Draft - Enhancing Forecast Reconciliation: A Study of Covariance Estimation

Vincent Su

2025-04-23

Notation

- Scalar y_t
- Vector $\boldsymbol{y}_t$
- Matrix $\boldsymbol{S}$
- Covariance matrix of in-sample 1-step-ahead base forecast errors $\hat{\boldsymbol{W}}_1$
- Its shrinkage estimator with diagonal target $\hat{\boldsymbol{W}}_{1,D}^{shr}$
- Its NOVELIST estimator $\hat{\boldsymbol{W}}_{1,thr}^{shr}$

Abstract

This is pasted from the Project Description

A collection of time series connected via a set of linear constraints is known as hierarchical time series. Forecasting these structures is challenging because the individual forecasts do not satisfy the linear constraints present in the data. To mitigate this issue, various forecast reconciliation approaches have been proposed in the literature, where the individual forecasts are adjusted by minimizing the mean squared reconciliation errors. Among these, the MinT (Minimum Trace) approach is widely used. However, this method requires an estimate of the covariance matrix of the base forecast errors. Although the sample covariance matrix is a natural choice, its performance is greatly impacted by the high-dimensional nature of hierarchical time series encountered in practice. Most published research uses shrinkage-type estimators that shrink the sample covariance matrix toward a diagonal matrix.

In this project, we aim to assess the forecasting performance of MinT when different covariance estimators are used, with a primary focus on the NOVELIST (NOVEL Integration of the Sample and Thresholded Covariance) estimator, which shrinks the sample covariance toward its thresholded version. A thorough literature review will also identify other high-dimensional covariance estimators relevant to this project.

1. Introduction

- **Context & Motivation:**

  Hierarchical (and grouped) time series are collections of time series data connected via linear constraints (linear combinations?). Forecast reconciliation ensures that forecasts satisfy these constraints.

  One of the widely used appoarch is the MinT approach. It minimises the trace of reconciled forecast errors covariance matrix (minimum variance) and depends critically on the covariance matrix of h-step ahead base forecast errors W_h.

- **Problem Statement:**
  The sample covariance matrix, although natural, suffers in high-dimensional settings. Especially when the number of series p is huge and larger than the time dimension T, the sample covariance matrix is non-positive definite (rank T if p>T).

  The shrinkage estimators come in to tackle this issue. The shrinkage estimator with diagonal target (often shrinking toward a diagonal matrix) is proven to produce a guaranteed PD matrix (Schäfer & Strimmer, 2005). However, as it shrinks the covariance matrix toward a diagonal one, it does not have flexibility and might neglect the prominent structure presented in the covariance matrix.

  An alternative approach is to perform shrinkage of the sample covariance towards its thresholded version, instead of a diagonal matrix. This is the NOVELIST (NOVEL Integration of the Sample and Thresholded covariance estimators) method proposed by Huang & Fryzlewicz (2019). They introduced thresholding functions applied only to off-diagonal elements, allowing for more flexibility in the estimation.

  ... can include more estimators ...

- **Research Aim:**
  This paper assesses the reconciled forecasting performance of MinT approach using various covariance estimators, with a focus on the NOVELIST estimator.

- **Paper Outline:**
  The paper is structured as follows:

  - A literature review of forecast reconciliation and covariance estimation.
  - A description of the methodology, including the NOVELIST estimator and its principal-component-adjusted variant.
  - An experimental design using both synthetic and real hierarchical time series.
  - Empirical results and discussion.
  - Conclusions and suggestions for future work.

2. Literature Review

2.1 Forecast Reconciliation in Hierarchical and Grouped Time Series

- Brief history of reconciliation methods (bottom-up, top-down, GLS, MinT).
- Statistical challenges and practical applications of MinT.

2.2 Covariance Estimation in High Dimensions

- Limitations of the sample covariance matrix.
- Estimators used by Wickramasuriya et al. (2019).
- Shrinkage estimators:

- Diagonal shrinkage (e.g., Schäfer & Strimmer, Ledoit & Wolf).
- NOVELIST estimator and its Cross-validation & PC-adjusted variant.
- ...

2.3 Relevance to Forecast Reconciliation

- Discuss how covariance estimation affects MinT performance.
- Identify research gaps.

3. Methodology

3.1 Theoretical Framework

Hierarchical and Grouped Time Series Structure

Hierarchical/Grouped time series satisfy a set of linear constraints, often represented as:

$$y_t = Sb_t,$$

where S is a summing matrix of order mxn which aggregates the bottom-level series b_t (n-vector) to the series at aggregation levels above. The m-vector y_t contains all observations at time t

Let $\hat{y}_t(h)$ be a vector of h-step-ahead base forecasts for each time series in the collection, using information up to and including time t, and stacked in the same order as y_t. Then we can rewrite the linear reconciliation forecast equation as:

$$\tilde{y}_t(h) = SP\hat{y}_t(h)$$

where P is an nxp matrix that maps the base forecasts into bottom-level disaggregated forecasts $\hat{b}_t$.

- Expand on this, include graphs, examples...

The Minimum Trace (MinT) Reconciliation

Lemma 1: ... the covariance matrix of the h-step-ahead reconciled forecast errors is given by

$$var[y_{t+h} - \tilde{y}_t(h)|F_t] = SPW_hP'S'$$

where $W_h = E[\hat{e}_t(h)\hat{e}'_t(h)|F_t]$ is the covariance matrix of the h-step-ahead base forecast errors.

The MinT reconciled forecast is obtained by minimizing the trace of the above covariance matrix:

$$\min \ tr(SPW_hP'S') \quad \text{subject to } PS = I.$$

which gives the optimal reconciliation matrix:

$$P = (S'W_h^{-1}S)^{-1}S'W_h^{-1}$$

3.2 Covariance Estimation Approaches

- W_h is challenging to estimate, when $h > 1$, we need alternative estimates.

For forecast errors $\hat{e}_{t|t-1}$, the unbiased sample covariance of in-sample one-step-ahead base forecast errors is:

$$\hat{W}_1 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{t|t-1}\hat{e}_{t|t-1}^{\top}.$$

We reconstruct estimator of W_h as $\hat{W}_h = k_h\, g(\hat{W}_1)$, where $k_h > 0$ simplifying the computation.

We will look into a few "sparse" + "non-sparse" approaches that apply on $\hat{W}_1$.

The Shrinkage Estimator

$$g(\hat{W}_1) = \hat{W}_{1,D}^{shr} = \lambda_D \hat{W}_{1,D} + (1-\lambda_D)\hat{W}_1$$

is the shrinkage estimator with diagonal target, proposed by Schäfer & Strimmer (2005). $\hat{W}_{1,D}$ is a diagonal matrix comprising the diagonal entries of $\hat{W}_1$. This approach will shrink the covariance matrix $\hat{W}_1$ towards its diagonal version, meaning the off-diagonal elements are shrunk towards zero while the diagonal ones remain unchanged.

They also proposed an optimal shrinkage intensity parameter λ_D for this setting, assuming the variances are constant:

$$\hat{\lambda}_D = \frac{\sum_{i \neq j} \widehat{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}}$$

- (need more explanation on how to obtain optimal param)
- Walkthrough the shrinkage estimator method

The NOVELIST Estimator

$$g(\hat{W}_1) = \hat{W}_{1,thr}^{shr} = \lambda_\delta \hat{W}_{1,\delta} + (1-\lambda_\delta)\hat{W}_1$$

is the NOVELIST shrinkage estimator, proposed by Huang & Fryzlewicz (2019). By convenient setting, we rewrite as sample correlation:

$$\hat{R}_{1,thr}^{shr} = \lambda_\delta \hat{R}_{1,\delta} + (1-\lambda_\delta)\hat{R}_1,$$

$\hat{R}_{1,\delta}$ is a thresholded correlation matrix, in which thresholding is applied only to each off-diagonal element. This approach will shrink the sample correlation matrix $\hat{R}_1$ towards its thresholded version. There are various choices for the thresholding function, in this work, we use the soft-thresholding operator, defined as:

$$\hat{r}_{1,ij}^{\delta} = \operatorname{sign}(\hat{r}_{1,ij})\,(|\hat{r}_{1,ij}| - \delta)_+.$$

After calculated the NOVELIST correlation matrix, we can re-obtain the covariance matrix $\hat{W}_{1,thr}^{shr} = \hat{D}_1^{1/2}\hat{R}_{1,thr}^{shr}\ \hat{D}_1^{1/2}$, where $\hat{D}_1 = diag(\hat{W}_1)$ is the diagonal matrix and the elements are given by the variances of the 1-step-ahead base forecast errors.

For a given threshold δ, the optimal shrinkage intensity parameter $\lambda(\delta)$ can be estimated as:

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

The formula derived using Ledoit-Wolf's lemma (Ledoit and Wolf, 2003) by Huang & Fryzlewicz. For the threshold parameter δ, we use a cross-validation to select the optimal value

- ALGORITHM WALKTHROUGH

Principal-Component Adjustment

When a factor structure is present, the procedure is:

Alternative Covariance Estimators

Brief overview of other high-dimensional estimators used in the literature for benchmarking.

Others:

- Principal Orthogonal complEment Thresholding (POET) by Fan et al. (2013)

  A Low-rank + Sparse method.
  It decompose the covariance matrix into a prominent principle components part (Low-rank) and a orthogonal complement part R_K. Then apply thresholding to R_K.
  This is similar to the NOVELIST estimator with PC-adjusted, but the difference is that POET apply thresholding to R_K, not a NOVELIST function.

4. Simulation

- Talk about description of the hierarchical time series data set (e.g., economic, financial, or synthetic data).
- Characteristics such as dimensionality, frequency, and hierarchical structure.

Experimental Design

- Design different cases of simulation studies and real-data experiments.
- Metrics: Forecast accuracy (e.g., RMSE, MAE), reconciliation error reduction, and matrix stability…
- Visualisation…

General Set-up I'm currently working on

The designed data generating process for bottom-level series is a stationary VAR(1) process, with the following structure:

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{e}_t,$$

where $\boldsymbol{A}$ is a $p \times p$ block diagonal matrix of autoregressive coefficients $\boldsymbol{A} = diag(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m)$, with each $\boldsymbol{A}_i$ being a $p_i \times p_i$ matrix. The block diagonal structure ensures that the time series are grouped into m groups, with each group having its own autoregressive coefficients. This aim to simulate the interdependencies between the time series within each group, where reconciliation will be better performed than the usual base forecasts.

The model is added with a Gaussian innovation process $\boldsymbol{e}_t$, with covariance matrix Σ. The covariance matrix Σ is generated specifically in the following way:

1. A compound symmetric correlation matrix is used for each block of size p_i in $\boldsymbol{A}_i$, where the coefficients are sampled from a uniform distribution.
2. The correlations between different blocks are imposed using the Algorithm 1 in Hardin, Garcia & Golan (2013).
3. The covariance matrix Σ is then constructed by uniform sampling of standard deviations for all p series.

We have an option to randomly flip the signs of the covariance elements, which will create a more realistic structure in the innovation process. This is also to simulate the real-world scenario where the observed covariance matrix is not necessarily positive definite.

5. Empirical Analysis

- Comparative analysis of forecast performance using different covariance estimators on real-life dataset.

6. Discussion and Conclusion

- Evaluate how covariance estimation impacts the MinT reconciliation.
- Advantages and limitations of the NOVELIST estimator (with different ways of choosing threshold parameter).
- Practical considerations: Computational efficiency, robustness, and ease of implementation.
- Limitations of the research

References