

# Probabilistic Forecasts for Hierarchical Time Series

Candidate:

Puwasala Gamakumara

Supervisors:

Rob J. Hyndman, George Athanasopoulos,  
Anastasios Panagiotelis

Department of Econometrics and Business Statistics

Monash University

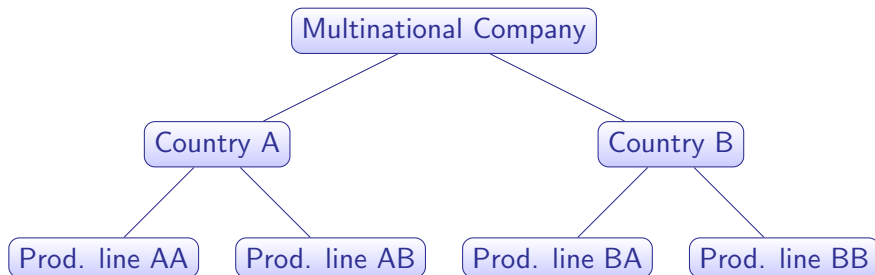
May 23, 2019

- 1 Project 1: Hierarchical Forecast Reconciliation: A Geometric View
- 2 Project 2: Probabilistic Forecasts for Hierarchical Time Series
  - Part 1: Definitions and A Parametric Approach
  - Part 2: A Non-parametric Bootstrap Approach
- 3 Project 3: Hierarchical forecasts for macroeconomic variables - An application to Australian GDP
- 4 Summary and time plan for completion

# Project 1: Hierarchical Forecast Reconciliation: A Geometric View

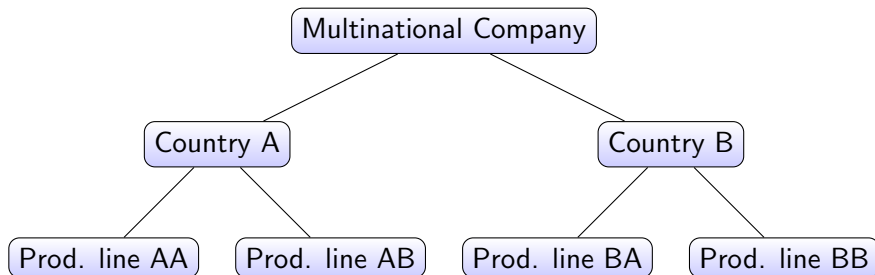
# Project 1: *Motivation and Contributions*

## ■ Example:



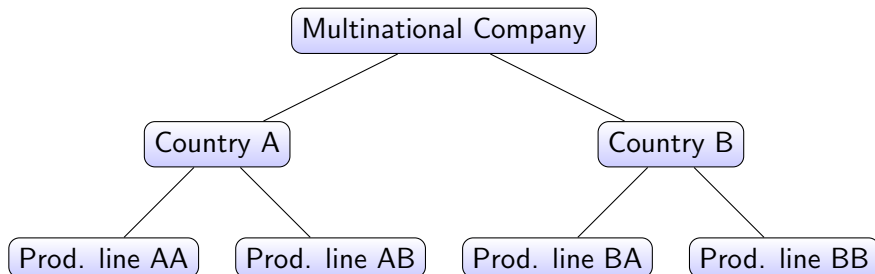
# Project 1: *Motivation and Contributions*

## ■ Example:



# Project 1: *Motivation and Contributions*

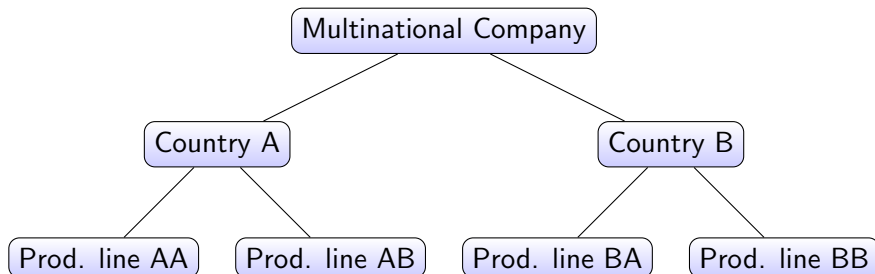
## ■ Example:



- **Hierarchical time series:** A collection of multiple time series that has an inherent aggregation structure.

# Project 1: *Motivation and Contributions*

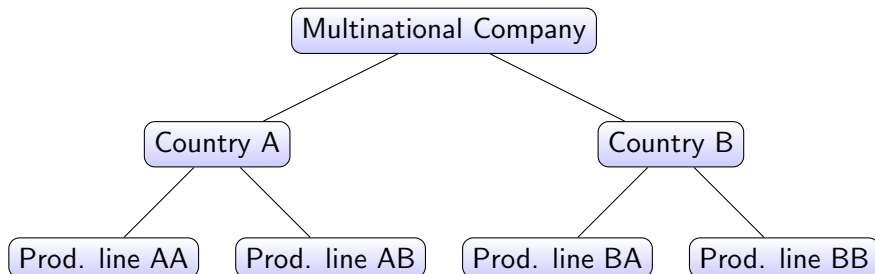
## ■ Example:



- **Hierarchical time series:** A collection of multiple time series that has an inherent aggregation structure.
- Forecasts should add up. We call these *coherent*.

# Project 1: *Motivation and Contributions*

## ■ Example:



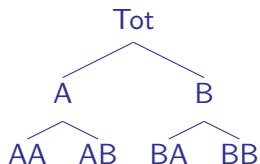
- **Hierarchical time series:** A collection of multiple time series that has an inherent aggregation structure.
- Forecasts should add up. We call these *coherent*.
- **Contribution:** Provide definitions for coherency and reconciliation of point forecasts in terms of geometric concepts.



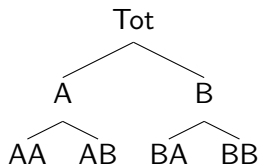
# Project 1: *Notations and Preliminaries*

# Project 1: *Notations and Preliminaries*

# Project 1: *Notations and Preliminaries*

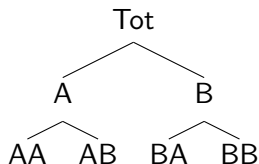


# Project 1: *Notations and Preliminaries*



# Project 1: *Notations and Preliminaries*

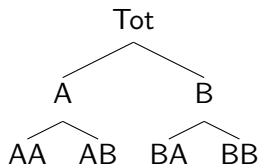
$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$



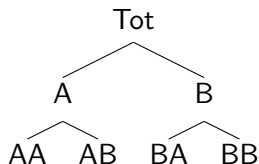
# Project 1: *Notations and Preliminaries*

$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$



# Project 1: *Notations and Preliminaries*

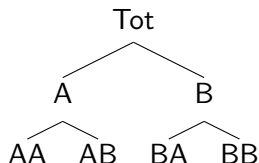


$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

# Project 1: *Notations and Preliminaries*



$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

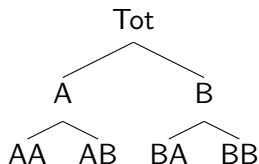
$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$



# Project 1: *Notations and Preliminaries*



$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

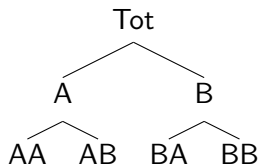
$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \mathbf{I}_{4 \times 4} \end{pmatrix}$$

# Project 1: Notations and Preliminaries



$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

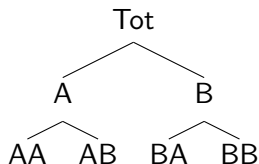
$$n = 7$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \mathbf{I}_{4 \times 4} \end{pmatrix}$$

- Due to the aggregation nature of the hierarchy we have,

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t.$$

# Project 1: Notations and Preliminaries



$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

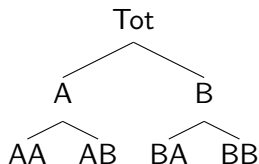
$$n = 7$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \mathbf{I}_{4 \times 4} \end{pmatrix}$$

- Due to the aggregation nature of the hierarchy we have,

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t.$$

# Project 1: Notations and Preliminaries



$$\mathbf{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\mathbf{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

$\mathbf{I}_{4 \times 4}$

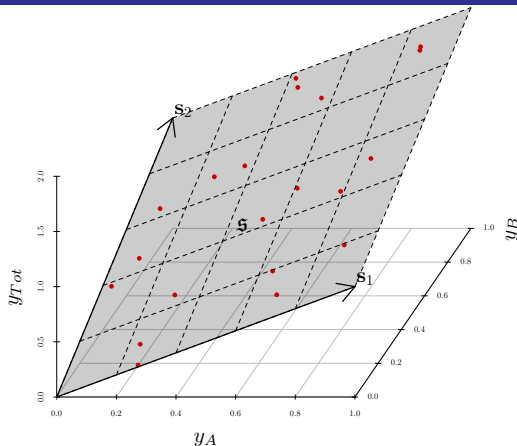
- Due to the aggregation nature of the hierarchy we have,

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t.$$

## Coherent subspace

The  $m$ -dimensional linear subspace  $\mathfrak{s} \subset \mathbb{R}^n$  that is spanned by the columns of  $\mathbf{S}$ , i.e.  $\mathfrak{s} = \text{span}(\mathbf{S})$ , is defined as the *coherent space*.

# Project 1: Coherent forecasts



- Three dimensional hierarchy,  $y_{Tot} = y_A + y_B$ .
- $\vec{s}_1 = (1, 1, 0)'$  and  $\vec{s}_2 = (1, 0, 1)'$  form a basis for  $\mathfrak{s}$ .

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$



# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

- If  $g(\cdot)$  is linear,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}.$$

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

- If  $g(\cdot)$  is linear,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}.$$

- Single level approaches:

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

- If  $g(\cdot)$  is linear,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}.$$

- Single level approaches:

- Bottom-up:  $\mathbf{G} = (\mathbf{0}_{(m \times n-m)} \quad \mathbf{I}_m)$

# Project 1: *Point forecast reconciliation*

- Let  $\hat{\mathbf{y}} \in \mathbb{R}^n$  be an incoherent forecast and  $g(\cdot)$  be a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

## Definition

**Point forecasts**  $\tilde{\mathbf{y}}$  are reconciled with respect to  $g(\cdot)$  iff

$$\tilde{\mathbf{y}} = \mathbf{S}g(\hat{\mathbf{y}})$$

- If  $g(\cdot)$  is linear,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}.$$

- Single level approaches:

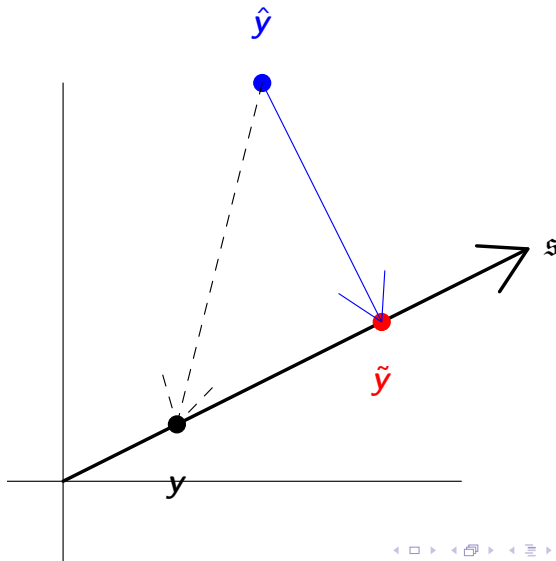
- Bottom-up:  $\mathbf{G} = \begin{pmatrix} \mathbf{0}_{(m \times n-m)} & \mathbf{I}_m \end{pmatrix}$
- Top-down:  $\mathbf{G} = \begin{pmatrix} \mathbf{p} & \mathbf{0}_{(m \times n-1)} \end{pmatrix}$

# Project 1: *Point forecast reconciliation - OLS*

$$\mathbf{S}\mathbf{G} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' \text{ (Hyndman et al., 2011)}$$

# Project 1: *Point forecast reconciliation - OLS*

$$\mathbf{S}\mathbf{G} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' \text{ (Hyndman et al., 2011)}$$



# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, 2018).

$$\mathbf{SG} = \mathbf{S} (\mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$

- **Geometry**



# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$\mathbf{SG} = \mathbf{S} (\mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $\mathbf{W}_{T+h}$  of  $\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$ .

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $W_{T+h}$  of  $y_{T+h} - \hat{y}_{T+h}$ .
- This can be estimated using in-sample forecast errors.

$$\hat{W}_{T+h} = \sum_{t=1}^T (y_t - \hat{y}_t)(y_t - \hat{y}_t)'$$

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$\mathbf{SG} = \mathbf{S} (\mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $\mathbf{W}_{T+h}$  of  $\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$ .
- This can be estimated using in-sample forecast errors.

$$\hat{\mathbf{W}}_{T+h} = \sum_{t=1}^T (\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)'$$

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $W_{T+h}$  of  $y_{T+h} - \hat{y}_{T+h}$ .
- This can be estimated using in-sample forecast errors.

$$\hat{W}_{T+h} = \sum_{t=1}^T (y_t - \hat{y}_t)(y_t - \hat{y}_t)'$$

- This provides information about the likely direction of the largest error variance.

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$\mathbf{SG} = \mathbf{S} (\mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $\mathbf{W}_{T+h}$  of  $\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$ .
- This can be estimated using in-sample forecast errors.

$$\hat{\mathbf{W}}_{T+h} = \sum_{t=1}^T (\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)'$$

- This provides information about the likely direction of the largest error variance.

# Project 1: *Point forecast reconciliation - MinT*

- Minimises the trace of mean squared reconciled forecast errors (Wickramasuriya, Athanasopoulos, and Hyndman, [2018](#)).

$$\mathbf{SG} = \mathbf{S} (\mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

## ■ Geometry

- Consider the covariance matrix  $\mathbf{W}_{T+h}$  of  $\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h}$ .
- This can be estimated using in-sample forecast errors.

$$\hat{\mathbf{W}}_{T+h} = \sum_{t=1}^T (\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)'$$

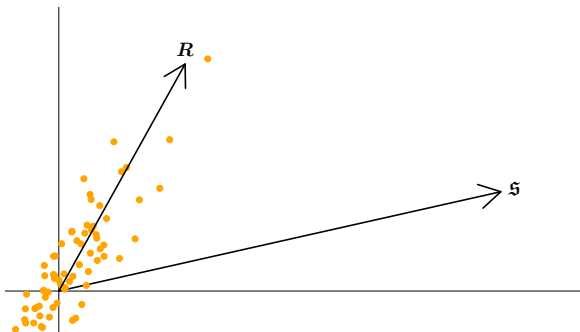
- This provides information about the likely direction of the largest error variance.
- Projecting along this direction is more likely to result in reconciled forecasts that are closer to the target.

# Project 1: *Point forecast reconciliation - MinT*

$$\mathbf{SG} = \mathbf{S} \left( \mathbf{S}' \mathbf{W}_{T+h}^{-1} \mathbf{S} \right)^{-1} \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

# Project 1: *Point forecast reconciliation - MinT*

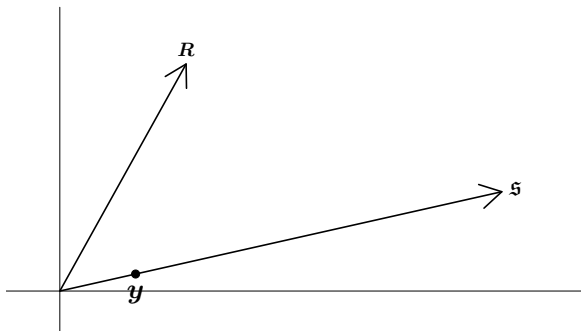
$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$





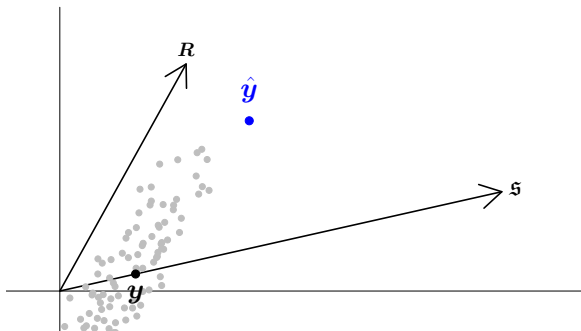
# Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



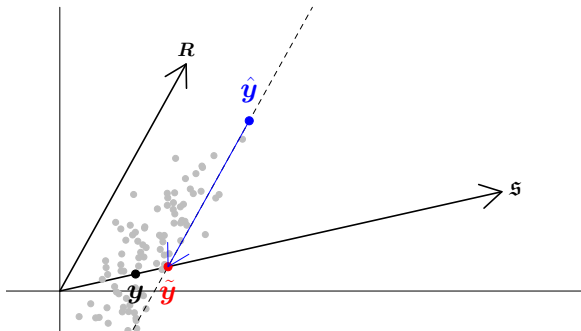
# Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



# Project 1: *Point forecast reconciliation - MinT*

$$SG = S (S' W_{T+h}^{-1} S)^{-1} S' W_{T+h}^{-1}$$



# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.

# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.

# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.
- What if the incoherent forecasts are biased?

# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.
- What if the incoherent forecasts are biased?

# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.
- What if the incoherent forecasts are biased?
- Can we bias correct and proceed with projections?



# Project 1: *Why projections?*

- Projections onto the coherent subspace preserve unbiasedness.
- What if the incoherent forecasts are biased?
- Can we bias correct and proceed with projections?  
We investigate this further.

# Project 2: Probabilistic Forecasts for Hierarchical Time Series

# Part 1: Definitions and A Parametric Approach

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. (2017)



## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.

# Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. (2017)
  - Jeon, Panagiotelis, and Petropoulos (2018)
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.
  - 2 A parametric framework for probabilistic forecast reconciliation.

# Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. ([2017](#))
  - Jeon, Panagiotelis, and Petropoulos ([2018](#))
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.
  - 2 A parametric framework for probabilistic forecast reconciliation.



## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. (2017)
  - Jeon, Panagiotelis, and Petropoulos (2018)
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.
  - 2 A parametric framework for probabilistic forecast reconciliation.
  - 3 A non-parametric framework for probabilistic forecast reconciliation.

## Project 2: *Motivation and Contributions*

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ★ Aggregation structure
  - ★ Correlation structure
- Lack of attention
  - Ben Taieb et al. (2017)
  - Jeon, Panagiotelis, and Petropoulos (2018)
- Extending “reconciliation” into a probabilistic framework.
- **Contributions:**
  - 1 Provide definitions for coherency and reconciliation for probabilistic forecasts.
  - 2 A parametric framework for probabilistic forecast reconciliation.
  - 3 A non-parametric framework for probabilistic forecast reconciliation.

## Project 2: *Coherent probabilistic forecasts*

Let  $(\mathbb{R}^m, \mathcal{F}_{\mathbb{R}^m}, \nu)$  and  $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \mu)$  be probability triples on  $m$ -dimensional space and the coherent subspace respectively.

### Definition

The probability measure  $\mu$  is coherent if

$$\nu(\mathcal{B}) = \mu(s(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}^m}$$

where  $s(\mathcal{B})$  is the image of  $\mathcal{B}$  under premultiplication by  $\mathbf{S}$

## Project 2: *Reconciled Probabilistic Forecast*

Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a function. Then

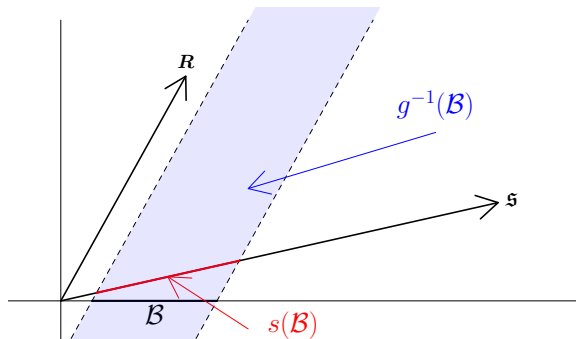
### Definition

The probability triple  $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \tilde{\nu})$  reconciles the probability triple  $(\mathbb{R}^n, \mathcal{F}_{\mathbb{R}^n}, \hat{\nu})$  with with respect to  $g$  iff

$$\tilde{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}) = \hat{\nu}(g^{-1}(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}^m}$$

where  $g^{-1}$  is the pre-image of  $g$ .

## Project 2: Geometry



$$\tilde{\nu}(s(\mathcal{B})) = \hat{\nu}(g^{-1}(\mathcal{B})) \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}_m}$$

## Project 2: *Analytically*

If we have an unreconciled density the reconciled density can be obtained by linear transformations and marginalisation.

$$\begin{aligned}\Pr(\hat{\mathbf{y}} \in g^{-1}(\mathcal{B})) &= \int_{g^{-1}(\mathcal{B})} f(\hat{\mathbf{y}}) d\hat{\mathbf{y}} \\ &= \int_{\mathcal{B}} \int f(\mathbf{S}\tilde{\mathbf{b}} + \mathbf{R}\tilde{\mathbf{a}}) |(\mathbf{S} \ \mathbf{R})| d\tilde{\mathbf{a}} d\tilde{\mathbf{b}} \\ &= \Pr(\tilde{\mathbf{b}} \in \mathcal{B})\end{aligned}$$

## Project 2: *Assuming Gaussian distribution*

A1

- Let  $\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h})$  be an incoherent forecast distribution at time  $T + h$  where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and  $\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is the incoherent variance

## Project 2: Assuming Gaussian distribution

A1

- Let  $\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h})$  be an incoherent forecast distribution at time  $T + h$  where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and  $\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is the incoherent variance
- The reconciled Gaussian distribution is given by,

$$\mathcal{N}(\mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}')$$



## Project 2: Assuming Gaussian distribution

A1

- Let  $\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h})$  be an incoherent forecast distribution at time  $T + h$  where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and  $\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is the incoherent variance
- The reconciled Gaussian distribution is given by,

$$\mathcal{N}(\mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}')$$

- $\mathbf{G} = (\mathbf{S}'\mathbf{W}_{T+h}^{-1}\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_{T+h}^{-1}$  minimizes the energy score in the limiting case

# Project 2: Assuming Gaussian distribution

A1

- Let  $\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h})$  be an incoherent forecast distribution at time  $T + h$  where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and  $\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is the incoherent variance
- The reconciled Gaussian distribution is given by,

$$\mathcal{N}(\mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}')$$

- $\mathbf{G} = (\mathbf{S}'\mathbf{W}_{T+h}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_{T+h}^{-1}$  minimizes the energy score in the limiting case
- Simulation study evidence for improved predictive performance in reconciled Gaussian forecast distributions.

## Project 2: *For elliptical distributions*

Consider the case where the base and true predictive distributions are elliptical.

### Theorem

*There exists a matrix  $\mathbf{G}$  such that the true predictive distribution can be recovered by linear reconciliation.*

This follows from the closure property of elliptical distributions under affine transformations and marginalisation.

## Project 2: *Evaluation*

- Scoring rules can be used to evaluate probabilistic forecasts A2

- Scoring rules can be used to evaluate probabilistic forecasts A2

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score



- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score

# Project 2: *Evaluation*

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score
- We may want to compare

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score
- We may want to compare
  - Coherent vs Incoherent

- Scoring rules can be used to evaluate probabilistic forecasts A2
  - Log Score
  - Energy Score
  - Variogram score
- We may want to compare
  - Coherent vs Incoherent
  - Coherent vs Coherent

## Project 2: *Evaluation - Coherent vs Incoherent*

When using log score

### Theorem

*Let  $f(\mathbf{y})$  be the true predictive density (on  $\mathfrak{s}$ ) and  $LS$  be the (negatively-oriented) log score. Then there exists an unreconciled density  $\hat{f}(\mathbf{y})$  on  $\mathbb{R}^n$  such that*

$$E_{\mathbf{y}} [LS(\hat{f}, \mathbf{y})] < E_{\mathbf{y}} [LS(f, \mathbf{y})]$$

The log score is not proper **in this context**.

# Today's talk

- A non-parametric bootstrap approach for probabilistic forecast reconciliation.

# Today's talk

- A non-parametric bootstrap approach for probabilistic forecast reconciliation.



# Today's talk

- A non-parametric bootstrap approach for probabilistic forecast reconciliation.
- Hierarchical forecasts for macroeconomic variables - An application to Australian GDP.

# Part 2: A non-parametric bootstrap approach for probabilistic forecast reconciliation

# Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.

# Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.

# Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.
- Suppose  $\hat{\mathbf{y}}_{T+h}^{[1]}, \dots, \hat{\mathbf{y}}_{T+h}^{[J]}$  is a sample from the incoherent predictive distribution.

# Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.
- Suppose  $\hat{\mathbf{y}}_{T+h}^{[1]}, \dots, \hat{\mathbf{y}}_{T+h}^{[J]}$  is a sample from the incoherent predictive distribution.

# Probabilistic forecast reconciliation: Non-parametric approach

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.
- Suppose  $\hat{\mathbf{y}}_{T+h}^{[1]}, \dots, \hat{\mathbf{y}}_{T+h}^{[J]}$  is a sample from the incoherent predictive distribution.
- Then setting  $\tilde{\mathbf{y}}_{T+h}^{[j]} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}^{[j]}$  produces a sample from the reconciled predictive distribution with respect to  $\mathbf{G}$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .



# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .
- 4 Generate  $h$ -step ahead sample paths from the fitted models incorporating  $\mathbf{\Gamma}^b$ . Denote these by  $\hat{\mathbf{y}}_{T+h}^b$ , for  $h = 1, \dots, H$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .
- 4 Generate  $h$ -step ahead sample paths from the fitted models incorporating  $\mathbf{\Gamma}^b$ . Denote these by  $\hat{\mathbf{y}}_{T+h}^b$ , for  $h = 1, \dots, H$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .
- 4 Generate  $h$ -step ahead sample paths from the fitted models incorporating  $\mathbf{\Gamma}^b$ . Denote these by  $\hat{\mathbf{y}}_{T+h}^b$ , for  $h = 1, \dots, H$ .
- 5 Repeat step 3 and 4 for  $b = 1, \dots, B$  times. Denote these as  $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$  for all  $h$ .



# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .
- 4 Generate  $h$ -step ahead sample paths from the fitted models incorporating  $\mathbf{\Gamma}^b$ . Denote these by  $\hat{\mathbf{y}}_{T+h}^b$ , for  $h = 1, \dots, H$ .
- 5 Repeat step 3 and 4 for  $b = 1, \dots, B$  times. Denote these as  $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$  for all  $h$ .

# A Non-parametric approach

- 1 Fit univariate models at each node using data up to time  $T$ .
- 2 Let  $\mathbf{\Gamma}_{(T \times n)} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_T)'$  be a matrix of in-sample residuals where  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ .
- 3 Let  $\mathbf{\Gamma}_{(H \times n)}^b = (\mathbf{e}_1^b, \dots, \mathbf{e}_H^b)'$  be a block bootstrap sample of size  $H$  from  $\mathbf{\Gamma}$ .
- 4 Generate  $h$ -step ahead sample paths from the fitted models incorporating  $\mathbf{\Gamma}^b$ . Denote these by  $\hat{\mathbf{y}}_{T+h}^b$ , for  $h = 1, \dots, H$ .
- 5 Repeat step 3 and 4 for  $b = 1, \dots, B$  times. Denote these as  $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$  for all  $h$ .
- 6 Setting  $\tilde{\mathbf{Y}}_{T+h} = \mathbf{SG} \hat{\mathbf{Y}}_{T+h}'$  produces a sample from the reconciled distribution.

# Optimal reconciliation of future paths

- We propose to find an optimal  $\mathbf{G}_h$  matrix by minimizing Energy score.

$$\operatorname{argmin}_{\mathbf{G}_h} E_Q[\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h})], \quad \tilde{\mathbf{F}} := \tilde{\mathbf{Y}}_{T+h} = \mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}'_{T+h}$$

where,

$$\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h}) = E_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha, \\ \alpha \in (0, 2]$$

# Optimal reconciliation of future paths

- We propose to find an optimal  $\mathbf{G}_h$  matrix by minimizing Energy score.

$$\operatorname{argmin}_{\mathbf{G}_h} E_Q[\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h})], \quad \tilde{\mathbf{F}} := \tilde{\mathbf{Y}}_{T+h} = \mathbf{S}\mathbf{G}_h\hat{\mathbf{Y}}'_{T+h}$$

where,

$$\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h}) = E_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha, \\ \alpha \in (0, 2]$$

- Monte-Carlo approximation to the above objective function is,

# Optimal reconciliation of future paths

- We propose to find an optimal  $\mathbf{G}_h$  matrix by minimizing Energy score.

$$\operatorname{argmin}_{\mathbf{G}_h} \mathbb{E}_Q[\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h})], \quad \tilde{\mathbf{F}} := \tilde{\mathbf{Y}}_{T+h} = \mathbf{S}\mathbf{G}_h \hat{\mathbf{Y}}'_{T+h}$$

where,

$$\text{eS}(\tilde{\mathbf{F}}, \mathbf{y}_{T+h}) = \mathbb{E}_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} \mathbb{E}_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha, \\ \alpha \in (0, 2]$$

- Monte-Carlo approximation to the above objective function is,

$$\operatorname{argmin}_{\mathbf{G}_h} \sum_{i=1}^N \left\{ \frac{1}{B} \sum_{b=1}^B \|\mathbf{S}\mathbf{G}_h \hat{\mathbf{y}}_{T+h,i}^b - \mathbf{y}_{T+h}\| - \frac{1}{2(B-1)} \sum_{b=1}^{B-1} \|\mathbf{S}\mathbf{G}_h (\hat{\mathbf{y}}_{T+h,i}^b - \hat{\mathbf{y}}_{T+h,i}^{b+1})\| \right\}$$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

## Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$



# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

**Method 3:** Optimising Cholesky of  $\mathbf{W}_h$  - restricted for scaling

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

**Method 3:** Optimising Cholesky of  $\mathbf{W}_h$  - restricted for scaling  
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  s.t.  $\mathbf{i}' \mathbf{W}_h \mathbf{i} = 1$  where  $\mathbf{i} = (1, 0, \dots, 0)'$

# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

**Method 3:** Optimising Cholesky of  $\mathbf{W}_h$  - restricted for scaling  
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  s.t.  $\mathbf{i}' \mathbf{W}_h \mathbf{i} = 1$  where  $\mathbf{i} = (1, 0, \dots, 0)'$



# Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the  $\mathbf{G}_h$  matrix

$$\mathbf{G}_h = (\mathbf{S}'\mathbf{W}_h\mathbf{S})^{-1} \mathbf{S}'\mathbf{W}_h \quad (1)$$

- We propose four methods to optimise  $\mathbf{G}_h$

**Method 1:** Optimising  $\mathbf{W}_h$

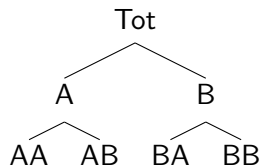
**Method 2:** Optimising Cholesky decomposition of  $\mathbf{W}_h$   
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  where  $\mathbf{R}_h$  is an upper triangular matrix

**Method 3:** Optimising Cholesky of  $\mathbf{W}_h$  - restricted for scaling  
 $\mathbf{W}_h = \mathbf{R}_h' \mathbf{R}_h$  s.t.  $\mathbf{i}' \mathbf{W}_h \mathbf{i} = 1$  where  $\mathbf{i} = (1, 0, \dots, 0)'$

**Method 4:** Optimising  $\mathbf{G}_h$  such that  $\mathbf{G}_h \mathbf{S} = \mathbf{I}$

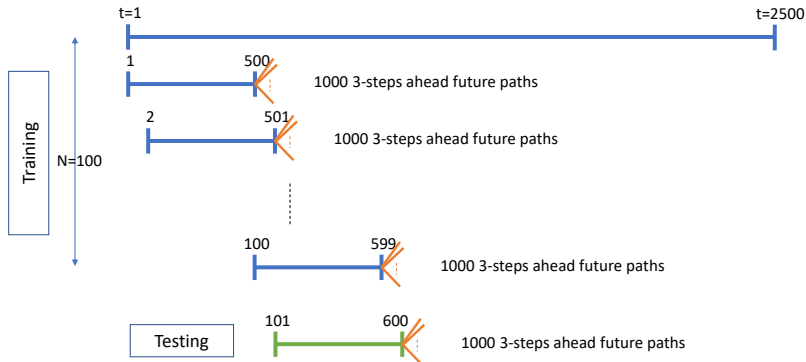
# Monte-Carlo Simulation

- Data generating process



- DGP was designed such that we have much noisier series in the bottom level.

# Monte-Carlo Simulation



# Monte-Carlo Simulation *Cont.*

- Simulation setup:
  - First 2500 observations were generated.

## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.

## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.
- $B = 1000$  of  $h = 1, 2, 3$  steps-ahead bootstrap future paths generated.

## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.
- $B = 1000$  of  $h = 1, 2, 3$  steps-ahead bootstrap future paths generated.
- Training window is rolled one observation ahead and process was repeated until  $N = 100$  incoherent,  $h = 1, 2, 3$  steps-ahead future paths were generated.

## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.
- $B = 1000$  of  $h = 1, 2, 3$  steps-ahead bootstrap future paths generated.
- Training window is rolled one observation ahead and process was repeated until  $N = 100$  incoherent,  $h = 1, 2, 3$  steps-ahead future paths were generated.
- We find the optimal  $\mathbf{G}_h$  for  $h = 1, 2, 3$  that reconciles  $h$ -step-ahead future paths giving minimal average Energy score.



## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.
- $B = 1000$  of  $h = 1, 2, 3$  steps-ahead bootstrap future paths generated.
- Training window is rolled one observation ahead and process was repeated until  $N = 100$  incoherent,  $h = 1, 2, 3$  steps-ahead future paths were generated.
- We find the optimal  $\mathbf{G}_h$  for  $h = 1, 2, 3$  that reconciles  $h$ -step-ahead future paths giving minimal average Energy score.
- This optimal  $\mathbf{G}_h$  is then used to reconcile the incoherent future paths for the test set.

## ■ Simulation setup:

- First 2500 observations were generated.
- Fit Univariate ARIMA models for a rolling window of 500 observations.
- $B = 1000$  of  $h = 1, 2, 3$  steps-ahead bootstrap future paths generated.
- Training window is rolled one observation ahead and process was repeated until  $N = 100$  incoherent,  $h = 1, 2, 3$  steps-ahead future paths were generated.
- We find the optimal  $\mathbf{G}_h$  for  $h = 1, 2, 3$  that reconciles  $h$ -step-ahead future paths giving minimal average Energy score.
- This optimal  $\mathbf{G}_h$  is then used to reconcile the incoherent future paths for the test set.
- The Process was repeated 1000 times and average scores were calculated for the test set.

# Monte-Carlo Simulation *Cont.*

Optimisation method	Hierarchy 1				Hierarchy 2			
	$h = 1$		$h = 3$		$h = 1$		$h = 3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Method 1 - Optimising <b><math>W</math></b>	2.48	0.11	2.75	0.11	5.36	1.21	5.83	1.38
Method 2 - Optimising <b><math>R</math></b>	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 3 - Optimising <b><math>R</math></b> (Restricted)	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 4 - Optimising <b><math>G</math></b>	2.48	0.11	2.75	0.11	5.38	1.21	5.83	1.38

- Parameterisation does not matter

# Monte-Carlo Simulation *Cont.*

## ■ Comparison with point forecast reconciliation methods.

Reconciliation method	Hierarchy 1				Hierarchy 2			
	$h = 1$		$h = 3$		$h = 1$		$h = 3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Optimal <b>G</b>	2.48*	0.106	2.75*	0.106	5.36*	1.21*	5.83*	1.38*
MinT(Shrink)	2.47*	0.105	2.74*	0.105	5.33*	1.19*	5.77*	1.34*
WLS	2.46*	0.105	2.74*	0.105	5.43*	1.23	5.98*	1.40*
OLS	2.54*	0.105	2.80*	0.105	5.51*	1.23	5.98*	1.40*
Base	2.67	0.105	2.94	0.105	5.71	1.28	6.27	1.49

"\*" indicates if the average score for a particular reconciliation method is significantly different from that of base forecasts.

## ■ Reconciliation methods perform better than Base forecasts.

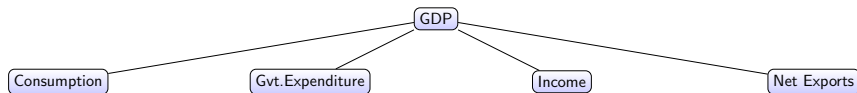
## ■ MinT(Shrink) is at least as good as Optimal method. Thus going forward with MinT projection.

# Project 3: Hierarchical forecasts for macroeconomic variables - An application to Australian GDP

# Macroeconomic forecasting



# Macroeconomic forecasting



- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.

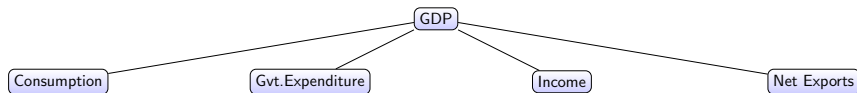
# Macroeconomic forecasting



- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.

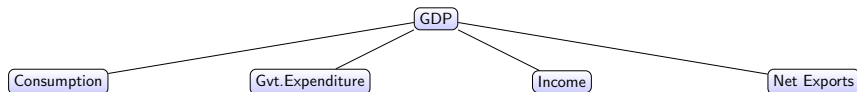


# Macroeconomic forecasting



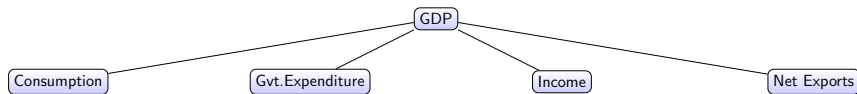
- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs.

# Macroeconomic forecasting



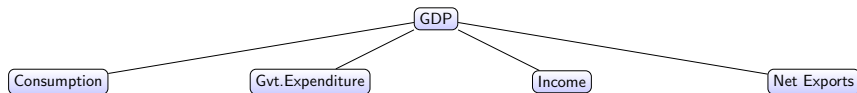
- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs.

# Macroeconomic forecasting



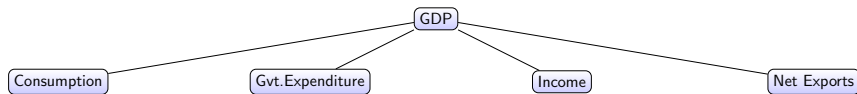
- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs.
- The predictors in these methods commonly include the components of the variables of interest.

# Macroeconomic forecasting



- Common forecasting approaches involves univariate methods or multivariate methods such as VAR.
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs.
- The predictors in these methods commonly include the components of the variables of interest.

# Macroeconomic forecasting



- Common forecasting approaches involve univariate methods or multivariate methods such as VAR.
- The era of big data led to the use of regularization and shrinkage methods - dynamic factor models, Lasso, Bayesian VARs.
- The predictors in these methods commonly include the components of the variables of interest.
- This might fail to reflect the deterministic relationship between macroeconomic variables in the forecasts.

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.



# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.
- Related literature: Only one application on point forecasting for inflation (Capistrán, Constandse, and Ramos-Francia, 2010; Weiss, 2018)

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.
- Related literature: Only one application on point forecasting for inflation (Capistrán, Constandse, and Ramos-Francia, [2010](#); Weiss, [2018](#))

# Macroeconomic forecasting

- Both aligned decision making and forecast accuracy are key concerns for economic agents and policy makers.
- Thus we propose to use hierarchical forecasting methods in macroeconomic forecasts.
- Related literature: Only one application on point forecasting for inflation (Capistrán, Constandse, and Ramos-Francia, [2010](#); Weiss, [2018](#))
- To the best of our knowledge we use hierarchical forecasting methods for point as well as probabilistic forecasts for the first time in macroeconomic literature.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.  
The final GDP figure is obtained as an average of these three figures.



# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.  
The final GDP figure is obtained as an average of these three figures.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.  
The final GDP figure is obtained as an average of these three figures.
- We restrict our attention to nominal, seasonally unadjusted data.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.  
The final GDP figure is obtained as an average of these three figures.
- We restrict our attention to nominal, seasonally unadjusted data.

# Australian GDP : *Data structures*

- We consider Gross Domestic Product (GDP) of Australia with quarterly data spanning the period 1984:Q4–2018:Q3.
- The Australian Bureau of Statistics (ABS) measures GDP using three main approaches - Production, Income and Expenditure.  
The final GDP figure is obtained as an average of these three figures.
- We restrict our attention to nominal, seasonally unadjusted data.
- Thus we concentrate on the Income and Expenditure approaches.

# Australian GDP : *Data structures*

## Income approach

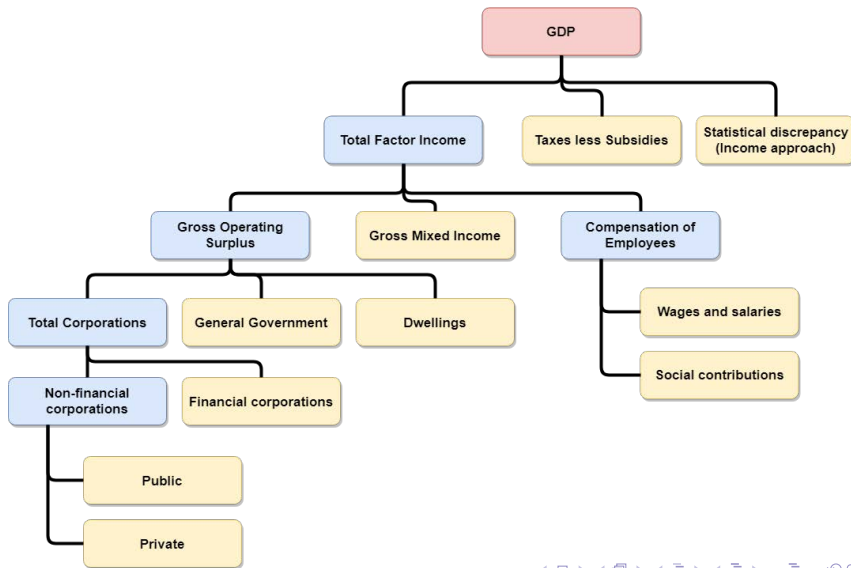
$$\begin{aligned} \text{GDP} = & \text{Gross operating surplus} + \text{Gross mixed income} \\ & + \text{Compensation of employees} \\ & + \text{Taxes less subsidies on production and imports} \\ & + \text{Statistical discrepancy (I)}. \end{aligned}$$

## Income approach

$$\begin{aligned} \text{GDP} = & \text{Gross operating surplus} + \text{Gross mixed income} \\ & + \text{Compensation of employees} \\ & + \text{Taxes less subsidies on production and imports} \\ & + \text{Statistical discrepancy (I)}. \end{aligned}$$

- The hierarchy has two levels of aggregation below the top-level, with a total of  $n = 16$  series and  $m = 10$  bottom level series.

# Australian GDP : *Data structures - Income approach*





# Australian GDP : *Data structures*

## Expenditure approach

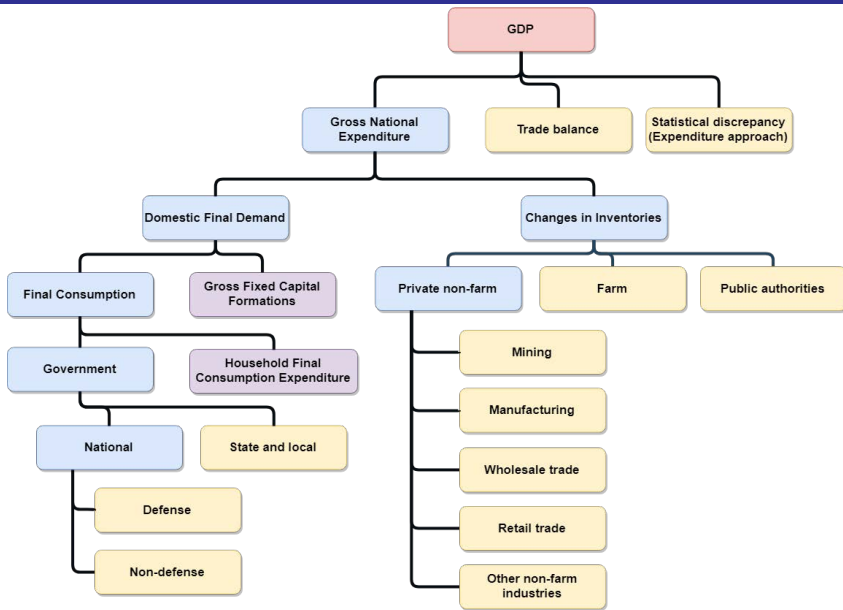
$$\begin{aligned} \text{GDP} = & \text{Final consumption expenditure} + \text{Gross fixed capital formation} \\ & + \text{Changes in inventories} + \text{Trade balance} \\ & + \text{Statistical discrepancy (E)}. \end{aligned}$$

## Expenditure approach

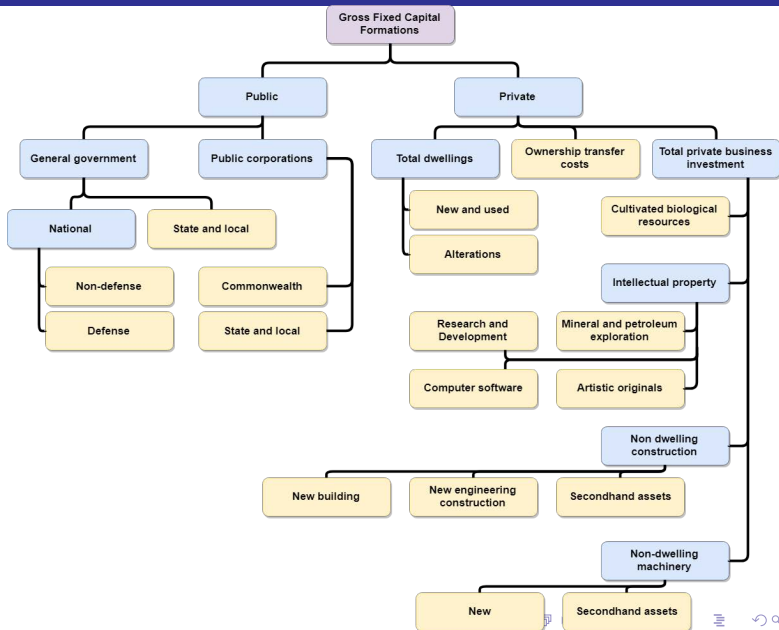
$$\begin{aligned} \text{GDP} = & \text{Final consumption expenditure} + \text{Gross fixed capital formation} \\ & + \text{Changes in inventories} + \text{Trade balance} \\ & + \text{Statistical discrepancy (E)}. \end{aligned}$$

- The hierarchy has three levels of aggregation below the top-level, with a total of  $n = 80$  series and  $m = 53$  series at the bottom level.

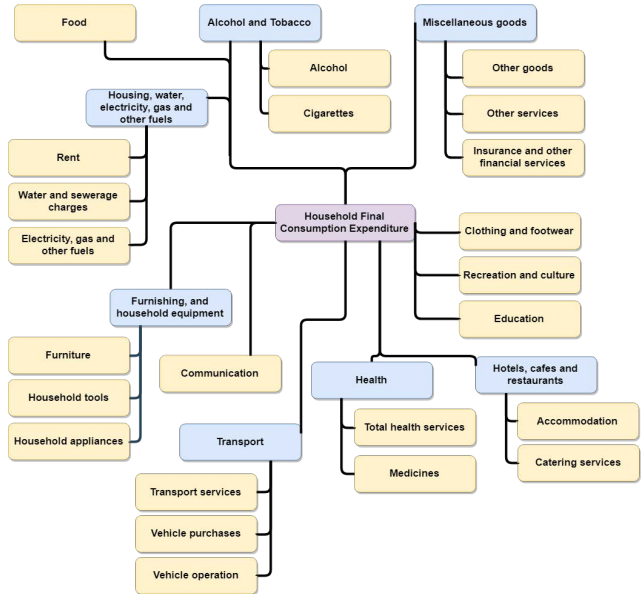
# Australian GDP : *Data structures - Expenditure approach*



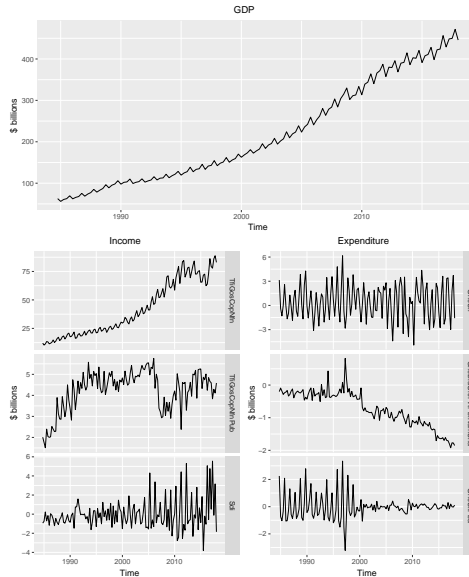
# Australian GDP : *Data structures - Expenditure approach*



## Australian GDP : *Data structures - Expenditure approach*



# Australian GDP : *Time plots for different levels*



## ■ Analysis set up:



## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.
- This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.
- This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.
- This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.

## ■ Base forecasting models:

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.
- This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.

## ■ Base forecasting models:

- Univariate ARIMA and ETS models were fitted for each training set.

## ■ Analysis set up:

- First training sample is set from 1984:Q4 to 1994:Q3 and forecasts produces for four quarters ahead (1994:Q4 to 1995:Q3).
- Then the training window is expanded by one quarter at a time.
- This leads to 94 1-step-ahead, 93 2-steps-ahead, 92 3-steps-ahead and 91 4-steps-ahead forecasts available for evaluation.

## ■ Base forecasting models:

- Univariate ARIMA and ETS models were fitted for each training set.
- $h = 1, 2, 3, 4$  steps ahead forecasts were generated using the fitted models.

# Point forecasts: *Base* vs *Seasonal Naïve*





# Point forecasts: *Reconciliation*

Reconciled forecasts are given by,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}$$

# Point forecasts: *Reconciliation*

Reconciled forecasts are given by,

$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}$$

---

$$\begin{aligned}\hat{\mathbf{W}}_{T+1}^{shr} &= \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam} \\ \hat{\mathbf{W}}_{T+1}^{wls} &= \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})\end{aligned}$$

---

# Point forecasts: *Reconciliation*

Reconciled forecasts are given by,

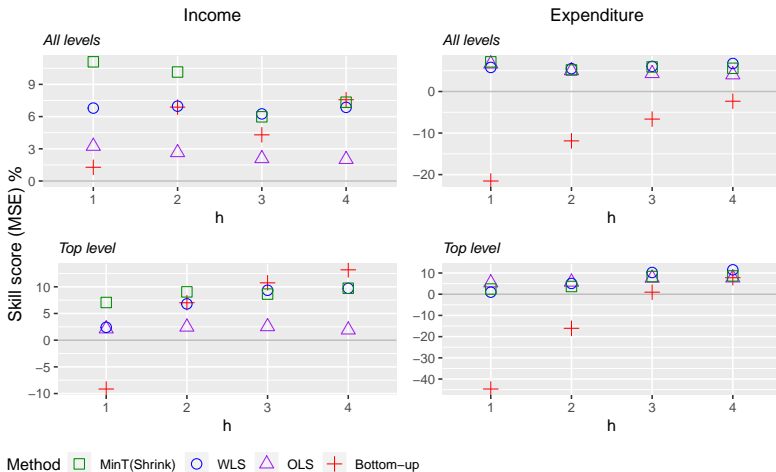
$$\tilde{\mathbf{y}}_{T+h} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h}$$

Method	$\mathbf{G}$
BU	$(\mathbf{0}_{m \times n-m} \quad \mathbf{I}_{m \times m})$
OLS	$(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$
WLS	$(\mathbf{S}'\hat{\mathbf{W}}_{T+1}^{wls}\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{T+1}^{wls}$
MinT(Shrink)	$(\mathbf{S}'\hat{\mathbf{W}}_{T+1}^{shr}\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{T+1}^{shr}$

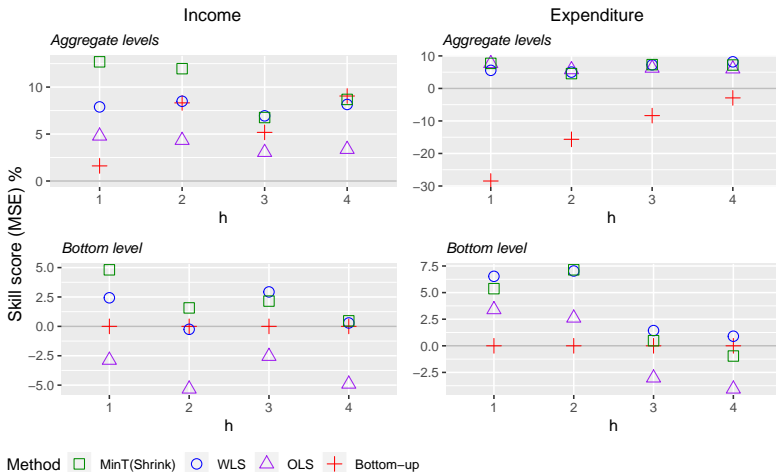
$$\hat{\mathbf{W}}_{T+1}^{shr} = \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam}$$

$$\hat{\mathbf{W}}_{T+1}^{wls} = \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})$$

# Reconciled Point forecasts - Results



# Reconciled Point forecasts - Results



- Gaussian approach :

$$\mathcal{N}(\mathbf{SG}\hat{\mathbf{y}}_{T+h}, \mathbf{SGW}_{T+h}\mathbf{G}'\mathbf{S}')$$

- Gaussian approach :

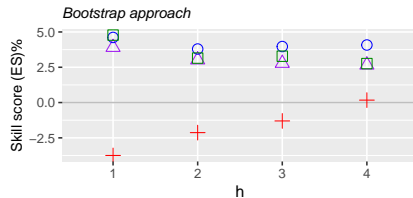
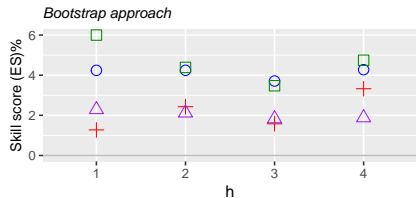
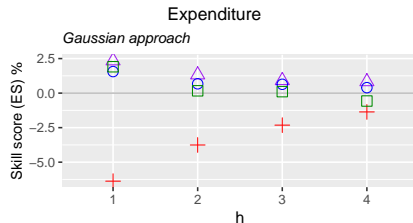
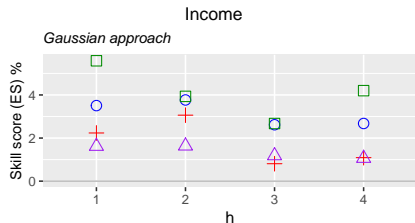
$$\mathcal{N}(\mathbf{SG}\hat{\mathbf{y}}_{T+h}, \mathbf{SGW}_{T+h}\mathbf{G}'\mathbf{S}')$$

- Non-parametric Bootstrap approach :

$$\tilde{\mathbf{Y}}_{T+h} = \mathbf{SG}\hat{\mathbf{Y}}'_{T+h}$$

where,  $\hat{\mathbf{Y}}_{T+h} = (\hat{\mathbf{y}}_{T+h}^1, \dots, \hat{\mathbf{y}}_{T+h}^B)'$

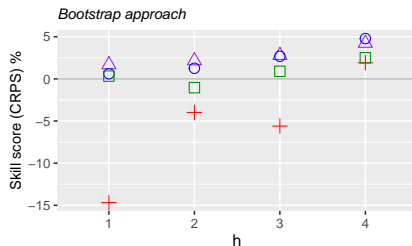
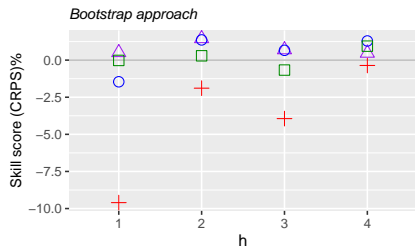
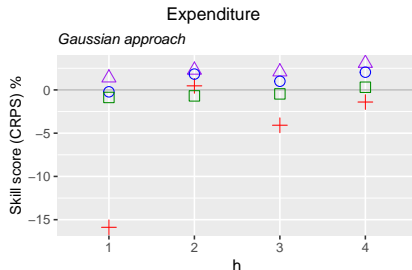
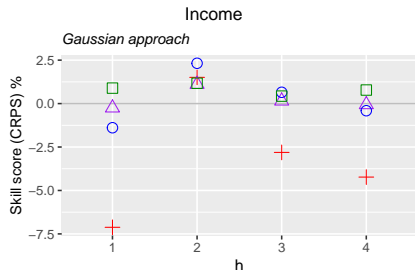
# Reconciled Probabilistic Forecasts



Method   MinT(Shrink)   WLS   OLS   OLS   Bottom-up



# Reconciled Probabilistic Forecasts



Method MinT(Shrink) WLS OLS Bottom-up

# Summary and time plan for completion

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts.

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts.

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts.
- Simulation study provides evidence that the optimal reconciliation with respect to energy score is equivalent to reconciling each sample path via MinT approach.

# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts.
- Simulation study provides evidence that the optimal reconciliation with respect to energy score is equivalent to reconciling each sample path via MinT approach.



# Summary

- We define point and probabilistic forecast reconciliation in geometric terms.
- We propose a parametric approach for probabilistic forecast reconciliation.
- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts.
- Simulation study provides evidence that the optimal reconciliation with respect to energy score is equivalent to reconciling each sample path via MinT approach.
- We apply hierarchical forecast reconciliation methods to forecast Australian GDP in point as well as probabilistic framework.

# Time plan for completion

	Thesis Chapter	Task description	Time duration	Progress
1 and 2.	Introduction and Background Review	Writing the chapter.	September/2019 - October/2019	40% complete
3.	Hierarchical forecast reconciliation in Geometric view	Bias correction and application	May/2019 - July/2019	75% Completed
4.	Probabilistic forecast reconciliation for hierarchical time series	Completing the paper.	June/2019 - August/2019	90% Completed
5.	Application	Forecasting Australian GDP		100% Completed

- Ben Taieb, S., R. Huser, R. J. Hyndman, and M. G. Genton (2017). Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid* **7**(5), 2448–2455.
- Capistrán, C., C. Constandse, and M. Ramos-Francia (2010). Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling* **27**(3), 666–677.
- Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**(477), 359–378.
- Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test* **17**(2), 211–235.

# References II

- Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis* **55**(9), 2579–2589.
- Jeon, J., A. Panagiotelis, and F. Petropoulos (2018). “Reconciliation of probabilistic forecasts with an application to wind power”.
- Scheuerer, M. and T. M. Hamill (2015). Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities \*. *Monthly Weather Review* **143**(4), 1321–1334.
- Weiss, C. (2018). “Essays in Hierarchical Time Series Forecasting and Forecast Combination”. PhD thesis. University of Cambridge.
- Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* **145**9, 1–45.

# Thank You!!

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

A1

- Assuming Gaussianity, let an incoherent forecast distribution at time  $T + h$  be given by

$$\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h}) \stackrel{d}{\leftrightarrow} \hat{\mathbf{f}}(.)$$

where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and

$\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is incoherent variance

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

A1

- Assuming Gaussianity, let an incoherent forecast distribution at time  $T + h$  be given by

$$\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h}) \stackrel{d}{\leftrightarrow} \hat{\mathbf{f}}(.)$$

where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and

$\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is incoherent variance

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

A1

- Assuming Gaussianity, let an incoherent forecast distribution at time  $T + h$  be given by

$$\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h}) \stackrel{d}{\leftrightarrow} \hat{\mathbf{f}}(.)$$

where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and

$\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is incoherent variance

- Gaussian densities are closed under affine transformation and marginalization



# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

A1

- Assuming Gaussianity, let an incoherent forecast distribution at time  $T + h$  be given by

$$\mathcal{N}(\hat{\mathbf{y}}_{T+h}, \mathbf{W}_{T+h}) \stackrel{d}{\leftrightarrow} \hat{\mathbf{f}}(\cdot)$$

where  $\hat{\mathbf{y}}_{T+h}$  is the incoherent mean and

$\mathbf{W}_{T+h} = E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]$  is incoherent variance

- Gaussian densities are closed under affine transformation and marginalization
- From the linear transformation

$$\mathbf{f}_B(\cdot) = \frac{\exp \left\{ -\frac{1}{2} \left( (\mathbf{S} : \mathbf{R}) \begin{pmatrix} \tilde{\mathbf{b}}_{T+h} \\ \tilde{\mathbf{a}}_{T+h} \end{pmatrix} - \hat{\mathbf{y}}_{T+h} \right)' \mathbf{W}_{T+h}^{-1} \left( (\mathbf{S} : \mathbf{R}) \begin{pmatrix} \tilde{\mathbf{b}}_{T+h} \\ \tilde{\mathbf{a}}_{T+h} \end{pmatrix} - \hat{\mathbf{y}}_{T+h} \right) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{W}_{T+h} \right|^{\frac{1}{2}} \left| (\mathbf{S} : \mathbf{R})^{-1} \right|}$$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

## ■ Marginalizing over $\mathbf{t}_{T+h}$ A1

$$\tilde{f}_b(\tilde{\mathbf{b}}_{T+h}) = \frac{\exp \left\{ -\frac{1}{2}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h})'(\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')^{-1}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h}) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}' \right|^{\frac{1}{2}}}$$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Marginalizing over  $\mathbf{t}_{T+h}$  A1

$$\tilde{\mathbf{f}}_b(\tilde{\mathbf{b}}_{T+h}) = \frac{\exp \left\{ -\frac{1}{2}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h})'(\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')^{-1}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h}) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}' \right|^{\frac{1}{2}}}$$

- This implies,  $\tilde{\mathbf{b}}_{T+h} \sim \mathcal{N}(\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')$  where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Marginalizing over  $\mathbf{t}_{T+h}$  A1

$$\tilde{\mathbf{f}}_b(\tilde{\mathbf{b}}_{T+h}) = \frac{\exp \left\{ -\frac{1}{2}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h})'(\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')^{-1}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h}) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}' \right|^{\frac{1}{2}}}$$

- This implies,  $\tilde{\mathbf{b}}_{T+h} \sim \mathcal{N}(\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')$  where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$
- Therefore the reconciled Gaussian distribution of the whole hierarchy is given by

$$\tilde{\mathbf{f}}(\tilde{\mathbf{y}}_{t+h}) = \tilde{\mathbf{f}}_b(\mathbf{S}\tilde{\mathbf{b}}_{T+h}) \stackrel{d}{\leftrightarrow} \mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}'),$$

where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Marginalizing over  $\mathbf{t}_{T+h}$  A1

$$\tilde{\mathbf{f}}_b(\tilde{\mathbf{b}}_{T+h}) = \frac{\exp \left\{ -\frac{1}{2}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h})'(\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')^{-1}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h}) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}' \right|^{\frac{1}{2}}}$$

- This implies,  $\tilde{\mathbf{b}}_{T+h} \sim \mathcal{N}(\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')$  where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$
- Therefore the reconciled Gaussian distribution of the whole hierarchy is given by

$$\tilde{\mathbf{f}}(\tilde{\mathbf{y}}_{t+h}) = \tilde{\mathbf{f}}_b(\mathbf{S}\tilde{\mathbf{b}}_{T+h}) \stackrel{d}{\leftrightarrow} \mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}'),$$

where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Marginalizing over  $\mathbf{t}_{T+h}$  A1

$$\tilde{\mathbf{f}}_b(\tilde{\mathbf{b}}_{T+h}) = \frac{\exp \left\{ -\frac{1}{2}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h})'(\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')^{-1}(\tilde{\mathbf{b}}_{T+h} - \mathbf{G}\hat{\mathbf{y}}_{T+h}) \right\}}{(2\pi)^{\frac{n}{2}} \left| \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}' \right|^{\frac{1}{2}}}$$

- This implies,  $\tilde{\mathbf{b}}_{T+h} \sim \mathcal{N}(\mathbf{G}\hat{\mathbf{y}}_{T+h}, \mathbf{G}\mathbf{W}_{T+h}\mathbf{G}')$  where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$
- Therefore the reconciled Gaussian distribution of the whole hierarchy is given by

$$\tilde{\mathbf{f}}(\tilde{\mathbf{y}}_{t+h}) = \tilde{\mathbf{f}}_b(\mathbf{S}\tilde{\mathbf{b}}_{T+h}) \stackrel{d}{\leftrightarrow} \mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{G}'\mathbf{S}'),$$

where  $\mathbf{G} = (\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}$

- **Goal:** Solve for  $\mathbf{G}$  by minimizing a proper scoring rule

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

where  $\tilde{\mathbf{Y}}_{T+h}$  and  $\tilde{\mathbf{Y}}_{T+h}^*$  are independent random variables from  $\mathcal{N}(\mathbf{SP}\hat{\mathbf{y}}_{T+h}, \mathbf{SPW}_{T+h}\mathbf{P}'\mathbf{S}')$ .



# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

where  $\tilde{\mathbf{Y}}_{T+h}$  and  $\tilde{\mathbf{Y}}_{T+h}^*$  are independent random variables from  $\mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{P}'\mathbf{S}')$ .

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

where  $\tilde{\mathbf{Y}}_{T+h}$  and  $\tilde{\mathbf{Y}}_{T+h}^*$  are independent random variables from  $\mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{P}'\mathbf{S}')$ .

- There is no closed form expression for  $ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h})$  for  $\alpha \in (0, 2)$  under the Gaussian predictive distribution

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

where  $\tilde{\mathbf{Y}}_{T+h}$  and  $\tilde{\mathbf{Y}}_{T+h}^*$  are independent random variables from  $\mathcal{N}(\mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}, \mathbf{S}\mathbf{P}\mathbf{W}_{T+h}\mathbf{P}'\mathbf{S}')$ .

- There is no closed form expression for  $ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h})$  for  $\alpha \in (0, 2)$  under the Gaussian predictive distribution

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- We use the **Energy Score**, which is a proper scoring rule

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} E_{\mathbf{Y}_{T+h}} \|\tilde{\mathbf{Y}}_{T+h} - \tilde{\mathbf{Y}}_{T+h}^*\|^\alpha,$$

for  $\alpha \in (0, 2]$ ,

where  $\tilde{\mathbf{Y}}_{T+h}$  and  $\tilde{\mathbf{Y}}_{T+h}^*$  are independent random variables from  $\mathcal{N}(\mathbf{SP}\hat{\mathbf{y}}_{T+h}, \mathbf{SPW}_{T+h}\mathbf{P}'\mathbf{S}')$ .

- There is no closed form expression for  $ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h})$  for  $\alpha \in (0, 2)$  under the Gaussian predictive distribution
- However in the upper limit of  $\alpha$ , i.e. when  $\alpha = 2$ ,

$$ES(\tilde{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = \|\mathbf{SP}\hat{\mathbf{y}}_{T+h} - \mathbf{y}_{T+h}\|^2$$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\arg \min_{\mathbf{G}} \text{Tr}\{E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]\},$$

where  $\mathcal{I}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\arg \min_{\mathbf{G}} \text{Tr}\{E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]\},$$

where  $\mathcal{I}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$

- Assuming unbiasedness for coherent forecasts

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\arg \min_{\mathbf{G}} \text{Tr}\{E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \mathbf{SP}\hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]\},$$

where  $\mathcal{I}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$

- Assuming unbiasedness for coherent forecasts

$$\arg \min_{\mathbf{G}} \text{Tr}\{\mathbf{SGW}_{T+h}\mathbf{G}^T\mathbf{S}^T\}$$

Subject to  $\mathbf{SPS} = \mathbf{S}$

# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\arg \min_{\mathbf{G}} \text{Tr}\{E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]\},$$

where  $\mathcal{I}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$

- Assuming unbiasedness for coherent forecasts

$$\arg \min_{\mathbf{G}} \text{Tr}\{\mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}^T\mathbf{S}^T\}$$

Subject to  $\mathbf{S}\mathbf{P}\mathbf{S} = \mathbf{S}$

- Wickramasuriya, Athanasopoulos, and Hyndman (2018) have shown that a unique solution to this optimization problem attain at

$$\mathbf{G} = (\mathbf{S}^T \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \mathbf{W}_{T+h}^{-1}.$$



# Appendix: Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\arg \min_{\mathbf{G}} \text{Tr}\{E_{\mathbf{y}_{T+h}}[(\mathbf{y}_{T+h} - \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h})^T | \mathcal{I}_T]\},$$

where  $\mathcal{I}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$

- Assuming unbiasedness for coherent forecasts

$$\arg \min_{\mathbf{G}} \text{Tr}\{\mathbf{S}\mathbf{G}\mathbf{W}_{T+h}\mathbf{G}^T\mathbf{S}^T\}$$

Subject to  $\mathbf{S}\mathbf{P}\mathbf{S} = \mathbf{S}$

- Wickramasuriya, Athanasopoulos, and Hyndman (2018) have shown that a unique solution to this optimization problem attain at

$$\mathbf{G} = (\mathbf{S}^T \mathbf{W}_{T+h}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \mathbf{W}_{T+h}^{-1}.$$

- Thus  $\mathbf{R}'_l = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$ . This is referred to as **MinT solution**

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \mathbf{W}_{T+h} \mathbf{R}_{\perp} (\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{S}']$ ,

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \mathbf{W}_{T+h} \mathbf{R}_{\perp} (\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \mathbf{W}_{T+h} \mathbf{R}_{\perp} (\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}' \mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{R}'_{\perp} \mathbf{W}_{T+h} \mathbf{R}_{\perp} (\mathbf{R}'_{\perp} \mathbf{S})^{-1} \mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}'\mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\mathbf{W}_{T+h}\mathbf{R}_{\perp}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

Method	Estimate of $\mathbf{W}_h$	Estimate of $\mathbf{R}'_{\perp}$
OLS	$\mathbf{I}$	$\mathbf{S}'$
MinT(Sample)	$\hat{\mathbf{W}}_{T+1}^{sam}$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{sam})^{-1}$
MinT(Shrink)	$\hat{\mathbf{W}}_{T+1}^{shr} = \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam}$ <span style="background-color: #d1c4e9; border-radius: 10px; padding: 2px;">A2</span>	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{shr})^{-1}$
MinT(WLS)	$\hat{\mathbf{W}}_{T+1}^{wls} = \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{wls})^{-1}$

# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}'\mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\mathbf{W}_{T+h}\mathbf{R}_{\perp}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

Method	Estimate of $\mathbf{W}_h$	Estimate of $\mathbf{R}'_{\perp}$
OLS	$\mathbf{I}$	$\mathbf{S}'$
MinT(Sample)	$\hat{\mathbf{W}}_{T+1}^{sam}$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{sam})^{-1}$
MinT(Shrink)	$\hat{\mathbf{W}}_{T+1}^{shr} = \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam}$ <span style="background-color: #d1c4e9; border-radius: 10px; padding: 2px;">A2</span>	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{shr})^{-1}$
MinT(WLS)	$\hat{\mathbf{W}}_{T+1}^{wls} = \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{wls})^{-1}$



# Appendix: Probabilistic forecasts in the Gaussian framework

$$\mathbf{R}'_{\perp} = \mathbf{S}'\mathbf{W}_{T+h}^{-1}$$

- The optimally reconciled Gaussian forecast density is given by  $\mathcal{N}[\mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\hat{\mathbf{y}}_{T+h}, \mathbf{S}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{R}'_{\perp}\mathbf{W}_{T+h}\mathbf{R}_{\perp}(\mathbf{R}'_{\perp}\mathbf{S})^{-1}\mathbf{S}']$ ,
- Different estimators of  $\mathbf{W}_{T+h}$  yield different  $\mathbf{R}'_{\perp}$  matrices

Method	Estimate of $\mathbf{W}_h$	Estimate of $\mathbf{R}'_{\perp}$
OLS	$\mathbf{I}$	$\mathbf{S}'$
MinT(Sample)	$\hat{\mathbf{W}}_{T+1}^{sam}$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{sam})^{-1}$
MinT(Shrink)	$\hat{\mathbf{W}}_{T+1}^{shr} = \tau \text{Diag}(\hat{\mathbf{W}}_{T+1}^{sam}) + (1 - \tau)\hat{\mathbf{W}}_{T+1}^{sam}$ <span style="background-color: #d1c4e9; border-radius: 10px; padding: 2px;">A2</span>	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{shr})^{-1}$
MinT(WLS)	$\hat{\mathbf{W}}_{T+1}^{wls} = \text{Diag}(\hat{\mathbf{W}}_{T+1}^{shr})$	$\mathbf{S}'(\hat{\mathbf{W}}_{T+1}^{wls})^{-1}$

- Predictive ability of the reconciled Gaussian densities from these methods will be evaluated in a simulation setting

# Appendix: Probabilistic forecasts evaluation

A2

**Energy score** (Gneiting et al., 2008)

$$\text{eS}(\check{\mathbf{Y}}_{T+h}, \mathbf{y}_{T+h}) = \mathbb{E}_{\check{\mathbf{F}}} \|\check{\mathbf{Y}}_{T+h} - \mathbf{y}_{T+h}\|^\alpha - \frac{1}{2} \mathbb{E}_{\check{\mathbf{F}}} \|\check{\mathbf{Y}}_{T+h} - \check{\mathbf{Y}}_{T+h}^*\|^\alpha, \quad \alpha \in (0, 2]$$

**Log score** (Gneiting and Raftery, 2007)

$$\text{LS}(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = -\log \check{\mathbf{f}}(\mathbf{y}_{T+h})$$

**Variogram score** (Scheuerer and Hamill, 2015)

$$\text{VS}(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left( |y_{T+h,i} - y_{T+h,j}|^p - \mathbb{E}_{\check{\mathbf{F}}} |\check{Y}_{T+h,i} - \check{Y}_{T+h,j}|^p \right)^2$$

**CRPS** (Gneiting and Raftery, 2007)

$$\text{CRPS}(\check{F}_i, y_{T+h,i}) = \mathbb{E}_{\check{F}_i} |\check{Y}_{T+h,i} - y_{T+h,i}| - \frac{1}{2} \mathbb{E}_{\check{F}_i} |\check{Y}_{T+h,i} - \check{Y}_{T+h,i}^*|$$

$\check{\mathbf{Y}}_{T+h}$  and  $\check{\mathbf{Y}}_{T+h}^*$  : Independent random vectors from the coherent forecast distribution  $\check{\mathbf{F}}$ .

$\mathbf{y}_{T+h}$  : Vector of realizations.

$\check{Y}_{T+h,i}$  and  $\check{Y}_{T+h,j}$  :  $i$ th and  $j$ th components of the vector  $\check{\mathbf{Y}}_{T+h}$

- ◀ A2 Shrinkage estimator for 1-step ahead base forecast errors

$$\hat{\Sigma}_{T+1}^{shr} = \tau \hat{\Sigma}_{T+1}^D + (1 - \tau) \hat{\Sigma}_{T+1},$$

where  $\hat{\Sigma}_{T+1}^D$  is the diagonal matrix comprising diagonal entries of  $\hat{\Sigma}_{T+1}$  and

$$\tau = \frac{\sum_{i \neq j} \text{Var}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}^2}$$

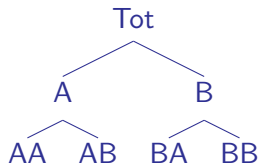
is a shrinkage parameter.  $\hat{r}_{ij}$  is the  $ij$ -th element of sample correlation matrix. In this estimation, the off-diagonal elements of 1-step ahead sample covariance matrix will be shrunk to zero depending on the sparsity.

# Monte-Carlo simulation

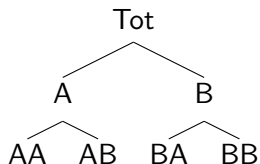
## ■ Data generating process ◀ A3

## ■ Data generating process ◀ A3

## ■ Data generating process ◀ A3

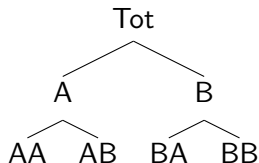


## ■ Data generating process ◀ A3



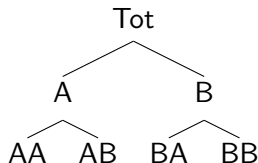
## ■ Data generating process ◀ A3

$$\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim \text{ARIMA}(p, d, q)$$



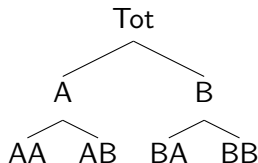


## ■ Data generating process ◀ A3



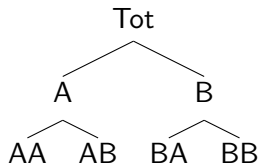
- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$

## ■ Data generating process ◀ A3



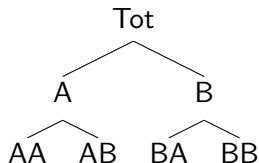
- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim \text{ARIMA}(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \Sigma)$

## ■ Data generating process ◀ A3



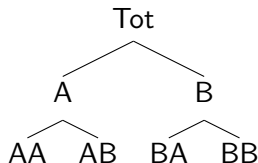
- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \Sigma)$
- Parameters for *AR* and *MA* components were randomly and uniformly generated from  $[0.3, 0.5]$  and  $[0.3, 0.7]$  respectively

## ■ Data generating process ◀ A3



- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \Sigma)$
- Parameters for *AR* and *MA* components were randomly and uniformly generated from  $[0.3, 0.5]$  and  $[0.3, 0.7]$  respectively

## ■ Data generating process ◀ A3

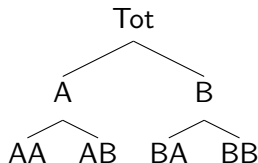


- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \Sigma)$
- Parameters for *AR* and *MA* components were randomly and uniformly generated from  $[0.3, 0.5]$  and  $[0.3, 0.7]$  respectively

■  $\mathbf{y}_t$  are then generated as follows

# Monte-Carlo simulation

## ■ Data generating process ◀ A3



- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$  and  $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \Sigma)$
- Parameters for *AR* and *MA* components were randomly and uniformly generated from  $[0.3, 0.5]$  and  $[0.3, 0.7]$  respectively

## ■ $y_t$ are then generated as follows

Bottom level	Aggregate level 1	Total
$y_{AA,t} = w_{AA,t} + u_t - 0.5v_t$	$y_{A,t} = w_{AA,t} + w_{AB,t} - v_t$	$y_{Tot,t} = w_{AA,t} + w_{AB,t} + w_{BA,t} + w_{BB,t}$
$y_{AB,t} = w_{AB,t} - u_t - 0.5v_t$	$y_{B,t} = w_{BA,t} + w_{BB,t} + v_t$	
$y_{BA,t} = w_{BA,t} + u_t + 0.5v_t$		
$y_{BB,t} = w_{BB,t} - u_t + 0.5v_t$		

- To get less noisier series at aggregate levels, we choose  $\Sigma, \sigma_u^2$  and  $\sigma_v^2$  such that,

$$\text{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} + \epsilon_{BA,t} + \epsilon_{BB,t}) \leq \text{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} - v_t) \leq \text{Var}(\epsilon_{AA,t} + u_t - 0.5v_t),$$

- To get less noisier series at aggregate levels, we choose  $\Sigma$ ,  $\sigma_u^2$  and  $\sigma_v^2$  such that,

$$\text{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} + \epsilon_{BA,t} + \epsilon_{BB,t}) \leq \text{Var}(\epsilon_{AA,t} + \epsilon_{AB,t} - v_t) \leq \text{Var}(\epsilon_{AA,t} + u_t - 0.5v_t),$$

- Thus we choose,  $\Sigma = \begin{pmatrix} 5.0 & 3.1 & 0.6 & 0.4 \\ 3.1 & 4.0 & 0.9 & 1.4 \\ 0.6 & 0.9 & 2.0 & 1.8 \\ 0.4 & 1.4 & 1.8 & 3.0 \end{pmatrix}$ ,  $\sigma_u^2 = 19$  and  $\sigma_v^2 = 18$ .



# Sample version of the scoring rules

- For a possible finite sample of size  $B$  from the multivariate forecast density  $\check{\mathbf{F}}$ , the variogram score is defined as,

$$VS(\check{\mathbf{F}}, \mathbf{y}_{T+h}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left( |y_{T+h,i} - y_{T+h,j}|^p - \frac{1}{B} \sum_{k=1}^B |\check{Y}_{T+h,i}^k - \check{Y}_{T+h,j}^k|^p \right)^2$$