

The Active Side of ESG Index Investing

WHO YOU EXCLUDE MATTERS MORE THAN HOW MUCH YOU EXCLUDE

Daria Plakhova, Durga Shankar, Gaston Siegelaer and Hendro Sugandi

Customizations driven by sustainability preferences are widely used to create indices that lie somewhere between two ends of investing style – active and passive. Previous research has shown that different sustainability-driven customizations lead to different active risk-return outcomes. This paper shows that not all exclusions are created equal. The impact of sustainability-driven exclusions depends not only on how much of the benchmark weight is removed, but also on which companies are excluded. Excluding a few benchmark-relevant large-cap companies can have a greater impact on tracking error than excluding a similar aggregate weight spread across many smaller companies. We further show that index methodology matters. The choice of index construction approach can materially affect benchmark replication effectiveness, highlighting the importance of research-driven index design choices, ongoing monitoring and regular evaluation of whether the methodology remains aligned with the index’s objectives.

Daria Plakhova CQF (l)
Quantitative researcher
within the Index
Solutions team at APG
Asset Management

Hendro Sugandi PhD (ml)
Quantitative researcher
within the Index
Solutions team at APG
Asset Management

Gaston Siegelaer PhD (mr)
Senior investment
specialist at APG Asset
Management

Durga Shankar CFA (r)
Quantitative researcher
within the Index
Solutions team at APG
Asset Management



PASSIVE INVESTING IN ESG INDICES HAS RAISED NEW QUESTIONS

In the past twenty years two trends characterized the development of the institutional investment industry: a trend towards more passive investing and the emergence of ESG investing. Cost awareness and increased cost transparency required by regulation drove pension funds towards larger passive mandates at the expense of actively managed mandates. At the same time, ESG investing policies of the pension funds required benchmarks for their mandates that aligned with those policies. In the early years, off-the-shelf ESG indices from index providers were being used as benchmarks. However, more sophisticated ESG policies triggered the need for tailor-made indices for their passive mandates. Those customized indices not only had to feature the elements of the pension fund’s ESG policy, but they also had to be fit for purpose for passive mandates.

The prudent-person rule for pension funds requires the investment portfolio to be sufficiently diversified. A typical way to address this fiduciary and regulatory requirement is to try to mimic a broad market index by using a specific index construction method for the customized ESG index. This broad market index, hereafter referred to as the parent index, serves as the basis for the customized ESG index. The customized ESG index is derived from the parent index through constituent exclusions or reweightings. Ideally, in the long run the investment performance of the customized ESG index is expected to be roughly similar to the performance of the parent index. The ex-ante tracking error of the customized index versus the parent index is a pragmatic indicator of the possible performance deviation between the two indices. The lower the ex-ante tracking error, the higher the likelihood that the customized index will perform in line with the parent index. One could expect the tracking error to be larger for an index which excludes a larger aggregate weight of companies, than for an index which excludes a lower aggregate weight of companies. However, this relationship is not linear and furthermore it also depends on the size, number and interrelatedness of excluded individual companies. The research question of this paper is: how do exclusion intensity and the size of the companies retained in the index affect the ability of a customized index to effectively replicate its parent index?

When investigating this research question, we consider various kinds of customization, both with regard to ESG policies and with regard to index construction. We consider two ESG policies. The first uses carbon intensity as a criterion for inclusion or exclusion. The second is based on MSCI ESG ratings and excludes companies with the lowest ESG ratings.

For index construction, we apply two different approaches. One approach uses pro rata reweighting of the companies in the customized index after taking out the weights of the excluded companies, such that the total weight sums up to 100% again. The second approach uses minimization of the ex-ante tracking error of the customized index versus the parent index. Other possible optimization approaches may use factor tilts or beta as

measure, but in this paper, we want to make an apples-to-apples comparison and therefore we use tracking error as the variable in scope. We will investigate how the tracking error behaves for different ESG policies leading to different weight patterns of exclusions and for different index construction approaches.

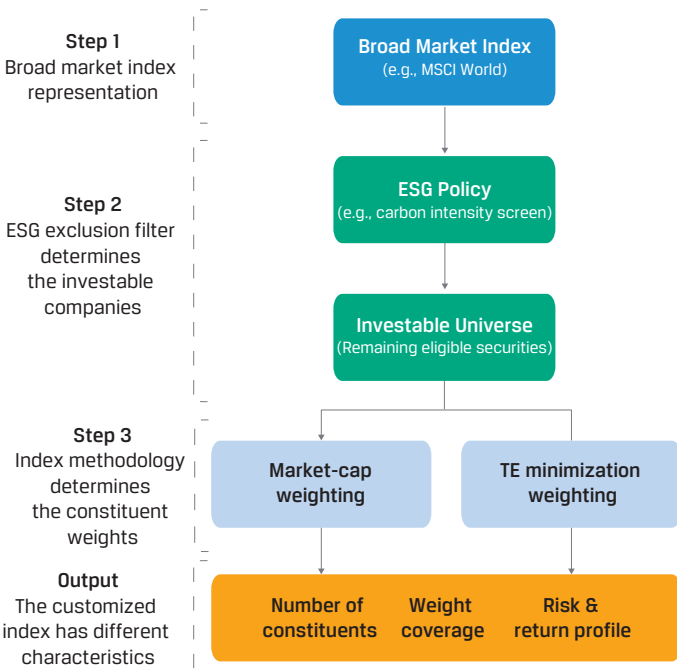
Other authors have analyzed the interaction between ESG policies and the investment risk-return profile (Huij et al., 2025) and have looked into the level of diversification (Brögger et al., 2025). Our paper adds to this literature by specifically focusing on the interaction of the size of excluded companies and index construction approaches.

Furthermore, we show in this paper how past performance of an optimized index can be compared with a pro rata reweighted index, which adds to the paper by Lammers et al. (2024) who analyzed a capped benchmark derived from an exclusion benchmark.

UNDERSTANDING THE IMPACT OF ESG EXCLUSIONS ON BROAD MARKET INDEX REPLICATION

As illustrated in Exhibit 1, the construction of a customized index can be viewed as a sequence of decisions. Starting from a broad market index, an ESG policy determines the investable universe by excluding companies that do not satisfy the investor’s sustainability criteria. An index methodology is then applied to assign weights to the remaining companies, resulting in the final customized index. Both decisions influence the benchmark’s characteristics, including its number of constituents, weight coverage, factor exposures, and risk-return profile.

Exhibit [1]
Conceptual Framework for ESG-Driven Index Customization



For investors seeking broad market exposure, an important question is how closely a customized index continues to represent its parent index. We use ex-ante and ex-post tracking error (TE) as the primary measure of benchmark representativeness. A lower tracking error indicates that the customized index is expected to exhibit risk and return characteristics similar to those of the parent index, whereas a higher tracking error implies greater deviations from the broad market benchmark.

Under a standard risk-model framework, tracking error is determined by the active weights relative to the parent index and the covariance structure of the underlying securities. Consequently, the impact of an ESG exclusion depends not only on the aggregate benchmark weight removed, but also on the characteristics of the excluded companies and the extent to which their exposures can be replicated using the remaining investable universe.

Our central hypothesis is that not all exclusions are created equal. Excluding large companies is expected to have a greater impact on benchmark replication than excluding a similar aggregate weight of relatively smaller sized companies. Large companies often represent economically important factor exposures and may contain idiosyncratic risk that cannot be fully replicated by the remaining universe. As a result, exclusions concentrated in a small number of large constituents may create structural gaps in systematic exposures and concentrated stock-specific risk. By contrast, exclusions spread across many smaller constituents are more likely to preserve the overall factor structure of the parent index and keep idiosyncratic risk diversified.

This distinction has important implications for index construction. When exclusions are broad and diversified, tracking-error minimization is expected to be more effective because the optimizer can replicate the parent index’s exposures using the remaining investable securities. Conversely, when exclusions are concentrated in a few large constituents, tracking-error minimization may be less effective because the missing exposures cannot be fully replicated through reweighting. For example, excluding a large constituent such as NVIDIA from a developed-markets benchmark may disproportionately increase tracking error because its sizeable weight and distinctive combination of semiconductor-industry risk and stock-specific risk are difficult to replicate.

EMPIRICAL ILLUSTRATION OF CUSTOMIZATION IMPACT

It is widely recognized that sustainability-driven customizations introduce an active component to index investing and lead to a different risk-return profile compared to the parent index. The magnitude of these deviations, however, need not be a surprise. The characteristics of an ESG policy – particularly the nature and size of its exclusions – already provide valuable information about the expected quality of benchmark replication and can, to a certain extent, even be controlled through smart choices in index construction.

To illustrate this effect, we construct two stylized ESG policies that differ primarily in the size characteristics of the excluded companies (Exhibit 2). The first policy is based on carbon-emission intensity. While approximately 35-40% of the parent-index weight is excluded, the exclusions are concentrated in smaller companies, leaving most large companies-particularly technology stocks-investable. The second policy is based on MSCI ESG ratings, excluding a similar share of the parent index weight, but primarily through large-cap names.

ESG policy is only one component of benchmark design. Other components of index methodology, such as the index objective and weighting scheme, are equally important. In this study, we examine how much information ESG exclusions reveal about the resulting benchmark and to what extent their effect on the relative risk can be mitigated through index construction. To illustrate this interaction, we compare two weighting schemes: traditional market-cap weighting and ex-ante tracking-error minimization. In the latter case, standard implementation constraints are applied, including constraints on active country and industry exposure relative to the parent index.

This analysis uses the MSCI World Index as the starting universe, providing broad representation of developed equity markets. The backtest covers the period from May 2022 to May 2026, with semi-annual rebalancing in May and November – a common schedule among index providers. As the objective is to illustrate the impact of different exclusion policies on benchmark replication, the selected period is intended as a recent and practically relevant case study rather than a long-run empirical analysis.

Combining the two ESG policies with the two weighting approaches results in four benchmark variants, shown in Exhibit 3. These cases allow us to isolate the effect of two key design choices: the size of the ESG exclusions and the index objective.

Exhibit [2] Illustrative ESG exclusion policies	Policy	ESG Policy	ESG Policy Description	Primary Impact
	1	Carbon Intensity	Excludes MSCI Laggards and companies with carbon-emission intensity (Scopes 1, 2, and estimated Scope 3 emissions) using Enterprise Value Including Cash (EVIC) above the median of the parent index	Exclusions driven primarily by smaller companies
	2	MSCI ESG Ratings	Excludes MSCI Laggards and the bottom quartile of the parent index based on MSCI ESG rating	Exclusions driven primarily by large-cap names

Exhibit [3]
Index Variants and
Results

Case	Description	Number of Constituents	MSCI Weight Coverage (%)	Relative Return (%)	Ex-post TE (%)	Ex-ante TE (%)
1	TE minimization, ESG Policy 1 (Carbon)	493	58	-0.6	1.6	1.3
2	TE minimization, ESG Policy 2 (MSCI ESG Scores)	704	59	-2.0	2.4	1.6
3	Market-cap weighting, ESG Policy 1 (Carbon)	678	62	0.7	2.6	2.4
4	Market-cap weighting, ESG Policy 2 (MSCI ESG Scores)	1025	67	-1.0	2.4	1.9

Notes: Number of Constituents and MSCI Weight Coverage (%) are reported as averages across rebalancing dates. Relative Return (%) denotes the annualized relative return of each case versus the MSCI World Index over the sample period. Ex-post Tracking Error (%) is the annualized standard deviation of daily active returns. Ex-ante Tracking Error (%) is the average monthly ex-ante tracking error estimate during the sample period, annualized. The sample period is from May 2022 to May 2026.

The results summarized can be analyzed from two perspectives. First, comparing the two ESG policies under the same weighting approach (Cases 1 and 2 and Cases 3 and 4), highlights the impact of the exclusion profile. Because the carbon-based policy primarily removes smaller companies, it results in fewer constituents than the ESG rating-based policy, regardless of the weighting approach. Some differences in excess returns are observed across the cases, although these are not the primary focus of the analysis. Both ex-post and ex-ante tracking error indicate higher relative risk under the carbon-based policy when no explicit risk controls are applied, as evidenced by Cases 3 and 4. The second perspective concerns the index objective. Comparing the two weighting approaches under the same ESG policy (Cases 1 and 3 and Cases 2 and 4), confirms, as expected, that tracking-error minimization reduces active risk relative to market-cap weighting. The more interesting observation, however, is that the effectiveness of optimization differs substantially between the two ESG policies.

Under the carbon-based policy, market-cap weighting results in an ex-ante tracking error of 2.4% (Case 3), while the introduction of the tracking error minimization allows the ex-ante tracking error to be reduced to just 1.3% (Case 1). The policy that appears most challenging from a benchmark replication perspective becomes the easiest to manage through optimization.

OPTIMIZATION HELPS STRIKE AN EFFECTIVE
BALANCE BETWEEN BENCHMARK
REPLICATION AND ESG POLICY OBJECTIVES

The opposite pattern emerges under the ESG rating-based policy. While optimization also reduces tracking error, the improvement is much smaller. Moving from market-cap weighting (Case 4) to tracking-error minimization (Case 2) lowers ex-ante tracking error from 1.9% to 1.6%, a considerably more modest reduction than observed under the carbon-based policy.

This leads to the central question of the study: to what extent can the quality of parent index replication be inferred from the size

of exclusions, and how much active risk can be mitigated through index design choices?

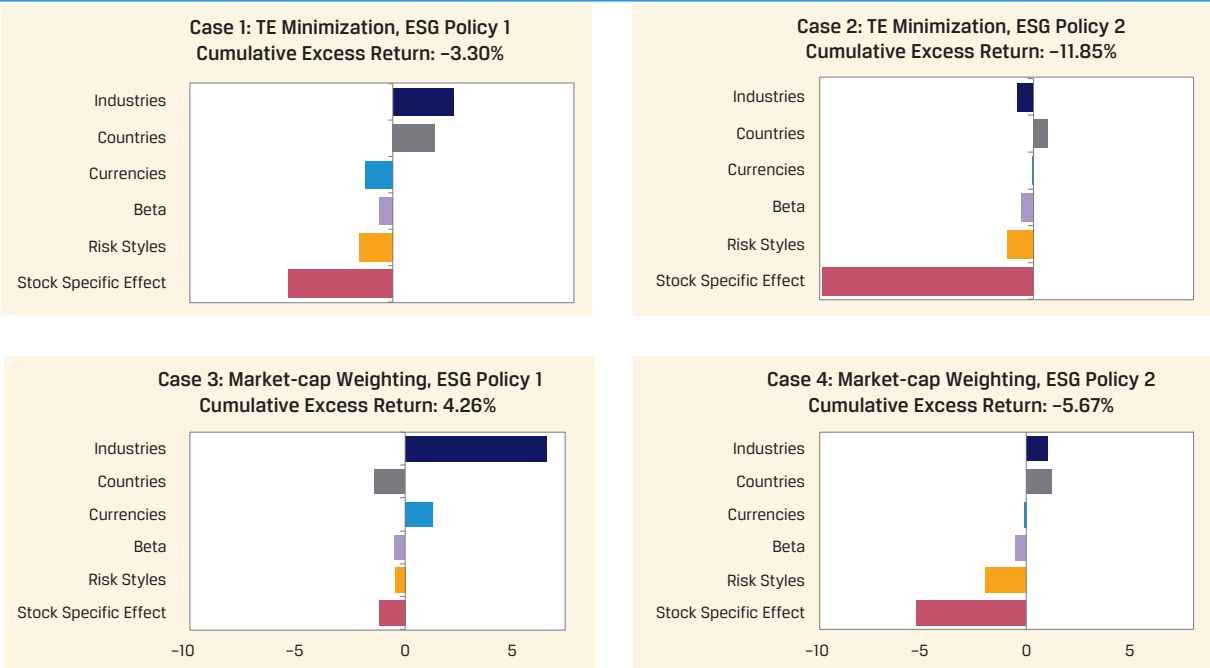
When comparing the ex-ante and ex-post tracking error, the difference between the two measures is expected, as the former is a model-based forecast while the latter reflects realized outcomes. However, the magnitude of this difference varies under the two ESG policies, regardless of the weighting approach. For example, in Case 2, where ex-ante tracking-error minimization is applied and the ESG rating-based policy excludes large benchmark-relevant names, the ex-post tracking error reaches 2.4% compared to an ex-ante estimate of 1.6%.

While discrepancies between ex-ante and ex-post tracking error can arise from changing market conditions, realized factor returns, and risk-model estimation error, the larger gap may also be attributable to idiosyncratic risk. Excluding large benchmark constituents removes stock-specific risk exposures that cannot be easily replicated. As a result, idiosyncratic risk accounts for a larger share of total tracking error. Because ex-post tracking error depends on the realized performance of the excluded stocks over the sample period, outcomes become more sensitive to stock-specific return surprises, increasing the likelihood that realized tracking error deviates materially from its ex-ante estimate.

UNDERSTANDING THE DRIVERS OF CUSTOMIZATION
IMPACT

To better understand how different ESG policies impact broad market replication, we first examine the results through performance attribution for the four customized index variants (Exhibit 4). A notable distinction between the two ESG policies is the contribution of the stock-specific effect. Comparing the two ESG policies under the same weighting scheme (Cases 1 and 2, and similarly Cases 3 and 4), a lower absolute contribution from the idiosyncratic factor is observed in the benchmarks based on the carbon-emission intensity policy as opposed to those based on the ESG rating-based policy, regardless of the weighting approach. This difference can be traced back to the nature of the exclusions. The ESG rating-based policy excludes a larger number of large-cap companies, making it more difficult to replicate the parent index and increasing exposure to idiosyncratic return drivers.

Exhibit [4]
Performance
Attribution Over
the Backtest Period



The choice of weighting scheme also has a visible impact on the sources of excess return. Moving from market-cap weighting to tracking-error minimization reduces the contribution of systematic risk factors such as industries, countries, and risk styles in most cases. For example, in Case 3, where the carbon-based policy is applied with the market-cap weighting scheme, the Industries factor makes a significant positive contribution. This is mainly driven by an overweight in the Semiconductors and Computers & Electronics industries, which delivered strong positive returns over the backtest period, combined with an underweight in the Oil and Gas Exploration & Production industry, which experienced negative returns. Excluding companies with higher carbon-emission intensity while retaining exposure to large technology companies resulted in a favorable industry exposure effect over the sample period. Under the same ESG policy but with tracking-error minimization (Case 1), the portfolio’s factor exposures are aligned more closely with those of the parent index. As a result, the contribution from the Industries factor declines substantially.

At the same time, the contribution of the stock-specific effect becomes more pronounced under the tracking-error minimization approach. While systematic factor exposures can be replicated relatively closely using the remaining constituents, stock-specific effects are unique to individual securities and may therefore be more difficult to replicate once those securities are excluded.

Performance attribution provides insight into the realized consequences of the exclusions, but it does not directly address future benchmark risk. For that purpose, ex-ante tracking error decomposition offers a relevant perspective (Exhibit 5). By separating tracking error into factor and stock-specific components, it allows us to identify which sources of risk are

introduced by the ESG policy and which can potentially be mitigated through index construction.

To isolate the effect of the exclusion policy itself, we begin with the market-cap-weighted cases (Cases 3 and 4). These portfolios involve minimal intervention through index design and therefore provide the clearest representation of the underlying ESG policies.

PERFORMANCE ATTRIBUTION AND TRACKING ERROR DECOMPOSITION REVEAL WHAT DRIVES INDEX RISK AND RETURN, CREATING A FEEDBACK LOOP THAT CAN IMPROVE INDEX DESIGN

In the market-cap-weighted cases, tracking error is higher under the carbon-based ESG policy, which retains most of the large names in the index and achieves exclusions primarily through relatively smaller companies. The key difference, similar to the performance attribution analysis, lies in the contribution of idiosyncratic risk. Under the carbon-based policy, stock-specific risk accounts for less than 50% of total ex-ante tracking error in most cases, with significant contributions coming from Industries (48% on average) and Risk Styles (12% on average). In contrast, under the ESG rating-based policy, excluding more large names, idiosyncratic risk is the dominant driver of tracking error, with its contribution increasing from 46% in May 2022 to 75% in May 2026.

This illustrates how sustainability-driven exclusions affect index composition. Excluding a large number of benchmark-relevant large companies introduces stock-specific risk that cannot be easily replicated. As a result, even though overall ex-ante tracking error is higher for the carbon-based ESG policy (Case 3) than for the ESG rating-based policy (Case 4), the contribution from idiosyncratic risk – arguably the most difficult risk component to manage and impossible to fully replicate in the presence of exclusions – is actually lower in Case 3.

A RECURRING QUESTION: DOES THE INDEX REMAIN FIT FOR PURPOSE OR WARRANT REFINEMENT?

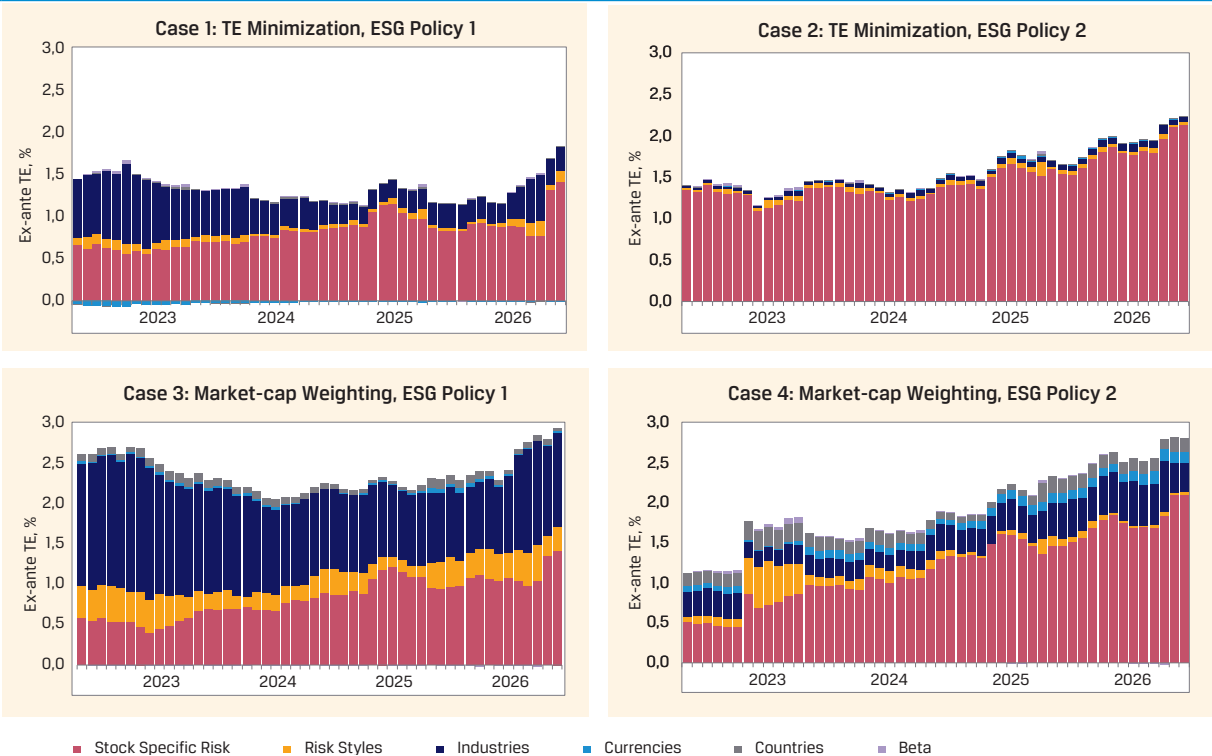
This observation helps explain why the tracking-error minimization proves more effective under the carbon-based policy, where exclusions are concentrated in relatively smaller companies, than under the ESG rating-based policy, where many large companies are excluded. Factor risk can be managed through tools such as tracking-error minimization as the optimization seeks to match factor exposures to those of the parent index as closely as possible. This results in a noticeable reduction in factor-risk contributions under both policies. However, the reduction is substantially larger under the carbon-based policy, as a greater share of the active risk originates from factor exposures. Under the ESG rating-based policy, a much larger proportion of active risk is idiosyncratic in nature and therefore cannot be effectively mitigated through optimization alone.

Another observation highlights the importance of every element of index design. Comparing the weighting schemes under the ESG rating-based policy (Cases 2 and 4), we observe that at the beginning of the backtest period there were several months when ex-ante tracking error under the optimized approach was higher than under market-cap weighting. At first glance, this may appear counterintuitive. However, it can be explained by the implementation constraints applied in the optimization, such as limits on active industry and country exposures.

Compared to the unconstrained market-cap-weighted case, optimization with constraints introduces safeguards intended to ensure close alignment of custom index characteristics with those of the parent index. However, when these constraints become binding, optimization can no longer achieve the unconstrained objective. This trade-off is important to keep in mind not only during index construction but also throughout the life of the index. The interaction between objectives and constraints may evolve over time as market conditions change, which was also evident from the performance attribution results discussed earlier.

A final observation relates to the importance of ongoing monitoring. Time-series tracking error decomposition reveals how the drivers of active risk evolve over time and can provide valuable information for investors. For example, under the ESG rating-based policy, ex-ante tracking error increases over the backtest period, primarily due to a rising contribution from stock-specific risk. Such a trend may warrant further investigation and illustrate the need for periodic evaluation of a custom benchmark.

Exhibit [5]
Time-Series of
Ex-Ante Tracking
Error Decomposition
Over the Backtest
Period



PRACTICAL IMPLICATIONS FOR INDEX DESIGN AND CONCLUSION

In this study, we show how the size characteristics of ESG exclusions provide important information about the expected quality of broad market index replication by a customized index. In general, the more large-cap companies are excluded, the more difficult it becomes to replicate the broad market index.

We explore this effect through the lens of ex-ante tracking error decomposition. Excluding benchmark-relevant large companies typically increases idiosyncratic risk, which cannot be effectively replicated through index construction techniques. As a result, tracking-error minimization is less effective in reducing active risk. By contrast, when sustainability-driven exclusions are mostly concentrated in relatively smaller companies, tracking-error minimization can significantly reduce active risk primarily by aligning the systematic factor exposures of the customized index with those of the parent index. Therefore, the size factor of exclusions interacts closely with index construction rules and can provide useful guidance on which approach is likely to be more effective under a given ESG policy. This is an important consideration for investors when integrating sustainability-driven exclusions into index design.

A further implication is that index design and management should be viewed as an ongoing process rather than a one-off exercise. Even when the ESG policy remains unchanged, the

drivers of active risk may evolve over time as market conditions and index characteristics change. For this reason, regular monitoring and periodic reassessment are essential to ensure that the customized index remains relevant and continues to function as intended. The objective is not only to ensure that the index continues to align with the intended investment objectives and represents the target asset class, but also that its risk-return characteristics remain consistent with expectations. From a practitioner's perspective, the recurring question is simple: is this index still fulfilling the purpose it was originally designed for?

References

- Huij, J., Laurs, D., & Van Zanten, J.A. (2025). *The Investment Implications of Sustainable Investing*. *Journal of International Money and Finance*, 159, December 2025.
- Brøgger, A., Koëter, J., & van Dijk, M. (2025). *FOMO in Equity Markets? Concentration Risk in (Sustainable) Investing*. SSRN Working Paper. Available at SSRN.
- Lammers, R., Radford, S., & Verkerk, V. (2024). *The Development of an ESG Attribution Model*. *VBA Journaal*, Winter 2024, pp. 15-24.

End Note

The views expressed in this paper are our own and do not necessarily reflect the views of APG Asset Management. We thank Hans van Westrienen and Ryan Chidley (APG) for their insightful comments. We thank the VBA editorial board and the reviewers for their feedback.