Hierarchical Time Series Forecasting with Robust Reconciliation

Anonymous authors
Paper under double-blind review

Abstract

This paper focuses on forecasting hierarchical time-series data, where each higher-level observation equals the sum of its corresponding lower-level time series. In such contexts, the forecast values should be coherent, meaning that the forecast value of each parent series exactly matches the sum of the forecast values of its child series. Existing hierarchical forecasting methods typically generate base forecasts independently for each series and then apply a reconciliation procedure to adjust them so that the resulting forecast values are coherent across the hierarchy. These methods generally derive an optimal reconciliation, using a covariance matrix of the forecast error. In practice, however, the true covariance matrix is unknown and has to be estimated from finite samples in advance. This gap between the true and estimated covariance matrix may degrade forecast performance. To address this issue, we propose a robust optimization framework for hierarchical reconciliation that accounts for uncertainty in the estimated covariance matrix. We first introduce an uncertainty set for the estimated covariance matrix and formulate a reconciliation problem that minimizes the worst-case expected squared error over this uncertainty set. We show that our problem can be cast as a semidefinite optimization problem. Numerical experiments demonstrate that the proposed robust reconciliation method achieved better forecast performance than existing hierarchical forecasting methods, which indicates the effectiveness of integrating uncertainty into the reconciliation process.

1 Introduction

Time series forecasting is indispensable across diverse fields, including sales planning and inventory management (Aviv, 2003; Ramos et al., 2015), energy supply planning (Suganthi & Samuel, 2012; Hernandez et al., 2014), and economic analysis and stock investment decision-making (Krollner et al., 2010). For instance, in the retail sector, accurate sales predictions based on historical data are crucial for optimizing inventory levels and preventing both overstocking and shortages. Similarly, for electric power companies, forecasting electricity consumption enables efficient facility operation and effective supply-demand balance management. Moreover, at both individual and national levels, leveraging forecasts of economic indicators and stock prices can contribute significantly to wealth generation. Conversely, low forecast accuracy can lead to substantial losses and missed opportunities for individuals, businesses, and society as a whole. Consequently, extensive research has been dedicated to developing various time series forecasting methods, and there remains a strong demand for more precise techniques (Mahalakshmi et al., 2016; Liu et al., 2021; Wen et al., 2022).

Many real-world datasets inherently possess hierarchical structures. Examples include sales data organized by region or demographic statistics categorized by gender or age group, which are commonly recorded across multiple levels of aggregation. In practice, the appropriate level of the hierarchy for forecasting depends on the specific application, and this choice can significantly influence the prediction outcomes. Generally, as one descends to lower hierarchical levels, the data becomes more granular but also more susceptible to individual variations and noise, leading to increased uncertainty. This often means that aggregated data at higher levels tends to be more stable and achieve greater forecast accuracy (Grunfeld & Griliches, 1960). However, higher-level aggregated data can obscure fine-grained patterns and individual variation factors.

Therefore, it has also been suggested that utilizing detailed data from lower levels, if appropriately modeled, can potentially yield superior forecast accuracy (Orcutt et al., 1968; Edwards & Orcutt, 1969).

Given this context, time series forecasting methods that explicitly account for hierarchical structures have garnered increasing attention (Athanasopoulos et al., 2009; Hyndman et al., 2011; Wickramasuriya et al., 2019; Panagiotelis et al., 2021; Hyndman & Athanasopoulos, 2021). These approaches aim to adjust forecasts across both lower and higher levels to ensure coherence when aggregating forecast values within the hierarchy. Such methods are expected to enhance forecast accuracy compared to conventional time series forecasting based on a single level.

Despite these advancements, existing hierarchical time series forecasting methods face certain challenges. Traditional approaches typically aim to minimize the expected forecast error at a given target time point, which necessitates a covariance matrix of the forecast errors. This matrix is commonly estimated from the residuals between observed and forecast values. However, if the underlying data trends shift or the forecasting model is not sufficiently accurate, discrepancies can emerge between the estimated and true covariance matrices. Thus, the estimated covariance matrix itself carries inherent uncertainty, which must be addressed. Prior work by Møller et al. (2024) focused on this issue, decomposing covariance matrix estimation into parameter estimation errors and stochastic irreducible errors to quantify uncertainty and improve forecast accuracy. Nevertheless, even with their method, the true covariance matrix cannot be perfectly determined, leaving room for further improvements in forecast accuracy.

To tackle the uncertainty of estimators, robust optimization has emerged as a powerful technique. This methodology is designed to yield solutions that remain effective even when the underlying data fluctuates within a defined uncertain range. Specifically, it involves establishing a range for uncertain parameters or data and then seeking an optimal solution that performs best under the worst-case scenario within that range. Since its inception by Ben-Tal & Nemirovski (1998), robust optimization has been extensively researched in both theoretical and applied domains (Bertsimas et al., 2011). Notably, models that incorporate covariance matrix uncertainty have been developed and applied to various problems, such as portfolio optimization (Lobo & Boyd, 2000; Halldórsson & Tütüncü, 2003; Tütüncü & Koenig, 2004).

In this paper, we propose a novel method that frames hierarchical time series forecasting as a robust optimization problem, specifically minimizing forecast error under uncertainty. We introduce an uncertainty set for the covariance matrix of forecast errors and solve an optimization problem that minimizes the forecast error under the worst-case scenario within this set. We demonstrate through duality that this robust optimization problem can be formulated as a semidefinite optimization problem, which is theoretically solvable efficiently. Furthermore, we present numerical experiments using five real-world datasets, showcasing that our proposed method achieves more accurate forecasting results compared to existing hierarchical time series forecasting techniques.

2 Hierarchical time series forecasting

2.1 Notation

This section defines the notation used throughout the paper, which is consistent with prior studies on hierarchical time series forecasting (Athanasopoulos et al., 2009; Hyndman et al., 2011; Wickramasuriya et al., 2019; Panagiotelis et al., 2021; Hyndman & Athanasopoulos, 2021).

A hierarchical structure is defined by a series of nested levels. Level 0 is the fully aggregated series. Level 1 consists of the series obtained by disaggregating the Level 0 series, and Level 2 contains the series that further disaggregate each Level 1 series. This process continues until the bottom-level, denoted as Level K, where its series can no longer be disaggregated.

Let $y_X^{(t)} \in \mathbb{R}$ denote the observation of a series X at time t. The label X is a series of labels representing the indices of each level. For example, a series X that belongs to series i at Level 1, series j at Level 2, and series k at Level 3 can be denoted by ijk. The series at Level 0 is simply written as $y^{(t)}$, without a series name X. A key property of hierarchical data is that, at any given time point, the value of a series at a specific level equals the sum of the values of the series nested directly below it:

$$y^{(t)} = \sum_i y_i^{(t)},\ y_i^{(t)} = \sum_j y_{ij}^{(t)},\ y_{ij}^{(t)} = \sum_k y_{ijk}^{(t)}.$$

To simplify the notation of the hierarchical structure, a matrix and vector expression is often used. Let n be the total number of series and m be the number of bottom-level series, which satisfy $n > m$. We denote the vector of all series observations at time t as $\boldsymbol{y}^{(t)} \in \mathbb{R}^n$ and the vector of bottom-level observations as $\boldsymbol{b}^{(t)} \in \mathbb{R}^m$. With the summing matrix $\boldsymbol{S} \in \mathbb{R}^{n \times m}$ that dictates the way in which the bottom-level series aggregate, the hierarchical structure can be written as:

$$\boldsymbol{y}^{(t)} = \boldsymbol{S}\boldsymbol{b}^{(t)}. \tag{1}$$

When Equation (1) holds for the values of all series and the bottom-level series at each time t, it is said the hierarchy is satisfied.

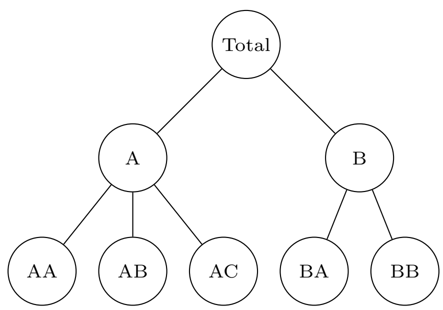

Figure 1: An example of a hierarchical structure

Figure 1 provides a simple example of a hierarchical structure. In this case, $K = 2$, $m = 5$, $n = 8$, and the following aggregation relationship holds:

$$y_{\text{A}}^{(t)} = y_{\text{AA}}^{(t)} + y_{\text{AB}}^{(t)} + y_{\text{AC}}^{(t)},\ y_{\text{B}}^{(t)} = y_{\text{BA}}^{(t)} + y_{\text{BB}}^{(t)},$$

$$y^{(t)} = y_{\text{A}}^{(t)} + y_{\text{B}}^{(t)} = y_{\text{AA}}^{(t)} + y_{\text{AB}}^{(t)} + y_{\text{AC}}^{(t)} + y_{\text{BA}}^{(t)} + y_{\text{BB}}^{(t)}.$$

The corresponding matrix and vector representation is defined by:

$$\boldsymbol{y}^{(t)} = \begin{bmatrix} y^{(t)}\ y_{\text{A}}^{(t)}\ y_{\text{B}}^{(t)}\ y_{\text{AA}}^{(t)}\ y_{\text{AB}}^{(t)}\ y_{\text{AC}}^{(t)}\ y_{\text{BA}}^{(t)}\ y_{\text{BB}}^{(t)} \end{bmatrix}^\top,$$

$$\boldsymbol{b}^{(t)} = \begin{bmatrix} y_{\text{AA}}^{(t)}\ y_{\text{AB}}^{(t)}\ y_{\text{AC}}^{(t)}\ y_{\text{BA}}^{(t)}\ y_{\text{BB}}^{(t)} \end{bmatrix}^\top,$$

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

With these definitions, the hierarchical structure is fully captured by Equation (1).

2.2 Reconciliation methods

Hierarchical time series forecasting is a process that adjusts, or "reconciles," forecast values to ensure they satisfy the hierarchical aggregation constraints. We assume that observations for each series are available for an observation period $t = 1, \ldots, T$, and the objective is to forecast values for the forecast period $\tau = T+1, \ldots, T+T'$.

Forecasting, ignoring any aggregation constraints, is called a base forecast. Let $\hat{\boldsymbol{y}}^{(\tau)} \in \mathbb{R}^n$ denote the vector of base forecasts for all series at time τ. Forecasting that satisfies the hierarchical aggregation constraints is called a coherent forecast. Let $\tilde{\boldsymbol{y}}^{(\tau)} \in \mathbb{R}^n$ denote the vector of coherent forecasts at time τ. To transform base forecasts into coherent forecasts, hierarchical forecasting methods estimate a reconciliation matrix $\boldsymbol{P} \in \mathbb{R}^{m\times n}$ that maps the base forecasts to the bottom-level. All hierarchical forecasting approaches can be expressed in the general form:

$$\tilde{\boldsymbol{y}}^{(\tau)} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(\tau)}, \tag{2}$$

where $\boldsymbol{S}$ is the summing matrix defined in Equation (1).

The bottom-up and top-down approaches are two traditional reconciliation methods. We again use the example from Figure 1 to illustrate these concepts. In the bottom-up approach, coherent forecasts are derived by summing the base forecasts of the bottom-level series. This corresponds to the reconciliation matrix:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Conversely, in the top-down approach, coherent forecasts are obtained by disaggregating the base forecast of the top-level series. If p_X is the proportion that allocates the total values to each bottom-level series X, the reconciliation matrix becomes:

$$\boldsymbol{P} = \begin{bmatrix} p_{\mathrm{AA}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{\mathrm{AB}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{\mathrm{AC}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{\mathrm{BA}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{\mathrm{BB}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A common way to determine these proportions is based on the average historical proportions of the data.

Because the bottom-up and top-down methods utilize base forecasts from only a single level of aggregation, they rely on limited information. To overcome this, subsequent research has proposed methods that use base forecasts from all series to estimate a reconciliation matrix, thus producing more comprehensive coherent forecasts.

Hyndman et al. (2011) proposed the generalized least squares (GLS) reconciliation, a regression-based approach. In this method, the reconciliation matrix is chosen to minimize errors between the base forecasts and the coherent forecasts. Specifically, consider the regression model for base forecasts

$$\hat{\boldsymbol{y}}^{(\tau)} = \boldsymbol{S}\boldsymbol{\beta}^{(\tau)} + \boldsymbol{\varepsilon}^{(\tau)}.$$

Let $\boldsymbol{\beta}^{(\tau)} = \mathbb{E}\left[\boldsymbol{b}^{(\tau)} \mid \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}\right] \in \mathbb{R}^m$ be the expectation of base forecasts at the bottom-level, and let the error term $\boldsymbol{\varepsilon}^{(\tau)}$ have zero mean with covariance matrix $\boldsymbol{\Sigma}^{(\tau)} = \mathrm{Var}\left(\boldsymbol{\varepsilon}^{(\tau)} \mid \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}\right) \in \mathbb{R}^{n\times n}$. If $\boldsymbol{\Sigma}^{(\tau)}$ were known, the minimum variance unbiased estimator of $\boldsymbol{\beta}^{(\tau)}$ would be the GLS estimator

$$\hat{\boldsymbol{\beta}}^{(\tau)} = \left(\boldsymbol{S}^\top \boldsymbol{\Sigma}^{(\tau)\dagger}\boldsymbol{S}\right)^{-1} \boldsymbol{S}^\top \boldsymbol{\Sigma}^{(\tau)\dagger}\hat{\boldsymbol{y}}^{(\tau)},$$

where $\boldsymbol{\Sigma}^{(\tau)\dagger}$ is the generalized inverse of $\boldsymbol{\Sigma}^{(\tau)}$. Comparing this expression with Equation (2) gives

$$\boldsymbol{P} = \left(\boldsymbol{S}^\top \boldsymbol{\Sigma}^{(\tau)\dagger}\boldsymbol{S}\right)^{-1} \boldsymbol{S}^\top \boldsymbol{\Sigma}^{(\tau)\dagger}.$$

In practice, however, the covariance matrix $\boldsymbol{\Sigma}^{(\tau)}$ is unknown and cannot be estimated. It represents the covariance of the reconciliation errors at time τ, but the errors are not observable until coherent forecasts have already been produced. Hyndman et al. (2011) shows that, under the strong assumption that the errors themselves satisfy the aggregation constraints, $\boldsymbol{\Sigma}^{(\tau)}$ can be replaced with the n-dimensional identity matrix $\boldsymbol{I}_n$. This replacement is equivalent to computing the ordinary least squares (OLS) estimator instead of the GLS estimator.

Wickramasuriya et al. (2019) proposed the minimum-trace (MinT) reconciliation, which does not rely on the assumptions required by Hyndman et al. (2011). MinT chooses the reconciliation matrix to minimize the variance of errors between the observation values and the coherent forecasts. Let

$$\boldsymbol{V}^{(\tau)} = \mathrm{Var}\left(\boldsymbol{y}^{(\tau)} - \tilde{\boldsymbol{y}}^{(\tau)} \mid \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}\right) \in \mathbb{R}^{n \times n}$$

denote the covariance matrix of those errors and define the covariance matrix of the errors between observation values and base forecasts as

$$\boldsymbol{W}^{(\tau)} = \mathrm{Var}\left(\boldsymbol{y}^{(\tau)} - \hat{\boldsymbol{y}}^{(\tau)} \mid \boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)}\right) \in \mathbb{R}^{n \times n}.$$

With these matrices, we can derive the relationship as follows:

$$\boldsymbol{V}^{(\tau)} = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}^{(\tau)}\boldsymbol{P}^\top\boldsymbol{S}^\top.$$

Then the optimal reconciliation matrix, which minimizes the total error variance $\mathrm{tr}\left(\boldsymbol{V}^{(\tau)}\right)$ is given by

$$\boldsymbol{P} = \left(\boldsymbol{S}^\top \left(\boldsymbol{W}^{(\tau)}\right)^{-1} \boldsymbol{S}\right)^{-1} \boldsymbol{S}^\top \left(\boldsymbol{W}^{(\tau)}\right)^{-1}.$$

Although the covariance matrix $\boldsymbol{W}^{(\tau)}$ is unknown, it can be estimated using the unbiased sample variance calculated from the observations and the base forecasts at times $t = 1, \ldots, T$. Because the same estimate is applied to every forecast period $\tau = T+1, \ldots, T+T'$, we denote it by $\boldsymbol{W}$ and set $\boldsymbol{W}^{(\tau)} = \boldsymbol{W}$ for all τ. The covariance matrix is always positive semidefinite but not always positive definite, and when it is not positive definite, it cannot be inverted. A common remedy is to apply a shrinkage approach (Schäfer & Strimmer, 2005) to obtain a well-conditioned matrix instead of the original covariance matrix.

Wickramasuriya et al. (2019) proposed MinT as an approach that minimizes the variance of the errors between the observation values and the coherent forecasts, while Panagiotelis et al. (2021) showed that MinT is equivalent to minimizing the expected value of the errors. Rewriting $\mathrm{tr}\left(\boldsymbol{V}^{(\tau)}\right)$ gives

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{V}^{(\tau)}\right) &= \mathrm{tr}\left(\mathbb{E}\left[\left(\boldsymbol{y}^{(\tau)} - \tilde{\boldsymbol{y}}^{(\tau)}\right)\boldsymbol{W}\left(\boldsymbol{y}^{(\tau)} - \tilde{\boldsymbol{y}}^{(\tau)}\right)^\top\right]\right) \\ &= \mathbb{E}\left[\left(\boldsymbol{y}^{(\tau)} - \tilde{\boldsymbol{y}}^{(\tau)}\right)^\top \boldsymbol{W}\left(\boldsymbol{y}^{(\tau)} - \tilde{\boldsymbol{y}}^{(\tau)}\right)\right].\end{aligned}$$

Consequently, computing the reconciliation matrix by minimizing $\mathrm{tr}\left(\boldsymbol{V}^{(\tau)}\right)$ is equivalent to solving the optimization problem

$$\min_{\boldsymbol{P}} \quad \mathbb{E}\left[\left(\boldsymbol{y}^{(\tau)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(\tau)}\right)^\top \boldsymbol{W}\left(\boldsymbol{y}^{(\tau)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(\tau)}\right)\right]. \tag{3}$$

3 Robust reconciliation

In MinT, an approach of hierarchical time series forecasting, the optimal solution of the optimization problem (3) is the reconciliation matrix. The covariance matrix $\boldsymbol{W}$ used in this problem is estimated from the errors between the observed values and the base forecasts for the observation period, and may not match the true covariance matrix for the forecast period. Therefore, the optimal solution of the problem (3) using the estimated covariance matrix does not necessarily minimize the expected value of the error in the forecast period. In this paper, we propose an approach to determine the reconciliation matrix by solving a robust optimization problem, considering the uncertainty in the covariance matrix. The robust optimization problem minimizes the forecast error between the observed values and the coherent forecasts under a covariance matrix that maximizes the forecast error among possible covariance matrices.

3.1 Formulation

Suppose that the expectation in the problem (3) is taken with respect to the empirical distribution constructed from the observation-period data. Then, the problem (3) can be written as the following optimization problem:

$$\min_{\boldsymbol{P}} \quad \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right)^{\top} \boldsymbol{W} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right), \tag{4}$$

where the $\hat{\boldsymbol{y}}^{(t)}$ in the observation period is the value of the forecast model used for the base forecasts. We introduce uncertainty to $\boldsymbol{W}$. Let $\mathcal{W} \subseteq \mathbb{R}^{n \times n}$ be an uncertainty set. A robust optimization problem that minimizes the forecast error in the worst-case situation of $\boldsymbol{W} \in \mathcal{W}$ is formulated as

$$\min_{\boldsymbol{P}} \max_{\boldsymbol{W} \in \mathcal{W}} \quad \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right)^{\top} \boldsymbol{W} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right),$$

where we omit the coefficient $1/T$ from the objective function, since it does not affect the optimal solution.

In robust optimization problems with uncertainty of the covariance matrix, a box uncertainty set is often used (Lobo & Boyd, 2000; Halldórsson & Tütüncü, 2003). The box uncertainty set places upper and lower bounds on the covariance matrix, so that

$$\mathcal{W} = \left\{ \boldsymbol{W} \mid \underline{\boldsymbol{W}} \leq \boldsymbol{W} \leq \overline{\boldsymbol{W}}, \ \boldsymbol{W} \succeq \boldsymbol{O} \right\},$$

where $\overline{\boldsymbol{W}}, \underline{\boldsymbol{W}} \in \mathbb{R}^{n \times n}$ are the upper and lower bounds of $\boldsymbol{W}$, respectively. Inequality for matrix implies that the inequality holds for each element of the matrix, i.e., for any $i, j = 1, \ldots, n$, $\underline{W}_{ij} \leq W_{ij} \leq \overline{W}_{ij}$. $\boldsymbol{W} \succeq \boldsymbol{O}$ denotes that $\boldsymbol{W}$ is positive semidefinite. Therefore, the proposed method determines the reconciliation matrix by solving the following robust optimization problem:

$$\begin{aligned} \min_{\boldsymbol{P}} \max_{\boldsymbol{W}} \quad & \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right)^{\top} \boldsymbol{W} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)} \right) \\ \text{s.t.} \quad & \underline{\boldsymbol{W}} \leq \boldsymbol{W} \leq \overline{\boldsymbol{W}}, \\ & \boldsymbol{W} \succeq \boldsymbol{O}. \end{aligned} \tag{5}$$

3.2 Equivalent reformulation

We show that the min-max problem (5) is equivalently reformulated as a semidefinite optimization problem following the approach of Lobo & Boyd (2000) and Halldórsson & Tütüncü (2003). Let $\boldsymbol{I}$ and $\boldsymbol{O}$ be the identity matrix and zero matrix of appropriate dimensions, respectively. For two symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same size, $\boldsymbol{A} \bullet \boldsymbol{B}$ denotes the standard inner product of the two matrices, defined as $\boldsymbol{A} \bullet \boldsymbol{B} := \operatorname{tr}\left(\boldsymbol{A}^{\top}\boldsymbol{B}\right)$, which is the sum of the element-wise product of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Proposition 1. *Assume that there exists a positive definite matrix $\boldsymbol{W}'$ satisfying $\underline{\boldsymbol{W}} < \boldsymbol{W}' < \overline{\boldsymbol{W}}$. Then, the problem (5) can be equivalently reformulated as the following semidefinite optimization problem:*

$$\begin{aligned} \min_{\overline{\boldsymbol{X}}, \underline{\boldsymbol{X}}, \boldsymbol{E}, \boldsymbol{P}} \quad & \overline{\boldsymbol{W}} \bullet \overline{\boldsymbol{X}} - \underline{\boldsymbol{W}} \bullet \underline{\boldsymbol{X}} \\ \text{s.t.} \quad & \begin{bmatrix} \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} & \boldsymbol{E} \\ \boldsymbol{E}^{\top} & \boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{O}, \\ & \boldsymbol{E} = \left[\boldsymbol{y}^{(1)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(1)} \ \cdots \ \boldsymbol{y}^{(T)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(T)} \right], \\ & \overline{\boldsymbol{X}}, \ \underline{\boldsymbol{X}} \geq \boldsymbol{O}, \end{aligned} \tag{6}$$

where $\overline{\boldsymbol{X}}$ and $\underline{\boldsymbol{X}}$ are n-dimensional symmetric matrix variables, and $\boldsymbol{E}$ is a $n \times T$-dimensional matrix variable whose columns are the error vectors between the observed values and coherent forecasts over the observation period.

Proof. Fix $\boldsymbol{P}$ and consider the inner maximization of the problem (5):

$$\begin{aligned} \max_{\boldsymbol{W}} \quad & \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right)^{\top} \boldsymbol{W} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right) \\ \text{s.t.} \quad & \underline{\boldsymbol{W}} \leq \boldsymbol{W} \leq \overline{\boldsymbol{W}}, \\ & \boldsymbol{W} \succeq \boldsymbol{O}. \end{aligned} \tag{7}$$

Its dual problem is formulated as

$$\begin{aligned} \min_{\overline{\boldsymbol{X}}, \underline{\boldsymbol{X}}} \quad & \overline{\boldsymbol{W}} \bullet \overline{\boldsymbol{X}} - \underline{\boldsymbol{W}} \bullet \underline{\boldsymbol{X}} \\ \text{s.t.} \quad & \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} - \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right) \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right)^{\top} \succeq \boldsymbol{O}, \\ & \overline{\boldsymbol{X}},\ \underline{\boldsymbol{X}} \geq \boldsymbol{O}, \end{aligned} \tag{8}$$

where $\overline{\boldsymbol{X}}$, $\underline{\boldsymbol{X}}$ are dual variables. Let us define the matrix $\boldsymbol{E} = \left[\boldsymbol{y}^{(1)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(1)} \ \cdots \ \boldsymbol{y}^{(T)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(T)}\right]$. Then, by the Schur complement, the semidefinite constraint in the problem (8) is equivalently transformed as follows:

$$\begin{aligned} \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} - \sum_{t=1}^{T} \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right) \left(\boldsymbol{y}^{(t)} - \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}^{(t)}\right)^{\top} \succeq \boldsymbol{O} & \iff \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} - \boldsymbol{E}\boldsymbol{E}^{\top} \succeq \boldsymbol{O} \\ & \iff \begin{bmatrix} \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} & \boldsymbol{E} \\ \boldsymbol{E}^{\top} & \boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{O}. \end{aligned} \tag{9}$$

Therefore, we can replace the semidefinite constraint of the problem (8) with Equation (9), which yields the following equivalent dual problem:

$$\begin{aligned} \min_{\overline{\boldsymbol{X}}, \underline{\boldsymbol{X}}} \quad & \overline{\boldsymbol{W}} \bullet \overline{\boldsymbol{X}} - \underline{\boldsymbol{W}} \bullet \underline{\boldsymbol{X}} \\ \text{s.t.} \quad & \begin{bmatrix} \overline{\boldsymbol{X}} - \underline{\boldsymbol{X}} & \boldsymbol{E} \\ \boldsymbol{E}^{\top} & \boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{O}, \\ & \overline{\boldsymbol{X}},\ \underline{\boldsymbol{X}} \geq \boldsymbol{O}. \end{aligned} \tag{10}$$

Finally, we show the dual problem (10) has a strictly feasible solution, which implies that the strong duality between the problems (7) and (10) holds. Let $\boldsymbol{J} \in \mathbb{R}^{n \times n}$ be a matrix whose elements are all one. For the fixed $\boldsymbol{P}$, we consider the following solution:

$$\overline{\boldsymbol{X}}' = \boldsymbol{E}\boldsymbol{E}^{\top} + \alpha\boldsymbol{J} + \boldsymbol{I}, \quad \underline{\boldsymbol{X}}' = \alpha\boldsymbol{J},$$

where $\alpha > 0$ is a sufficiently large constant such that $\boldsymbol{E}\boldsymbol{E}^{\top} + \alpha\boldsymbol{J} > \boldsymbol{O}$. Since we take $\alpha > 0$ sufficiently large, $\overline{\boldsymbol{X}}', \underline{\boldsymbol{X}}' > \boldsymbol{O}$. Focusing on the semidefinite constraint of the problem (10), we have

$$\begin{bmatrix} \overline{\boldsymbol{X}}' - \underline{\boldsymbol{X}}' & \boldsymbol{E} \\ \boldsymbol{E}^{\top} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{E}\boldsymbol{E}^{\top} + \boldsymbol{I} & \boldsymbol{E} \\ \boldsymbol{E}^{\top} & \boldsymbol{I} \end{bmatrix} \succ \boldsymbol{O},$$

thus $\left(\overline{\boldsymbol{X}}', \underline{\boldsymbol{X}}'\right)$ is strictly feasible. Since we assume that the primal problem (7) is strictly feasible, the strong duality theorem holds and the optimal values of the two problems (7) and (10) are equal (Vandenberghe & Boyd, 1996). As we took $\boldsymbol{P}$ arbitrarily, the above argument holds for all $\boldsymbol{P}$. Therefore, the min-max problem (5) is equivalently reformulated as the semidefinite optimization problem (6). $\square$

Hence, the approach of the proposed method is to solve the semidefinite optimization problem (6) and use the optimal solution for $\boldsymbol{P}$ as the reconciliation matrix for hierarchical time series forecasts.

3.3 Uncertainty set

In order to deal with the robust optimization problem described in the previous sections, it is necessary to determine the uncertainty set, i.e., $\overline{\boldsymbol{W}}$ and $\underline{\boldsymbol{W}}$ in advance. We set the upper and lower bounds of the uncertainty set from the observation period data using bootstrap with reference to Bertsimas et al. (2018).

We summarize the method for determining the uncertainty set in Algorithm 1. In the first step, calculate a parameter λ for the shrinkage approach, similar to the existing hierarchical time series forecasting described in Section 2.2. This shrinkage approach is applied to the covariance matrix estimated by unbiased variance from the observation period data. In the next step, estimate the covariance matrix of each sample obtained by sampling the data of the observation period. For sampling, we select the same number of time points as the observation period T with replacement. Then, using the shrinkage intensity parameter λ obtained in the first step, the shrinkage approach is applied to the sampled covariance matrix. This sampling is repeated $N_B \geq 1$ times to obtain N_B covariance matrices. In the last step, determine upper and lower bounds from each element of the sampled covariance matrices. Let $0 < \alpha \leq 1$, then set the width of the uncertainty set to be α times the width of the maximum and minimum sample values of the covariance matrix.

Algorithm 1 Designing uncertainty set

Input: $n, T, \{\boldsymbol{y}^{(t)}\}_{t=1}^{T}, \{\hat{\boldsymbol{y}}^{t)}\}_{t=1}^{T}, N_B, \alpha$
Output: $\overline{\boldsymbol{W}}, \underline{\boldsymbol{W}}$
  $\boldsymbol{W} \leftarrow$ estimate covariance matrix from $\{\boldsymbol{y}^{(t)}\}_{t=1}^{T}$ and $\{\hat{\boldsymbol{y}}^{(t)}\}_{t=1}^{T}$
  $\lambda \leftarrow$ calculate shrinkage intensity parameter of $\boldsymbol{W}$
  for $s = 1, \ldots, N_B$ **do**
    $\{\boldsymbol{y}^{(t')}\}_{t'=1}^{T}, \{\hat{\boldsymbol{y}}^{(t')}\}_{t'=1}^{T} \leftarrow$ sample T data points from $\{\boldsymbol{y}^{(t)}\}_{t=1}^{T}$ and $\{\hat{\boldsymbol{y}}^{(t)}\}_{t=1}^{T}$ with replacement
    $\boldsymbol{W}_{(s)} \leftarrow$ estimate covariance matrix from $\{\boldsymbol{y}^{(t')}\}_{t'=1}^{T}$ and $\{\hat{\boldsymbol{y}}^{(t')}\}_{t'=1}^{T}$
    $\boldsymbol{W}_{(s)} \leftarrow$ update $\boldsymbol{W}_{(s)}$ by shrinkage estimator with λ
  end for
  for $i = 1, \ldots, n,\ j = 1, \ldots, n$ **do**
    $\overline{W}_{ij} \leftarrow$ calculate $100 \cdot (1 + \alpha)/2$-percentile of $\{W_{(s)ij}\}_{s=1}^{N_B}$
    $\underline{W}_{ij} \leftarrow$ calculate $100 \cdot (1 - \alpha)/2$-percentile of $\{W_{(s)ij}\}_{s=1}^{N_B}$
  end for
  return $\overline{\boldsymbol{W}}, \underline{\boldsymbol{W}}$

4 Numerical experiments

To verify the effectiveness of the proposed method, we compared its forecast accuracy with that of existing hierarchical time series forecasting methods across multiple real-world datasets.

4.1 Datasets

The numerical experiments used four datasets from prior studies. An overview of each dataset is provided below, including the number of hierarchical levels and series, as well as the lengths of the observation and forecast periods. These four datasets allowed us to assess the proposed method's performance under various real-world conditions, including hierarchies with different scales of levels and series.

The first dataset is the Australian births data (Hyndman & Athanasopoulos, 2021). It records the number of births in Australia every month from January 1975 to December 2022. For this experiment, the first 516 months (January 1975 to December 2017) served as the observation period, while the following 60 months (January 2018 to November 2022) comprised the forecast period. As summarized in Table 1, the hierarchy consisted of a single disaggregation level ($K = 1$) with nine bottom-level series ($m = 9$) and ten series in total ($n = 10$). Level 1, the bottom level, disaggregated the national series into nine states and territories: ACT, AUS, NSW, NT, QLD, SA, TAS, VIC, and WA.

Table 1: Hierarchy for Australian births data

Level	Number of series	Total series per level
total	1	1
state	9	9

The second dataset is the Australian tourism data (Athanasopoulos et al., 2009), which records the number of domestic travelers every quarter from Q1 1998 to Q4 2017. Here, the observation period consisted of the first 68 quarters (Q1 1998 to Q4 2014), and the forecast period covered the subsequent 12 quarters (Q1 2015 to Q4 2017). As Table 2 shows, the hierarchy was structured with $K = 2$, $m = 27$, and $n = 35$. Level 1 disaggregated the national total into seven states: NSW, NT, QLD, SA, TAS, VIC, and WA. Note that this state grouping differed from the one used for the births dataset, as we followed prior work. Level 2, the bottom level, further breaked down each state into finer geographic zones.

Table 2: Hierarchy for Australian tourism data

Level	Number of series	Total series per level
total	1	1
state	7	7
zone	6-2-4-4-3-5-3	27

The third dataset is the U.S. Walmart sales data (Mancuso et al., 2021). It tracks weekly sales from January 3, 2011, to May 29, 2016. In this experiment, we used the first 261 weeks (January 3, 2011, to January 3, 2016) as the observation period and the next 21 weeks (January 4, 2016, to May 29, 2016) for the forecast period. Table 3 shows the hierarchy as $K = 2$, $m = 10$, and $n = 14$. Level 1 splited the national total into three states: CA, TX, and WI. Level 2, the bottom level, further subdivided each state into its constituent stores: four in CA, three in TX, and three in WI.

Table 3: Hierarchy for Walmart sales data

Level	Number of series	Total series per level
total	1	1
state	3	3
store	4-3-3	10

The fourth dataset is the Swiss electricity demand data (Nespoli et al., 2020), recording electricity supply every ten minutes from January 13, 2018, to January 19, 2019. After converting the data to a daily unit, we used the first 353 days (January 13, 2018, to December 31, 2018) for the observation period and the subsequent 19 days (January 1, 2019, to January 19, 2019) for the forecast period. As outlined in Table 4, the hierarchy was $K = 3$, $m = 24$, and $n = 31$. Level 1 disaggregated the grid into two synthetic meter aggregations: S1 and S2. Level 2 further subdivided each aggregation into two synthetic sub-aggregations: S11 and S12 for S1, and S21 and S22 for S2. Level 3, the bottom level, then separated each sub-aggregation into six individual meters.

4.2 Experimental setup

This subsection details the benchmark methods, our proposed robust methods, and evaluation metrics.

Hierarchical forecasting first requires a set of base forecasts (**Base**), which do not account for the hierarchical structure. We generated these base forecasts using Prophet (Taylor & Letham, 2018), an open-source time series library from Meta, with its default settings. As a preliminary experiment, we also performed base

Table 4: Hierarchy for Swiss electricity demand data

Level	Number of series	Total series per level
grid	1	1
agg.	2	2
sub-agg.	2-2	4
meter	6-6-6-6	24

forecasts using other time series forecasting methods such as ARIMA and LightGBM, but Prophet showed the best results. Our comparative methods included the bottom-up (**BU**), top-down (**TD**), GLS reconciliation (**GLS**), and MinT reconciliation (**MinT**) approaches introduced in Section 2.

For our proposed method (**Robust**), the optimization problem (6) was solved using the MOSEK solver (MOSEK ApS, 2025). Moreover, we needed to decide two parameters for bootstrap to design uncertainty set, which was the number of sampleing N_B and a width of the uncertainty set α. In this experiments, N_B was fixed at 5000 and α was determined by validation from among the candidates specified in advance. For validation, we divided the observation period data into train and validation data in a ratio of 9:1, and used the α with the smallest RMSE in the validation data. The candidates for α were 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.

As evaluation metrics, we used the mean absolute error (MAE) and the root mean squared error (RMSE) for each series. For a given series X, these are defined as:

$$\mathrm{MAE} = \frac{1}{T'} \sum_{\tau=T+1}^{T+T'} \left| y_X^{(\tau)} - \tilde{y}_X^{(\tau)} \right|, \ \mathrm{RMSE} = \sqrt{\frac{1}{T'} \sum_{\tau=T+1}^{T+T'} \left(y_X^{(\tau)} - \tilde{y}_X^{(\tau)} \right)^2}$$

Note that RMSE gives a harsh evaluation when the prediction deviates significantly from the MAE.

4.3 Results and discussion

The experimental results for all datasets are summarized in Table 5–8. To make the results easier to understand, we calculated the ratio of MAE and RMSE for each hierarchical time series forecsting method when the MAE and RMSE of base forecast are set to 1, and then calculated the mean and standard deviation for series included in the same hierarchical level. We defined these metrics as relative MAE and relative RMSE, respectively. The original MAE and RMSE of all series are listed in the appendix. In the tables, the format is "mean ± standard deviation." Since top level of hierarchical structure (Level 0) includes only one series, the standard deviation for the series is not shown. The underlined values indicate the best forecast accuracy in terms of the mean for the corresponding hierarchical level.

For the Australian births dataset, the validation to decide a parameter for bootstrap resulted in $\alpha = 1.0$. The proposed method achieved the highest accuracy in most cases. In the top level, our proposed method was the only approach that achieved higer accuracy than base forecast. In the bottom level, considering the standard deviation, some series showed significant improvement in forecast accuracy with the proposed method.

On the Australian tourism dataset, through the validation process, the bootstrap parameter was determined to be $\alpha = 0.5$. Our proposed method yielded lower forecast accuracy than the existing GLS and MinT reconciliation methods except for the top level. This underperformance seemed to be due to the small discrepancy between the forecast error covariance matrix estimated from the observation period and the true covariance matrix in the forecast period. This reduced the benefit of explicitly accounting for covariance uncertainty, so our method was not the most effective for this type of data.

For the Walmart sales, the validation procedure for selecting the bootstrap parameter yielded a value of $\alpha = 0.9$. Our method achieved the highest forecast accuracy for all levels, with substantial performance gains over existing methods.

Table 5: Forecast accuracy for Australian births data

(a) Relative MAE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	1.027	1.000	1.003	1.003	0.929
state	1.000	1.000 ± 0.000	1.518 ± 1.224	0.963 ± 0.104	0.985 ± 0.031	0.949 ± 0.146

(b) Relative RMSE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	1.011	1.000	1.001	1.001	0.970
state	1.000	1.000 ± 0.000	1.300 ± 0.781	0.976 ± 0.069	0.992 ± 0.019	0.976 ± 0.091

Table 6: Forecast accuracy for Australian tourism data

(a) Relative MAE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	2.315	1.000	1.125	0.663	0.647
state	1.000	1.277 ± 0.229	0.968 ± 0.371	0.635 ± 0.168	0.573 ± 0.125	0.779 ± 0.415
zone	1.000	1.000 ± 0.000	0.925 ± 0.373	0.623 ± 0.163	0.568 ± 0.155	0.840 ± 0.396

(b) Relative RMSE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	2.132	1.000	1.105	0.698	0.638
state	1.000	1.227 ± 0.179	0.989 ± 0.360	0.680 ± 0.171	0.617 ± 0.125	0.789 ± 0.377
zone	1.000	1.000 ± 0.000	0.959 ± 0.360	0.657 ± 0.151	0.606 ± 0.138	0.841 ± 0.349

Table 7: Forecast accuracy for Walmart sales data

(a) Relative MAE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	1.141	1.000	1.012	0.964	0.884
state	1.000	1.132 ± 0.110	1.260 ± 0.800	1.000 ± 0.027	0.953 ± 0.045	0.864 ± 0.027
store	1.000	1.000 ± 0.000	1.601 ± 0.929	0.878 ± 0.071	0.859 ± 0.085	0.791 ± 0.145

(b) Relative RMSE

	Base	BU	TD	GLS	MinT	Robust
total	1.000	1.136	1.000	1.012	0.965	0.887
state	1.000	1.128 ± 0.110	1.237 ± 0.770	0.999 ± 0.027	0.952 ± 0.046	0.866 ± 0.030
store	1.000	1.000 ± 0.000	1.564 ± 0.877	0.880 ± 0.070	0.863 ± 0.081	0.798 ± 0.135

On the Swiss electricity demand datasets, as determined by validation, the bootstrap parameter was set to $\alpha = 0.7$. Our proposed method achieved the highest forecast accuracy at the upper levels, but its accuracy deteriorated at the bottom level compared to the base forecast. In other words, in order to improve the forecast accuracy of the upper levels, the forecast accuracy of the bottom level was being sacrificed. Similar to the Australian tourism data, the gap between the true covariance matrix and the estimated covariance matrix was small, so the effect of robustification was considered to be small.

Table 8: Forecast accuracy for Swiss electricity demand data

(a) Relative MAE

	Base	BU	TD	GLS	MinT	Robust
grid	1.000	1.050	1.000	1.020	1.119	$\underline{0.893}$
agg.	1.000	1.052 ± 0.038	0.990 ± 0.294	0.988 ± 0.009	1.004 ± 0.004	$\underline{0.817 \pm 0.187}$
sub-agg.	1.000	1.072 ± 0.066	1.393 ± 0.485	0.996 ± 0.013	1.016 ± 0.086	$\underline{0.871 \pm 0.200}$
meter	1.000	1.000 ± 0.000	1.809 ± 0.895	$\underline{0.984 \pm 0.102}$	1.023 ± 0.171	1.039 ± 0.265

(b) Relative RMSE

	Base	BU	TD	GLS	MinT	Robust
grid	1.000	1.025	1.000	1.011	1.052	$\underline{0.948}$
agg.	1.000	1.020 ± 0.037	0.976 ± 0.148	0.990 ± 0.003	1.012 ± 0.013	$\underline{0.902 \pm 0.115}$
sub-agg.	1.000	1.033 ± 0.055	1.169 ± 0.312	0.992 ± 0.004	1.026 ± 0.063	$\underline{0.931 \pm 0.137}$
meter	1.000	1.000 ± 0.000	1.713 ± 0.757	$\underline{0.978 \pm 0.088}$	1.020 ± 0.137	1.015 ± 0.217

The numerical experiments on these four datasets demonstrated the effectiveness of our proposed method for hierarchical time series forecasting. Across most target series and hierarchical levels, the proposed approach consistently achieved higher forecast accuracy than existing techniques. However, its accuracy was slightly lower than that of the GLS and MinT reconciliation methods on a few datasets. As mentioned in the paragraph on the results of the Australian tourism data, when the gap between the true covariance matrix and the estimated covariance matrix is small, the proposed method is not considered superior. In fact, while validation yielded a large α for Australian births data and Walmart sales data where the proposed method was highly effective, it yielded a small α for Australian tourism data and Swiss electricity demand data where the method was less effective. Therefore, if you want to know in advance whether the proposed method is effective for the data, we recommend obtaining the parameter α from the observation period data and checking whether the α is large.

5 Conclusion

In this paper, we proposed a robust hierarchical time series forecasting method. This approach introduces an uncertainty set for the covariance matrix of forecast errors and minimizes the forecast error between the observation values and the coherent forecasts. Through numerical experiments, we demonstrated that the proposed method often provides more accurate forecasts than existing hierarchical time series forecasting methods, although its performance can vary across different datasets.

The limitations of our method are twofold: its accuracy may not always surpass existing methods, and its scalability is limited due to the computational demands of the optimization problem. The first issue is as stated in the discussion of the experimental results. It arises when the discrepancy between the true and estimated covariance matrices is small, in which case the robust approach offers little advantage. The second limitation concerns the computational time required to obtain a reconciliation matrix, as our method relies on solving a semidefinite optimization problem. The size of the optimization problem depends on the total number of series and the length of observation periods. While a solution was achievable in tens of seconds for the dataset sizes used in our experiments, this approach would not be practical for very large-scale predictions.

References

George Athanasopoulos, Roman A Ahmed, and Rob J Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166, 2009.

Yossi Aviv. A time-series framework for supply-chain inventory management. *Operations Research*, 51(2): 210–227, 2003.

Aharon Ben-Tal and Arkadi Nemirovski. Robust convex optimization. *Mathematics of operations research*, 23(4):769–805, 1998.

Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. Data-driven robust optimization. *Mathematical Programming*, 167:235–292, 2018.

John B Edwards and Guy H Orcutt. Should aggregation prior to estimation be the rule? *The Review of Economics and Statistics*, pp. 409–420, 1969.

Yehuda Grunfeld and Zvi Griliches. Is aggregation necessarily bad? *The review of economics and statistics*, pp. 1–13, 1960.

Bjarni V Halldórsson and Reha H Tütüncü. An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications*, 116:559–590, 2003.

Luis Hernandez, Carlos Baladron, Javier M Aguiar, Belen Carro, Antonio J Sanchez-Esguevillas, Jaime Lloret, and Joaquim Massana. A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings. *IEEE Communications Surveys & Tutorials*, 16(3):1460–1495, 2014.

Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and practice*. OTexts, Melbourne, Australia, 3rd edition, 2021. URL `https://OTexts.com/fpp3`. Accessed on 2025-05-06.

Rob J Hyndman, Roman A Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9):2579–2589, 2011.

Bjoern Krollner, Bruce Vanstone, and Gavin Finnie. Financial time series forecasting with machine learning techniques: A survey. In *European Symposium on Artificial Neural Networks: Computational Intelligence and Machine Learning*, pp. 25–30, 2010.

Zhenyu Liu, Zhengtong Zhu, Jing Gao, and Cheng Xu. Forecast methods for time series data: A survey. *IEEE Access*, 9:91896–91912, 2021.

Miguel Sousa Lobo and Stephen Boyd. The worst-case risk of a portfolio. Technical report, 2000. URL `https://web.stanford.edu/~boyd/papers/pdf/risk_bnd.pdf`. Accessed on 2025-05-06.

Ganapathy Mahalakshmi, S Sridevi, and Shyamsundar Rajaram. A survey on forecasting of time series data. In *2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16)*, pp. 1–8. IEEE, 2016.

Paolo Mancuso, Veronica Piccialli, and Antonio M Sudoso. A machine learning approach for forecasting hierarchical time series. *Expert Systems with Applications*, 182, 2021.

Jan Kloppenborg Møller, Peter Nystrup, Poul G Hjorth, and Henrik Madsen. Optimal forecast reconciliation with uncertainty quantification. *arXiv preprint arXiv:2402.06480*, 2024.

MOSEK ApS. *The MOSEK Python Fusion API manual. Version 11.0.*, 2025. URL `https://docs.mosek.com/latest/pythonfusion/index.html`. Accessed on 2025-05-06.

Lorenzo Nespoli, Vasco Medici, Kristijan Lopatichki, and Fabrizio Sossan. Hierarchical demand forecasting benchmark for the distribution grid. *Electric Power Systems Research*, 189:106755, 2020.

Guy H Orcutt, Harold W Watts, and John B Edwards. Data aggregation and information loss. *The American Economic Review*, 58(4):773–787, 1968.

Anastasios Panagiotelis, George Athanasopoulos, Puwasala Gamakumara, and Rob J Hyndman. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, 37(1):343–359, 2021.

Patrícia Ramos, Nicolau Santos, and Rui Rebelo. Performance of state space and arima models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing*, 34:151–163, 2015.

Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

L Suganthi and Anand A Samuel. Energy models for demand forecasting—a review. *Renewable and sustainable energy reviews*, 16(2):1223–1240, 2012.

Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

Reha H Tütüncü and Mark Koenig. Robust asset allocation. *Annals of Operations Research*, 132:157–187, 2004.

Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*, 2022.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819, 2019.

A Full results of numerical experiments

This section reports the original MAE and RMSE for all series from the numerical experiments that could not be included in Section 4.3. Table 9 and Table 10 present the MAE and RMSE for the Australian births data, respectively. Similarly, Table 11 and Table 12 report the forecast accuracy for the Australian tourism data, Table 13 and Table 14 for the Walmart sales data, and Table 15 and Table 16 for the Swiss electricity demand data. Within each table, series separated by horizontal lines correspond to the same hierarchical level.

Table 9: MAE for Australian births data

	Base	BU	TD	GLS	MinT	Robust
total	3716.98	3818.08	3716.98	3726.75	3728.20	3453.87
ACT	79.19	79.19	23.98	68.19	77.37	70.47
AUS	1864.44	1864.44	1858.23	1854.67	1834.54	1727.37
NSW	430.46	430.46	937.69	422.76	414.65	376.57
NT	41.22	41.22	43.60	31.95	39.70	36.66
QLD	221.98	221.98	230.76	218.12	219.48	204.42
SA	106.12	106.12	328.78	97.96	100.94	85.08
TAS	40.78	40.78	163.65	48.17	43.43	54.79
VIC	711.02	711.02	432.63	701.14	702.36	677.66
WA	483.61	483.61	179.59	472.60	474.25	455.24

Table 10: RMSE for Australian births data

	Base	BU	TD	GLS	MinT	Robust
total	6433.27	6504.39	6433.27	6440.33	6441.38	<u>6238.75</u>
ACT	93.86	93.86	<u>44.96</u>	84.08	92.21	86.18
AUS	3221.01	3221.01	3216.54	3213.95	3199.22	<u>3119.71</u>
NSW	779.88	779.88	1168.37	773.80	767.51	<u>739.04</u>
NT	54.88	54.88	57.73	<u>46.72</u>	53.51	50.69
QLD	451.99	451.99	462.24	447.91	449.34	<u>435.25</u>
SA	165.37	165.37	354.97	158.39	160.96	<u>147.93</u>
TAS	<u>55.31</u>	<u>55.31</u>	169.64	61.64	57.54	67.60
VIC	1215.67	1215.67	<u>1005.35</u>	1208.26	1209.18	1191.31
WA	630.09	630.09	<u>383.28</u>	620.83	622.21	606.17

Table 11: MAE for Australian tourism data

	Base	BU	TD	GLS	MinT	Robust
total	1507.73	3490.86	1507.73	1696.30	999.98	<u>975.44</u>
NSW	706.38	1056.55	<u>336.42</u>	594.78	445.01	680.28
NT	122.88	122.99	151.28	57.50	82.99	<u>49.88</u>
QLD	565.88	626.18	503.91	427.08	<u>206.32</u>	210.33
SA	213.29	278.99	90.94	82.64	83.73	<u>82.31</u>
TAS	132.18	132.50	158.77	<u>67.01</u>	84.52	79.48
VIC	509.36	828.63	532.14	416.27	<u>347.01</u>	694.33
WA	331.08	462.09	498.96	222.41	<u>207.11</u>	450.89
NSW_ACT	131.54	131.54	68.63	<u>63.37</u>	68.05	70.95
NSW_Nth	139.01	139.01	92.79	75.93	82.49	<u>53.79</u>
NSW_Sth	131.20	131.20	115.03	<u>77.85</u>	94.87	168.27
NSW_Metro	313.69	313.69	161.47	241.31	<u>142.86</u>	285.20
NSW_Nth_Coast	237.56	237.56	<u>119.50</u>	160.60	130.42	179.47
NSW_Sth_Coast	118.41	118.41	136.72	53.19	<u>44.83</u>	119.64
NT_Central	72.85	72.85	84.78	<u>37.56</u>	51.82	43.39
NT_Nth_Coast	50.14	50.14	67.35	<u>30.65</u>	31.72	70.02
QLD_Metro	360.78	360.78	209.51	318.68	197.90	<u>176.87</u>
QLD_Central_Coast	53.20	53.20	69.45	<u>33.78</u>	38.48	36.41
QLD_Inland	150.29	150.29	172.18	107.54	<u>64.82</u>	78.66
QLD_Nth_Coast	80.10	80.10	140.75	<u>58.20</u>	63.47	113.08
SA_Metro	127.55	127.55	<u>34.11</u>	76.69	36.13	35.48
SA_Inland	78.95	78.95	<u>30.53</u>	34.33	32.78	32.01
SA_West_Coast	27.14	27.14	24.60	31.01	<u>16.42</u>	19.22
SA_Sth_Coast	52.22	52.22	70.85	<u>37.05</u>	40.02	77.66
TAS_Nth_East	50.76	50.76	61.78	<u>29.15</u>	32.34	34.77
TAS_Sth	53.34	53.34	70.71	<u>39.37</u>	41.03	47.04
TAS_Nth_West	39.50	39.50	33.94	<u>23.07</u>	25.23	61.00
VIC_Nth_West	100.44	100.44	64.15	53.47	<u>53.19</u>	65.82
VIC_Nth_East	209.01	209.01	<u>86.61</u>	127.61	98.84	107.58
VIC_Metro	335.23	335.23	339.66	250.59	<u>176.29</u>	352.00
VIC_East_Coast	113.59	113.59	126.28	<u>73.29</u>	79.66	157.49
VIC_West_Coast	85.57	85.57	93.60	<u>39.71</u>	43.26	47.45
WA_West_Coast	237.14	237.14	326.91	<u>171.87</u>	199.24	363.23
WA_Sth	96.10	96.10	56.24	<u>30.70</u>	30.76	62.60
WA_Nth	130.13	130.13	115.82	50.24	<u>35.39</u>	47.26

Table 12: RMSE for Australian tourism data

	Base	BU	TD	GLS	MinT	Robust
total	1726.07	3680.00	1726.07	1906.48	1204.56	1101.69
NSW	814.62	1153.92	390.45	709.89	556.17	765.66
NT	135.97	136.13	177.71	62.33	101.17	61.77
QLD	633.93	691.06	581.55	503.91	264.39	262.21
SA	226.56	291.26	105.01	98.52	99.75	100.02
TAS	148.29	148.68	185.51	86.92	96.55	95.52
VIC	642.69	938.29	697.81	553.19	475.28	775.65
WA	362.15	481.64	513.88	272.77	233.58	514.91
NSW_ACT	146.72	146.72	86.05	79.23	83.41	88.06
NSW_Nth	158.32	158.32	110.01	93.05	99.64	67.53
NSW_Sth	145.49	145.49	128.61	90.52	104.94	186.33
NSW_Metro	370.82	370.82	192.91	300.52	178.58	329.89
NSW_Nth_Coast	265.32	265.32	145.86	193.64	162.47	211.55
NSW_Sth_Coast	131.95	131.95	165.04	64.79	54.05	130.65
NT_Central	87.08	87.08	102.50	48.88	69.71	56.96
NT_Nth_Coast	58.57	58.57	81.15	36.06	43.98	73.95
QLD_Metro	432.75	432.75	278.40	389.27	243.99	211.76
QLD_Central_Coast	62.53	62.53	78.83	42.89	48.47	45.56
QLD_Inland	174.46	174.46	194.94	131.45	78.24	98.68
QLD_Nth_Coast	101.77	101.77	183.26	73.57	70.65	127.87
SA_Metro	134.08	134.08	41.14	84.94	45.04	40.84
SA_Inland	87.19	87.19	36.99	43.59	42.36	38.44
SA_West_Coast	31.63	31.63	33.36	35.80	21.76	25.81
SA_Sth_Coast	66.34	66.34	82.00	42.47	46.66	92.30
TAS_Nth_East	58.96	58.96	70.44	35.35	39.60	43.19
TAS_Sth	61.49	61.49	84.42	47.51	48.14	53.70
TAS_Nth_West	44.88	44.88	42.50	30.26	33.01	66.44
VIC_Nth_West	124.89	124.89	86.11	76.51	74.90	87.18
VIC_Nth_East	232.47	232.47	108.39	153.51	116.29	132.61
VIC_Metro	385.23	385.23	397.61	307.86	238.76	386.30
VIC_East_Coast	142.78	142.78	164.21	87.36	97.22	180.77
VIC_West_Coast	101.05	101.05	122.62	43.97	50.62	56.05
WA_West_Coast	275.32	275.32	364.37	214.80	225.39	419.57
WA_Sth	101.13	101.13	63.93	41.77	42.26	70.69
WA_Nth	135.48	135.48	134.23	63.10	48.77	54.08

Table 13: MAE for Walmart sales data

	Base	BU	TD	GLS	MinT	Robust
total	2341.30	2672.04	2341.30	2369.42	2257.15	2069.26
CA	1235.18	1390.02	473.17	1243.65	1192.38	1115.32
TX	588.46	588.94	1363.35	567.46	525.10	496.01
WI	551.08	700.38	594.97	567.29	551.10	466.53
CA_1	218.13	218.13	509.39	183.49	196.72	188.12
CA_2	702.68	702.68	626.25	666.09	555.67	493.65
CA_3	202.95	202.95	371.52	166.36	192.27	192.00
CA_4	266.87	266.87	153.50	230.28	248.95	241.55
TX_1	144.27	144.27	495.96	137.11	122.65	110.52
TX_2	205.94	205.94	504.96	198.77	187.25	180.26
TX_3	239.98	239.98	362.43	232.81	215.85	206.92
WI_1	270.88	270.88	59.44	226.51	192.03	142.00
WI_2	181.19	181.19	171.11	137.77	129.79	98.80
WI_3	251.52	251.52	456.19	206.62	233.63	231.29

Table 14: RMSE for Walmart sales data

	Base	BU	TD	GLS	MinT	Robust
total	2476.17	2813.26	2476.17	2505.82	2390.48	2195.19
CA	1312.65	1459.66	500.67	1318.29	1264.27	1190.99
TX	616.51	617.08	1385.58	594.22	549.17	517.10
WI	585.71	743.75	634.64	602.89	586.51	498.61
CA_1	229.50	229.50	516.40	193.23	206.75	197.65
CA_2	756.67	756.67	631.13	723.15	621.26	573.69
CA_3	209.95	209.95	376.76	172.81	198.88	197.87
CA_4	277.57	277.57	163.07	240.85	259.08	250.52
TX_1	152.46	152.46	503.13	144.98	129.80	116.76
TX_2	214.92	214.92	511.64	207.35	195.32	187.26
TX_3	251.75	251.75	371.40	244.08	226.39	215.27
WI_1	289.03	289.03	90.64	241.61	205.28	152.04
WI_2	194.00	194.00	183.29	149.06	141.13	112.22
WI_3	264.39	264.39	464.58	217.38	245.68	242.46

Table 15: MAE for Swiss electricity demand data

	Base	BU	TD	GLS	MinT	Robust
grid	39.615	41.587	39.615	40.410	44.313	35.362
S1	18.900	19.164	24.268	18.844	19.059	18.992
S2	32.251	35.129	22.439	31.545	32.255	20.321
S11	7.743	8.514	17.003	7.852	7.641	7.629
S12	11.359	11.358	14.274	11.361	11.499	11.363
S21	25.096	25.621	22.450	24.633	22.938	13.200
S22	8.353	9.748	10.231	8.242	9.617	8.137
S11_1	1.093	1.093	1.867	0.789	0.485	0.491
S11_2	2.261	2.261	4.386	2.620	2.610	2.024
S11_3	1.756	1.756	2.318	2.123	2.183	1.930
S11_4	3.007	3.007	3.340	2.966	2.908	2.991
S11_5	5.959	5.959	15.835	5.688	5.730	5.201
S11_6	2.624	2.624	3.617	2.733	2.774	2.617
S12_1	4.556	4.556	8.824	4.590	4.742	5.093
S12_2	1.776	1.776	1.667	1.738	2.038	2.181
S12_3	4.502	4.502	7.805	4.454	5.275	4.640
S12_4	4.230	4.230	6.363	4.264	4.272	3.132
S12_5	4.332	4.332	4.246	4.370	3.950	5.448
S12_6	2.504	2.504	2.156	2.461	2.889	2.669
S21_1	1.339	1.339	2.927	1.442	1.172	1.438
S21_2	1.121	1.121	1.132	1.173	1.109	1.265
S21_3	0.876	0.876	2.786	0.849	0.844	1.361
S21_4	2.748	2.748	10.603	2.810	2.712	3.279
S21_5	6.577	6.577	4.226	6.413	7.174	5.530
S21_6	17.532	17.532	18.858	17.367	13.743	9.232
S22_1	2.160	2.160	5.757	1.823	2.475	2.315
S22_2	1.965	1.965	7.783	1.696	2.161	2.122
S22_3	3.879	3.879	9.010	3.688	3.720	4.363
S22_4	1.099	1.099	1.980	1.207	1.097	1.101
S22_5	1.220	1.220	1.179	1.079	1.262	1.072
S22_6	1.190	1.190	2.002	0.995	1.612	2.043

Table 16: RMSE for Swiss electricity demand data

	Base	BU	TD	GLS	MinT	Robust
grid	64.441	66.057	64.441	65.156	67.813	61.076
S1	30.936	30.400	34.750	30.722	31.714	31.483
S2	39.334	41.592	32.570	38.789	39.311	30.959
S11	12.054	11.937	20.505	11.968	12.354	12.442
S12	19.044	19.041	17.878	18.984	19.535	19.208
S21	28.566	28.995	27.339	28.181	26.775	19.874
S22	11.779	13.269	12.713	11.676	13.129	11.617
S11_1	1.172	1.172	2.092	0.883	0.655	0.657
S11_2	2.795	2.795	5.071	3.111	3.103	2.552
S11_3	2.009	2.009	2.772	2.347	2.403	2.160
S11_4	3.464	3.464	4.340	3.408	3.607	3.858
S11_5	6.308	6.308	16.369	6.017	6.062	5.514
S11_6	4.689	4.689	5.593	4.832	4.864	4.669
S12_1	6.568	6.568	12.232	6.569	6.564	6.622
S12_2	2.366	2.366	2.104	2.332	2.599	2.717
S12_3	5.022	5.022	8.321	4.978	5.736	5.140
S12_4	4.704	4.704	7.391	4.728	4.733	4.185
S12_5	5.151	5.151	5.078	5.174	4.954	5.972
S12_6	2.743	2.743	2.983	2.699	3.104	2.899
S21_1	1.726	1.726	3.776	1.785	1.654	1.782
S21_2	1.564	1.564	1.613	1.590	1.566	1.631
S21_3	1.028	1.028	3.708	0.998	0.992	1.602
S21_4	4.410	4.410	11.256	4.364	4.455	4.228
S21_5	7.195	7.195	5.089	7.047	7.737	6.277
S21_6	18.540	18.540	19.633	18.392	15.176	11.590
S22_1	2.547	2.547	6.196	2.209	2.841	2.690
S22_2	2.387	2.387	8.250	2.055	2.640	2.593
S22_3	6.343	6.343	10.262	6.134	5.684	5.568
S22_4	1.324	1.324	2.462	1.454	1.308	1.318
S22_5	1.713	1.713	1.498	1.500	1.753	1.495
S22_6	1.455	1.455	2.422	1.226	1.869	2.287