Stochastic Coherency in Forecasting Reconciliation

Quarterly Forecasting Forum

International Institute of Forecasting

Kandrika Pritularga
Ivan Svetunkov
Nikolaos Kourentzes

Manchester, 6th March 2020

Marketing Analytics & Forecasting

Lancaster University Management School

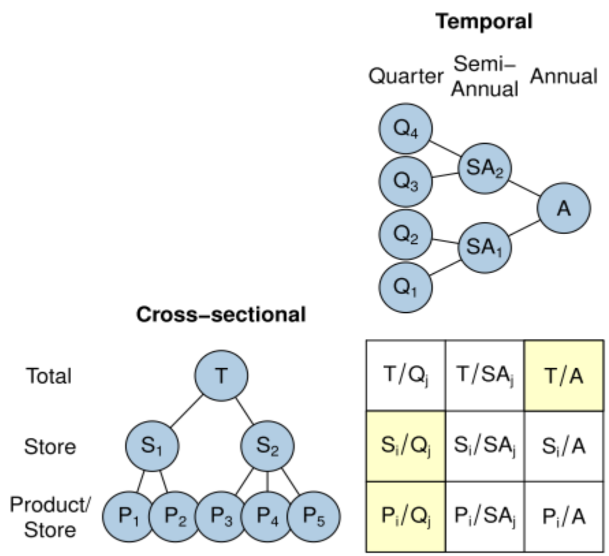
Outline

1. Introduction
2. Hierarchical Forecasting
3. Stochastic Coherency
4. Forecast Reconciliation under Different Model Specification
5. Simulation Studies
6. A Case Study: Scandinavian unemployment monthly data
7. Discussion
8. Conclusion

INTRODUCTION

Introduction

- Forecasting is essential for supporting decisions
- Such decisions are structured in hierarchies
  - Product categories, market segments
  - Weekly, monthly, quarterly, semi-annually, yearly
- Hierarchies can be structured in temporal, cross-sectional, or both



Athanasopoulos & Kourentzes (2020)

$$\boldsymbol{S} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \boldsymbol{y}_b = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}$$

Hierarchical Forecasting

- Conventional methods: Bottom-up and top-down
  - Both ignore important information in the hierarchy
- Current method: MinT Reconciliation
  - Generate forecast for each time series independently
  - Combine the forecast linearly to produce adjusted bottom-level forecasts
  - Aggregate the adjusted forecasts into a complete hierarchy

$$\widetilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}\widehat{\boldsymbol{y}}_{t+h|t}$$
$$\boldsymbol{G} = (\boldsymbol{S}^T \boldsymbol{W}_{t+h|t}^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{W}_{t+h|t}^{-1}$$

Hierarchical Forecasting

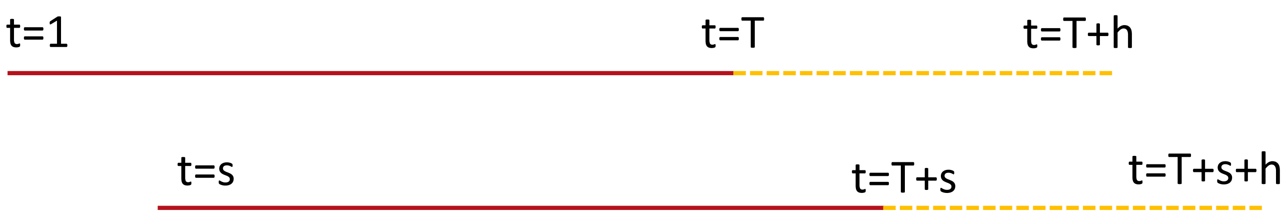
- Formulation (Hyndman et al 2011, Wickramasuriya et al. 2018):

$$\tilde{\boldsymbol{y}}_{t+h|t} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{t+h|t}$$
$$\boldsymbol{G} = (\boldsymbol{S}^T \boldsymbol{W}_{t+h|t}^{-1} \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{W}_{t+h|t}^{-1}$$

$\tilde{\boldsymbol{y}}_{t+h|t}$ is the h-step ahead reconciled forecast; $\hat{\boldsymbol{y}}_{t+h|t}$ is the h-step ahead unreconciled based forecast; $\boldsymbol{S}$ is a summation matrix; $\boldsymbol{G}$ is the hierarchical reconciliation weight matrix; $\boldsymbol{W}_{t+h|t}$ is the covariance matrix of h-step ahead base forecasts

Hierarchical Forecasting

- $\boldsymbol{W}_{t+h|t}$ is difficult to estimate
- Literature approximates $\boldsymbol{W}_{t+h|t}$ with the covariance matrix of one-step ahead in-sample base forecast error ($\widehat{\boldsymbol{W}}$)
- Suppose we have additional sample and re-estimate $\boldsymbol{G}$ given $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{G}}$ will vary from the previous $\widehat{\boldsymbol{W}}$



$$\widehat{\boldsymbol{G}}_T = f(\boldsymbol{S}, \widehat{\boldsymbol{W}}_T)$$

$$\widehat{\boldsymbol{G}}_{T+s} = f(\boldsymbol{S}, \widehat{\boldsymbol{W}}_{T+s})$$

Is $\widehat{\boldsymbol{G}}$ uncertain?

$$\widehat{\boldsymbol{G}}_T \neq \widehat{\boldsymbol{G}}_{T+s}$$

$$\boldsymbol{S}\widehat{\boldsymbol{G}}_T\widehat{\boldsymbol{y}}_{T+h|T} \neq \boldsymbol{S}\widehat{\boldsymbol{G}}_{T+s}\widehat{\boldsymbol{y}}_{T+s+h|T+s}$$

Classical Coherency

- Most literature in hierarchical assume that the data, or in our words, the observed data is coherent
- **Classical coherency** is defined as
$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{y}_{b,t}$$
- where the observed time series in the complete hierarchy $\boldsymbol{y}_t$ is an aggregation from the observed bottom-level time series, $\boldsymbol{y}_{b,t}$
- This definition does not allow us to capture uncertainty because it tends to treat the process as deterministic

Research Problems

- Not only uncertainty from the base forecast, but $\boldsymbol{G}$ is also subject to uncertainty
- It is due to uncertainties in forecasting models, namely parameter and model uncertainty
- We propose to redefine coherency from deterministic (observational-focused) to stochastic
- **Stochastic coherency** ensures that the time series are coherent in expectations
- It enables us to trace back the sources of the uncertainty

STOCHASTIC COHERENCY

Stochastic Coherency

- Suppose that the bottom-level process is shown as $\boldsymbol{y}_{b,t} = \boldsymbol{\mu}_{b,t} + \boldsymbol{\varepsilon}_{b,t}$ where $\mathrm{E}(\boldsymbol{y}_{b,t}|\boldsymbol{I}_t) = \boldsymbol{\mu}_{b,t}$, $\mathrm{E}(\boldsymbol{\varepsilon}_{b,t}|\boldsymbol{I}_t) = \mathbf{0}$, and $\boldsymbol{\varepsilon}_{b,t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$, then due to aggregation,

$$\boxed{\begin{gathered}\mathrm{E}(\boldsymbol{y}_t|\boldsymbol{I}_t) = \mathrm{E}(\boldsymbol{S}\boldsymbol{y}_{b,t}|\boldsymbol{I}_t) \\ \boldsymbol{\mu}_t = \boldsymbol{S}\boldsymbol{\mu}_{b,t}\end{gathered}}$$

- Coherency is maintained in the expectation level instead of the observational level

- $\boldsymbol{\mu}_{b,t}$ is the true mean of the bottom-level data generating process, $\boldsymbol{\varepsilon}_{b,t}$ is the innovation in the bottom-level time series, $\boldsymbol{\mu}_t$ is the true mean of the complete hierarchy, and $\boldsymbol{I}_t$ is the information at t.

Stochastic Coherency in Forecast Reconciliation

$$\mathrm{E}(\boldsymbol{y}_t|\boldsymbol{I}_t) = \mathrm{E}(\boldsymbol{S}\boldsymbol{y}_{b,t}|\boldsymbol{I}_t)$$
$$\boldsymbol{\mu}_t = \boldsymbol{S}\boldsymbol{\mu}_{b,t}$$

- Now it depends on how well the forecasting models understand the data generating process (DGP)
- We illustrate three scenarios

Source of uncertainties	Perfect model specification	Well-specified models	Mis-specified models
Model	Perfectly known	Known	Wrong models
Parameter	Perfectly known	Estimated	Estimated

Perfect Model Specification

- In this case, all information is known, e.g. $\boldsymbol{\Sigma}_b$, and the forecasting model is

$$\widehat{\boldsymbol{y}}_t = \boldsymbol{\mu}_t$$

- And the h-step ahead base forecast error is

$$\hat{\boldsymbol{e}}_{t+h|t} = \boldsymbol{\varepsilon}_{t+h}$$

- As $\boldsymbol{\varepsilon}_{t+h} = \boldsymbol{S}\boldsymbol{\varepsilon}_{b,t+h}$, consequently $\boldsymbol{\Sigma} = \boldsymbol{S}\boldsymbol{\Sigma}_b\boldsymbol{S}^T$
- $\boldsymbol{G} = (\boldsymbol{S}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{\Sigma}^{-1}$
    - $\boldsymbol{G}$ is deterministic as everything is known
- Hence, $\widetilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\boldsymbol{G}\widehat{\boldsymbol{\mu}}_{t+h}$ and Var($\widetilde{\boldsymbol{y}}_{t+h}$)=Var($\boldsymbol{S}\boldsymbol{G}\widehat{\boldsymbol{\mu}}_{t+h}$) which is constant

Well- v.s. Mis-specified Models

Criteria	Well-specified Models	Mis-specified Models
DGP	$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$	$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$
Forecasting Model	$\widehat{\boldsymbol{y}}_{t+h} = \mathrm{E}(\widehat{\boldsymbol{y}}_{t+h\|t}\|\boldsymbol{I}_t) + \boldsymbol{v}_{t+h\|t}$	$\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t} = \mathrm{E}(\widehat{\boldsymbol{y}}_{t+h\|t}\|\boldsymbol{I}_t) + \boldsymbol{v}^{\dagger}_{t+h\|t}$
Forecast Error	$\hat{\boldsymbol{e}}_{t+h\|t} = \boldsymbol{\varepsilon}_{t+h} - \boldsymbol{v}_{t+h\|t}$	$\hat{\boldsymbol{e}}^{\dagger}_{t+h\|t} = \boldsymbol{\varepsilon}_{t+h} - \boldsymbol{v}^{\dagger}_{t+h\|t}$
$\boldsymbol{v}_{t+h\|t}$	As $t \to \infty$, $\boldsymbol{v}_{t+h\|t} \to \mathbf{0}$, $\boldsymbol{V} \to \mathbf{0}$	$\mathrm{E}(\boldsymbol{v}^{\dagger}_{t+h\|t}) = \boldsymbol{v}$ and $\|\boldsymbol{v}_{t+h\|t}\| \leq \|\boldsymbol{v}^{\dagger}_{t+h\|t}\|$
Covariance Matrix	$\boldsymbol{W}_{t+h\|t} = \boldsymbol{\Sigma} + \boldsymbol{V} - 2\mathrm{cov}(\boldsymbol{\varepsilon}_{t+h}, \boldsymbol{v}_{t+h\|t})$	$\boldsymbol{W}^{\dagger}_{t+h\|t} = \boldsymbol{\Sigma} + \boldsymbol{V}^{\dagger} - 2\mathrm{cov}(\boldsymbol{\varepsilon}_{t+h}, \boldsymbol{v}^{\dagger}_{t+h\|t})$
$\widehat{\boldsymbol{G}}$	$\widehat{\boldsymbol{G}} = (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}_{t+h\|t}{}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}_{t+h\|t}{}^{-1}$	$\widehat{\boldsymbol{G}}^{\dagger} = (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}^{*}_{t+h\|t}{}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}^{\dagger}_{t+h\|t}{}^{-1}$
$\widetilde{\boldsymbol{y}}_{t+h}$	$\widetilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{y}}_{t+h}$	$\widetilde{\boldsymbol{y}}^{*}_{t+h\|t} = \boldsymbol{S}\widehat{\boldsymbol{G}}^{\dagger}\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t}$
Var($\widetilde{\boldsymbol{y}}_{t+h}$)	Var($\boldsymbol{S}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{y}}_{t+h}$)	Var($\boldsymbol{S}\widehat{\boldsymbol{G}}^{\dagger}\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t}$)

Note:

- $\boldsymbol{\varepsilon}_{t+h}$ is the innovation of the time series across the hierarchy
- $\boldsymbol{V}$ is the covariance matrix of $\boldsymbol{v}_{t+h|t}$
- $\boldsymbol{V}^{*}$ is the covariance matrix of $\boldsymbol{v}^{*}_{t+h|t}$
- Variables with star denote that those are generated from mis-specified models

Well- v.s. Mis-specified Models

Criteria	Well-specified Models	Mis-specified Models
DGP	$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$	$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$
Forecasting Model	$\widehat{\boldsymbol{y}}_{t+h} = \mathrm{E}(\widehat{\boldsymbol{y}}_{t+h\|t}\|\boldsymbol{I}_t) + \boldsymbol{v}_{t+h\|t}$	$\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t} = \mathrm{E}(\widehat{\boldsymbol{y}}_{t+h\|t}\|\boldsymbol{I}_t) + \boldsymbol{v}^{\dagger}_{t+h\|t}$
Forecast Error	$\hat{\boldsymbol{e}}_{t+h\|t} = \boldsymbol{\varepsilon}_{t+h} - \boldsymbol{v}_{t+h\|t}$	$\hat{\boldsymbol{e}}^{\dagger}_{t+h\|t} = \boldsymbol{\varepsilon}_{t+h} - \boldsymbol{v}^{\dagger}_{t+h\|t}$
$\boldsymbol{v}_{t+h\|t}$	As $t \to \infty$, $\boldsymbol{v}_{t+h\|t} \to \boldsymbol{0}$, $\boldsymbol{V} \to \boldsymbol{0}$	$\mathrm{E}(\boldsymbol{v}^{\dagger}_{t+h\|t}) = \boldsymbol{v}$ and $\|\boldsymbol{v}_{t+h\|t}\| \leq \|\boldsymbol{v}^{\dagger}_{t+h\|t}\|$
Covariance Matrix	$\boldsymbol{W}_{t+h\|t} = \boldsymbol{\Sigma} + \boldsymbol{V} - 2\mathrm{cov}(\boldsymbol{\varepsilon}_{t+h}, \boldsymbol{v}_{t+h\|t})$	$\boldsymbol{W}^{\dagger}_{t+h\|t} = \boldsymbol{\Sigma} + \boldsymbol{V}^{\dagger} - 2\mathrm{cov}(\boldsymbol{\varepsilon}_{t+h}, \boldsymbol{v}^{\dagger}_{t+h\|t})$
$\widehat{\boldsymbol{G}}$	$\widehat{\boldsymbol{G}} = (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}_{t+h\|t}{}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}_{t+h\|t}{}^{-1}$	$\widehat{\boldsymbol{G}}^{\dagger} = (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}^{\dagger}_{t+h\|t}{}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{W}^{\dagger}_{t+h\|t}{}^{-1}$
$\widetilde{\boldsymbol{y}}_{t+h}$	$\widetilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{y}}_{t+h}$	$\widetilde{\boldsymbol{y}}^{\dagger}_{t+h\|t} = \boldsymbol{S}\widehat{\boldsymbol{G}}^{\dagger}\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t}$
Var($\widetilde{\boldsymbol{y}}_{t+h}$)	Var($\boldsymbol{S}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{y}}_{t+h}$)	Var($\boldsymbol{S}\widehat{\boldsymbol{G}}^{\dagger}\widehat{\boldsymbol{y}}^{\dagger}_{t+h\|t}$)

Note:

- $\boldsymbol{\varepsilon}_{t+h}$ is the innovation of the time series across the hierarchy
- $\boldsymbol{V}$ is the covariance matrix of $\boldsymbol{v}_{t+h|t}$
- $\boldsymbol{V}^*$ is the covariance matrix of $\boldsymbol{v}^*_{t+h|t}$
- Variables with star denote that those are generated from mis-specified models

Accuracy and Variance in Forecast Reconciliation

- Uncertainties in forecast reconciliation originates from uncertainties in the base forecasts
- Analogous to forecast combination literature
  - Estimated weights in forecast combination increase the variance of the combined forecast, while fixed weights, even if suboptimal, often provide better results
- Further complication: $\boldsymbol{W}_{t+h|t}$ is difficult to estimate
  - Need approximations, usually based on $\widehat{\boldsymbol{W}}$
  - The estimated covariance with shrinkage is believed to be the best approximation among all
- The difference between two coherencies lies on the variance instead of the point forecasts

SIMULATION STUDIES

Experiment Design

- Small hierarchy, simple data generating process
  - 4 bottom-level time series, 7 time series in the complete hierarchy
  - DGP of the bottom-level: AR(1), with the parameter of 0.4, the correlated innovation is generated
  - Sample size = 24, 120; forecast horizon = 1 and 6-step
- Large hierarchy, more complicated DGP
  - 50 bottom-level time series with two levels in the upper-level of the hierarchy
  - DGP of the bottom-level: ARIMA, AR and MA orders are sampled from 0 and 3, and the integration is from 0 and 1
  - Sample size = 30, 150

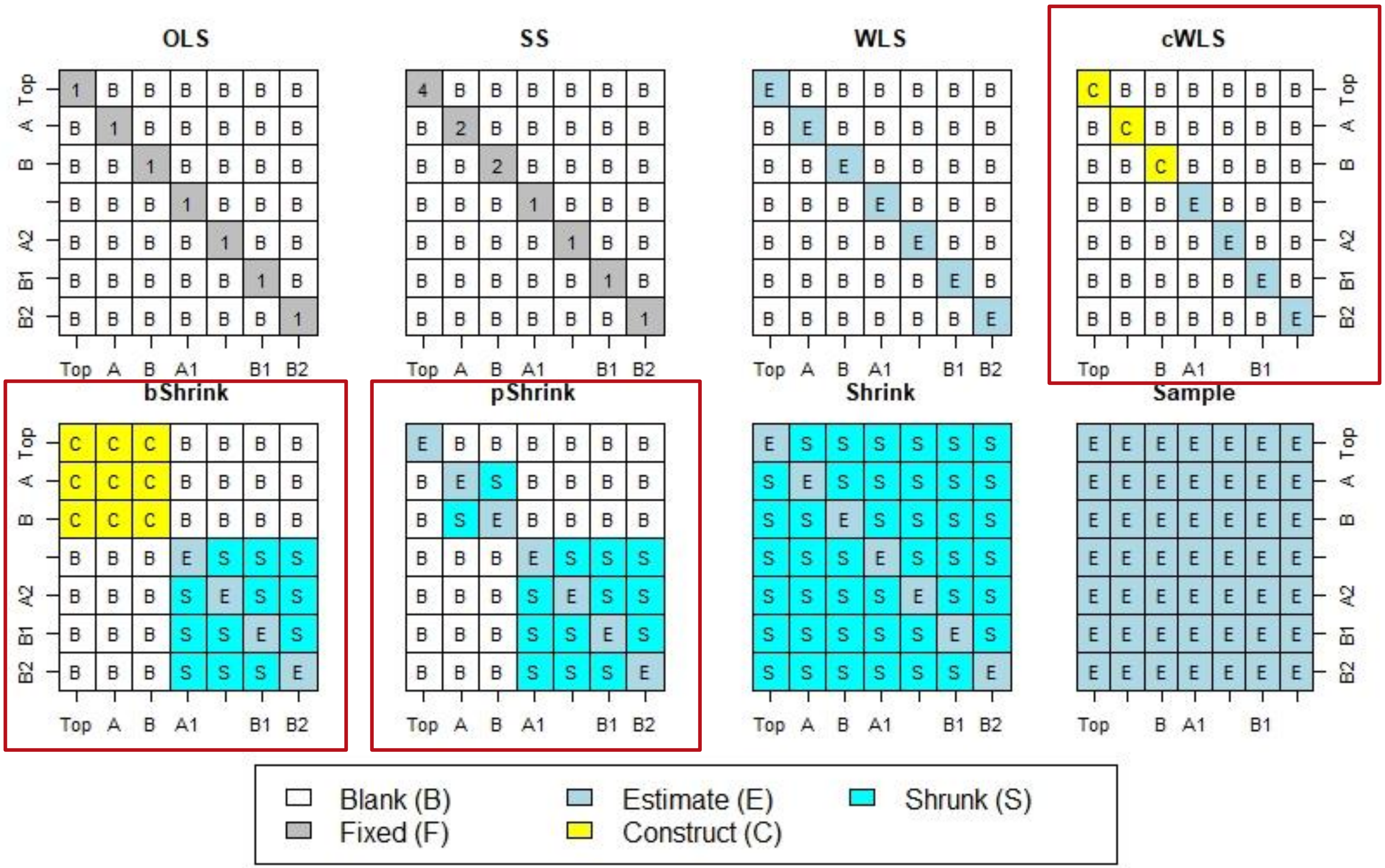
Models

- Small hierarchy
  - Four levels of model specification
    - Perfect AR(1)
    - Estimated AR(1)
    - AutoARIMA: automatic selection (include AR(1)
    - ETS(AAN): wrong models (does not include AR(1))
- Large hierarchy
  - We employ AutoARIMA and ETS(ZZZ)

Error Metrics

- Measure the overall performance across the hierarchy (Gamakumara et al. 2018)
  - Follow the theory but little benefits in practice
  - $RelTotSE_h = \sum_{i=1}^{N} MSE_{ih,recon} / \sum_{i=1}^{N} MSE_{ih,base}$
  - "Loss function error metrics"
- Measure the performance of individual time series then summarise them for a complete hierarchy (Athanasopoulos & Kourentzes 2020, Davydenko & Fildes 2013)
  - Does not follow the loss function but practically beneficial
  - $AvgRelMSE_h = \left(\prod_{i=1}^{N} (MSE_{ih,recon} / MSE_{ih,base})\right)^{1/N}$
  - N is the number of time series in the hierarchy
  - "Decision error metrics"

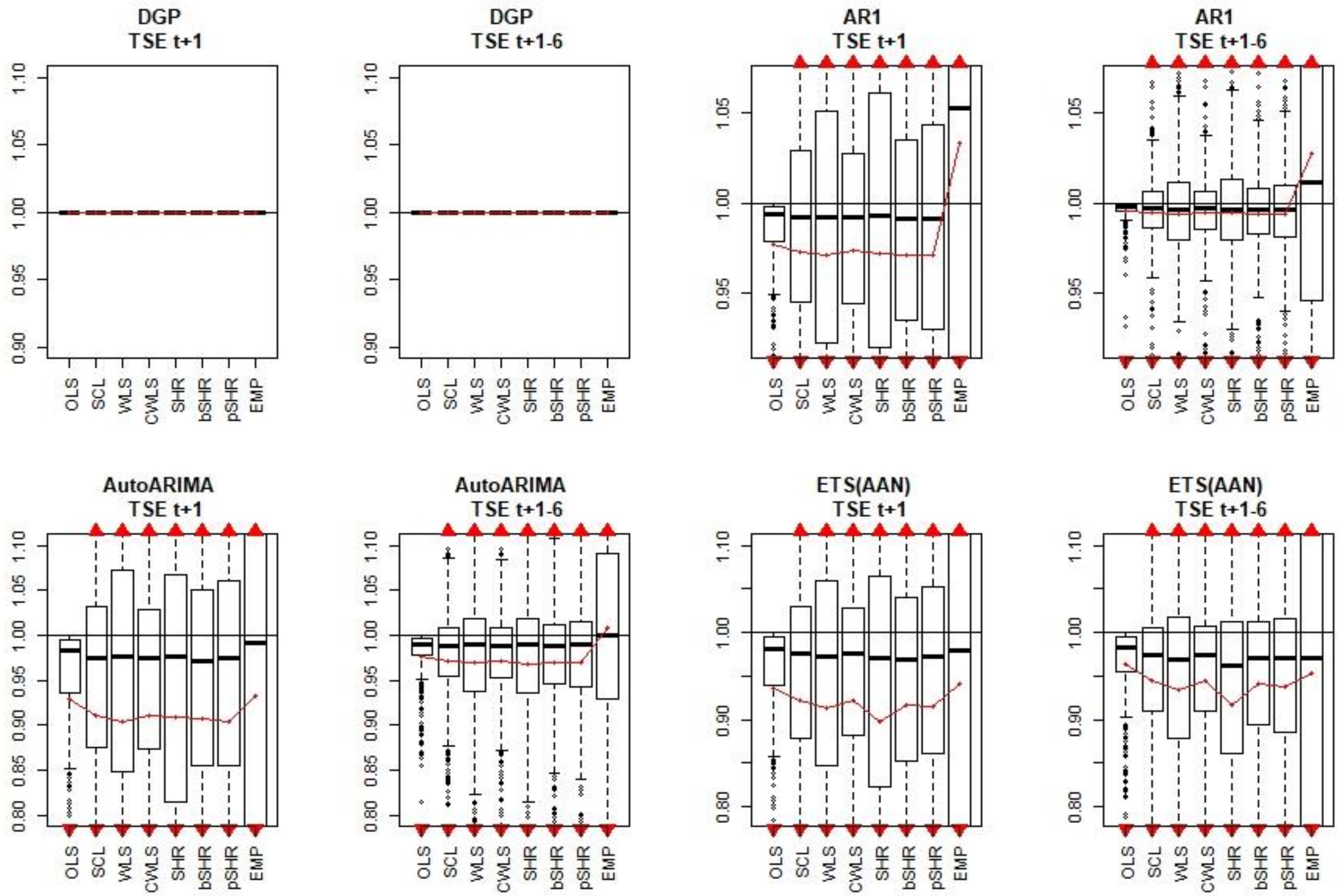
Covariance Matrix Approximations



pSHR: pShrink; bSHR: bShrink; SHR: MinT-Shrink; EMP: MinT-Sample; our proposed covariance matrices are inside the red boxes

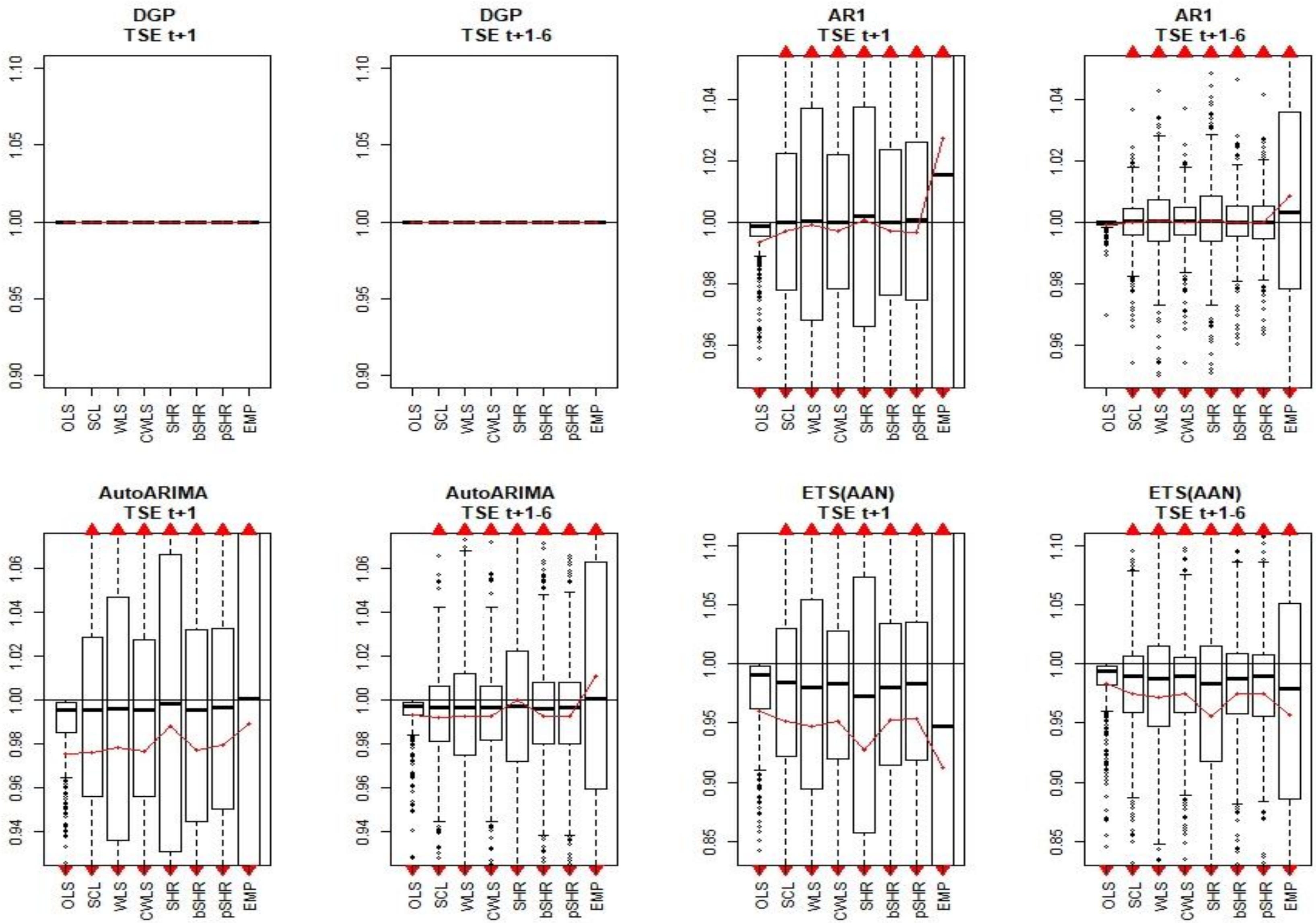
FINDINGS: SIMULATION

Findings: Small Hierarchy, Sample = 24



Red lines: geometric mean; red triangles: omitting some part of the distribution from the plots; TSE means RelTotSE

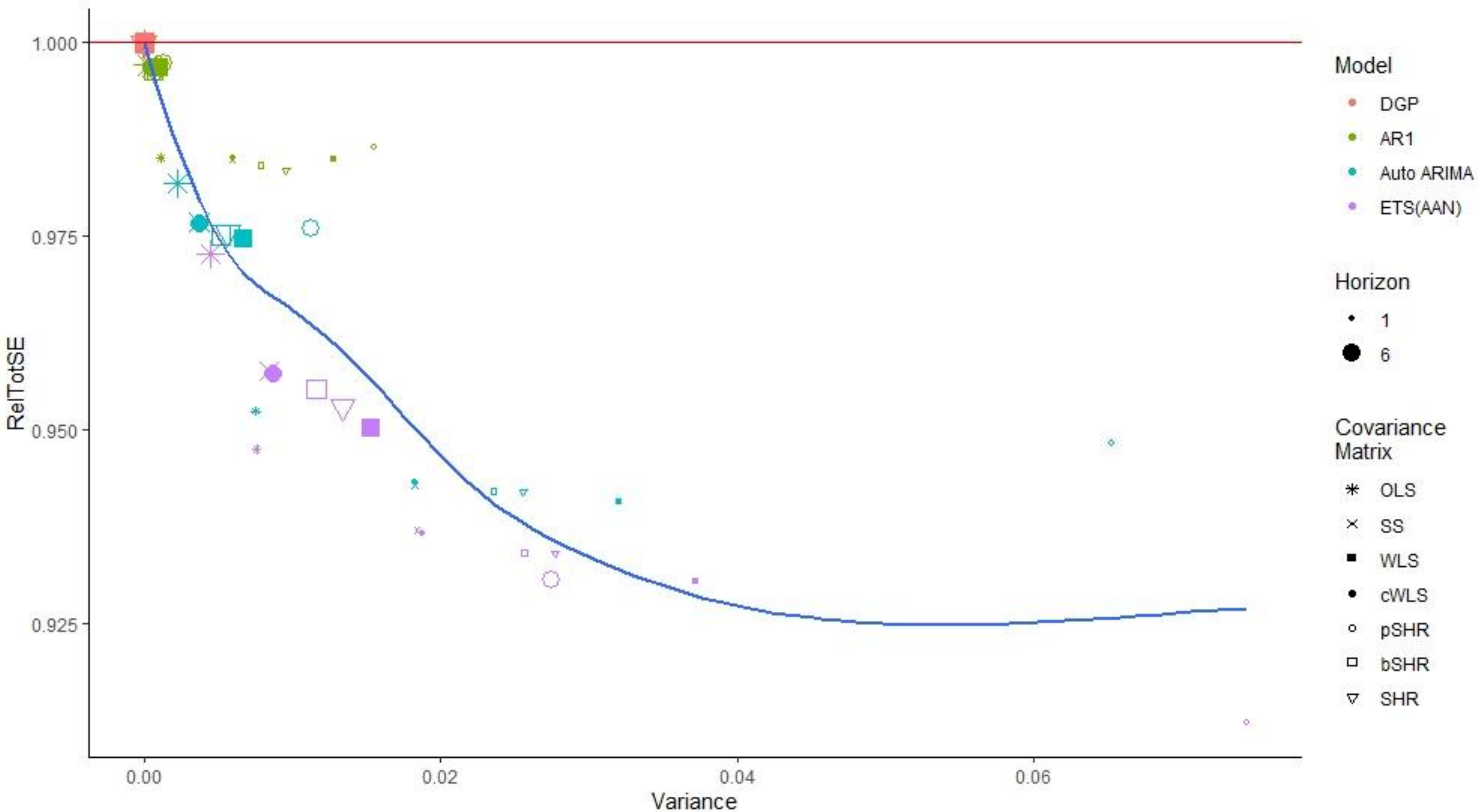
Findings: Small Hierarchy, Sample = 120



Findings: Small Hierarchy

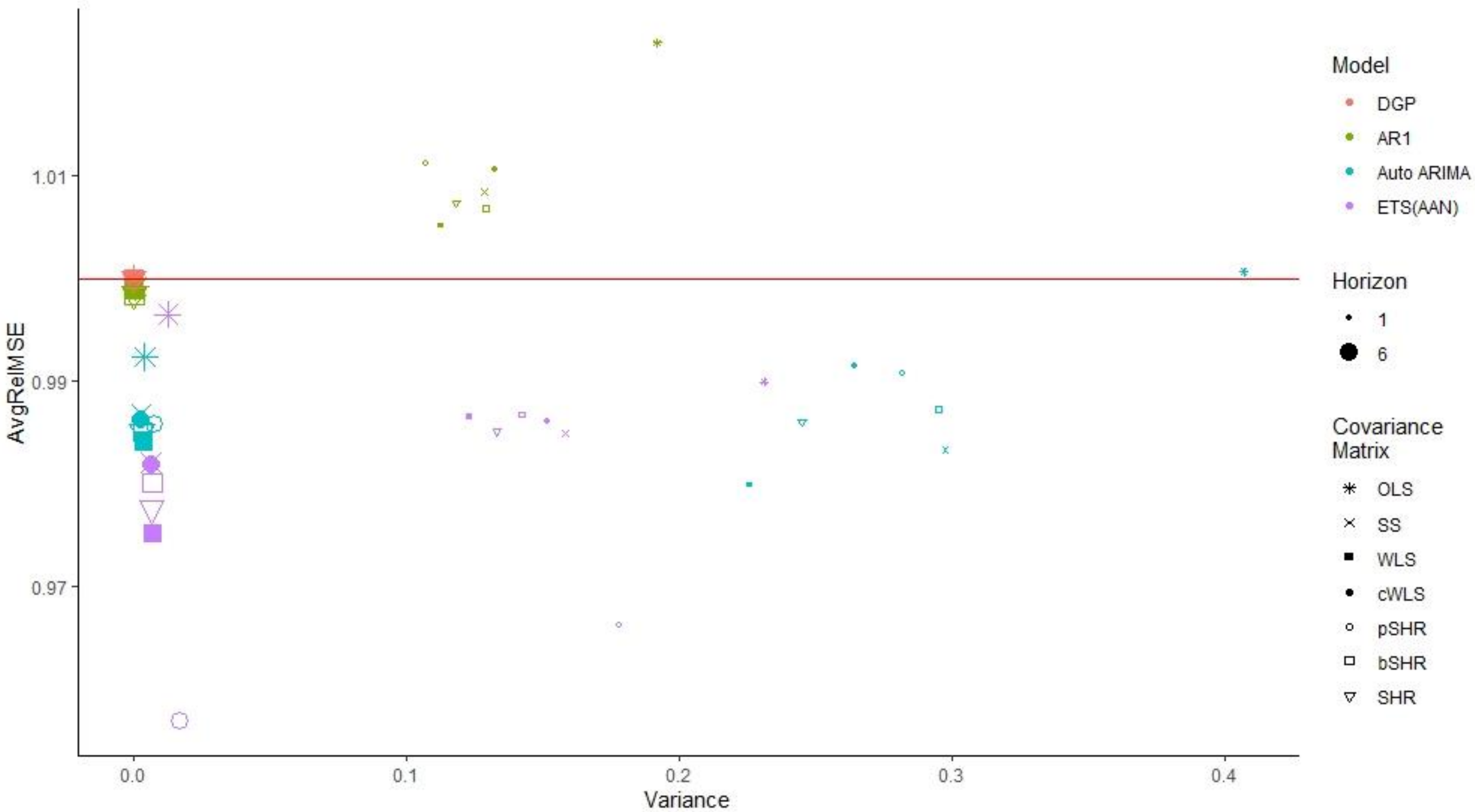
- When we estimate the model with certainty, i.e. perfectly estimate the parameters and the structure of the model, no benefit of reconciliation
- The benefits will be larger when we have mis-specified models and more complex covariance matrix at the cost of variability
- However, The gain of reconciliation becomes limited when we have larger sample sizes and longer forecast horizons
- OLS: reconciled forecasts are never worse than base forecasts → align to Theorem 3.1. in Gamakumara et al 2018

Accuracy vs Variance (RelTotSE)



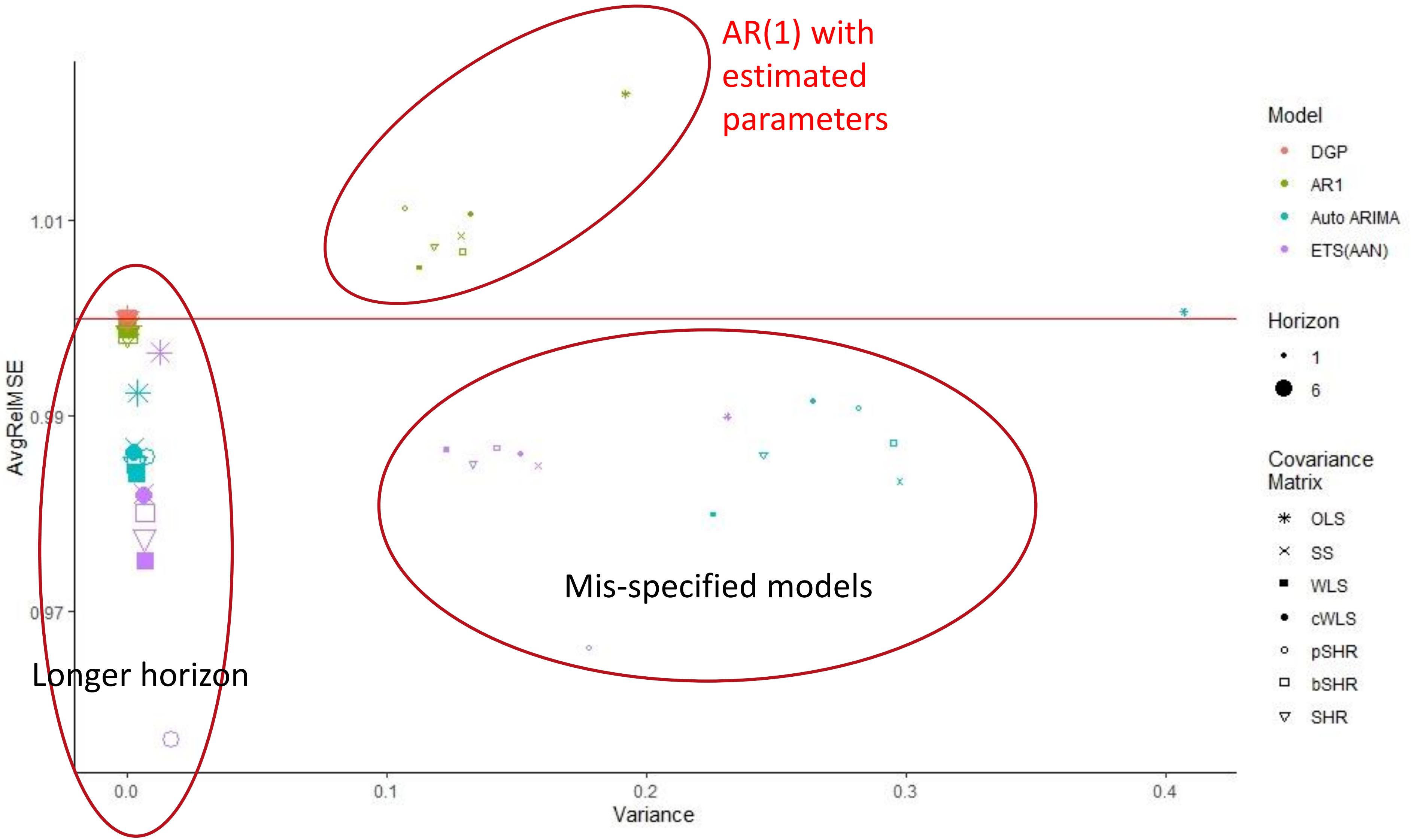
MinT-Sample (EMP) is omitted; the red line is the threshold when the reconciled forecast is worse than the base; the results are from the sample size of 24; we calculate the variance from 500 runs in the simulation, for each scenario.

Accuracy vs Variance (AvgRelMSE)



MinT-Sample (EMP) is omitted; the red line is the threshold when the reconciled forecast is worse than the base; the results are from the sample size of 24; we calculate the variance from 500 runs in the simulation, for each scenario.

Accuracy vs Variance (AvgRelMSE)

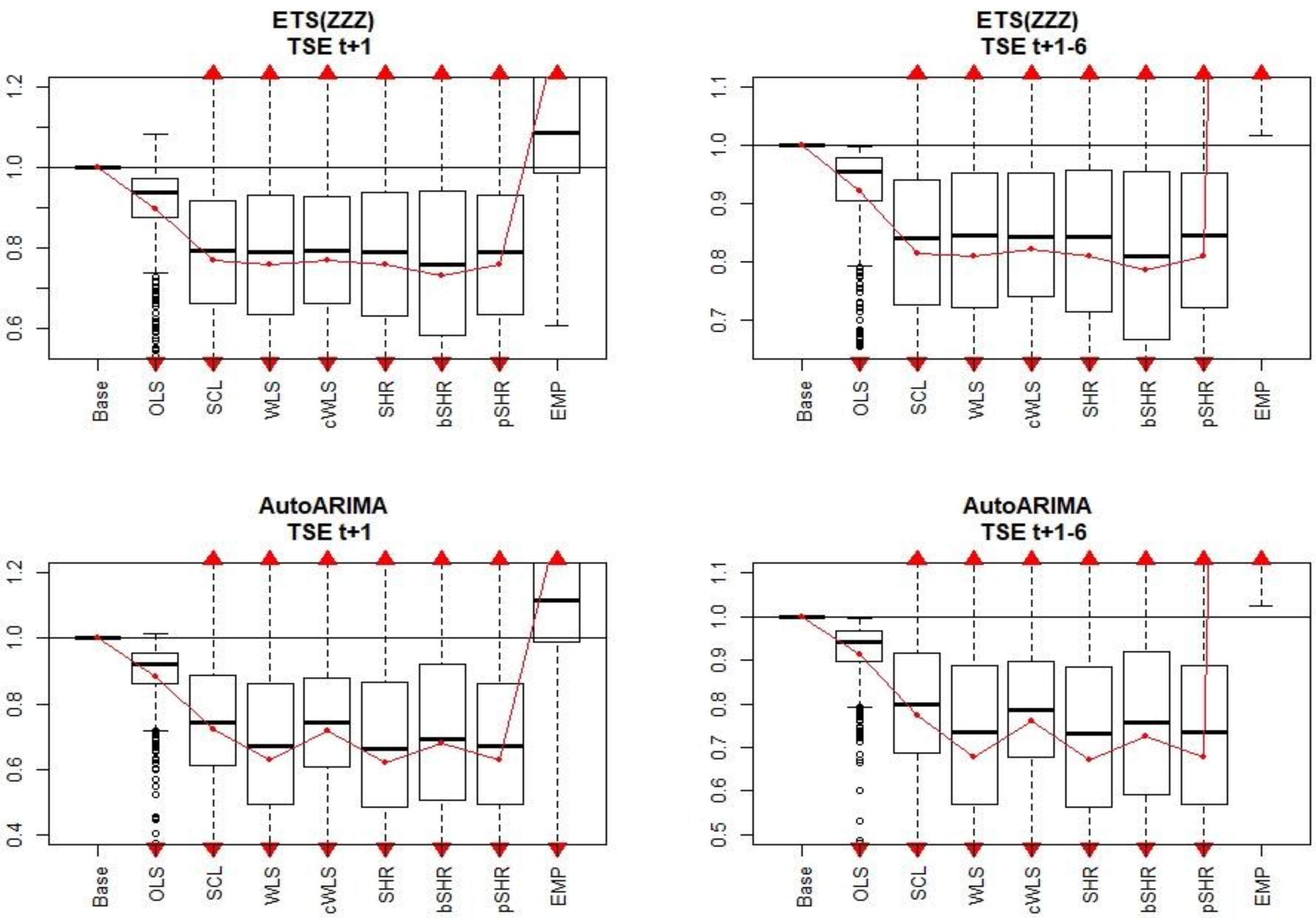


MinT-Sample (EMP) is omitted; the red line is the threshold when the reconciled forecast is worse than the base; the results are from the sample size of 24; we calculate the variance from 500 runs in the simulation, for each scenario.

Findings: Error Metrics

- The choice of error metrics is important
- Different error metrics tell different stories
  - RelTotSE: overall performance on average
  - AvgRelMSE: a summary of individual performance
- Using RelTotSE confirms the theories in Gamakumara et al 2018
- Little benefits are shown from AvgRelMSE
- <u>NONETHELESS,</u> we observe variability regardless the error metrics

Findings: Large Hierarchy, Sample = 30



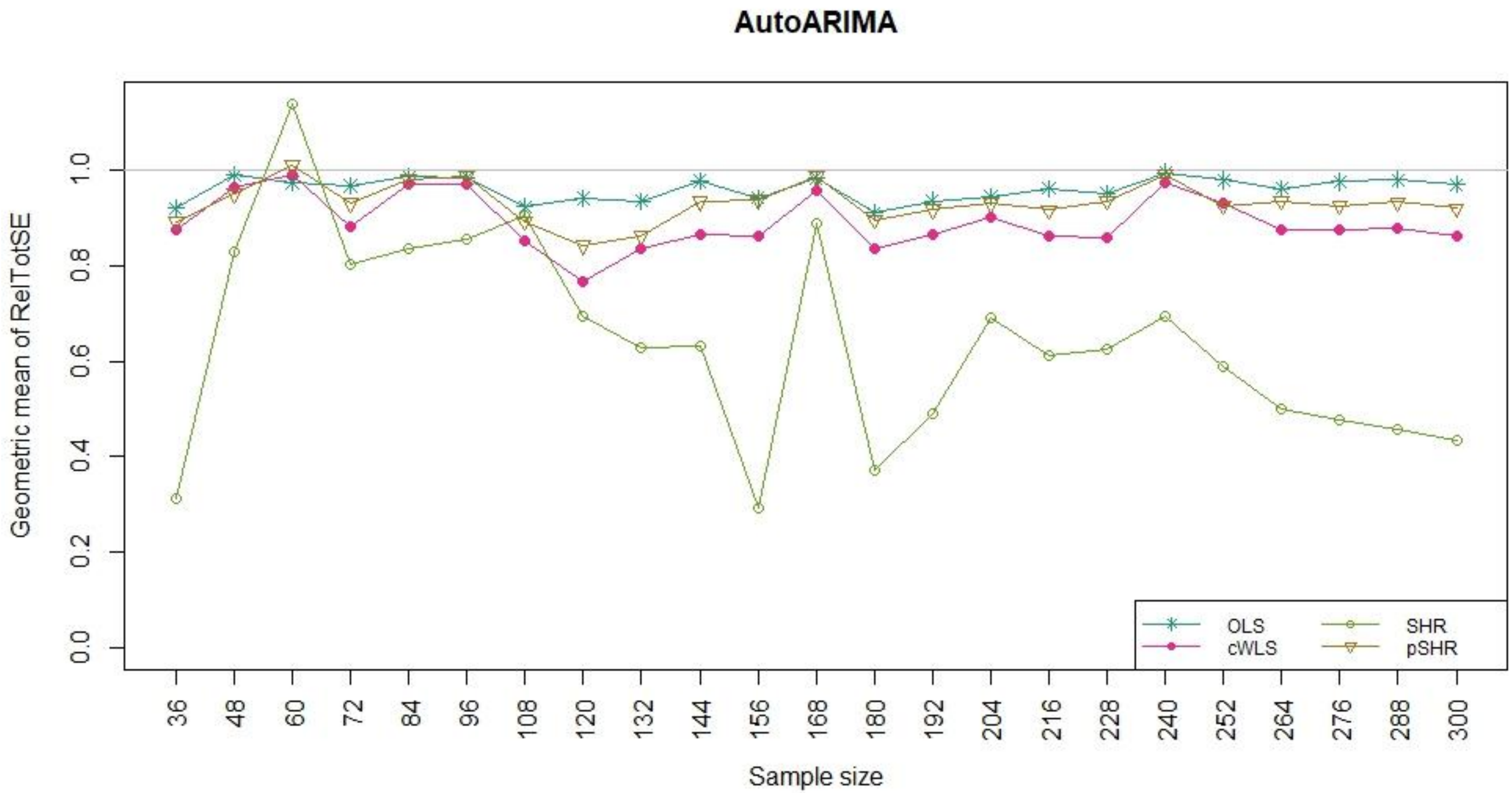
Larger sample sizes and hierarchy sizes confirm the findings

A CASE STUDY: SCANDINAVIAN UNEMPLOYMENT MONTHLY DATA

Scandinavian Unemployment

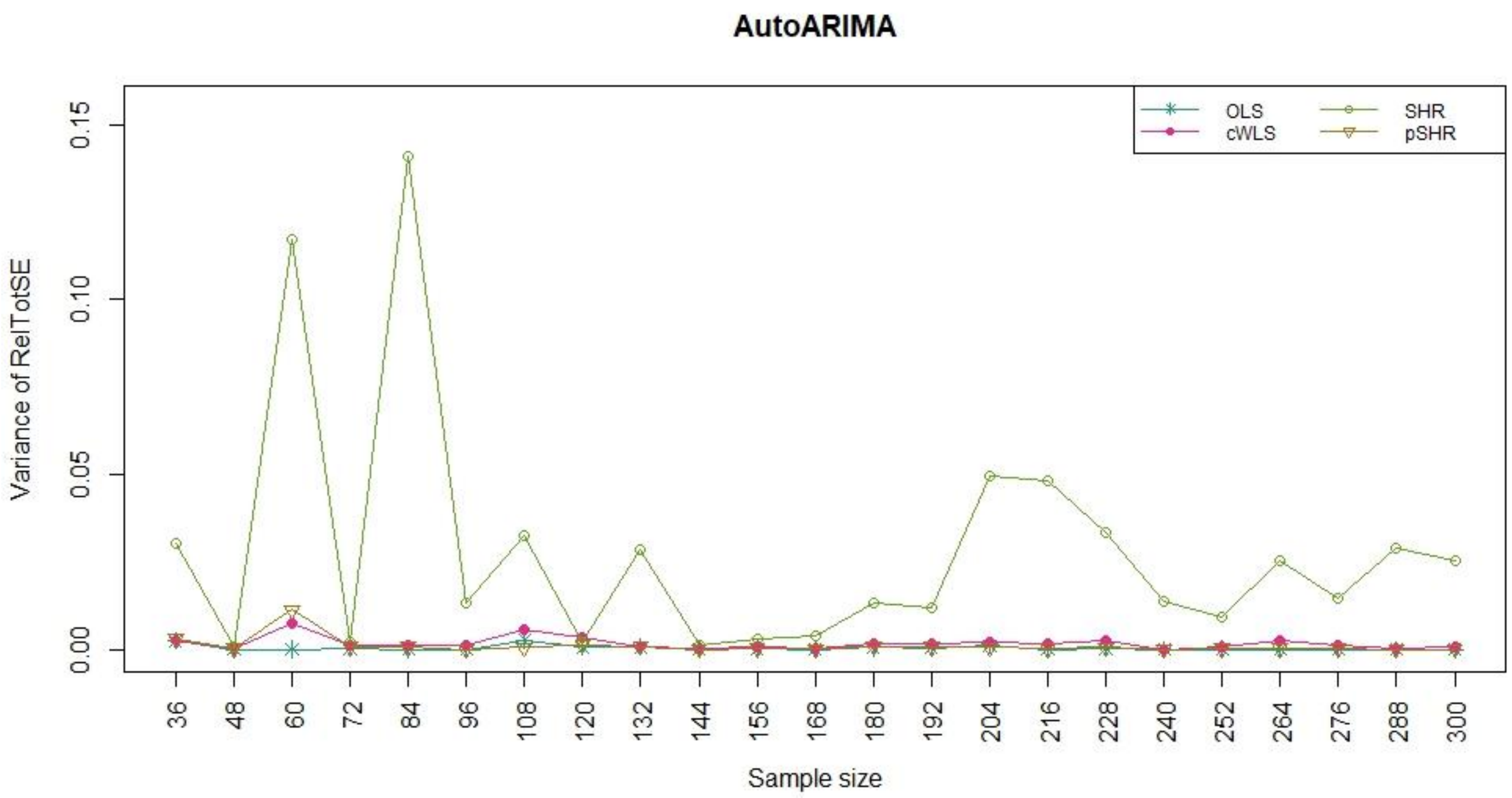
- Data: Scandinavian unemployment monthly data from January 1989 to December 2014
  - Structured based on age, gender, and countries
  - Age: 15-24 years old and over 25 years old
  - Gender: male and female
  - Countries: Denmark, Sweden, Finland, Norway
- Sample size: 36 – 300, with a step of 12
- Forecasting models: AutoARIMA and ETS with automatic selection
- Generate from one- to twelve-step ahead forecasts
- Implement 8 covariance matrix approximations

Scandinavian Unemployment



The geometric means are calculated across forecast horizons

Scandinavian Unemployment



The variances are calculated across forecast horizons

Finding: Scandinavian Unemployment

- The findings confirm the simulation results
- MinT-Shrink performs the best among all at the cost of very volatile results across the forecast horizons.
- The second best is WLS, which ignores the off-diagonal information, and it is competing with pSHR
- Exploiting information in the off-diagonal of the covariance matrix improves the accuracy but increases its variability

DISCUSSION

Discussion: Source of Uncertainties

- Stochastic coherency shifts our paradigm to a non-deterministic forecast reconciliation
- Two sources of uncertainties in forecast reconciliation
  - Base forecast uncertainty: the main source
  - Reconciliation weight uncertainty: amplified the former
- Forecast reconciliation improves the accuracy, on average
  - It depends on the quality of the base forecasts and of the reconciliation weights

Discussion: Error Metrics

- The evidence shows that the MinT Reconciliation works as the theories tell us, given that we use RelTotSE or “loss function error metric”
- However, the decision error metric does not show significant benefits of reconciliation, clearly.
  - AvgRelMSE does not align the loss function of MinT Reconciliation
- However, we observe uncertainty from the two measures → uncertainty in forecast reconciliation is inevitable regardless the metrics

Conclusion

- We propose stochastic coherency to overcome a limitation in the conventional coherency in hierarchical forecasting
- We observe two sources of uncertainties in forecast reconciliation, namely base forecast and reconciliation weight uncertainty
- Practical framework is proposed to mitigate uncertainty, i.e. comparing the benefits and the costs given the base model specification and the covariance matrix approximation

Our homework?

- It is not finished yet!
- We see the evidence of the uncertainty in the projection matrix but we have not yet understood the behaviour of the projection matrix given any approximation of the covariance matrix
- If we find a way to understand the projection matrix, then we may suggest a selection framework in forecast reconciliation

Thank you for your attention!

Kandrika Pritularga

k.pritularga@lancaster.ac.uk

Marketing Analytics & Forecasting | Lancaster University Management School

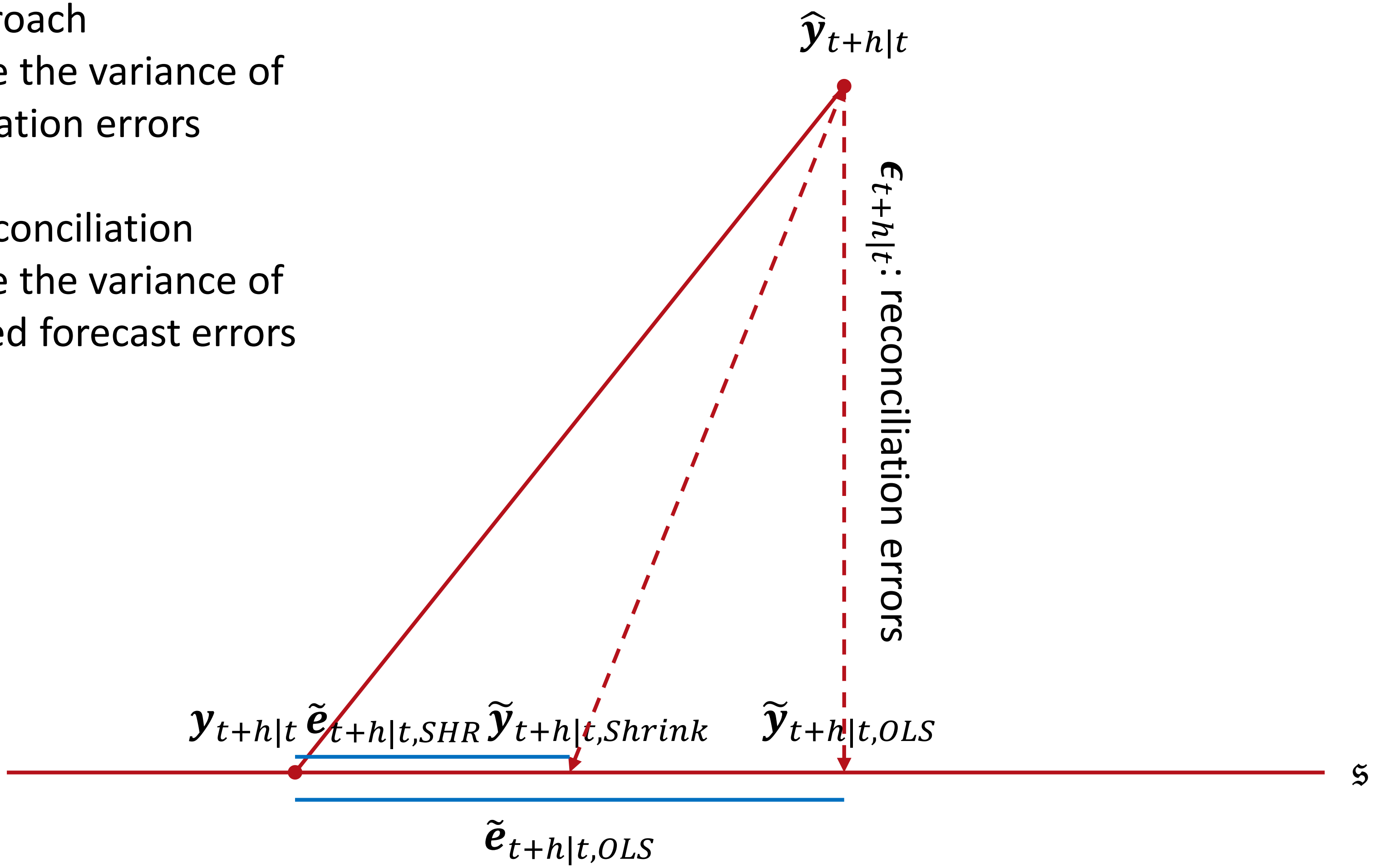
References

- Athanasopoulos, G., Kourentzes, N., 2020. On the evaluation of hierarchical forecasts. doi:10.1017/s0950184300800014.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. International Journal of Forecasting, 29(3), 510-522.
- Gamakumara, P., Panagiotelis, A., Athanasopoulos, G., Hyndman, R.J., 2018. Probabilistic Forecasts in Hierarchical Time Series.
- Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., Shang, H.L., 2011. Optimal combination forecasts for hierarchical time series. Computational Statistics and Data Analysis doi:10.1016/j.csda.2011.03.006.
- Panagiotelis, A., Gamakumara, P., Athanasopoulos, G. & Hyndman, R.J. (2019). Forecast reconciliation A geometric view with new insights on bias correction (Department of Econometrics and Business Statistics Working Paper Series 18/19).
- Wickramasuriya, S.L., Athanasopoulos, G., Hyndman, R.J., 2018. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. Journal of the American Statistical Association 1459. doi:10.1080/01621459.2018.1448825.

Geometric Presentation of Forecast Reconciliation

GLS approach
Minimise the variance of reconciliation errors

MinT Reconciliation
Minimise the variance of reconciled forecast errors



$\mathfrak{s}$ is the coherency space, defined by Gamakumara et al 2018 and Panagiotelis et al 2019

Example of The Case Study Data