

Brinson and Factors: A Unified Framework for Complete Portfolio Return Attribution

July 12, 2026

Abstract

Treating performance attribution as an estimation problem yields structured solutions to three common challenges: reallocating interaction terms in Brinson attribution, reallocating residual returns in factor attribution, and reconciling arithmetic attribution with compounded portfolio returns.

All three problems adjust attribution components to achieve a cleaner and more interpretable presentation of results. Such adjustments imply uncertainty in the underlying attribution estimates. We propose a regression-based framework that makes this uncertainty explicit and measurable. We then direct the required adjustments toward the least precise estimates.

In all three settings, we implement the adjustments as restricted generalized least squares updates to the original estimates. This approach provides a statistically principled alternative to ad hoc reallocations of residuals, interaction effects, and multi-period smoothing adjustments, while preserving exactness and internal consistency of the attribution results.

Contents

1	Introduction	1
2	Factor-Based Attribution	2
2.1	Classical decomposition	2
2.2	Regression formulation	3
2.3	Reallocating residuals	4
3	Brinson Attribution	6
3.1	Classical decomposition	6
3.2	Regression formulation	6
3.3	Reallocating interaction effects	8
3.4	Active-exactness as a restriction on both regressions	8
4	Illustrative Examples	12
4.1	Balanced precision	12
4.2	Spherical specific risk and optimal interaction splits	14
5	Multi-Period Attribution	17
5.1	The weighting problem	17
5.2	A precision-weighted GLS choice of period weights	18
5.3	Construction of the covariance matrix	19
5.4	Entry-specific reconciliation	20
6	Summary	20
7	References	22
	Appendices	23
A	Group-Level Active-Exactness Restrictions	23
A.1	Multiple active-exactness constraints	23
A.2	Restricted GLS formulation	24
A.3	Interpretation and practical remarks	24
B	Long-Short and Market-Neutral Portfolios	24
C	Smoothing and Linking as GLS Selection Rules	25
C.1	The under-identified weighting problem	25
C.2	GLS formulation	26
C.3	Recovering a given admissible weight vector	26
C.4	Special cases	27
C.5	Relation to entry-specific reconciliation	28
C.6	Interpretation	28
D	Sample Code	29

Acknowledgements

I am grateful for helpful comments from Tony Elavia, Harvey Fram, and Migene Kim.

1 Introduction

Portfolio return attribution is a central tool in performance evaluation for investment funds. The classical allocation-selection framework of Brinson and Fachler (1985) and Brinson, Hood, and Beebower (1986) decomposes active returns into allocation, selection, and interaction effects. Alternatively, factor-based performance attribution uses regression techniques to explain portfolio returns through exposures to systematic factors and residual idiosyncratic returns (see, for example, Grinold (2006)).

Spaulding (2003) argues that a complete attribution should satisfy two “laws”: it must attribute performance to the manager’s actual decisions and exhaust all returns without residuals. We formalize these principles as linear restrictions within a regression framework, ensuring that all portfolio returns are assigned to identifiable investment decisions in a statistically efficient and internally consistent way.

Standard factor attribution separates portfolio returns into factor components and unidentified residuals. These residuals are an inevitable part of factor regressions but represent a nuisance for performance attribution. We reallocate these residuals to factor returns using a restricted generalized least squares (GLS) update that enforces completeness while minimizing statistical distortion. Because the restriction adds only one degree of constraint, the adjustments are typically small and fall on the least precisely estimated factor returns.

The same logic extends to allocation-selection attribution, where the interaction term plays a role analogous to residuals in factor models. In Brinson, Hood, and Beebower (1986), portfolio returns are decomposed into allocation across groups, selection within groups, and an interaction term measuring how well selection worked in overweighted groups. This interaction is usually treated as a nuisance and reallocated to allocation or selection components through ad hoc rules.¹ Although Brinson-style attribution is commonly treated as a calculation, we reinterpret this framework statistically, allowing interaction effects to be redistributed across allocation and selection components according to their relative estimation precision, in a way that minimizes the statistical distortion to both.

We also apply the same perspective to multi-period attribution. Per-period restricted regressions produce attribution tables that add up exactly to realized period returns, but the arithmetic aggregation of these rows does not in general match the compounded portfolio return over a longer horizon.

¹Spaulding (2008) describes some common static reallocation rules and proposes a set of conditional reallocation rules.

The linking and smoothing literature resolves this mismatch by selecting per-period weights that reconcile arithmetic and geometric returns. We show how to select such weights using the covariance information implied by the underlying regressions, choosing the least-distorting reconciliation subject to the compounding constraint. This yields a precision-aware alternative to standard linking and smoothing rules and extends naturally to entry-specific reconciliation when attribution precision varies across components as well as across time.

Practical guides such as Bacon (2004) have made these attribution approaches mainstream, but they often leave residual or interaction terms or rely on heuristic reconciliation rules that can obscure the manager's true contribution. We unify these settings within a single regression-based framework that provides a principled statistical foundation for imposing adding-up restrictions and reallocating nuisance components.

Section 2 introduces the restricted-regression method for reallocating residual return components in factor attribution. Section 3 derives the Brinson, Hood, and Beebower (1986) allocation-selection decomposition from benchmark- and portfolio-weighted regressions and applies the same restriction to eliminate interaction effects. Section 4 presents illustrative examples showing how the reallocation depends on relative estimation precision. Section 5 extends the framework to multi-period attribution by constructing a cumulative attribution row that matches the compounded portfolio return using precision-aware GLS reconciliation. Section 6 concludes. The appendix discusses how classical linking and smoothing rules arise as special cases under restrictive choices of the GLS penalty metric.

2 Factor-Based Attribution

Suppose we have N assets and write the N -dimensional vectors $\mathbf{r}$ for returns, $\mathbf{w}_P$ for portfolio weights, and $\mathbf{w}_B$ for benchmark weights. Let the $N \times K$ matrix $\mathbf{X}$ contain the assets' exposures to K factors, which may include both risk factors and alpha factors. (See Connor (1995) and Grinold (2006), for example.) In broader practice, such factor structures often trace back to empirical asset pricing models such as Fama and French (1993). Throughout, bold symbols denote vectors and matrices; scalars are not bold.

2.1 Classical decomposition

Factor attribution starts from the cross-sectional linear model for returns

$$\mathbf{r} = \mathbf{X}\mathbf{f} + \mathbf{u}, \tag{1}$$

where f are factor returns and u are idiosyncratic (security-specific) residuals. Asset returns, factor exposures, factor returns, and residual returns vary over time. For simplicity, we suppress time subscripts on all variables.

The portfolio's active return decomposes as

$$\begin{aligned} r_A &= w_P' r - w_B' r \\ &= (w_P - w_B)' X f + (w_P - w_B)' u \\ &= w_A' X f + w_A' u, \end{aligned} \quad (2)$$

where $w_A \equiv w_P - w_B$ are the active portfolio weights.

The first term is the factor-driven return contribution. The second term is the unexplained residual return contribution.

If the factors are generic risk factors, investors often refer to the residual contribution as “security selection.” While this may be technically correct, it obscures essentially all investment decisions made by the portfolio manager.

A clearer approach includes alpha factors in X and f that drive the main investment decisions. The return contributions from these alpha factors lay bare the success or failures of the individual investment decisions. In this case, however, interpreting the residual return contributions as “security selection” is no longer correct. The residual returns now capture deviations from the factors. Such deviations may arise because the factor model does not fully capture all of the factors or because the portfolio construction leads to deviations from the alpha factor portfolios.

Especially in systematic investment processes, the portfolio manager generally tries to minimize deviations from the alpha factor portfolios. As a result, we should expect the residual return contributions to be relatively small. Even when they are small, however, the residual return contributions can be a nuisance in performance reporting.

2.2 Regression formulation

We generally estimate factor returns via a cross-sectional regression, possibly with a weight matrix W_f , which yields the factor return estimates

$$\hat{f} = (X' W_f X)^{-1} X' W_f r. \quad (3)$$

Common choices for the weighting matrix are $W_f = I$ (for OLS), $W_f = \text{diag}(\text{market cap})$ (for market cap weighting), or $W_f = \text{diag}(\text{precision})$ (for inverse-variance weighting). Different W_f reflect different views about which securities should anchor the cross-sectional “factor plane.” Our qualitative results are not sensitive to reasonable choices of W_f .

2.3 Reallocating residuals

To avoid a residual line item at the portfolio level, we can estimate restricted factor returns $\hat{f}^*$ that satisfy

$$w'_A X \hat{f}^* = w'_A r. \quad (4)$$

This restriction implies

$$w'_A u = 0, \quad (5)$$

so that we reallocate what would otherwise appear as a residual return contribution to the factor return contributions.

Although Greene and Seaks (1991) argue that solving the associated Lagrangian system is numerically preferable we also present the analytical solution because it shows exactly how the restriction reallocates residuals across factors.

Analytical solution

In our case, we estimate the factor returns using GLS with weighting matrix W_f but impose a linear constraint using the unweighted, raw asset returns r . We focus on the raw returns since they are used in attribution.

Define

$$\Omega = (X' W_f X)^{-1}. \quad (6)$$

to be the usual covariance of the estimated factor returns.

Theil (1971) derives the restricted GLS estimate of the factor returns, which we can write as a single-constraint update of the unconstrained factor return estimates $\hat{f}$,

$$\hat{f}^* = \hat{f} + \Omega X' w_A (w'_A X \Omega X' w_A)^{-1} w'_A (r - X \hat{f}). \quad (7)$$

This is a textbook result for GLS estimates subject to linear restrictions on the coefficients.²

This solution adjusts the unconstrained factor returns $\hat{f}$ just enough to satisfy the restriction $w'_A (r - X \hat{f}^*) = 0$, so that the portfolio-weighted

² See, for example, equation (8.5) in Theil (1971). Let the restriction be $R\beta = q$ with $R \equiv w'_A X$ and $q \equiv w'_A r$. Theil's formula is $\hat{\beta}^* = \hat{\beta} + \Sigma_\beta R' (R \Sigma_\beta R')^{-1} (q - R \hat{\beta})$. After substituting $\hat{\beta} = \hat{f}$ and $\Sigma_\beta = \Omega$, we find $\hat{f}^* = \hat{f} + \Omega X' w_A (w'_A X \Omega X' w_A)^{-1} w'_A (r - X \hat{f})$. Alternatively, the result follows from the more common textbook form of the restricted OLS estimator after replacing the covariance of the unrestricted estimates, $(X'X)^{-1}$, with the GLS covariance, Ω . See Greene and Seaks (1991), for example.

residuals vanish and the active portfolio return is fully explained by the factor contributions,

$$r_A = w'_A X \hat{f}^*. \quad (8)$$

The overall adjustment is proportional to Ω , so the restricted estimate makes larger changes to those factor returns that are estimated with greater uncertainty. Sneddon (2021) highlights that factor attribution commonly contains estimation noise. Here, we allocate the residual return contributions to the factor return contributions in a manner that takes the best advantage of the flexibility offered by estimation noise.

The adjustment term $\Omega X' w_A (w'_A X \Omega X' w_A)^{-1}$ can be interpreted as the regression coefficient of the factor vector on the fitted active return. Since $\Omega X' w_A = \text{Cov}(\hat{f}, \hat{r}_A)$ and $w'_A X \Omega X' w_A = \text{Var}(\hat{r}_A)$, the update $\hat{f}^* - \hat{f} = \beta_{f|\hat{r}_A} w'_A (r - X \hat{f})$ adjusts factor estimates in proportion to their beta with the fitted active return and in proportion to the distance from the constraint. Factors that are more relevant, those with higher beta to the active fitted return, absorb a larger share of the adjustment, and factors with larger estimation variance move farther because such shifts incur a smaller statistical cost.

In statistical terms, the restriction redistributes the unexplained returns across factors in the least-distorting way, satisfying the active-exactness constraint with minimal loss of likelihood.

Discussion

The restriction ensures that the portfolio has no residual return contributions. This can be extremely convenient when the factors include the primary alpha factors that drive the portfolio weights. Although the portfolio most likely incurred some residual returns, these returns are a reporting nuisance, even when they represent a small part of the total returns.

Investors often use ad hoc methods to distribute the residual returns to the factor returns in order to avoid itemizing them.

The restricted regression statistically reallocates the residual returns to the factor returns in a way that has the smallest impact on the overall fit of the factor regression. When the estimation universe contains a substantial number of securities N , this single linear restriction generally has minimal impact on the regression fit. The reallocation mostly changes the factor returns that are estimated with the least precision.

It is important, however, that the portfolio manager remains aware of the residual returns. Especially when residual returns are material, they may warrant the attention of the portfolio manager. For internal analysis,

it probably is wise to consider both the unrestricted attribution and the restricted attribution.

Next, we apply the same logic to Brinson-style attribution: We interpret the attribution in regression terms and impose a single linear restriction in order to remove the interaction term by the smallest statistically justified adjustment.

3 Brinson Attribution

Brinson and Fachler (1985) and Brinson, Hood, and Beebower (1986) describe attribution that partitions active returns by groups (e.g., sectors). Each asset belongs to a group $g(i) \in 1, \dots, G$, and the attribution divides portfolio return contributions into allocation decisions across groups, selection decisions within groups, and an interaction term that ensures that the decomposition exactly matches the portfolio's active return.

3.1 Classical decomposition

The classical Brinson decomposition of the active return r_A is

$$\begin{aligned}
 r_A = & \sum_g (w_{P,g} - w_{B,g})(\bar{r}_{B,g} - r_B) + \\
 & \sum_g w_{B,g}(\bar{r}_{P,g} - \bar{r}_{B,g}) + \\
 & \sum_g (w_{P,g} - w_{B,g})(\bar{r}_{P,g} - \bar{r}_{B,g}).
 \end{aligned} \tag{9}$$

Here, $w_{P,g}$ and $w_{B,g}$ are the portfolio and benchmark exposures to group g , $\bar{r}_{P,g}$ and $\bar{r}_{B,g}$ are the corresponding within-group returns, and r_B is the total benchmark return.

The three terms represent allocation, selection, and interaction effects, respectively. The interaction term has a natural covariance interpretation: It is positive when the portfolio has positive allocation effects in groups where it also has positive selection effects. Nonetheless, the interaction term is generally treated as a nuisance: necessary for arithmetic completeness but not easily interpreted on its own. As a result, investors commonly reallocate it to the allocation or selection components using ad hoc rules.

3.2 Regression formulation

The attribution in equation (9) is usually interpreted as a mechanical arithmetic identity. But this view obscures the structure that makes the Brinson results consistent with standard regression-based attributions. We can express

the same relationships in regression form without changing any of the components. Doing so clarifies the estimation precision of the allocation and selection terms and provides a statistical foundation for a principled reallocation of the interaction term. This may seem like a conceptual leap from the Brinson arithmetic, but it relies only on standard results from weighted least squares regression.

We can view the within-group averages $\bar{r}_{p,g}$ and $\bar{r}_{B,g}$ as estimates of group mean returns obtained from two weighted least squares (WLS) regressions. Let D be an $(N \times G)$ dummy matrix indicating each asset's group membership. The benchmark- and portfolio-weighted regressions are

$$r = D\mu_B + \varepsilon_B, \quad \hat{\mu}_B = (D'W_B D)^{-1} D'W_B r, \quad (10)$$

and

$$r = D\mu_P + \varepsilon_P, \quad \hat{\mu}_P = (D'W_P D)^{-1} D'W_P r, \quad (11)$$

with $W_B = \text{diag}(w_B)$ and $W_P = \text{diag}(w_P)$. Each regression simply computes the weighted average return in each group. This is exactly what the classical Brinson arithmetic does.

These regressions correspond to the unique factor models that reproduce Brinson-style attribution exactly. Each model contains only the group-dummy factors, and the weighting matrices must be the benchmark and portfolio weights, W_B and W_P , respectively. Using any other factors or weights would produce a different decomposition and therefore depart from the classical Brinson results. In contrast, elsewhere in this paper our generalized GLS framework allows flexible factor sets and weighting matrices. Here, however, the dummy-only specification with these specific weights is the only one consistent with the standard Brinson components.

Thus, the two mean estimates $\hat{\mu}_{B,g} = (\sum_{i \in g} w_{B,i} r_i) / (\sum_{i \in g} w_{B,i}) = \bar{r}_{B,g}$ and $\hat{\mu}_{P,g} = (\sum_{i \in g} w_{P,i} r_i) / (\sum_{i \in g} w_{P,i}) = \bar{r}_{P,g}$ are the familiar group average returns for the benchmark and portfolio, respectively. The fitted values

$$\hat{r}_B = D\hat{\mu}_B, \quad \hat{r}_P = D\hat{\mu}_P, \quad (12)$$

reproduce the benchmark's and portfolio's group-level fitted returns. We then have $\hat{r}_{B,g} = \bar{r}_{B,g}$ and $\hat{r}_{P,g} = \bar{r}_{P,g}$. To emphasize the regression source of these values, we will use the "hat" notation for these fitted returns in the remainder.

The difference between the fitted returns, $\hat{r}_P - \hat{r}_B = D(\hat{\mu}_P - \hat{\mu}_B)$, represents

the change in estimated group means when moving from benchmark to portfolio weighting.

Expressed in this notation, the Brinson decomposition is

$$\begin{aligned} r_A = & \mathbf{w}'_A (\widehat{\mathbf{r}}_B - r_B \mathbf{1}) + \\ & \mathbf{w}'_B (\widehat{\mathbf{r}}_P - \widehat{\mathbf{r}}_B) + \\ & \mathbf{w}'_A (\widehat{\mathbf{r}}_P - \widehat{\mathbf{r}}_B), \end{aligned} \quad (13)$$

where $\mathbf{1}$ is an N -vector of ones and $\mathbf{w}_A = \mathbf{w}_P - \mathbf{w}_B$. The three terms correspond exactly to allocation, selection, and interaction effects, as in the arithmetic decomposition of equation (9). The regression representation simply provides a statistical framework for reasoning about their sampling variability, covariance, and optimal reallocation.

3.3 Reallocating interaction effects

Having cast Brinson attribution as two parallel WLS regressions, we now apply the same single active-exactness restriction used for factors (section 2.3) to the stacked group-mean estimates. Imposing the linear constraint that the difference between the estimates is zero eliminates the interaction term and, via restricted GLS, adjusts $\widehat{\boldsymbol{\mu}}_B$ and $\widehat{\boldsymbol{\mu}}_P$ by the smallest GLS distance.

Rather than ad hoc splits, the restricted-GLS update redistributes the interaction across allocation and selection according to their relative imprecision encoded in the covariance matrix of the attribution components: precise components move little, noisier ones move more. Equivalently, the share absorbed by each side is proportional to its covariance (beta) with the fitted active return (see section 3.4). When the two regressions have similar precision, the split is near 50/50; if the portfolio-weighted regression is less precise (e.g., because of greater portfolio concentration), more of the interaction is assigned to allocation, and conversely when it is more precise.

3.4 Active-exactness as a restriction on both regressions

Our goal is to reallocate the interaction term by imposing a single linear restriction that makes the active return equal the sum of allocation and selection effects. We seek adjusted group means $\widehat{\boldsymbol{\mu}}_B^*$ and $\widehat{\boldsymbol{\mu}}_P^*$ such that, using the active portfolio weights $\mathbf{w}_A$, the portfolio-weighted difference in group means is zero

$$\mathbf{w}'_A D(\widehat{\boldsymbol{\mu}}_P^* - \widehat{\boldsymbol{\mu}}_B^*) = 0. \quad (14)$$

This “active-exactness” restriction ensures that the total active return r_A is fully explained by the adjusted allocation and selection effects, with no residual or interaction term.

To express this compactly, define the stacked vector of group-mean estimates

$$\widehat{\mathbf{m}} \equiv \begin{bmatrix} \widehat{\boldsymbol{\mu}}_B \\ \widehat{\boldsymbol{\mu}}_P \end{bmatrix}, \quad (15)$$

and let $\mathbf{H}$ extract the difference between the two halves,

$$\mathbf{H} \equiv \begin{bmatrix} -\mathbf{I}_G & \mathbf{I}_G \end{bmatrix}, \quad \text{so that} \quad \mathbf{H}\widehat{\mathbf{m}} = \widehat{\boldsymbol{\mu}}_P - \widehat{\boldsymbol{\mu}}_B. \quad (16)$$

Then the active-exactness condition in equation (14) can be written as a single linear restriction on the stacked means

$$\mathbf{w}'_A \mathbf{D} \mathbf{H} \widehat{\mathbf{m}} = 0. \quad (17)$$

Among all adjusted means $\mathbf{m}^*$ that satisfy this restriction, we want the one that is statistically most plausible, the one closest to the original estimates in the GLS sense. This leads to the restricted GLS problem

$$\min_{\mathbf{m}} \frac{1}{2} (\mathbf{m} - \widehat{\mathbf{m}})' \boldsymbol{\Gamma}^{-1} (\mathbf{m} - \widehat{\mathbf{m}}) \quad \text{s.t.} \quad \mathbf{w}'_A \mathbf{D} \mathbf{H} \mathbf{m} = 0, \quad (18)$$

where $\boldsymbol{\Gamma}$ is the joint covariance of the stacked estimates,

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Sigma}_B & \boldsymbol{\Sigma}_{B,P} \\ \boldsymbol{\Sigma}_{P,B} & \boldsymbol{\Sigma}_P \end{bmatrix}. \quad (19)$$

This is a standard GLS problem with a single linear restriction. As for the restricted factor returns, we can follow Theil (1971) and write the restricted estimate of the group means as

$$\widehat{\mathbf{m}}^* = \widehat{\mathbf{m}} + \boldsymbol{\Gamma} \mathbf{H}' \mathbf{D}' \mathbf{w}_A (\mathbf{w}'_A \mathbf{D} \mathbf{H} \boldsymbol{\Gamma} \mathbf{H}' \mathbf{D}' \mathbf{w}_A)^{-1} (0 - \mathbf{w}'_A \mathbf{D} \mathbf{H} \widehat{\mathbf{m}}). \quad (20)$$

This makes small adjustments to the group means in $\widehat{\boldsymbol{\mu}}_B$ and $\widehat{\boldsymbol{\mu}}_P$ in order to just satisfy the constraint that the interaction term in the attribution is 0.

The size of the adjustment depends on two components of this expression: the covariance matrix $\boldsymbol{\Gamma}$, which captures the statistical uncertainty of the group-mean estimates, and the deviation from the restriction, $(0 - \mathbf{w}'_A \mathbf{D} \mathbf{H} \widehat{\mathbf{m}})$, which measures how far the unrestricted estimates violate active-exactness. The

covariance governs the relative size of the adjustments. Group means that are more uncertain or more inconsistent with the restriction receive proportionally larger adjustments, while precise or already-consistent estimates remain largely unchanged. The deviation from the restriction governs the total size of the adjustment.

Unstacking $\widehat{m}^*$ into its benchmark and portfolio components yields the adjusted group means

$$\widehat{m}^* = \begin{bmatrix} \widehat{\mu}_B^* \\ \widehat{\mu}_P^* \end{bmatrix}, \quad \text{so that} \quad \widehat{r}_B^* = D\widehat{\mu}_B^*, \quad \widehat{r}_P^* = D\widehat{\mu}_P^*. \quad (21)$$

By construction, the restriction $w'_A D(\widehat{\mu}_P^* - \widehat{\mu}_B^*) = 0$ is exactly satisfied, so that the interaction component of the Brinson decomposition vanishes while both the benchmark- and portfolio-weighted regressions remain as close as possible (in GLS distance) to their unconstrained estimates.

With these adjusted fitted vectors, the complete attribution without an interaction term is

$$r_A = A^* + S^*, \quad A^* = w'_A (\widehat{r}_B^* - r_B \mathbf{1}), \quad S^* = w'_B (\widehat{r}_P^* - \widehat{r}_B^*). \quad (22)$$

The constraint ensures that the total active return equals the sum of allocation and selection effects, while the GLS weighting ensures that the adjustments to $\widehat{r}_B$ and $\widehat{r}_P$ are statistically optimal – larger for the noisier estimates, smaller for the precise ones.

Precision inputs for the stacked restriction

When a reliable full return covariance Σ is available, for example from a risk model, we obtain the regression residual covariance under generic weights W from the wls hat matrix in raw return space,

$$P_W = D(D'WD)^{-1}D'W, \quad M_W = I - P_W, \quad \Delta = M_W \Sigma M'_W. \quad (23)$$

In practice we set $W = W_B$ for consistency with benchmark-relative reporting, but any reasonable choice (cap or precision weights) only affects precision weighting, not the active-exactness identity.³

³In principle, one could compute separate residual covariances $\Delta_B = M_{W_B} \Sigma M'_{W_B}$ and $\Delta_P = M_{W_P} \Sigma M'_{W_P}$ to reflect the distinct weighting structures of the benchmark- and portfolio-weighted regressions. In practice, however, using a single residual covariance Δ , typically based on W_B for consistency with benchmark-relative reporting, yields nearly identical results for benchmark-aware portfolios.

Because the estimates $\widehat{\mu}_B$ and $\widehat{\mu}_P$ are linear in r , we write

$$\widehat{\mu}_B = D' A_B r, \quad D' A_B \equiv (D' W_B D)^{-1} D' W_B, \quad (24)$$

$$\widehat{\mu}_P = D' A_P r, \quad D' A_P \equiv (D' W_P D)^{-1} D' W_P. \quad (25)$$

and obtain the needed covariances by propagation

$$\Sigma_B = D' A_B \Delta D' A'_B, \quad \Sigma_P = D' A_P \Delta D' A'_P, \quad \Sigma_{B,P} = D' A_B \Delta D' A'_P. \quad (26)$$

These blocks define Γ used in the restricted GLS update (20).

If Σ is unavailable, we can use a shrinkage estimator of the sample covariance, as suggested by Ledoit and Wolf (2004). As fallbacks, the diagonal sample variances $\Delta = \text{diag}(\widehat{\sigma}_1^2, \dots, \widehat{\sigma}_N^2)$ or even $\Delta = I$ are acceptable. These only affect the precision weights. The constraint guarantees exactness under any reasonable covariance matrix.

Process summary

In summary, we compute complete Brinson-style attribution with zero interaction terms as follows.

1. Run the two unrestricted WLS return regressions under benchmark and portfolio weights.
2. Obtain the return covariance $\widehat{\Sigma}$ and construct the residual covariance $\widehat{\Delta} = M_W \widehat{\Sigma} M'_W$.
3. Propagate $\widehat{\Delta}$ to the group-mean covariances Σ_B , Σ_P , and $\Sigma_{B,P}$, and assemble the joint covariance Γ .
4. Apply the restricted GLS update (20) to obtain the adjusted group means $(\widehat{\mu}_B^*, \widehat{\mu}_P^*)$ that satisfy $w'_A D(\widehat{\mu}_P^* - \widehat{\mu}_B^*) = 0$.
5. Compute the adjusted attribution components $A^* = w'_A(\widehat{r}_B^* - r_{Bt})$ and $S^* = w'_B(\widehat{r}_P^* - \widehat{r}_B^*) = S + I - (A^* - A)$, so that $r_A = A^* + S^*$ with no interaction term.

The procedure can be viewed as a two-step solution: first estimate the benchmark- and portfolio-weighted group means as in the standard Brinson decomposition, then adjust both by restricted GLS so that the active return identity holds exactly. The result preserves the familiar structure of Brinson attribution while reallocating the interaction term in a statistically optimal, least-distorting manner.

Next, we show that for similar portfolio and benchmark weights, the GLS weights approach equality and the interaction is reallocated about 50/50 across allocation and selection components. When the portfolio is substantially more concentrated (or more diversified) than the benchmark, the GLS

weighting tilts toward the more precise regression, reallocating more of the interaction toward the allocation (or selection) component, respectively.

4 Illustrative Examples

Consider $G = 2$ groups with benchmark weights $w_{B,1} = w_{B,2} = 0.5$ and portfolio weights $w_{P,1} = 0.6$, $w_{P,2} = 0.4$, so $w_{A,1} = 0.1$ and $w_{A,2} = -0.1$. Let the (unrestricted) group means be

$$\mu_{B,1} = 2.0\%, \quad \mu_{B,2} = -1.0\%, \quad \mu_{P,1} = 2.4\%, \quad \mu_{P,2} = -1.2\%.$$

Then

$$r_B = 0.5(2.0\%) + 0.5(-1.0\%) = 0.50\%,$$

$$r_P = 0.6(2.4\%) + 0.4(-1.2\%) = 0.96\%,$$

so $r_A = r_P - r_B = 0.46\%$. The classical Brinson components are

$$A = \sum_g (w_{P,g} - w_{B,g})(\mu_{B,g} - r_B) = 0.30\%,$$

$$S = \sum_g w_{B,g}(\mu_{P,g} - \mu_{B,g}) = 0.10\%,$$

$$I = \sum_g (w_{P,g} - w_{B,g})(\mu_{P,g} - \mu_{B,g}) = 0.06\%,$$

so $A + S + I = r_A$, as required.

4.1 Balanced precision

If the two regressions have balanced precision, Γ is approximately block-diagonal with equal blocks, $\Gamma = \text{diag}(\Sigma, \Sigma)$, and negligible cross-covariance. When both estimates are equally precise, the interaction effect should be split evenly between allocation and selection. We can verify that the restricted GLS update indeed produces this outcome.

In this case,

$$H \Gamma H' = 2\Sigma, \quad \Gamma H' = \begin{bmatrix} -\Sigma \\ \Sigma \end{bmatrix}$$

with $H = [-I_G \ I_G]$.

Write the interaction scalar I as the deviation from the constraint

$$I \equiv w_A' D(\hat{\mu}_P - \hat{\mu}_B) = w_A' D H \hat{m}.$$

We can plug these into the update equation

$$\widehat{m}^* = \widehat{m} - \Gamma H' D' w_A (w_A' D H \Gamma H' D' w_A)^{-1} w_A' D H \widehat{m}$$

and notice that the central term $w_A' D H \Gamma H' D' w_A = 2 w_A' D \Sigma D' w_A$ is a scalar. This gives the block adjustments

$$\begin{aligned} \Delta \mu_B &\equiv \widehat{\mu}_B^* - \widehat{\mu}_B = \frac{\Sigma D' w_A}{2 w_A' D \Sigma D' w_A} I, \\ \Delta \mu_P &\equiv \widehat{\mu}_P^* - \widehat{\mu}_P = -\frac{\Sigma D' w_A}{2 w_A' D \Sigma D' w_A} I. \end{aligned}$$

The term $\Sigma D' w_A (2 w_A' D \Sigma D' w_A)^{-1}$ plays the same role as the regression coefficient of the group-mean vector on the fitted active return. The numerator $\Sigma D' w_A$ is the covariance between the estimated group means and the active fitted return, and the denominator is twice its variance. Hence, the update to each block can be viewed as a regression-style adjustment: the group means move in proportion to their beta with the active fitted return and in proportion to the distance from the active-exact constraint. Group-mean components that are estimated with higher uncertainty (larger variance in Γ) or that covary more with the active return absorb a larger share of the adjustment, just as in a standard GLS update.

The updated allocation effect is $A^* = w_A' \widehat{r}_B^* = w_A' D \widehat{\mu}_B^*$, and the update to the allocation component is

$$\Delta A = w_A' D \Delta \mu_B = \frac{w_A' D \Sigma D' w_A}{2 w_A' D \Sigma D' w_A} I = \frac{1}{2} I.$$

By exactness, after the update $A^* + S^* = A + S + I$, so

$$\Delta S = I - \Delta A = \frac{1}{2} I.$$

Therefore, under balanced precision the restricted GLS update splits the interaction approximately 50/50,

$$A^* = A + \frac{1}{2} I = 0.33\%, \quad S^* = S + \frac{1}{2} I = 0.13\%,$$

with $I^* = 0$ by construction. This produces complete attribution with $A^* + S^* = r_A = 0.46\%$.⁴

⁴If the portfolio and benchmark employ different leverage, $w_A' \iota \neq 0$, the updated allocation effect includes a funding/leverage adjustment, $\Delta A = \frac{1}{2} I - (w_A' \iota) \Delta r_B$, where

If the benchmark and portfolio mean estimates have unequal precision or non-negligible cross-covariance, the same algebra yields unequal shares that are determined by the blocks of Γ (see section 3.4).

4.2 Spherical specific risk and optimal interaction splits

For cases where the reallocation may not be equal, we now consider cases where regression precision is determined mainly by the dispersion of portfolio weights. This occurs when returns have equal residual risk but portfolios differ in concentration. In this setting, differences in diversification alone determine the relative precision of the benchmark- and portfolio-weighted regressions, which in turn determines the optimal reallocation of the interaction term. The spherical risk case extends the balanced-precision logic by allowing the benchmark and portfolio regressions to differ in effective sampling precision because of unequal diversification.

Under spherical residual risk, $\Sigma = \sigma^2 I_N$, each group-level regression has sampling variance that depends only on the dispersion of its weights. The corresponding benchmark- and portfolio-weighted group mean estimates have block-diagonal covariance

$$\Gamma = \begin{bmatrix} \Sigma_B & \mathbf{0} \\ \mathbf{0} & \Sigma_P \end{bmatrix},$$

with $\Sigma_B = \sigma^2 \text{diag}(s_{B,1}, \dots, s_{B,G})$ and $\Sigma_P = \sigma^2 \text{diag}(s_{P,1}, \dots, s_{P,G})$, where $s_{B,g} \propto \sum_{i \in g} w_{B,i}^2$ and $s_{P,g} \propto \sum_{i \in g} w_{P,i}^2$ measure the effective concentration of the benchmark and portfolio within group g .

Starting from the restricted GLS update,

$$\hat{m}^* = \hat{m} - \Gamma H' D' w_A (w_A' D H \Gamma H' D' w_A)^{-1} w_A' D H \hat{m},$$

with $H = [-I_G \ I_G]$ and $\hat{m}' = [\hat{\mu}_B', \hat{\mu}_P']$, as before. Define the scalar interaction as the deviation from the restriction, $I = w_A' D (\hat{\mu}_P - \hat{\mu}_B)$. Using $\Gamma H' = [-\Sigma_B; \Sigma_P]$ and $H \Gamma H' = \Sigma_B + \Sigma_P$, we obtain the block updates

$$\begin{aligned} \Delta \mu_B &= \frac{\Sigma_B D' w_A}{w_A' D (\Sigma_B + \Sigma_P) D' w_A} I, \\ \Delta \mu_P &= -\frac{\Sigma_P D' w_A}{w_A' D (\Sigma_B + \Sigma_P) D' w_A} I. \end{aligned}$$

$\Delta r_B = (w_B' D \Sigma D' w_A) / (2 w_A' D \Sigma D' w_A) I$. Many reports show active attribution on a leverage-matched basis, in which case this term is zero by construction.

The corresponding adjustments to the allocation and selection effects are

$$A^* = A + \lambda_A I, \quad S^* = S + \lambda_S I, \quad \lambda_A + \lambda_S = 1,$$

where the optimal reallocation weights follow directly from the formulas above as

$$\lambda_A = \frac{\mathbf{w}'_A D \Sigma_B D' \mathbf{w}_A}{\mathbf{w}'_A D (\Sigma_B + \Sigma_P) D' \mathbf{w}_A}, \quad \lambda_S = \frac{\mathbf{w}'_A D \Sigma_P D' \mathbf{w}_A}{\mathbf{w}'_A D (\Sigma_B + \Sigma_P) D' \mathbf{w}_A}.$$

If the portfolio has uniform relative concentration $s_{P,g} = k s_{B,g}$ across groups, the ratio

$$k \equiv \frac{\mathbf{w}'_P \mathbf{w}_P}{\mathbf{w}'_B \mathbf{w}_B}$$

fully characterizes the relative sampling precision of the two regressions.⁵ The weight expressions simplify to

$$\lambda_S = \frac{1}{1+k}, \quad \lambda_A = \frac{k}{1+k}.$$

Under spherical return risk, the concentration ratio k measures the relative sampling variance of the portfolio- and benchmark-weighted regressions. When $k > 1$, the portfolio is more concentrated and its group means are estimated less precisely; when $k < 1$, the portfolio is more diversified and its group means are estimated more precisely. Because the selection effect depends on the difference between the two regressions, while the allocation effect depends only on the benchmark, changes in portfolio concentration affect the noise in selection more strongly than in allocation. As concentration rises, the GLS update assigns a larger share of the interaction to allocation; as concentration falls, it assigns a larger share to selection.

?? shows this relationship for several concentration levels. When $k = 1$ (equal diversification), the split is 50/50. Even moderate concentration tilts the optimal redistribution substantially: if the portfolio holds only half the benchmark's names per sector, the interaction shifts from a 50/50 split to roughly one-third selection and two-thirds allocation. This statistical asymmetry grows with relative portfolio concentration, reflecting that more

⁵ The Brinson regressions must use the portfolio's actual (fully levered) weights to reproduce the correct group-wise means. The relative concentration index k appropriately reflects leverage: higher gross exposure increases the sampling variance of the portfolio-weighted regression and thus tilts the optimal reallocation.

concentrated portfolios provide less precise within-group evidence and therefore merit smaller selection adjustments. However, it requires extreme differences in portfolio concentration to make a 100% reallocation of the interaction effect to either allocation or selection statistically optimal.

To illustrate numerically, suppose the portfolio is approximately twice as concentrated as the benchmark ($k = 2$). The GLS weights are then $\lambda_S = 1/3$ and $\lambda_A = 2/3$, implying

$$\begin{aligned} A^* &= A + \frac{2}{3}I = 0.34\%, \\ S^* &= S + \frac{1}{3}I = 0.12\%. \end{aligned}$$

The adjusted effects again sum to $r_A = 0.46\%$, but the reallocation now places greater emphasis on the allocation component. Such asymmetry arises automatically from the relative statistical precision of the two regressions, without any arbitrary manual adjustment. The same logic applies symmetrically when the portfolio is more diversified than the benchmark, in which case the selection component absorbs a larger share of the interaction.

When concentration ratios differ across groups, the restricted GLS update reallocates the interaction term heterogeneously across them. Now, each group g has its own effective precision ratio

$$k_g = \frac{s_{P,g}}{s_{B,g}} = \frac{\sum_{i \in g} w_{P,i}^2}{\sum_{i \in g} w_{B,i}^2},$$

so that

$$\lambda_{S,g} = \frac{1}{1 + k_g}, \quad \lambda_{A,g} = \frac{k_g}{1 + k_g}.$$

Groups with more concentrated portfolio weights ($k_g > 1$) receive a larger reallocation toward allocation and a smaller one toward selection, while more diversified groups ($k_g < 1$) receive the opposite treatment. If we define $\mathbf{a}_g = D' \mathbf{w}_{A,g}$, the overall portfolio-level split is

$$\lambda_S = \frac{\sum_g \mathbf{a}_g' \Sigma_{P,g} \mathbf{a}_g}{\sum_g \mathbf{a}_g' (\Sigma_{B,g} + \Sigma_{P,g}) \mathbf{a}_g}, \quad \lambda_A = 1 - \lambda_S,$$

which is a weighted average of these group-level shares, determined by each group's contribution to active risk. This formulation highlights that the GLS adjustment acts locally within groups, with each group's share of the interaction depending on its own relative precision. Note, however, that

the residual interaction effects generally remain nonzero within each group. Elimination of interaction terms for all groups requires active-exactness constraints for each group rather than the single portfolio-wide restriction we use here. Appendix A discuss the group-specific constraints.

5 Multi-Period Attribution

The procedures developed in the previous sections produce a *per-period attribution table* from cross-sectional regressions. For each period $t = 1, \dots, T$ and each attribution component $k = 1, \dots, K$, we obtain an estimate $c_{k,t}$ such that

$$\sum_{k=1}^K c_{k,t} = r_{A,t}, \quad (27)$$

where $r_{A,t}$ is the realized *active* portfolio return in period t . By construction, each row of the table adds up exactly to the corresponding period's active return. This per-period exactness follows directly from the restricted regression framework and plays a central role in the interpretation of the attribution results.

For multi-period reporting, investors typically augment this table with a final row of cumulative attribution totals. Let

$$R_A \equiv \prod_{t=1}^T (1 + r_{P,t}) - \prod_{t=1}^T (1 + r_{B,t}) \quad (28)$$

denote the compounded active return over the full horizon, where $r_{P,t}$ and $r_{B,t}$ are the portfolio and benchmark returns in period t , respectively. We seek cumulative attribution totals $C = (C_1, \dots, C_K)$ that satisfy

$$\sum_{k=1}^K C_k = R_A. \quad (29)$$

In general, the simple column sums $\sum_t c_{k,t}$ do not satisfy this identity, because $\sum_t r_{A,t} \neq R_A$ whenever compounding effects are material.

5.1 The weighting problem

A common approach constructs cumulative attribution totals as weighted sums of the per-period entries. If we restrict attention to a single vector of per-period weights $w = (w_1, \dots, w_T)$ applied uniformly across columns, we

write

$$C_k = \sum_{t=1}^T w_t c_{k,t}. \quad (30)$$

Requiring the cumulative row to add up to the compounded active return imposes the single linear constraint

$$\mathbf{r}'_A \mathbf{w} = R_A, \quad (31)$$

where $\mathbf{r}_A = (r_{A,1}, \dots, r_{A,T})'$.

Equation (31) places one restriction on T unknown weights. As a result, an infinite set of weight vectors satisfies the compounding identity. Most of these solutions assign extreme or counterintuitive weights to individual periods and lack economic or statistical appeal. Any practical multi-period attribution method must therefore supply an additional criterion that selects a particular solution from this admissible set.

The mismatch between arithmetic aggregation and compounded returns has long been recognized in the performance attribution literature. Linking and smoothing methods, such as those of Cariño (1999), Menchero (2000), and Frongello (2002), resolve this under-identification by selecting particular weight vectors based on economic or regularity heuristics. Although these approaches differ in motivation, they share a common structure: each chooses one admissible solution without explicitly appealing to the statistical precision of the underlying attribution estimates.

5.2 A precision-weighted GLS choice of period weights

Because the attribution entries $c_{k,t}$ arise from regressions, we interpret them as noisy estimates rather than exact quantities. The cross-sectional analysis developed earlier provides not only point estimates but also information about their sampling precision. We use this information to select reconciliation weights that minimize statistically meaningful distortion subject to the compounding constraint.

Let $\bar{\mathbf{w}}$ denote a baseline vector of per-period weights. In the absence of strong prior preferences, a natural choice is $\bar{\mathbf{w}} = \mathbf{1}$, the vector of ones corresponding to simple summation over time. We can accommodate alternative baselines, such as Carino, Menchero, or standard linking weights, without changing the analysis.

We choose $\mathbf{w}$ by solving the restricted generalized least squares problem

$$\min_{\mathbf{w}} \frac{1}{2}(\mathbf{w} - \bar{\mathbf{w}})' \mathbf{V}^{-1}(\mathbf{w} - \bar{\mathbf{w}}) \quad \text{s.t.} \quad \mathbf{r}'_A \mathbf{w} = R_A, \quad (32)$$

where $\mathbf{V}$ is a positive-definite matrix defining the penalty metric. The solution has the closed form

$$\mathbf{w}^* = \bar{\mathbf{w}} + \mathbf{V} \mathbf{r}_A (\mathbf{r}'_A \mathbf{V} \mathbf{r}_A)^{-1} (R_A - \mathbf{r}'_A \bar{\mathbf{w}}). \quad (33)$$

This choice selects the weight vector closest to the baseline in a precision-weighted sense while exactly satisfying the compounding identity. When $\mathbf{V}$ is proportional to the identity matrix and $\bar{\mathbf{w}} = \mathbf{1}$, the solution reduces to a minimum-variation rule closely related to the smoothing approaches of Carino and Menchero. More generally, the GLS adjustment respects the chosen baseline to the extent that it is consistent with the covariance structure implied by the attribution regressions, and departs from it only when precision considerations warrant.

5.3 Construction of the covariance matrix

To implement period-only reconciliation, we must specify the covariance matrix $\mathbf{V}$ governing deviations of the period weights from the baseline. Here, $\mathbf{V}$ is a $T \times T$ matrix defined over periods, not over attribution components.

For each period t , the vector $\mathbf{c}_t = (c_{1,t}, \dots, c_{K,t})$ is a linear function of the estimated regression coefficients. Using standard covariance propagation, we compute

$$\mathbf{\Omega}_t = \text{Cov}(\mathbf{c}_t). \quad (34)$$

To map component-level uncertainty to the period level, let $\mathbf{a} = (1, \dots, 1)'$ denote a K -vector of ones so that $\mathbf{a}' \mathbf{c}_t = r_{A,t}$. A simple and effective diagonal choice for $\mathbf{V}$ is then

$$\mathbf{V} = \text{diag}(v_1, \dots, v_T), \quad v_t = \mathbf{a}' \mathbf{\Omega}_t \mathbf{a}, \quad (35)$$

so that periods whose total active attribution is estimated with greater uncertainty absorb a larger share of the compounding adjustment. More general constructions can incorporate cross-period covariance when such information is available; the algebra of the GLS solution remains unchanged.

5.4 Entry-specific reconciliation

Restricting attention to a single vector of period weights implicitly assumes that all attribution components within a period share the same reliability. In practice, this assumption often fails. Some components may exhibit substantially greater estimation uncertainty than others, either persistently or episodically.

We therefore consider a fully general reconciliation that allows adjustments to differ across both periods and components. Let $c = vec(c_{k,t})$ denote the stacked vector of all attribution entries and let $\Sigma = Cov(c)$ be its covariance matrix, constructed by stacking the period-specific covariances Ω_t (and, when appropriate, cross-period covariances). We choose adjusted entries c^* by solving

$$\min_{c^*} \frac{1}{2}(c^* - c)' \Sigma^{-1}(c^* - c) \quad \text{s.t.} \quad \iota' c^* = R_A, \quad (36)$$

where ι is a conformable vector of ones. The solution is

$$c^* = c + \Sigma \iota (\iota' \Sigma \iota)^{-1} (R_A - \iota' c). \quad (37)$$

This reconciliation allocates the compounding adjustment toward the noisiest entries in the attribution table, leaving more precisely estimated components largely unchanged. Period-only weighting and classical smoothing methods arise as special cases under restrictive covariance assumptions. The general solution exploits the full precision information generated by the regression framework and delivers a statistically principled, least-distorting cumulative attribution that adds exactly to the compounded active return.

6 Summary

This paper expresses both factor attribution and Brinson-style allocation-selection attribution in common regression-based representation. In each case, the familiar attribution components arise from linear regressions, while residual or interaction terms appear because the attribution problem is under-identified. By imposing a small number of economically meaningful linear restrictions, we eliminate these nuisance terms and obtain attribution results that are exact, complete, and interpretable. Restricted generalized least squares provides a natural mechanism for implementing these restrictions while minimizing statistically meaningful distortion.

The resulting framework unifies several attribution problems that are usually treated separately. In factor attribution, the restriction reallocates

residual returns across factors. In Brinson attribution, the same logic real-locates interaction effects across allocation and selection components. In both settings, the adjustment depends on estimation precision: components estimated with greater uncertainty absorb larger adjustments, while precisely estimated components remain largely unchanged. This statistical structure clarifies when common heuristics are appropriate and when they are not. For example, when portfolio and benchmark regressions have similar precision, splitting the interaction term approximately evenly between allocation and selection is near optimal. When portfolio weights are substantially more concentrated than benchmark weights, the reallocation tilts toward allocation, reflecting greater uncertainty in within-group selection estimates. When the portfolio is more diversified or departs from the benchmark universe, the opposite tilt arises naturally.

We extend the same perspective to multi-period attribution. Per-period attribution tables constructed from cross-sectional regressions add up exactly to realized period returns but do not, in general, aggregate to the compounded portfolio return over time. This mismatch reflects the same under-identification problem that arises in cross-sectional attribution. Existing linking and smoothing methods resolve this problem by selecting particular weighting schemes based on economic or regularity heuristics. In contrast, the framework developed here treats all attribution entries as noisy estimates and selects reconciliation weights by minimizing statistically meaningful distortion subject to the compounding constraint. This approach nests standard linking and smoothing methods as special cases under restrictive covariance assumptions and yields more general solutions when attribution precision varies across periods or components.

Taken together, these results show that residual reallocation, interaction reallocation, and multi-period smoothing are not distinct problems but manifestations of a single principle: exact attribution requires selecting one solution from an under-identified linear system, and statistical precision provides a coherent criterion for doing so. Restricted regression methods offer a unified and extensible foundation for portfolio return attribution across assets, groups, and time.

7 References

- Bacon, Carl R., 2004, *Practical Portfolio Performance Measurement and Attribution* (John Wiley & Sons, Chichester, England).
- Brinson, Gary P., and Nimrod Fachler, 1985, Measuring non-US equity portfolio performance, *Journal of Portfolio Management* 11 (3), 73–76.
- Brinson, Gary P., L. Randolph Hood, and Gilbert L. Beebower, 1986, Determinants of portfolio performance, *Financial Analysts Journal* 42 (4), 39–44.
- Cariño, David R., 1999, Combining attribution effects over time, *Journal of Performance Measurement* (Summer), 5–14.
- Connor, Gregory, 1995, The three types of factor models: A comparison of their explanatory power, *Financial Analysts Journal* 51 (3), 42–46.
- 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.
- Frongello, Andrew S.B., 2002, Linking single period attribution results, *Journal of Performance Measurement* (Spring), 10–22.
- Greene, William H., and Terry G. Seaks, 1991, The restricted least squares estimator: A pedagogical note, *Review of Economics and Statistics* 73 (3), 563–567.
- Grinold, Richard C., 2006, Attribution, *Journal of Portfolio Management* (Winter), 9–22.
- Ledoit, Olivier, and Michael Wolf, 2004, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* 88 (2), 365–411.
- Menchero, Jose G., 2000, An optimized approach to linking attribution effects over time, *Journal of Performance Measurement* (Fall), 36–42.
- Sneddon, Leigh, 2021, Hidden errors in regression-based attribution, *Journal of Performance Measurement* (Winter), 10–17.
- Spaulding, David, 2003, *Investment Performance Attribution* (McGraw-Hill, New York, NY).
- Spaulding, David, 2008, Should the interaction effect be allocated? A “black box” approach to interaction, *Journal of Performance Measurement* (Spring), 23–30.
- Theil, Henri, 1971, *Principles of Econometrics* (John Wiley & Sons, New York, NY).

A Group-Level Active-Exactness Restrictions

In the main text we imposed a single *active-exactness* restriction that jointly adjusts the benchmark- and portfolio-weighted regressions so that the portfolio-weighted active residuals vanish:

$$w_A' D(\hat{\mu}_P^* - \hat{\mu}_B^*) = 0. \quad (38)$$

This ensures that the portfolio's active return equals the sum of allocation and selection effects, with no residual or interaction term. In some reporting frameworks, investors also display allocation, selection, and interaction effects by sector. This appendix generalizes the single restriction to a set of group-level restrictions that eliminate interaction terms within each group.

A.1 Multiple active-exactness constraints

Let $g = 1, \dots, G$ index groups (e.g., sectors), and define the diagonal selector S_g that isolates assets belonging to group g . Each S_g is an $N \times N$ diagonal matrix with ones for assets in group g and zeros elsewhere, so that $S_g r$ retains only the returns of assets in g . This choice preserves consistent vector dimensions across all groups.

We continue to define $w_A = w_P - w_B$ as the vector of active portfolio weights and

$$\hat{m} \equiv \begin{bmatrix} \hat{\mu}_B \\ \hat{\mu}_P \end{bmatrix}, \quad H \equiv \begin{bmatrix} -I_G & I_G \end{bmatrix}, \quad (39)$$

so that $H\hat{m} = \hat{\mu}_P - \hat{\mu}_B$. For each group g , define $c_g = S_g w_A$ as the active-weight vector restricted to assets in that group, and let $C = [c_1, \dots, c_G]$ stack these vectors. The group-level active-exactness condition requires the portfolio-weighted residual in each group to vanish:

$$c_g' D H \hat{m}^* = 0, \quad g = 1, \dots, G. \quad (40)$$

Stacking these restrictions yields

$$C' D H \hat{m}^* = 0. \quad (41)$$

If a group has zero active weight ($w_{P,g} = w_{B,g}$), the corresponding column of C is zero and can be omitted. Summing all group constraints recovers the single portfolio-level restriction used in the main text.

A.2 Restricted GLS formulation

Let $\widehat{\mathbf{m}}$ denote the unrestricted stacked group-mean estimates obtained from the two separate regressions. To satisfy the G_c group-level active-exactness constraints $\mathbf{C}'\mathbf{D}\mathbf{H}\widehat{\mathbf{m}}^* = \mathbf{0}$ while minimally disturbing the unrestricted estimates, we solve the restricted GLS problem

$$\min_{\mathbf{m}} \frac{1}{2}(\mathbf{m} - \widehat{\mathbf{m}})' \mathbf{\Gamma}^{-1}(\mathbf{m} - \widehat{\mathbf{m}}) \quad \text{s.t.} \quad \mathbf{C}'\mathbf{D}\mathbf{H}\mathbf{m} = \mathbf{0}, \quad (42)$$

where $\mathbf{\Gamma} = \text{Var}(\widehat{\mathbf{m}})$ is the stacked covariance matrix of the unrestricted group-mean estimates, constructed exactly as in section 3.4 from the residual covariance $\Delta = \mathbf{M}_W \mathbf{\Sigma} \mathbf{M}_W'$.

Once again, the standard restricted GLS solution is

$$\widehat{\mathbf{m}}^* = \widehat{\mathbf{m}} - \mathbf{\Gamma}(\mathbf{C}'\mathbf{D}\mathbf{H})'[\mathbf{C}'\mathbf{D}\mathbf{H}\mathbf{\Gamma}(\mathbf{C}'\mathbf{D}\mathbf{H})']^{-1}\mathbf{C}'\mathbf{D}\mathbf{H}\widehat{\mathbf{m}}. \quad (43)$$

The updated stacked estimates $\widehat{\mathbf{m}}^* = (\widehat{\mu}_B^*, \widehat{\mu}_P^*)'$ jointly satisfy

$$\mathbf{C}'\mathbf{D}\mathbf{H}\widehat{\mathbf{m}}^* = \mathbf{0}, \quad (44)$$

so that every constrained group has zero interaction term. Because the portfolio total is the sum of all groups, the overall portfolio-level interaction term also vanishes.

A.3 Interpretation and practical remarks

This formulation generalizes the single portfolio-level active-exactness restriction in a natural way. The one-constraint system of the main text is recovered when $\mathbf{C} = \mathbf{c}$. Adding further constraints enforces zero interaction effects within each group while preserving the total active return $r_A = A + S$. Although these additional restrictions introduce slightly more distortion relative to the unrestricted estimates $\widehat{\mathbf{m}}$, the impact is typically small when $G \ll N$ and most groups contain many securities. The system (43) or its equivalent KKT form is symmetric indefinite and can be solved efficiently via an LDL' factorization, as recommended by Greene and Seaks (1991). This generalization is therefore both conceptually straightforward and practically useful for attribution systems that report sector-level effects.

B Long-Short and Market-Neutral Portfolios

Factor-based attribution extends naturally to long-short and market-neutral portfolios. Negative portfolio or benchmark weights cause no difficulties in computing factor exposures or attributing returns. For market-neutral

portfolios, we can use notional allocations to define portfolio weights and proceed exactly as before. Once the portfolio weights are well defined, factor-based return attribution, whether unconstrained or under the linear restriction introduced in the main text, remains valid. Using a cash benchmark for a market-neutral portfolio is economically sensible and poses no technical problems for factor-based attribution.

Brinson-style attribution, however, requires additional care when portfolio allocations are zero or when the portfolio is market-neutral. In this case, the natural cash benchmark has zero exposure to all risky assets and produces zero returns in each group, so allocation and selection effects for risky assets vanish by construction. To recover economically meaningful attributions, many practitioners split market-neutral portfolios into separate long and short partitions. Standard attribution is then performed on the long side, and again on the absolute value of the short positions, both relative to the same long-only benchmark; the results are then combined (long minus short). In this formulation, a long-only benchmark provides a natural reference for both sides.

This split-book approach can also be used within the restricted regression framework developed in the main text. Each sleeve (long and short) can be analyzed separately using the same active-exactness restriction, and the resulting effects recombined to yield a coherent, additive attribution for the overall market-neutral portfolio.

C Smoothing and Linking as GLS Selection Rules

This appendix clarifies the relationship between classical multi-period linking and smoothing rules and the precision-weighted generalized least squares (GLS) reconciliation developed in the main text. The purpose of the appendix is not to propose alternative estimators of sampling variance, but to show that commonly used smoothing rules can be interpreted as solutions to constrained quadratic programs under particular choices of penalty metrics. This perspective positions the GLS-based approach as a principled generalization rather than as a competing heuristic.

C.1 The under-identified weighting problem

Let r_t denote the realized portfolio return in period t and let

$$R \equiv \prod_{t=1}^T (1 + r_t) - 1 \quad (45)$$

be the compounded return over the full horizon. Consider cumulative attribution totals constructed using per-period weights $\mathbf{w} = (w_1, \dots, w_T)$,

$$C_k = \sum_{t=1}^T w_t c_{k,t}, \quad (46)$$

where the per-period attribution entries satisfy $\sum_k c_{k,t} = r_t$. Requiring the cumulative row to add up to the compounded return imposes the single linear constraint

$$\mathbf{r}'\mathbf{w} = R. \quad (47)$$

Equation (47) defines an affine subspace of dimension $T - 1$. Any practical linking or smoothing rule therefore amounts to selecting one element of this admissible set by imposing an additional optimality criterion.

C.2 GLS formulation

We write the generic GLS selection problem as

$$\min_{\mathbf{w}} \frac{1}{2}(\mathbf{w} - \boldsymbol{\iota})' \mathbf{V}^{-1}(\mathbf{w} - \boldsymbol{\iota}) \quad \text{s.t.} \quad \mathbf{r}'\mathbf{w} = R, \quad (48)$$

where $\boldsymbol{\iota}$ denotes the vector of unit weights and $\mathbf{V}$ is a positive-definite matrix that defines the penalty metric (not necessarily a literal sampling covariance). The solution is

$$\mathbf{w}^* = \boldsymbol{\iota} + \mathbf{V}\mathbf{r}(\mathbf{r}'\mathbf{V}\mathbf{r})^{-1}(R - \boldsymbol{\iota}'\mathbf{r}). \quad (49)$$

Thus, the shape of the selected weights depends entirely on the vector $\mathbf{V}\mathbf{r}$. Different choices of $\mathbf{V}$ correspond to different selection principles.

C.3 Recovering a given admissible weight vector

Let $\bar{\mathbf{w}}$ be any admissible weight vector satisfying $\mathbf{r}'\bar{\mathbf{w}} = R$, and define $\mathbf{u} \equiv \bar{\mathbf{w}} - \boldsymbol{\iota}$. Then $\mathbf{r}'\mathbf{u} = R - \boldsymbol{\iota}'\mathbf{r}$. If the penalty matrix $\mathbf{V}$ satisfies

$$\mathbf{V}\mathbf{r} = \alpha\mathbf{u} \quad \text{for some } \alpha > 0, \quad (50)$$

the GLS solution (49) coincides exactly with $\bar{\mathbf{w}}$. This observation shows that any admissible smoothing or linking rule can be represented as the solution of a GLS problem for a suitable choice of metric $\mathbf{V}$. The representation is not unique: many positive-definite matrices $\mathbf{V}$ satisfy (50).

A simple constructive example is

$$V_\varepsilon = \varepsilon I + \frac{uu'}{r'u'}, \quad \varepsilon > 0, \quad (51)$$

for which $V_\varepsilon r = \varepsilon r + u$. As $\varepsilon \rightarrow 0$, the GLS solution converges to $\bar{w}$. This construction emphasizes that V should be interpreted as a selection metric rather than as a literal sampling covariance.

C.4 Special cases

Homogeneous precision (Menchero-type smoothing)

If V is proportional to the identity matrix,

$$V \propto I, \quad (52)$$

then (49) yields

$$w^* = \iota + r \frac{R - \iota' r}{r' r}. \quad (53)$$

This solution minimizes the Euclidean distance from uniform weights subject to the compounding constraint and corresponds to a minimum-variation selection rule. Up to normalization conventions, this criterion aligns with the smoothing philosophy emphasized by Menchero, which treats all periods as equally reliable.

Smoothness-based penalties (Carino-type criteria)

If the penalty matrix takes the form

$$V^{-1} \propto L' L, \quad (54)$$

where L is a first-difference or higher-order difference operator, the GLS problem (48) selects weights that satisfy the compounding constraint while minimizing variation over time. This choice penalizes rapid changes in weights across adjacent periods and corresponds to a smoothness-based selection criterion similar in spirit to that advocated by Carino. Because $L' L$ is singular, practical implementations require a small ridge adjustment to ensure invertibility.

Standard linking weights

Standard linking rules assign per-period weights

$$w_t^{\text{link}} = \prod_{s=t+1}^T (1 + r_s), \quad (55)$$

which satisfy the admissibility condition $\mathbf{r}'\mathbf{w} = R$ by construction.

Within the GLS framework, linking corresponds to a penalty metric V such that

$$V\mathbf{r} \propto \mathbf{w}^{\text{link}} - \mathbf{1}. \quad (56)$$

This metric places relatively small penalties on deviations aligned with linking weights and larger penalties on orthogonal deviations.

Interpreted statistically, this choice treats early-period attribution entries as substantially more precise than later ones. Such an ordering is exact for buy-and-hold portfolios without rebalancing, where early returns permanently scale the capital base. For rebalanced or actively traded portfolios, however, this implicit precision structure does not arise naturally from the cross-sectional regressions used to construct per-period attribution.

From the GLS perspective, standard linking therefore represents a strong prior on the relative reliability of periods rather than a covariance structure implied by the data. When the regression-based covariance resembles this structure, the GLS solution approximates linking; when it does not, the solution deviates in a direction justified by statistical precision.

C.5 Relation to entry-specific reconciliation

The period-weight formulation above restricts attention to a single vector of weights applied uniformly across attribution components. The main text considers a more general reconciliation that allows adjustments to differ across both periods and components. In that setting, the role of V is played by the full covariance matrix $\Sigma = \text{Cov}(c)$ of the stacked attribution table, and the reconciliation adjustment is proportional to $\Sigma\mathbf{1}$. The same logic applies: any admissible cumulative attribution vector can be generated by a suitable choice of metric, but the GLS approach fixes this metric using the precision information implied by the underlying regressions.

C.6 Interpretation

Classical linking and smoothing rules and the GLS-based approach share a common mathematical structure: all resolve an under-identified linear constraint by selecting a solution that is optimal under a particular quadratic criterion. Classical methods impose simple, typically homogeneous or smoothness-based metrics. The GLS approach replaces these ad hoc choices with a metric grounded in the estimated precision of the attribution entries. Under restrictive covariance assumptions, the GLS solution reproduces familiar

rules; under more general conditions, it yields precision-aware adjustments that classical methods cannot capture.

D Sample Code

This section shows example Python code to compute the primary attribution objects described in the main text.

Code Listing 1: Sample Code for Regression Attribution

```
import numpy as np
import pandas as pd
from scipy.linalg import solve

EPS = 1e-12

def _ascol(x):
    x = np.asarray(x)
    return x.reshape(-1, 1)

def restricted_gls_1c(theta_hat, Sigma, a, b, ridge=1e-10):
    """
    One linear restriction: a' theta = b.
    theta_hat: (m,) or (m,1)
    Sigma: (m,m) covariance (or GLS metric)
    a: (m,) or (m,1)
    b: scalar
    """
    th = _ascol(theta_hat)
    a = _ascol(a)
    S = np.asarray(Sigma)
    if ridge:
        S = S + ridge * np.eye(S.shape[0])

    Sa = S @ a                # (m,1)
    denom = float(a.T @ Sa)    # scalar
    if abs(denom) < EPS:
        raise np.linalg.LinAlgError("a' Sigma a near zero; check covariance/constraint.")
    adj = Sa * ((b - float(a.T @ th)) / denom)
    return (th + adj).ravel()

# -----
# Factor attribution (restricted)
```

```

# -----

def factor_attrib_restricted(r, X, wP, wB, Wf=None, ridge=1e-10):
    """
    r: (N,) returns
    X: (N,K) factor exposures
    wP, wB: (N,) portfolio & benchmark weights
    Wf: (N,) regression weights, default equal weights
    Returns a dict with f_hat, f_star, active_return, contrib_hat,
        contrib_star.
    """
    r = np.asarray(r).reshape(-1)
    X = np.asarray(X)
    wA = np.asarray(wP).reshape(-1) - np.asarray(wB).reshape(-1)

    if Wf is None:
        W = np.eye(X.shape[0])
    else:
        W = np.diag(np.asarray(Wf).reshape(-1))

    XtWX = X.T @ W @ X
    XtWr = X.T @ W @ r
    # f_hat = (XtWX)^-1 XtWr
    f_hat = solve(XtWX + ridge*np.eye(XtWX.shape[0]), XtWr,
        assume_a="sym")

    Omega = np.linalg.inv(XtWX + ridge*np.eye(XtWX.shape[0])) #
        covariance/metric
    a = X.T @ wA
    b = float(wA @ r)
    f_star = restricted_gls_1c(f_hat, Omega, a, b, ridge=0.0)

    # Active contributions (K,)
    xA = wA @ X
    contrib_hat = xA * f_hat
    contrib_star = xA * f_star
    return {
        "f_hat": f_hat,
        "f_star": f_star,
        "active_return": b,
        "contrib_hat": contrib_hat,
        "contrib_star": contrib_star,
    }

```

```

# -----
# Brinson attribution (restricted)
# -----

def _group_means(r, groups, w):
    """Weighted mean return per group."""
    df = pd.DataFrame({"r": r, "g": groups, "w": w})
    num = df.groupby("g").apply(lambda x: (x["w"] * x["r"]).sum())
    den = df.groupby("g")["w"].sum()
    mu = (num / den).sort_index()
    return mu

def brinson_restricted(r, groups, wP, wB, Sigma=None, ridge=1e-10):
    """
    r: (N,) returns
    groups: (N,) group labels
    wP, wB: (N,) weights
    Sigma: (N,N) asset return covariance (optional). If None, use
        diag(var) proxy.
    Produces adjusted allocation/selection with zero interaction
        (portfolio-wide).
    """
    r = np.asarray(r).reshape(-1)
    wP = np.asarray(wP).reshape(-1)
    wB = np.asarray(wB).reshape(-1)
    wA = wP - wB

    # group means from WLS identities
    muB = _group_means(r, groups, wB)
    muP = _group_means(r, groups, wP)
    G = len(muB)
    idx = muB.index

    # Build D (N,G) dummy matrix aligned to idx
    g_codes = pd.Categorical(pd.Series(groups), categories=idx).codes
    D = np.zeros((len(r), G))
    D[np.arange(len(r)), g_codes] = 1.0

    # Residual covariance proxy Delta
    if Sigma is None:
        # simple diagonal proxy from returns; replace with risk-model
        # Sigma if available

```

```

s2 = np.var(r, ddof=1)
Sigma = np.eye(len(r)) * s2
Sigma = np.asarray(Sigma)

# Choose W for residual-maker; W_B keeps benchmark reporting
consistent
WB = np.diag(wB)
P = D @ np.linalg.inv(D.T @ WB @ D + ridge*np.eye(G)) @ D.T @ WB
M = np.eye(len(r)) - P
Delta = M @ Sigma @ M.T

# A_B and A_P maps r -> mu (GxN) with mu = A r
AB = np.linalg.inv(D.T @ WB @ D + ridge*np.eye(G)) @ D.T @ WB
WP = np.diag(wP)
AP = np.linalg.inv(D.T @ WP @ D + ridge*np.eye(G)) @ D.T @ WP

SigB = AB @ Delta @ AB.T
SigP = AP @ Delta @ AP.T
SigBP = AB @ Delta @ AP.T

# Stack m = [muB; muP], Gamma = [[SigB, SigBP],[SigBP', SigP]]
m_hat = np.r_[muB.values, muP.values]
Gamma = np.block([[SigB, SigBP], [SigBP.T, SigP]])

# Constraint: wA' D (muP - muB) = 0 <=> a' m = 0 with a = H' D' wA
# H = [-I, I] so H' = [-I; I]
d_wA = D.T @ wA # (G,)
a = np.r_[ -d_wA, d_wA]
m_star = restricted_gls_1c(m_hat, Gamma, a, 0.0, ridge=1e-12)

muB_star = pd.Series(m_star[:G], index=idx)
muP_star = pd.Series(m_star[G:], index=idx)

# Attribution components (using your notation)
rB = float(wB @ r)
# fitted vectors
rB_hat_star = D @ muB_star.values
rP_hat_star = D @ muP_star.values

A_star = float(wA @ (rB_hat_star - rB))
S_star = float(wB @ (rP_hat_star - rB_hat_star))
rA = float((wP - wB) @ r)

```

```

    return {
        "muB": muB, "muP": muP,
        "muB_star": muB_star, "muP_star": muP_star,
        "A_star": A_star, "S_star": S_star,
        "active_return": rA,
        "check_AplusS": A_star + S_star,
    }

# -----
# Multi-period attribution: One weight per period
# -----

def compounded_return(r: pd.Series) -> float:
    """R = prod(1+r_t) - 1."""
    r = pd.Series(r).astype(float)
    return float(np.prod(1.0 + r.values) - 1.0)

def gls_period_weights(r: pd.Series, R: float, V: np.ndarray = None,
    ridge: float = 1e-12) -> pd.Series:
    """
    Choose period weights w to satisfy  $r'w = R$  with minimal GLS
    distortion from 1:
    
$$\min 0.5 (w-1)' V^{-1} (w-1) \text{ s.t. } r'w = R$$

    Closed form:  $w = 1 + V r * (R - 1'r) / (r' V r)$ 

    r: (T,) realized period portfolio returns
    V: (T,T) metric/covariance for weights; if None -> identity
    """
    r = pd.Series(r).astype(float)
    T = len(r)
    rv = r.values.reshape(-1, 1)

    if V is None:
        V = np.eye(T)
    V = np.asarray(V, dtype=float)

    # ridge for stability
    V = V + ridge * np.eye(T)

    one = np.ones((T, 1))
    gap = float(R - float(one.T @ rv)) # R - sum r_t
    Vr = V @ rv
    denom = float(rv.T @ Vr) # r' V r

```

```

if abs(denom) < EPS:
    raise np.linalg.LinAlgError("r' V r near zero; check V or r.")

w = one + Vr * (gap / denom)
w = w.ravel()
return pd.Series(w, index=r.index, name="w_gls")

# -----
# Multi-period attribution: Full weighting matrix
# -----

def gls_reconcile_entries(
    C: pd.DataFrame,
    R: float,
    Sigma: np.ndarray,
    ridge: float = 1e-12,
) -> pd.DataFrame:
    """
    Entry-specific GLS reconciliation:
    min 0.5 (c*-c)' Sigma^{-1} (c*-c) s.t. 1' c* = R
    Solution:
    c* = c + Sigma 1 * (R - 1'c) / (1' Sigma 1)

    C: (T,K) per-period attribution table (rows=periods,
        cols=components)
    Sigma: (TK, TK) covariance/metric of vec(C) in row-major order (t
        then k)
    """
    C = C.copy()
    T, K = C.shape
    c = C.to_numpy(dtype=float).reshape(-1, 1, order="C") # vec by
        rows

    Sigma = np.asarray(Sigma, dtype=float)
    if Sigma.shape != (T * K, T * K):
        raise ValueError(f"Sigma must be {(T*K, T*K)}; got
            {Sigma.shape}.")

    # Ridge for numerical stability
    Sigma = Sigma + ridge * np.eye(T * K)

```

```

one = np.ones((T * K, 1))
gap = float(R - float(one.T @ c))      # R - sum_{t,k} c_{k,t}
S1 = Sigma @ one
denom = float(one.T @ S1)              # 1' Sigma 1

if abs(denom) < EPS:
    raise np.linalg.LinAlgError("1' Sigma 1 near zero; check
                                Sigma.")

c_star = c + S1 * (gap / denom)
C_star = c_star.reshape((T, K), order="C")
return pd.DataFrame(C_star, index=C.index, columns=C.columns)

# -----
# Reporting table
# -----

def make_attrib_report(
    C: pd.DataFrame,
    r: pd.Series = None,
    method: str = "period_weights", # "period_weights" or "entry_gls"
    V: np.ndarray = None,
    Sigma: np.ndarray = None,
    label_cum: str = "Cumulative",
) -> pd.DataFrame:
    """
    Build an attribution report table with a final cumulative row
        that sums to compounded return.

    Inputs
    -----
    C: (T,K) per-period attribution entries, rows=periods,
        cols=components
        Each row should sum to realized period return (or close, if
        rounding).
    r: optional (T,) realized period returns. If None, use row sums
        of C.
    method:
        - "period_weights": Cumulative = sum_t w_t * c_{k,t}, weights
            chosen so sum_t w_t r_t = R
        - "entry_gls": First reconcile the full table via GLS so
            sum_{t,k} c*_{k,t} = R,
            then cumulative row is column sums of c*
    """

```

V: (T,T) metric for period_weights (optional; default identity)
 Sigma: (TK,TK) covariance for entry_gls (required for entry_gls)

Examples

1) Period-only weights

```
tbl1 = make_attrib_report(C, r=r, method="period_weights",
                          V=None)
tbl1.attrs["weights"] # the w_t
```

2) Full GLS table reconciliation

```
tbl2 = make_attrib_report(C, r=r, method="entry_gls",
                          Sigma=Sigma_full)
"""
C = C.copy()
if r is None:
    r = C.sum(axis=1)
r = pd.Series(r, index=C.index).astype(float)

# Basic per-row check
row_gap = (C.sum(axis=1) - r)
# keep this diagnostic; don't hard-fail (users often have
# rounding)
diag = {
    "max_abs_row_gap": float(np.max(np.abs(row_gap.values))),
}

R = compounded_return(r)

if method == "period_weights":
    w = gls_period_weights(r, R, V=V)
    C_cum = (C.mul(w, axis=0)).sum(axis=0) # (K,)
    total_cum = float((w * r).sum())
    # table itself remains per-period exact; only cumulative is
    # weighted
    C_used = C
elif method == "entry_gls":
    if Sigma is None:
        raise ValueError("Sigma is required for
                           method='entry_gls'.")
    C_used = gls_reconcile_entries(C, R, Sigma=Sigma)
    C_cum = C_used.sum(axis=0)
    total_cum = float(C_used.to_numpy().sum())
```

```

        w = None
    else:
        raise ValueError("method must be 'period_weights' or
                          'entry_gls'.")

    report = C_used.copy()
    report["Total"] = report.sum(axis=1)
    cum_row = pd.Series(C_cum, index=C.columns, name=label_cum)
    cum_row["Total"] = total_cum
    report = pd.concat([report, cum_row.to_frame().T], axis=0)

    # Add a couple diagnostics as attributes (handy in notebooks)
    report.attrs["compounded_return_R"] = R
    report.attrs["cumulative_total"] = total_cum
    report.attrs["max_abs_row_gap"] = diag["max_abs_row_gap"]
    if w is not None:
        report.attrs["weights"] = w

    return report

# -----
# Covariance construction
# -----

def V_from_row_se(row_se: pd.Series, index=None, floor: float =
                  1e-12) -> np.ndarray:
    """
    row_se: (T,) standard error proxy per period. Larger -> more
            adjustment allowed.
    Returns diagonal V with entries proportional to row_se^2.
    """
    s = pd.Series(row_se).astype(float).clip(lower=floor)
    if index is not None:
        s = s.reindex(index)
    return np.diag((s.values) ** 2)

def Sigma_blockdiag(Omegas: list[np.ndarray], ridge: float = 1e-12)
-> np.ndarray:
    """
    Omegas: list of (K,K) per-period covariance matrices Omega_t =
            cov(c_t)
    Returns block-diagonal Sigma of shape (T*K, T*K), with tiny ridge
            on each block.

```

```
"""
blocks = []
for Om in Omegas:
    Om = np.asarray(Om, dtype=float)
    K = Om.shape[0]
    blocks.append(Om + ridge * np.eye(K))
return np.block([[blocks[i] if i == j else
    np.zeros_like(blocks[0])
    for j in range(len(blocks))]
    for i in range(len(blocks))])
```