MONASH University

Enhancing Forecast Reconciliation: A Study of Alternative Covariance Estimators

Vincent Su

B.Com. (Hons), Monash University

Supervised by

Shanika Wickramasuriya
and George Athanasopoulos

A thesis submitted for the degree of
Bachelor of Commerce (Honours)
at Monash University
Department of Econometrics and Business Statistics

25 October 2025

Abstract

A collection of time series connected via a set of linear constraints is known as hierarchical time series. Forecasting these series without respecting the hierarchical nature of the data can lead to incoherent forecasts across aggregation levels and, in practice, reduced accuracy. Forecast reconciliation corrects this by adjusting base forecasts to satisfy such constraints. Among modern reconciliation methods, Minimum Trace (MinT) is widely used, however, it requires a good estimate of the forecast error covariance matrix. The current practice uses linear shrinkage towards a diagonal target. Furthermore, the covariance estimate is based on 1-step-ahead residuals, then proportionally scale it to approximate h-step-ahead covariance matrix. This leaves a question of whether this method is appropriate for all real-world applications. We study the shortcomings of current practice and propose alternative covariance estimators, including the NOVELIST estimator (shrinkage towards a soft-thresholded target), PC-adjusted shrinkage (which utilises latent factor structures), and horizon-specific estimators that relax proportional scaling. We evaluate MinT using these covariance estimates for both point and probabilistic reconciliation, and demonstrate their effectiveness and improvements over the shrinkage estimator in a complex, large-hierarchy dataset.

Keywords: *Coherent forecasts, Hierarchical time series, Covariance estimation, Australian tourism, Aggregation.*

Table of contents

1 Introduction

In time series forecasting, aggregation occurs in a variety of settings. For example, Starbucks Corporation operates in many countries, and each country has multiple cities where they have outlets. The sales data is structured *hierarchically*: the top level is the company's total sales, which disaggregates into sales by country, then sales by city within each country, and finally down to sales by individual outlet within each city. As a result, there are over 50,000 sales series across all *aggregation levels*, and decision makers need forecasts at each level to manage inventory and plan marketing strategies effectively. The hierarchy can be even more complex if we consider the sales of different categories of products (e.g., beverage, food, etc.) and the sales of product within each category (e.g., latte, cappuccino, etc.) at each aggregation level. In this case, the structure is called a *grouped structure*, where the aggregation paths are not unique. Such structures also arises in many other decision-making contexts, from supply chains (**?**; **?**) and energy planning (**?**), to macroeconomics (**?**; **?**) and tourism analysis (**?**). Stakeholders in these settings need forecasts at several aggregation levels to allocate resources and manage risk.

In practice, when forecasts are produced for all series (often called *base forecasts*), they typically violate the aggregation constraints observed in the data (e.g., the sum of all countries' sales forecasts does not equal the total sales forecast). Such forecasts are called *incoherent*. Incoherence undermines downstream decisions that require internal consistency and can degrade forecasting performance. To tackle this problem, forecast reconciliation was introduced. Forecast reconciliation, a post-processing step, utilises the information from the hierarchical structure and data to adjust the initially produced base forecasts, so that the resulting *reconciled forecasts* are *coherent* (i.e., respecting the aggregation constraints). It was first introduced by Hyndman et al. (**?**), and later developed by Erven & Cugliari (**?**), Hyndman et al. (**?**), Ben Taieb & Koo (**?**), Wickramasuriya et al. (**?**), Wickramasuriya et al. (**?**), and others that focus point forecast reconciliation (more accurately, reconciling the means of predictive distributions). Recognising the important of probabilistic forecasting, probabilistic forecast reconciliation was studied by Shang & Hyndman (**?**), Jeon et al. (**?**), and Ben Taieb et al. (**?**), and later formally defined by Panagiotelis et al. (**?**). Athanasopoulos et al. (**?**) provide a comprehensive review of the literature on forecast reconciliation.

Among the modern methods, the Min Trace (MinT) approach developed by Wickramasuriya et al. (**?**) is is widely used due to its strong theoretical properties for minimising total reconciled forecast error variance, computational efficiency, and robust empirical performance. MinT is later extended to probabilistic reconciliation by Wickramasuriya (**?**), showing that it also minimises the negative log score of the reconciled distribution under Gaussian assumptions.

Wickramasuriya et al. (?) also argued that modelling spatial autocorrelations directly from the start would be challenging as in this case of a large collection of time series. Post-processing reconciliation has the advantage to implicitly model this spatial autocorrelation structure, especially true for MinT. MinT is implemented in popular R and Python software ecosystems (?; ?).

A central difficulty for MinT is estimating the covariance matrix of base-forecast errors, particularly beyond one-step forecast horizons. This is a high-dimensional estimation problem in which the number of series often exceeds the time dimension. A common practice, following Wickramasuriya et al. (?), is to estimate the 1-step-ahead covariance from the residuals, using linear shrinkage toward a diagonal target (?). Then, proportionally scale this estimate to approximate the multi-step-ahead covariance matrix. While convenient and guaranteed to produce a positive-definite estimate, this practice has three important shortcomings. First, the shrinkage is uniform across off-diagonals, applying a single penalty that may over-shrink genuine dependence and under-shrink noise. Second, many hierarchical data sets exhibit strong latent low-rank structures that can be explicitly exploited. Third, the proportionality relationship between the h-step-ahead and 1-step-ahead forecast error covariance matrices might not hold when error dynamics change with horizon.

These limitations matter because the theoretical advantages of MinT depend on the quality of the covariance estimate available in finite samples. Despite its central role, there has been limited work on tailored covariance estimation for reconciliation. An exception is the recent double-shrinkage proposal in Carrara et al. (?), which introduces an additional target designed to encode conditional dependence suggested by the hierarchy. This line of work still remains at an early stage and does not fully address the aforementioned limitations.

Meanwhile, there has been substantial progress in high-dimensional covariance estimation that can be leveraged for MinT. Building on shrinkage, Huang & Fryzlewicz (?) proposed NOVELIST, which shrinks toward a soft-thresholded target rather than a diagonal one, improving flexibility to sparse targets. Extending beyond linear shrinkage, Ledoit & Wolf (?) (and further developed by Ledoit & Wolf (?)) introduced nonlinear shrinkage that replaces sample eigenvalues with data-driven nonlinear transformation of themselves. In a complementary direction, Fan et al. (?) introduce factor-based estimators that explicitly preserve latent low-rank structure while thresholding the idiosyncratic component. Related thresholding approaches, including hard thresholding (?), generalised thresholding (?), and adaptive thresholding (?), have also been developed. This growing toolkit provides multiple pathways to maximise the performance of MinT.

This paper focuses on covariance estimation for MinT. We examine the shortcomings of the standard shrinkage plus scaling practice and introduce alternative estimators that address these issues individually and in combination. Specifically, we study NOVELIST as a more adaptive target-based shrinkage; principal-component-adjusted (PC-adjusted) estimators that exploit latent factor structures; and horizon-specific estimators that relax proportional scaling, including direct estimation from multi-step residuals. We evaluate MinT under these alternatives for both point and probabilistic reconciliation. In a large, complex real-world hierarchy, our findings reveal three main insights. First, NOVELIST improves probabilistic performance relative to linear shrinkage, including higher empirical coverage. Second, PC-adjusted estimators consistently outperform other estimators when common components are strong, for both point and probabilistic metrics. Third, proportional scaling of the one-step covariance remains a competitive baseline, indicating that the assumption may not be universally invalid.

The remainder of the paper is organised as follows. Section **??** sets out the framework for hierarchical time series, forecast reconciliation, and MinT, and motivates the need for improved covariance estimation. Section **??** presents the covariance estimators considered, outlining their strengths and limitations in the reconciliation context. Section **??** defines the evaluation metrics for point and probabilistic reconciliation. Section **??** details the simulation design and examines the performance of NOVELIST compared to shrinkage. Section **??** dives into a real-world application with all proposed methods and highlights behaviours not observed in the simplified simulations.

2 Theoretical Framework

2.1 Hierarchical Time Series and Reconciliation

Hierarchical time series are multivariate time series $\boldsymbol{y}_t \in \mathbb{R}^n$ organised in a structure where the series adheres to linear constraints. Figure **??** illustrates a simple two-level hierarchical structure with one top-level series $y_{Tot,t}$, disaggregating down to two level-1 series $(y_{A,t}, y_{B,t})'$, and to four bottom-level series $(y_{A1,t}, y_{A2,t}, y_{B1,t}, y_{B2,t})'$. Here, the *aggregation constraints* imply that $y_{Tot,t} = y_{A,t} + y_{B,t}$, $y_{A,t} = y_{A1,t} + y_{A2,t}$, and $y_{B,t} = y_{B1,t} + y_{B2,t}$, for all time points t.

The bottom-level (or most disaggregated) series are denoted as $\boldsymbol{b}_t \in \mathbb{R}^{n_b}$. Thus, the full vector of all series in the hierarchy can be represented as:

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t,$$

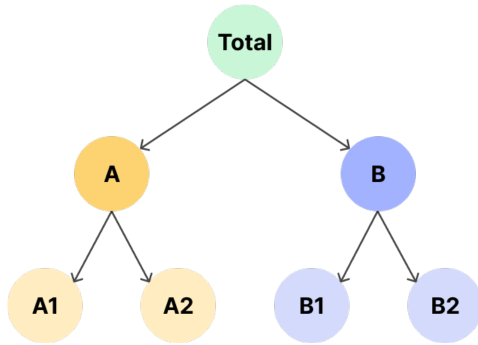

Figure 1: A 2-level hierarchical tree structure

where $\boldsymbol{S} \in \mathbb{R}^{n \times n_b}$ is a summing matrix that aggregates the bottom-level to all-level series. The summing matrix $\boldsymbol{S}$ for the tree structure in Figure **??** is:

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I_4} & & \end{bmatrix},$$

where $\boldsymbol{I_4}$ is the 4×4 identity matrix. The matrix $\boldsymbol{S}$ encodes the aggregation constraints implied by the structure. Hence, the columns of $\boldsymbol{S}$ span a linear subspace. Any observation $\boldsymbol{y}_t$ that lies inside this subspace is called *coherent*, while those outside are *incoherent*. We refer to the subspace spanned by $\boldsymbol{S}$ as the *coherent subspace* $\mathfrak{s} \in \mathbb{R}^{n_b}$.

This representation extends beyond hierarchical (nested) structures. When attributes of interest are crossed, such as the company sales at any aggregation level (company-wise, city-wise, or outlet-wise) is also considered by kinds of products, the structure is described as a *grouped structure*. In grouped systems, as illustrated in Figure **??**, aggregation and disaggregation paths are not unique, but the linear constraints can still be written compactly through a summing matrix $\boldsymbol{S}$. For simplicity, we refer to both structures as hierarchical structure and distinguish between them when needed.

In practice, when we produce h-step-ahead forecasts for each individual series, referred to as *base forecasts* $\hat{\boldsymbol{y}}_{t+h|t}$, they generally violate the aggregation constraints, and thus are incoherent. Coherency can be restored by linearly projecting the base forecasts onto the coherent subspace $\mathfrak{s}$ using a projection matrix $\boldsymbol{P}$: $\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{P}\hat{\boldsymbol{y}}_{t+h|t}$, where $\tilde{\boldsymbol{y}}_{t+h|t}$ are h-step-ahead *reconciled forecasts*.

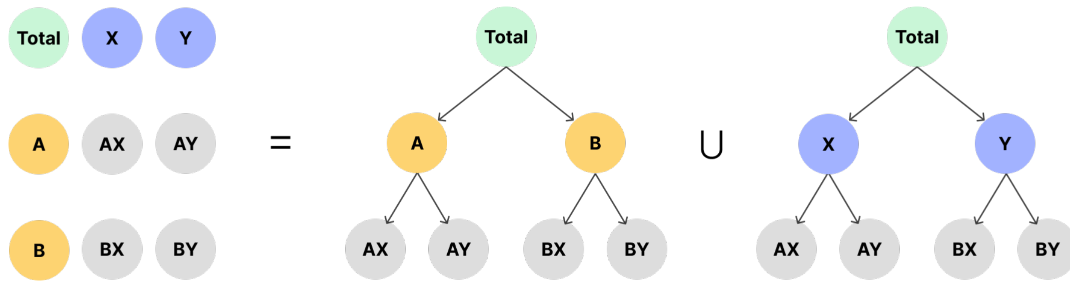

Figure 2: A 2-level grouped structure, which can be considered as the union of two hierarchical trees with common top and bottom level series

Many existing reconciliation methods including the OLS from Hyndman et al. (**?**), WLS from Hyndman et al. (**?**), and MinT from Wickramasuriya et al. (**?**) express the projection matrix as $\boldsymbol{P} = \boldsymbol{SG}$, for a suitable $n_b \times n$ mapping matrix $\boldsymbol{G}$. The role of this matrix is to map the base forecasts of all levels $\hat{\boldsymbol{y}}_{t+h|t}$ down into the bottom level, which is then aggregated to the higher levels by premultiplying with $\boldsymbol{S}$. Since the projection matrix $\boldsymbol{P}$ is idempotent, $\boldsymbol{G}$ must satisfy the condition $\boldsymbol{SGS} = \boldsymbol{S}$. Within this class, a broad family of mapping matrices is given by: $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{M}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{M}^{-1}$, for some positive definite matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ (**?**).

When setting $\boldsymbol{M} = \boldsymbol{I}_n$, the identity matrix, the $\boldsymbol{P}$ matrix reduces to the OLS reconciliation, which corresponds to an orthogonal projection onto $\mathfrak{s}$. For a general positive definite matrix $\boldsymbol{M}$, $\boldsymbol{P}$ is idempotent but not necessarily symmetric, resulting in an oblique projection. Thus, the choice of $\boldsymbol{M}$ determines the projection direction and, in turn, how disagreements among levels are resolved.

This projection view extends seamlessly from point to probabilistic forecast reconciliation. Let $\hat{\boldsymbol{Y}}_{t+h|t}$ denote the random vector representing the h-step-ahead base forecasts. The random reconciled forecast vector is obtained similarly via projection: $\tilde{\boldsymbol{Y}}_{t+h|t} = \boldsymbol{P}\hat{\boldsymbol{Y}}_{t+h|t}$. Consequently, point forecasts (which we refer to as means of the predictive distributions) reconcile as $\mathbb{E}(\tilde{\boldsymbol{Y}}_{t+h|t}) = \mathbb{E}(\boldsymbol{P}\hat{\boldsymbol{Y}}_{t+h|t}) = \boldsymbol{P}\hat{\boldsymbol{y}}_{t+h|t} = \tilde{\boldsymbol{y}}_{t+h|t}$, and the covariance of the predictive distribution transforms linearly as $\mathrm{Var}(\tilde{\boldsymbol{Y}}_{t+h|t}) = \mathrm{Var}(\boldsymbol{P}\hat{\boldsymbol{Y}}_{t+h|t}) = \boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'$, where $\boldsymbol{W}_h = \mathrm{Var}(\hat{\boldsymbol{Y}}_{t+h|t})$.

Figure **??** schematically illustrates this projection. Reconciliation takes the (possibly elliptical) base forecast distribution and “pushes” the entire mass linearly onto the coherent subspace, resulting in the reconciled forecast distribution.

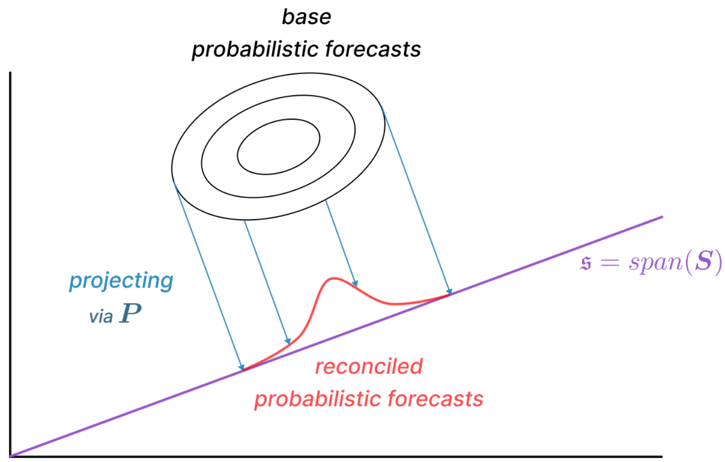

Figure 3: Geometry of probabilistic forecast reconciliation. The base forecast distribution is projected orthogonally onto the coherent subspace (purple line), resulting in the reconciled forecast distribution (red). The projection is defined by the projection matrix $\boldsymbol{P}$. Note that this figure is a schematic since most applications are high-dimensional.

2.2 The Minimum Trace Reconciliation

Wickramasuriya et al. (**?**) showed that by setting $\boldsymbol{M} = \boldsymbol{W}_h = \mathbb{E}(\hat{\boldsymbol{e}}_{t+h|t}\ \hat{\boldsymbol{e}}'_{t+h|t})$, the covariance matrix of the h-step-ahead base forecast errors $\hat{\boldsymbol{e}}_{t+h|t} = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t}$, we essentially minimise the total variance of the reconciled forecast errors across all series. Equivalently, MinT is the unique linear-unbiased reconciler that minimises the trace of $\mathrm{Var}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h|t}] = \boldsymbol{S}\boldsymbol{G}_h\boldsymbol{W}_h\boldsymbol{G}'_h\boldsymbol{S}'$. This method is thus called Minimum Trace (MinT) reconciliation. The matrix $\boldsymbol{G}_h$ is thus given by:

$$\boldsymbol{G}_h = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1},$$

provided that $\boldsymbol{W}_h$ is positive definite.

Although originally developed for point reconciliation, Wickramasuriya (**?**) showed that MinT extends naturally to the probabilistic setting when base predictive distributions are Gaussian. If h-step-ahead base forecast distribution is $\mathcal{N}(\hat{\boldsymbol{y}}_{t+h|t}, \boldsymbol{W}_h)$, then reconciliation gives $\tilde{\boldsymbol{Y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}_h\hat{\boldsymbol{Y}}_{t+h|t} \sim \mathcal{N}(\tilde{\boldsymbol{y}}_{t+h|t}, \boldsymbol{S}\boldsymbol{G}_h\boldsymbol{W}_h\boldsymbol{G}'_h\boldsymbol{S}')$. Within the class of linear-unbiased reconcilers, MinT also minimises the negative log score (equivalently, the Gaussian log predictive score) of the reconciled distribution.

In this paper we adopt the Gaussian framework to compare the impacts of different covariance estimators for MinT in both point and probabilistic reconciliation. It is also worth to mention that the methods can be extended to non-Gaussian settings by bootstrapping (**?**; **?**), which is beyond the scope of this paper.

2.3 Shrinkage Estimator for MinT

The performance of MinT hinges on a reliable, positive-definite estimate of $\boldsymbol{W}_h$, which comes in both the mapping matrix $\boldsymbol{G}_h$ and the reconciled forecast error variance $\boldsymbol{S}\boldsymbol{G}_h\boldsymbol{W}_h\boldsymbol{G}_h'\boldsymbol{S}'$.

However, the covariance matrix $\boldsymbol{W}_h$ is often not available in closed-form, and is challenging to estimate in high-dimensional setting where the number of series n is larger than the time dimension T. To tackle this issue, the original paper by Wickramasuriya et al. (**?**) assumed a proportionality relationship $\hat{\boldsymbol{W}}_h^g = k_h g(\hat{\boldsymbol{W}}_1)$, where $\hat{\boldsymbol{W}}_1$ is the sample covariance matrix of the in-sample 1-step-ahead base forecast errors (to approximate $\boldsymbol{W}_1$) and $k_h > 0$ is a scaling constant (which will be algebraically cancelled out in point-forecast reconciliation). The function $g(.)$ is a covariance estimator that produces a positive-definite matrix, the main focus of this paper.

The recommended choice for $g(.)$ in the original work is the shrinkage estimator with diagonal target from Schäfer & Strimmer (**?**):

$$\hat{\boldsymbol{W}}_1^S = \lambda_S \boldsymbol{D}_1 + (1 - \lambda_S)\hat{\boldsymbol{W}}_1 \,,$$

where $\boldsymbol{D}_1 = \text{diag}(\hat{\boldsymbol{W}}_1)$ is the diagonal matrix comprising the diagonal elements of $\hat{\boldsymbol{W}}_1$ (i.e., the sample variances). We refer to any $\lambda_S \in [0, 1]$ as the shrinkage intensity of the shrinkage estimator. This approach shrinks the covariance matrix $\hat{\boldsymbol{W}}_1$ towards the diagonal target $\text{diag}(\hat{\boldsymbol{W}}_1)$, meaning the off-diagonal elements are shrunk towards zero while the diagonal ones remain unchanged.

Schäfer & Strimmer (**?**) also provided a closed-form estimate of the optimal shrinkage intensity parameter λ_S:

$$\hat{\lambda}_S = \frac{\sum_{i \neq j} \widehat{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}^2} \,,$$

where $\hat{r}_{ij}$ is the i, j-th element of $\hat{\boldsymbol{R}}_1$, the 1-step-ahead sample correlation matrix (obtained from $\hat{\boldsymbol{W}}_1$). The optimal estimate is obtained by minimising $MSE(\hat{\boldsymbol{W}}_1) = Bias(\hat{\boldsymbol{W}}_1)^2 +$

$Var(\hat{\boldsymbol{W}}_1)$. More specifically, we trade a small bias for a substantial variance reduction, which is especially valuable in high dimension.

Despite its simplicity and guaranteed positive-definiteness, MinT coupled with diagonal-target shrinkage presents three important limitations.

Problem 1: Uniform shrinkage

Linear shrinkage shrinks all off-diagonal elements of $\hat{\boldsymbol{W}}_1$ towards zeros with equal weights $1-\lambda_S$. The resulting penalty is global and non-adaptive: strong, genuinely systematic correlations are shrunk at the same rate as weak, noisy ones. In hierarchical and grouped systems, this can be problematic. Aggregation naturally induces stronger dependence among related nodes (for example, a region and a neighbouring region, or a region and its parent state), while many unrelated pairs exhibit near-zero correlations. A uniform penalty may over-shrink informative co-movements or under-shrink idiosyncratic noise, reducing reconciliation efficiency.

Problem 2: Latent factors

Many real-world hierarchical data sets may exhibit a prominent low-rank (factor) structure. A clear example is the Australian domestic overnight trips (**?**), where the national trips are disaggregated into states and territories, and further into regions. The tourism activities might be driven by a few common factors, such as economic conditions, fuel prices, or major events, which might be left in the forecast errors after fitting the base models. Figure **??** illustrates the largest eigenvalues from the one-step forecast error correlation matrix $\hat{\boldsymbol{R}}_1$, typically showing a marked elbow after a few principal components (largest eigenvalues), indicating strong latent structure. Diagonal-target shrinkage is factor-unaware: it neither preserves the low-rank common subspace nor differentially shrinks the idiosyncratic remainder, and so can be inefficient when common factors dominate.

Problem 3: Proportional scaling

The proportionality relationship $\hat{\boldsymbol{W}}_h^g = k_h g(\hat{\boldsymbol{W}}_1)$ enforces horizon-invariant cross-series dependence of the forecast errors up to a scalar factor. In many applications, the structure of the error covariance might change with horizon due to evolving error dynamics or horizon-specific interactions among series. Proportional scaling may therefore misrepresent multi-step dependence.

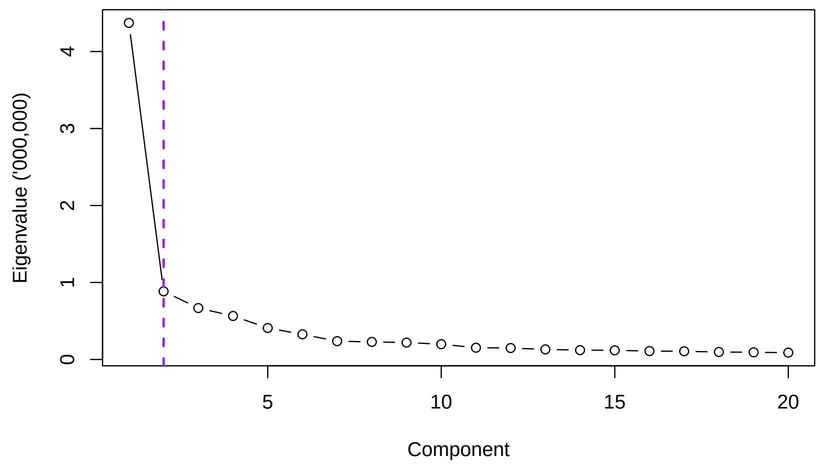

Figure 4: Scree plot of the eigenvalues of the 1-step-ahead forecast error covariance matrix from the Australian domestic tourism dataset. The point of inflection occurs around the second largest eigenvalue.

Additionally, while the scalar factor cancels in point reconciliation, it directly controls dispersion in probabilistic reconciliation and can lead to miscalibrated predictive distributions. However, we will not explore this issue in this paper, and leave it for future research.

In the next sections, we introduce alternative covariance estimators that address these limitations individually and in combination.

3 Covariance Estimation Approaches

3.1 NOVELIST Estimator

To tackle the Problem 1 of uniform shrinkage, we consider the NOVELIST estimator from Huang & Fryzlewicz (**?**). The NOVELIST (NOVEL Integration of the Sample and Thresholded Covariance) estimator introduces extra parameter, the threshold δ, to control the sparsity level in the target toward which the sample covariance matrix is shrunk. By soft-thresholding small correlations before shrinkage, NOVELIST preserves strong signals while attenuating weak, noisy ones, addressing the uniform, non-adaptive nature of diagonal-target shrinkage.

The construction proceeds in two steps. First, apply elementwise soft-thresholding to the sample correlation matrix; second, shrink the sample correlation toward this thresholded target.

The NOVELIST estimator for covariance matrix is given by:

$$\hat{\boldsymbol{W}}_1^N = \lambda_\delta \hat{\boldsymbol{W}}_{1,\delta} + (1-\lambda_\delta)\hat{\boldsymbol{W}}_1,$$

where $\hat{\boldsymbol{W}}_{1,\delta}$ is the thresholded covariance matrix of $\hat{\boldsymbol{W}}_1$. As working with correlations avoids rescaling issues and keeps diagonal entries at one, we rewrite it in terms of correlations:

$$\hat{\boldsymbol{R}}_1^N = \lambda_\delta \hat{\boldsymbol{R}}_{1,\delta} + (1-\lambda_\delta)\hat{\boldsymbol{R}}_1, \tag{1}$$

where, $\hat{\boldsymbol{R}}_{1,\delta}$ is the thresholded correlation matrix, where each element is regularised by:

$$\hat{r}_{1,ij}^{\delta} = \mathrm{sign}(\hat{r}_{1,ij}) \max(|\hat{r}_{1,ij}| - \delta,\ 0), \tag{2}$$

where $\delta \in [0,1]$. For a given threshold δ, Huang & Fryzlewicz (**?**) derived an analytical expression for the optimal shrinkage intensity parameter $\lambda(\delta)$, following similar logic to Schäfer & Strimmer (**?**). It can be computed as:

$$\hat{\lambda}(\delta) = \frac{\sum_{i\neq j} \widehat{Var}(\hat{r}_{1,ij})\ \mathbf{1}(|\hat{r}_{1,ij}| \leq \delta)}{\sum_{i\neq j}(\hat{r}_{1,ij} - \hat{r}_{1,ij}^{\delta})^2}, \tag{3}$$

where $\mathbf{1}(.)$ is the indicator function.

On the other hand, the optimal threshold $\hat{\delta}$ does not have a closed-form solution, and is obtained by rolling-window cross-validation procedure. The idea is to find the threshold δ^*, with the corresponding λ^* and $\hat{\boldsymbol{R}}_1^N(\delta^*, \lambda^*)$, that minimises the average out-of-sample 1-step-ahead mean squared reconciled forecast errors over all windows. The formal algorithm is given in Algorithm **??**.

Note that when $\delta \in \big[\max_{i\neq j}|\hat{r}_{1,ij}|,\ 1\big]$, the NOVELIST estimator collapses to the shrinkage estimator, and when $\delta = 0$, it becomes the sample covariance matrix. An additional concern is that the estimator is not guaranteed to be positive definite, but we can use Higham (**?**) algorithm to compute the nearest positive definite matrix if needed.

Algorithm 1 Cross-validation procedure

Input: Observations and fitted values $\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t \in \mathbb{R}^n$ for $t = 1, \dots, T$ [1], set of threshold candidates Δ, window size v.
$\hat{\boldsymbol{e}}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$ for $t = 1, \dots, T$
for $i = v : T-1$ **do**
    $j = i - v + 1$
    $\hat{\boldsymbol{W}}_1 = \frac{1}{v} \sum_{t=j}^{i} \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t'$
    $\hat{\boldsymbol{D}} = \mathrm{diag}(\hat{\boldsymbol{W}}_1)$
    $\hat{\boldsymbol{R}}_1 = \hat{\boldsymbol{D}}^{-1/2} \hat{\boldsymbol{W}}_1 \hat{\boldsymbol{D}}^{-1/2}$
    for $\delta \in \Delta$ **do**
        Compute thresholded correlation $\hat{\boldsymbol{R}}_{1,\delta}$ using Equation 2
        Compute shrinkage parameter $\hat{\lambda}_\delta$ using Equation 3
        Compute $\hat{\boldsymbol{R}}^N_{1,\delta}$ using Equation 1
        $\hat{\boldsymbol{W}}^N_{1,\delta} = \hat{\boldsymbol{D}}^{1/2} \hat{\boldsymbol{R}}^N_{1,\delta} \hat{\boldsymbol{D}}^{1/2}$
        $\boldsymbol{G} = (\boldsymbol{S}' \hat{\boldsymbol{W}}^{N^{-1}}_{1,\delta} \boldsymbol{S})^{-1} \boldsymbol{S}' \hat{\boldsymbol{W}}^{N^{-1}}_{1,\delta}$
        Reconciled forecasts $\tilde{\boldsymbol{y}}_{i+1} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{i+1}$
        $\tilde{\boldsymbol{e}}_{i+1,\delta} = \boldsymbol{y}_{i+1} - \tilde{\boldsymbol{y}}_{i+1}$
    end for
end for
$\mathrm{MSE}_\delta = \frac{1}{T-v} \sum_{i=v}^{T-1} (\tilde{\boldsymbol{e}}_{i+1,\delta})^2$ for each $\delta \in \Delta$
$\hat{\delta} = \arg\min_{\delta \in \Delta} \mathrm{MSE}_\delta$
Output: Estimate of optimal threshold parameter $\hat{\delta}$

Remark. Minimising a multi-step-ahead forecast error metric in the cross-validation procedure oftern yields a $\hat{\delta}$ that is close to the 1-step-ahead case.

3.2 PC-adjusted Estimator

To address Problem 2 and exploit latent common components explicitly, the PC-adjusted (Principal-Component-adjusted) method takes the latent factors directly into its construction. It starts by decomposing the covariance matrix $\hat{\boldsymbol{W}}_1$ into a prominent principle components part (low-rank) and a orthogonal complement part $\hat{\Upsilon}$ (the covariance matrix after removing the first K principal components). Then we can apply either shrinkage or NOVELIST estimator to $\hat{\boldsymbol{W}}^K_1$:

[1] It is only required to fit the base models once on the whole training data $\{\boldsymbol{y}_t\}_{t=1}^T$, and obtain the in-sample fitted values $\{\hat{\boldsymbol{y}}_t\}_{t=1}^T$.