Prentice Hall.
Grinold, R. 1989. The Fundamental Law of Active Management. *Journal of Portfolio Management* 15, no. 3 (spring): 30–37.
———. 2006. Attribution. *Journal of Portfolio Management* 32, no. 2 (winter): 9–22.
———. 2011. The Description of Portfolios. *Journal of Portfolio Management* 37, no. 2 (winter): 15–30.
Grinold, R., and R. Kahn. 2000. *Active Portfolio Management*, 2nd ed. New York: McGraw-Hill.
Gu, S., B. Kelly, and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33, no. 5 (May): 2,223–2,273.
Hong, H., T. Lim, and J. C. Stein. 2000. Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. *Journal of Finance* 55, no. 1 (February): 265–295.
Jagannathan, R., and T. Ma. 2003. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance* 58, no. 4 (August): 1,651–1,684.
Lee, J.-H., and D. Stefek. 2008. Do Risk Factors Eat Alphas? *Journal of Portfolio Management* 34, no. 4 (summer): 12–25.
Menchero, J., and V. Poduri. 2008. Custom Factor Attribution, *Financial Analysts Journal* 64, no. 2 (March–April): 81–92.
Saxena, A., C. Martin, and R. A. Stubbs. 2013. Constraints in Quantitative Strategies: An Alignment Perspective. *Journal of Asset Management* 14, no. 5 (October): 278–292.
Sivaramakrishnan, K., and R. A. Stubbs. 2013. Improving the Investment Process with a Custom Risk Model: A Case Study of the GLER Model. *Journal of Investing* 22, no. 4 (winter): 129–146.
Stubbs, R. A., and V. Jeet. 2016. Adjusted Factor-Based Performance Attribution. *Journal of Portfolio Management* 42, no. 5 (Special Quality Equity Strategies Issue): 67–78.
Stubbs, R. A., and D. Vandenbussche. 2010. Constraint Attribution. *Journal of Portfolio Management* 36, no. 4 (summer): 48–59.
Vandenbussche, D. 2016. Long-only Factor Mimicking Portfolios. Presented at the Fifth Annual Axioma Quant Forum, November 1, 2016. New York, New York.

5619 DTC Parkway, Suite 500
Greenwood Village, CO 80111
Phone: +1 303-770-3377
www.investmentsandwealth.org