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

March 21, 2018

Overview

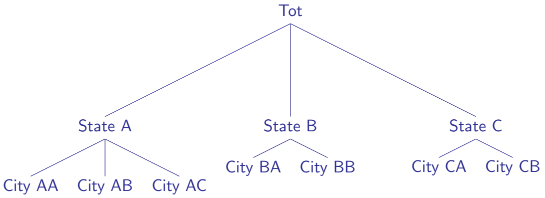
Introduction

Introduction

- A collection of multiple time series that has an inherent aggregation structure is called hierarchical time series

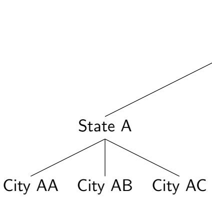
Introduction

- A collection of multiple time series that has an inherent aggregation structure is called hierarchical time series
- **Example:** Electricity demand of a country can be disaggregated into demand of states, cities and households



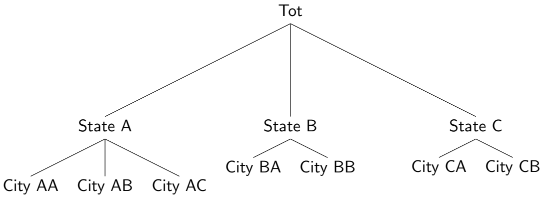
Introduction

- A collection of multiple time series that has an inherent aggregation structure is called hierarchical time series
- **Example:** Electricity demand of a country can be disaggregated into demand of states, cities and households



Introduction

- A collection of multiple time series that has an inherent aggregation structure is called hierarchical time series
- **Example:** Electricity demand of a country can be disaggregated into demand of states, cities and households



- Forecasts of these hierarchical time series should be *coherent*

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available

- Point forecast reconciliation (Hyndman et al., 2011)

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available
- Point forecast reconciliation (Hyndman et al., 2011)
  - OLS

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available
- Point forecast reconciliation (Hyndman et al., 2011)
  - OLS
  - MinT (Wickramasuriya, Athanasopoulos, and Hyndman, 2018)

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available
- Point forecast reconciliation (Hyndman et al., 2011)
  - OLS
  - MinT (Wickramasuriya, Athanasopoulos, and Hyndman, 2018)

Introduction *Cont..*

- Traditional approaches for coherent point forecasts:
  - Bottom-up
  - Top-down
  - Middle-out
- Use only part of the information available
- Point forecast reconciliation (Hyndman et al., 2011)
  - OLS
  - MinT (Wickramasuriya, Athanasopoulos, and Hyndman, 2018)
- Traditional methods are not reconciliation methods

Motivation of the study

- Importance of probabilistic forecasts in time series

Motivation of the study

- Importance of probabilistic forecasts in time series

Motivation of the study

- Importance of probabilistic forecasts in time series

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

Motivation of the study

- Importance of probabilistic forecasts in time series
- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure

Motivation of the study

- Importance of probabilistic forecasts in time series

- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure

Motivation of the study

- Importance of probabilistic forecasts in time series
- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature

Motivation of the study

- Importance of probabilistic forecasts in time series
- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature
  - Only relevant paper "Hierarchical probabilistic forecasting of electricity demand with smart meter data" (Ben Taieb et al., 2017)

Motivation of the study

- Importance of probabilistic forecasts in time series

- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature
  - Only relevant paper "Hierarchical probabilistic forecasting of electricity demand with smart meter data" (Ben Taieb et al., 2017)

Motivation of the study

- Importance of probabilistic forecasts in time series
- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature
  - Only relevant paper "Hierarchical probabilistic forecasting of electricity demand with smart meter data" (Ben Taieb et al., 2017)
- Extending the "reconciliation" method into probabilistic framework

Motivation of the study

- Importance of probabilistic forecasts in time series

- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature
  - Only relevant paper "Hierarchical probabilistic forecasting of electricity demand with smart meter data" (Ben Taieb et al., 2017)

- Extending the "reconciliation" method into probabilistic framework

Motivation of the study

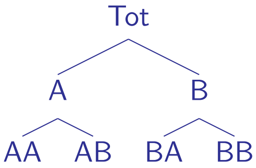
- Importance of probabilistic forecasts in time series
- Probabilistic forecasts of hierarchical time series should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure
- Lack of attention on probabilistic forecasting in hierarchical literature
  - Only relevant paper "Hierarchical probabilistic forecasting of electricity demand with smart meter data" (Ben Taieb et al., 2017)
- Extending the "reconciliation" method into probabilistic framework
- **Objective:** Reconciliation of probabilistic forecasts for hierarchical time series such that they preserve the inherent properties of the hierarchical nature

Probabilistic forecast reconciliation for hierarchical time series

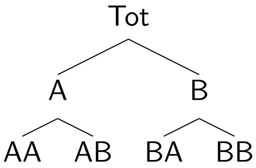
Notations in hierarchical time series and forecasting

Notations in hierarchical time series and forecasting

Notations in hierarchical time series and forecasting

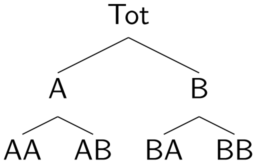


Notations in hierarchical time series and forecasting



Notations in hierarchical time series and forecasting

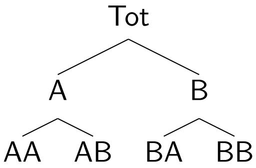
$$\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$



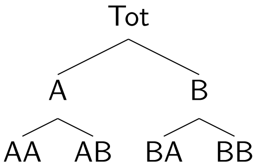
Notations in hierarchical time series and forecasting

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

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



Notations in hierarchical time series and forecasting

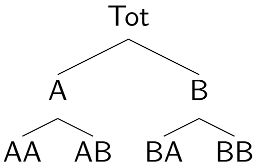


$$\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

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

$$m = 4$$

Notations in hierarchical time series and forecasting



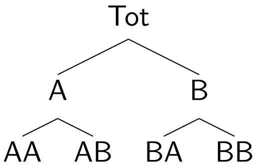
$\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$

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

$m = 4$

$n = 7$

Notations in hierarchical time series and forecasting



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

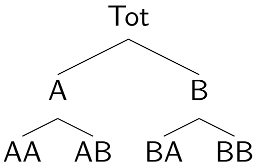
$$\boldsymbol{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$

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

Notations in hierarchical time series and forecasting



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

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

$$m = 4$$

$$n = 7$$

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

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

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

Coherent forecasts

Coherent forecasts

Definition 1: Coherent subspace

Suppose an n-dimensional time series $\boldsymbol{y}_t \in \mathbb{R}^n$ is subject to the linear aggregation constraint $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{b}_t \in \mathbb{R}^m$ and $\boldsymbol{S}$ is an $n \times m$ constant matrix. Let $\mathbb{C}^m$ be an m-dimensional subspace of $\mathbb{R}^n$, where $\mathbb{C}^m \subset \mathbb{R}^n$. Then $\mathbb{C}^m$ is said to be a coherent space if it is spanned by the columns of $\boldsymbol{S}$.

Coherent forecasts

Definition 1: Coherent subspace

Suppose an n-dimensional time series $\boldsymbol{y}_t \in \mathbb{R}^n$ is subject to the linear aggregation constraint $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{b}_t \in \mathbb{R}^m$ and $\boldsymbol{S}$ is an $n \times m$ constant matrix. Let $\mathbb{C}^m$ be an m-dimensional subspace of $\mathbb{R}^n$, where $\mathbb{C}^m \subset \mathbb{R}^n$. Then $\mathbb{C}^m$ is said to be a coherent space if it is spanned by the columns of $\boldsymbol{S}$.

- The null space of $\mathbb{C}^m$ is denoted by $\mathbb{N}^{n-m}$

Coherent forecasts

Definition 1: Coherent subspace

Suppose an n-dimensional time series $\boldsymbol{y}_t \in \mathbb{R}^n$ is subject to the linear aggregation constraint $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{b}_t \in \mathbb{R}^m$ and $\boldsymbol{S}$ is an $n \times m$ constant matrix. Let $\mathbb{C}^m$ be an m-dimensional subspace of $\mathbb{R}^n$, where $\mathbb{C}^m \subset \mathbb{R}^n$. Then $\mathbb{C}^m$ is said to be a coherent space if it is spanned by the columns of $\boldsymbol{S}$.

- The null space of $\mathbb{C}^m$ is denoted by $\mathbb{N}^{n-m}$

Coherent forecasts

Definition 1: Coherent subspace

Suppose an n-dimensional time series $\boldsymbol{y}_t \in \mathbb{R}^n$ is subject to the linear aggregation constraint $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$ where $\boldsymbol{b}_t \in \mathbb{R}^m$ and $\boldsymbol{S}$ is an $n \times m$ constant matrix. Let $\mathbb{C}^m$ be an m-dimensional subspace of $\mathbb{R}^n$, where $\mathbb{C}^m \subset \mathbb{R}^n$. Then $\mathbb{C}^m$ is said to be a coherent space if it is spanned by the columns of $\boldsymbol{S}$.

- The null space of $\mathbb{C}^m$ is denoted by $\mathbb{N}^{n-m}$

Definition 2: Coherent point forecasts

Suppose $\breve{\boldsymbol{y}}_{t+h|t} \in \mathbb{R}^n$ denotes point forecasts of each series in the hierarchy at time $t+h$. Then $\breve{\boldsymbol{y}}_{t+h|t}$ is *coherent* if $\breve{\boldsymbol{y}}_{t+h|t} \in \mathbb{C}^m$.

Coherent forecasts *cont.*

Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.

Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.

Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.

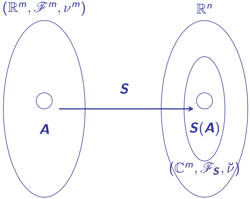
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$

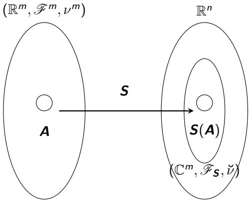
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$

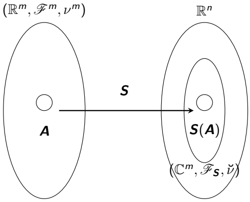
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

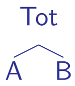
Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$



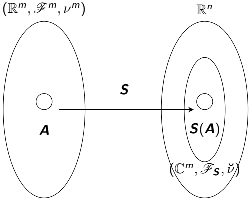
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

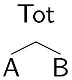
Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$



$$\breve{\boldsymbol{y}}_{t+h} = [\breve{y}_{Tot,t+h}, \breve{y}_{A,t+h}, \breve{y}_{B,t+h}]$$

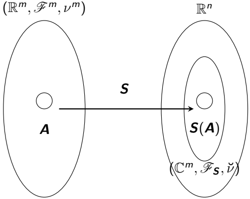
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

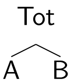
Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$



$$\breve{\boldsymbol{y}}_{t+h} = [\breve{y}_{Tot,t+h}, \breve{y}_{A,t+h}, \breve{y}_{B,t+h}]$$

$$\breve{y}_{Tot,t+h} = \breve{y}_{A,t+h} + \breve{y}_{B,t+h}$$

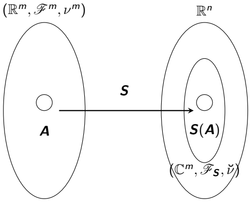
Coherent forecasts *cont.*

Definition 3: Coherent probabilistic forecasts

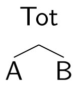
Let $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ be a probability triple where $\mathscr{F}^m$ is a σ-algebra on $\mathbb{R}^m$. Then, $(\mathbb{C}^m, \mathscr{F}_{\boldsymbol{S}}, \breve{\nu})$ is said to be a coherent probability triple iff

$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A}) \quad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S}(\boldsymbol{A})$ denotes the image of subset $\boldsymbol{A}$ under $\boldsymbol{S}$.



$$\breve{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \nu^m(\boldsymbol{A})$$



$$\breve{\boldsymbol{y}}_{t+h} = [\breve{y}_{Tot,t+h}, \breve{y}_{A,t+h}, \breve{y}_{B,t+h}]$$

$$\breve{y}_{Tot,t+h} = \breve{y}_{A,t+h} + \breve{y}_{B,t+h}$$

$$f(\breve{\boldsymbol{y}}_{t+h}) = 0 \quad \forall \breve{\boldsymbol{y}}_{t+h} \in \mathbb{N}^{n-m}$$

Point forecast reconciliation

Point forecast reconciliation

Definition 4: Point forecast reconciliation

Let $\hat{\boldsymbol{y}}_{t+h} \in \mathbb{R}^n$ be any set of incoherent forecasts at time $t+h$, and let

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S} \circ \boldsymbol{g}(\hat{\boldsymbol{y}}_{t+h}),$$

where $\boldsymbol{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\boldsymbol{S} \circ \boldsymbol{g}(\cdot)$ is a mapping of $\boldsymbol{g}(\cdot)$ onto $\mathbb{C}^m$. Then $\tilde{\boldsymbol{y}}_{t+h}$ is said to be "reconciled" if $\tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m$.

Point forecast reconciliation

Definition 4: Point forecast reconciliation

Let $\hat{\boldsymbol{y}}_{t+h} \in \mathbb{R}^n$ be any set of incoherent forecasts at time $t+h$, and let

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S} \circ \boldsymbol{g}(\hat{\boldsymbol{y}}_{t+h}),$$

where $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ and $\boldsymbol{S} \circ \boldsymbol{g}(\cdot)$ is a mapping of $\boldsymbol{g}(\cdot)$ onto $\mathbb{C}^m$. Then $\tilde{\boldsymbol{y}}_{t+h}$ is said to be "reconciled" if $\tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m$.

- Definition 4 holds for linear as well as non-linear reconciliation

Point forecast reconciliation

Definition 4: Point forecast reconciliation

Let $\hat{\boldsymbol{y}}_{t+h} \in \mathbb{R}^n$ be any set of incoherent forecasts at time $t+h$, and let

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S} \circ \boldsymbol{g}(\hat{\boldsymbol{y}}_{t+h}),$$

where $\boldsymbol{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\boldsymbol{S} \circ \boldsymbol{g}(\cdot)$ is a mapping of $\boldsymbol{g}(\cdot)$ onto $\mathbb{C}^m$. Then $\tilde{\boldsymbol{y}}_{t+h}$ is said to be "reconciled" if $\tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m$.

- Definition 4 holds for linear as well as non-linear reconciliation
- Previous studies have only focused on linear case,

$$\boldsymbol{g}(\hat{\boldsymbol{y}}) = \boldsymbol{P}\hat{\boldsymbol{y}},$$

where $\boldsymbol{P}$ is a $m \times n$ matrix. Thus giving

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{t+h}$$

Point forecast reconciliation *cont..*

- Suppose $\boldsymbol{S} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_m\}$ spans $\mathbb{C}^m$, $\boldsymbol{R} = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{N}^{n-m}$. Then $(\boldsymbol{S} \vdots \boldsymbol{R}) = \boldsymbol{B} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_m, \boldsymbol{r}_1, \dots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{R}^n$.

Point forecast reconciliation *cont..*

- Suppose $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ spans $\mathbb{C}^m$, $\boldsymbol{R} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{N}^{n-m}$. Then $(\boldsymbol{S} \vdots \boldsymbol{R}) = \boldsymbol{B} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{R}^n$.

Point forecast reconciliation *cont..*

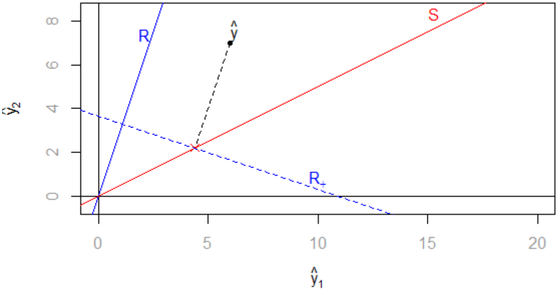
- Suppose $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ spans $\mathbb{C}^m$, $\boldsymbol{R} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{N}^{n-m}$. Then $(\boldsymbol{S} \vdots \boldsymbol{R}) = \boldsymbol{B} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{R}^n$.

- Schematic in $\mathbb{R}^2$

Point forecast reconciliation *cont..*

- Suppose $\boldsymbol{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ spans $\mathbb{C}^m$, $\boldsymbol{R} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{N}^{n-m}$. Then $(\boldsymbol{S} \vdots \boldsymbol{R}) = \boldsymbol{B} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-m}\}$ spans $\mathbb{R}^n$.

- Schematic in $\mathbb{R}^2$



Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$

Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

$(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$

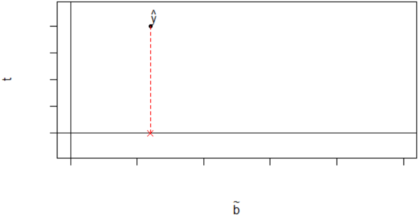
Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$

- **Step 1: Change of coordinates (linear transformation step)**

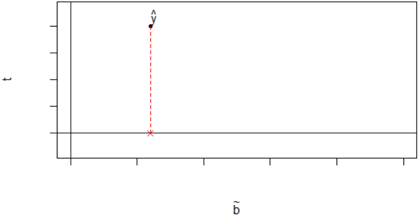
  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$



Point forecast reconciliation *cont..*

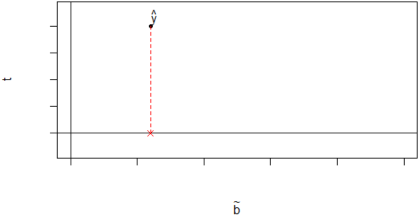
- **Step 1: Change of coordinates (linear transformation step)**

$(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$



- **Step 1: Change of coordinates (linear transformation step)**

  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$

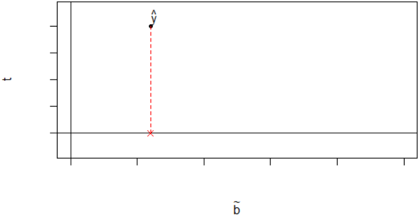


$$(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h},$$

Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

$(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$



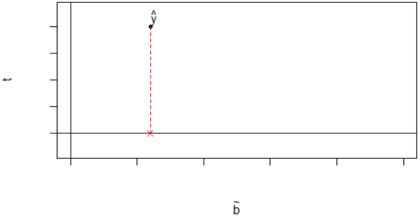
$$(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h},$$

$$(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = (\boldsymbol{S} \vdots \boldsymbol{R})^{-1}\hat{\boldsymbol{y}}_{t+h}.$$

Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$



$$(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h},$$

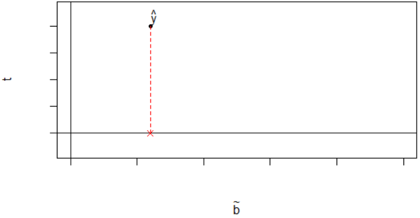
$$(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = (\boldsymbol{S} \vdots \boldsymbol{R})^{-1}\hat{\boldsymbol{y}}_{t+h}.$$

Since $(\boldsymbol{S} \vdots \boldsymbol{R})^{-1} = \begin{pmatrix} (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp} \\ \cdots \\ (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp} \end{pmatrix}$ we have,

Point forecast reconciliation *cont..*

- **Step 1: Change of coordinates (linear transformation step)**

  $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$



$$(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h},$$

$$(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = (\boldsymbol{S} \vdots \boldsymbol{R})^{-1}\hat{\boldsymbol{y}}_{t+h}.$$

Since $(\boldsymbol{S} \vdots \boldsymbol{R})^{-1} = \begin{pmatrix} (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp} \\ \cdots \\ (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp} \end{pmatrix}$ we have,

$$\begin{pmatrix} \tilde{\boldsymbol{b}}_{t+h} \\ \cdots \\ \boldsymbol{t}_{t+h} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp} \\ \cdots \\ (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp} \end{pmatrix} \hat{\boldsymbol{y}}_{t+h}.$$

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}. \tag{1}$$

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}. \tag{1}$$

- **Step 3: Reconciled point forecasts for the whole hierarchy**

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}. \tag{1}$$

- **Step 3: Reconciled point forecasts for the whole hierarchy**
  - Mapping $\tilde{\boldsymbol{b}}_{t+h} \in \mathbb{R}^m$ to $\mathbb{C}^m$ through $\boldsymbol{S}$ we get

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}. \tag{1}$$

- **Step 3: Reconciled point forecasts for the whole hierarchy**
  - Mapping $\tilde{\boldsymbol{b}}_{t+h} \in \mathbb{R}^m$ to $\mathbb{C}^m$ through $\boldsymbol{S}$ we get

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}(\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}, \qquad \tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m \subset \mathbb{R}^n. \tag{2}$$

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}. \tag{1}$$

- **Step 3: Reconciled point forecasts for the whole hierarchy**
    - Mapping $\tilde{\boldsymbol{b}}_{t+h} \in \mathbb{R}^m$ to $\mathbb{C}^m$ through $\boldsymbol{S}$ we get

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}(\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}\hat{\boldsymbol{y}}_{t+h}, \qquad \tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m \subset \mathbb{R}^n. \tag{2}$$

- Previous definitions on point forecast reconciliation coincides if $\boldsymbol{P} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}$

Point forecast reconciliation *cont..*

- **Step 2: Ignoring the null coordinates**

$$\tilde{\boldsymbol{b}}_{t+h} = (\boldsymbol{R}'_\perp \boldsymbol{S})^{-1} \boldsymbol{R}'_\perp \hat{\boldsymbol{y}}_{t+h}. \quad (1)$$

- **Step 3: Reconciled point forecasts for the whole hierarchy**
  - Mapping $\tilde{\boldsymbol{b}}_{t+h} \in \mathbb{R}^m$ to $\mathbb{C}^m$ through $\boldsymbol{S}$ we get

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}(\boldsymbol{R}'_\perp \boldsymbol{S})^{-1} \boldsymbol{R}'_\perp \hat{\boldsymbol{y}}_{t+h}, \qquad \tilde{\boldsymbol{y}}_{t+h} \in \mathbb{C}^m \subset \mathbb{R}^n. \quad (2)$$

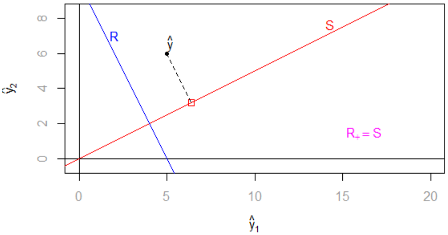
- Previous definitions on point forecast reconciliation coincides if $\boldsymbol{P} = (\boldsymbol{R}'_\perp \boldsymbol{S})^{-1} \boldsymbol{R}'_\perp$
- Further we find $\boldsymbol{R}_\perp$ s.t $(\boldsymbol{R}'_\perp \boldsymbol{S})^{-1} \boldsymbol{R}'_\perp \boldsymbol{S} = \boldsymbol{I}$. This coincides with the unbiased condition $\boldsymbol{SPS} = \boldsymbol{S}$ (Hyndman et al., 2011)

Point forecast reconciliation *cont..*

- OLS Reconciliation

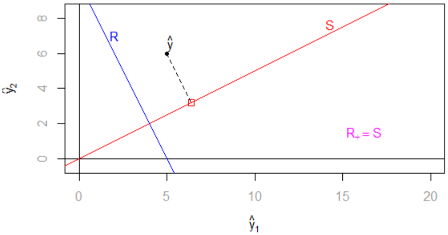
Point forecast reconciliation *cont..*

- OLS Reconciliation



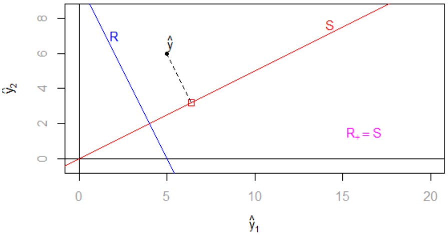
Point forecast reconciliation *cont..*

- OLS Reconciliation



Point forecast reconciliation *cont..*

- OLS Reconciliation



Method	Reconciled point forecasts	$R'_{\perp}$
OLS	$\tilde{\boldsymbol{y}}^{OLS}_{t+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_{t+h}$	$\boldsymbol{S}'$
MinT	$\tilde{\boldsymbol{y}}^{MinT}_{t+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_{t+h}$	$\boldsymbol{S}'\boldsymbol{W}_h^{-1}$

Probabilistic forecast reconciliation

Probabilistic forecast reconciliation

Definition 5: Probabilistic forecast reconciliation

Suppose $(\mathbb{R}^n, \mathscr{F}^n, \hat{\nu})$ is an incoherent probability triple and $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ is a probability triple defined on $\mathbb{R}^m$. Let $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$. Then the probability measure on the reconciled bottom levels is such that

$$\nu^m(\boldsymbol{A}) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})), \quad \forall \quad \boldsymbol{A} \in \mathscr{F}^m.$$

Further the probability measure of the whole reconciled hierarchy is given by

$$\tilde{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})) \qquad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S} : \mathbb{R}^m \to \mathbb{C}^m$, $\tilde{\nu}(\cdot)$ is the probability measure on the measure space $(\mathbb{C}^m, \mathscr{F}_S)$ and $\boldsymbol{g}^{-1}(\boldsymbol{A})$ is the pre-image of $\boldsymbol{A}$ in $\mathbb{R}^n$

Probabilistic forecast reconciliation

Definition 5: Probabilistic forecast reconciliation

Suppose $(\mathbb{R}^n, \mathscr{F}^n, \hat{\nu})$ is an incoherent probability triple and $(\mathbb{R}^m, \mathscr{F}^m, \nu^m)$ is a probability triple defined on $\mathbb{R}^m$. Let $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$. Then the probability measure on the reconciled bottom levels is such that

$$\nu^m(\boldsymbol{A}) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})), \quad \forall \quad \boldsymbol{A} \in \mathscr{F}^m.$$

Further the probability measure of the whole reconciled hierarchy is given by

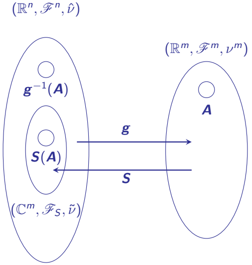
$$\tilde{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})) \qquad \forall \boldsymbol{A} \in \mathscr{F}^m,$$

where $\boldsymbol{S} : \mathbb{R}^m \to \mathbb{C}^m$, $\tilde{\nu}(\cdot)$ is the probability measure on the measure space $(\mathbb{C}^m, \mathscr{F}_S)$ and $\boldsymbol{g}^{-1}(\boldsymbol{A})$ is the pre-image of $\boldsymbol{A}$ in $\mathbb{R}^n$

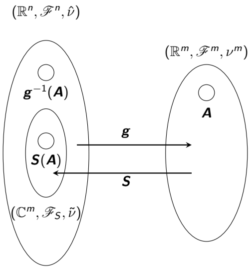
Probabilistic forecast reconciliation *cont..*

Probabilistic forecast reconciliation *cont..*

Probabilistic forecast reconciliation *cont..*



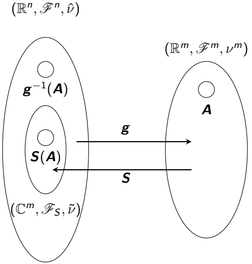
Probabilistic forecast reconciliation *cont..*



$$\nu^m(\boldsymbol{A}) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})), \quad \forall \boldsymbol{A} \in \mathscr{F}^m.$$

$$\tilde{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})) \qquad \forall \boldsymbol{A} \in \mathscr{F}^m.$$

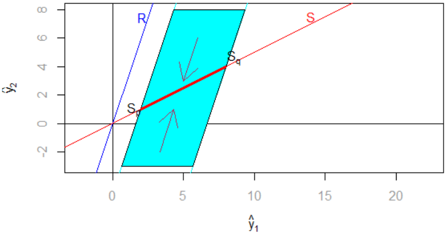
Probabilistic forecast reconciliation *cont..*



$$\nu^m(\boldsymbol{A}) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})), \quad \forall \boldsymbol{A} \in \mathscr{F}^m.$$

$$\tilde{\nu}(\boldsymbol{S}(\boldsymbol{A})) = \hat{\nu}(\boldsymbol{g}^{-1}(\boldsymbol{A})) \qquad \forall \boldsymbol{A} \in \mathscr{F}^m.$$

- Back to 2-dimensional schematic



Probabilistic forecast reconciliation *cont..*

- **Step 1: Change of coordinates** (linear transformation step)

Probabilistic forecast reconciliation *cont..*

- **Step 1: Change of coordinates** (linear transformation step)

  Recall $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$.

Probabilistic forecast reconciliation *cont..*

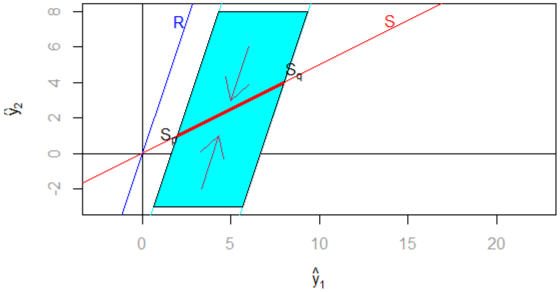
- **Step 1: Change of coordinates** (linear transformation step)

  Recall $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$.

Probabilistic forecast reconciliation *cont..*

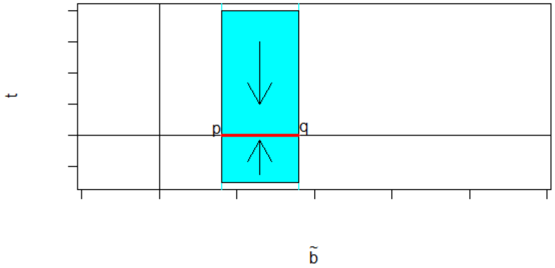
- **Step 1: Change of coordinates** (linear transformation step)

  Recall $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$.



- **Step 1: Change of coordinates** (linear transformation step)

Recall $(\tilde{\boldsymbol{b}}'_{t+h}, \boldsymbol{t}'_{t+h})'$: coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to $(\boldsymbol{S} \vdots \boldsymbol{R})$.



Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f}_{\boldsymbol{B}}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f_B}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$
- Recall that $(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h}, \ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h}$, and $\hat{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}$

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f_B}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$
- Recall that $(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h}$, and $\hat{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}$
  Then we have

$$\boldsymbol{f_B}(\tilde{\boldsymbol{b}}_{t+h}, \boldsymbol{t}_{t+h}) = \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left|\boldsymbol{S} \vdots \boldsymbol{R}\right|$$

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f_B}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$
- Recall that $(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h}$, and $\hat{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}$ Then we have

$$\boldsymbol{f_B}(\tilde{\boldsymbol{b}}_{t+h}, \boldsymbol{t}_{t+h}) = \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left|\boldsymbol{S} \vdots \boldsymbol{R}\right|$$

- **Step 2: Marginalizing over the null space**

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f_B}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$
- Recall that $(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h}$, and $\hat{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}$
  Then we have

$$\boldsymbol{f_B}(\tilde{\boldsymbol{b}}_{t+h}, \boldsymbol{t}_{t+h}) = \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left|\boldsymbol{S} \vdots \boldsymbol{R}\right|$$

- **Step 2: Marginalizing over the null space**
  i.e.,

$$\tilde{\boldsymbol{f}}_b(\tilde{\boldsymbol{b}}_{t+h}) = \int \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left|\boldsymbol{S} \vdots \boldsymbol{R}\right| d\boldsymbol{t}_{t+h}$$

Probabilistic forecast reconciliation *cont..*

- Let $\hat{\boldsymbol{f}}(\cdot)$ be the probability density of $\hat{\boldsymbol{y}}_{t+h}$
- Let $\boldsymbol{f_B}(\cdot)$ be the density of the coordinates of $\hat{\boldsymbol{y}}_{t+h}$ with respect to basis $\boldsymbol{B}$
- Recall that $(\boldsymbol{S} \vdots \boldsymbol{R})(\tilde{\boldsymbol{b}}'_{t+h},\ \boldsymbol{t}'_{t+h})' = \hat{\boldsymbol{y}}_{t+h}$, and $\hat{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}$
Then we have

$$\boldsymbol{f_B}(\tilde{\boldsymbol{b}}_{t+h}, \boldsymbol{t}_{t+h}) = \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left| \boldsymbol{S} \vdots \boldsymbol{R} \right|$$

- **Step 2: Marginalizing over the null space**
i.e.,

$$\tilde{\boldsymbol{f}}_b(\tilde{\boldsymbol{b}}_{t+h}) = \int \hat{\boldsymbol{f}}(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h} + \boldsymbol{R}\boldsymbol{t}_{t+h}) \left| \boldsymbol{S} \vdots \boldsymbol{R} \right| d\boldsymbol{t}_{t+h}$$

- **Step 3: Reconciled density of the whole hierarchy**

$$\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{y}}_{t+h}) = \tilde{\boldsymbol{f}}_b(\boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h}), \quad [\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{S}\tilde{\boldsymbol{b}}_{t+h}]$$

Probabilistic forecasts reconciliation assuming Gaussianity

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

$$\mathscr{N}(\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{W}_{T+h}) \overset{d}{\leftrightarrow} \hat{\boldsymbol{f}}(.)$$

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

- Gaussian densities are closed under affine transformation and marginalization

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

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

$$\boldsymbol{f_B}(\cdot) = \frac{\exp\left\{-\frac{1}{2}\left((\boldsymbol{S} \vdots \boldsymbol{R})\begin{pmatrix}\tilde{\boldsymbol{b}}_{T+h} \\ \boldsymbol{t}_{T+h}\end{pmatrix} - \hat{\boldsymbol{y}}_{T+h}\right)' \boldsymbol{W}_{\boldsymbol{T+h}}^{-1}\left((\boldsymbol{S} \vdots \boldsymbol{R})\begin{pmatrix}\tilde{\boldsymbol{b}}_{T+h} \\ \boldsymbol{t}_{T+h}\end{pmatrix} - \hat{\boldsymbol{y}}_{T+h}\right)\right\}}{(2\pi)^{\frac{n}{2}} \left|\boldsymbol{W}_{T+h}\right|^{\frac{1}{2}} \left|(\boldsymbol{S} \vdots \boldsymbol{R})^{-1}\right|}$$

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

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

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

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

$$\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{y}}_{t+h}) = \tilde{\boldsymbol{f}}_b(\boldsymbol{S}\tilde{\boldsymbol{b}}_{T+h}) \overset{d}{\leftrightarrow} \mathscr{N}(\boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{T+h},\ \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_{T+h}\boldsymbol{P}'\boldsymbol{S}'),$$

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

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

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

for $\alpha \in (0, 2]$,
where $\tilde{\boldsymbol{Y}}_{T+h}$ and $\tilde{\boldsymbol{Y}}^{*}_{T+h}$ are independent random variables from $\mathcal{N}(\boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{SPW}_{T+h}\boldsymbol{P'S'})$.

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

for $\alpha \in (0, 2]$,
where $\tilde{\boldsymbol{Y}}_{T+h}$ and $\tilde{\boldsymbol{Y}}^{*}_{T+h}$ are independent random variables from $\mathcal{N}(\boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{SPW}_{T+h}\boldsymbol{P'S'})$.

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

  for $\alpha \in (0, 2]$,
  where $\tilde{\boldsymbol{Y}}_{T+h}$ and $\tilde{\boldsymbol{Y}}^{*}_{T+h}$ are independent random variables from $\mathcal{N}(\boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{SPW}_{T+h}\boldsymbol{P'S'})$.

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

  for $\alpha \in (0, 2]$,
  where $\boldsymbol{\tilde{Y}}_{T+h}$ and $\boldsymbol{\tilde{Y}}^{*}_{T+h}$ are independent random variables from $\mathcal{N}(\boldsymbol{SP\hat{y}}_{T+h}, \boldsymbol{SPW}_{T+h}\boldsymbol{P'S'})$.

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

Probabilistic forecasts reconciliation assuming Gaussianity

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

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

  for $\alpha \in (0, 2]$,
  where $\tilde{\boldsymbol{Y}}_{T+h}$ and $\tilde{\boldsymbol{Y}}^{*}_{T+h}$ are independent random variables from $\mathcal{N}(\boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h}, \boldsymbol{SPW}_{T+h}\boldsymbol{P'S'})$.

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

- However in the upper limit of α, i.e. when $\alpha = 2$,

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

Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

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

where $\mathcal{I}_T = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$

Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{E_{\boldsymbol{y}_{T+h}}[(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})^T|\mathcal{I}_T]\},$$

  where $\boldsymbol{\mathcal{I}}_T = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$
- Assuming unbiasedness for coherent forecasts

Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{E_{\boldsymbol{y}_{T+h}}[(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})^T|\mathcal{I}_T]\},$$

where $\boldsymbol{\mathcal{I}}_T = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$

- Assuming unbiasedness for coherent forecasts

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{\boldsymbol{SPW}_{T+h}\boldsymbol{P}^T\boldsymbol{S}^T\}$$

$$\text{Subject to} \quad \boldsymbol{SPS} = \boldsymbol{S}$$

Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{E_{\boldsymbol{y}_{T+h}}[(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})^T|\mathcal{I}_T]\},$$

  where $\mathcal{I}_T = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$
- Assuming unbiasedness for coherent forecasts

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{\boldsymbol{SPW}_{T+h}\boldsymbol{P}^T\boldsymbol{S}^T\}$$

$$\text{Subject to} \quad \boldsymbol{SPS} = \boldsymbol{S}$$

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

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

Probabilistic forecasts reconciliation assuming Gaussianity

- Then our objective function is

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{E_{\boldsymbol{y}_{T+h}}[(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})(\boldsymbol{y}_{T+h} - \boldsymbol{SP}\hat{\boldsymbol{y}}_{T+h})^T|\mathcal{I}_T]\},$$

where $\mathcal{I}_T = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_T\}$

- Assuming unbiasedness for coherent forecasts

$$\underset{\boldsymbol{P}}{\arg\min}\ Tr\{\boldsymbol{SPW}_{T+h}\boldsymbol{P}^T\boldsymbol{S}^T\}$$

$$\text{Subject to} \quad \boldsymbol{SPS} = \boldsymbol{S}$$

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

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

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

Probabilistic forecasts in the Gaussian framework

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

Probabilistic forecasts in the Gaussian framework

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

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

Probabilistic forecasts in the Gaussian framework

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

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

Probabilistic forecasts in the Gaussian framework

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

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

Probabilistic forecasts in the Gaussian framework

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

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

Probabilistic forecasts in the Gaussian framework

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

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

Method	**Estimate of $\boldsymbol{W}_h$**	**Estimate of $\boldsymbol{R}'_{\perp}$**
OLS	$\boldsymbol{I}$	$\boldsymbol{S}'$
MinT(Sample)	$\hat{\boldsymbol{W}}^{sam}_{T+1}$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{sam}_{T+1})^{-1}$
MinT(Shrink)	$\hat{\boldsymbol{W}}^{shr}_{T+1} = \tau \text{Diag}(\hat{\boldsymbol{W}}^{sam}_{T+1}) + (1-\tau)\hat{\boldsymbol{W}}^{sam}_{T+1}$ A2	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{shr}_{T+1})^{-1}$
MinT(WLS)	$\hat{\boldsymbol{W}}^{wls}_{T+1} = \text{Diag}(\hat{\boldsymbol{W}}^{shr}_{T+1})$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{wls}_{T+1})^{-1}$

Probabilistic forecasts in the Gaussian framework

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

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

Method	**Estimate of $\boldsymbol{W}_h$**	**Estimate of $\boldsymbol{R}'_{\perp}$**
OLS	$\boldsymbol{I}$	$\boldsymbol{S}'$
MinT(Sample)	$\hat{\boldsymbol{W}}^{sam}_{T+1}$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{sam}_{T+1})^{-1}$
MinT(Shrink)	$\hat{\boldsymbol{W}}^{shr}_{T+1} = \tau\text{Diag}(\hat{\boldsymbol{W}}^{sam}_{T+1}) + (1-\tau)\hat{\boldsymbol{W}}^{sam}_{T+1}$ A2	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{shr}_{T+1})^{-1}$
MinT(WLS)	$\hat{\boldsymbol{W}}^{wls}_{T+1} = \text{Diag}(\hat{\boldsymbol{W}}^{shr}_{T+1})$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}^{wls}_{T+1})^{-1}$

Probabilistic forecasts in the Gaussian framework

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

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

Method	**Estimate of $\boldsymbol{W}_h$**	**Estimate of $\boldsymbol{R}'_{\perp}$**
OLS	$\boldsymbol{I}$	$\boldsymbol{S}'$
MinT(Sample)	$\hat{\boldsymbol{W}}_{T+1}^{sam}$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}_{T+1}^{sam})^{-1}$
MinT(Shrink)	$\hat{\boldsymbol{W}}_{T+1}^{shr} = \tau\text{Diag}(\hat{\boldsymbol{W}}_{T+1}^{sam}) + (1-\tau)\hat{\boldsymbol{W}}_{T+1}^{sam}$ A2	$\boldsymbol{S}'(\hat{\boldsymbol{W}}_{T+1}^{shr})^{-1}$
MinT(WLS)	$\hat{\boldsymbol{W}}_{T+1}^{wls} = \text{Diag}(\hat{\boldsymbol{W}}_{T+1}^{shr})$	$\boldsymbol{S}'(\hat{\boldsymbol{W}}_{T+1}^{wls})^{-1}$

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

Evaluating probabilistic forecasts

Energy score (Gneiting et al., 2008)

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

Log score (Gneiting and Raftery, 2007)

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

Variogram score (**SCHEUERER˙2015**)

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

CRPS (Gneiting and Raftery, 2007)

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

$\breve{\boldsymbol{Y}}_{T+h}$ and $\breve{\boldsymbol{Y}}^{*}_{T+h}$	:	Independent random vectors from the coherent forecast distribution $\breve{\boldsymbol{F}}$.
$\boldsymbol{y}_{T+h}$	:	Vector of realizations.
$\breve{Y}_{T+h,i}$ and $\breve{Y}_{T+h,j}$	:	ith and jth components of the vector $\breve{\boldsymbol{Y}}_{T+h}$
w_{ij}	:	Non-negative weights

Evaluating probabilistic forecasts *Cont..*

- To compare two competitive forecasting methods, we use Skill score (Gneiting and Raftery, 2007)

Evaluating probabilistic forecasts *Cont..*

- To compare two competitive forecasting methods, we use Skill score (Gneiting and Raftery, 2007)

Evaluating probabilistic forecasts *Cont..*

- To compare two competitive forecasting methods, we use Skill score (Gneiting and Raftery, 2007)

- For a given forecasting method, evaluated by a particular scoring rule $S(\cdot)$, the skill score is given by

$$SS[S_B(\cdot)] = \frac{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}} - S_B(\check{\boldsymbol{Y}}, \boldsymbol{y})}{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}} \times 100\%,$$

where

$S_B(\cdot)$	:	Average score over B samples
$S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}$	:	Average score of the reference forecasting methods

Evaluating probabilistic forecasts *Cont..*

- To compare two competitive forecasting methods, we use Skill score (Gneiting and Raftery, 2007)

- For a given forecasting method, evaluated by a particular scoring rule $S(\cdot)$, the skill score is given by

$$SS[S_B(\cdot)] = \frac{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}} - S_B(\check{\boldsymbol{Y}}, \boldsymbol{y})}{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}} \times 100\%,$$

where

$S_B(\cdot)$	:	Average score over B samples
$S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}$	:	Average score of the reference forecasting methods

Evaluating probabilistic forecasts *Cont..*

- To compare two competitive forecasting methods, we use Skill score (Gneiting and Raftery, 2007)

- For a given forecasting method, evaluated by a particular scoring rule $S(\cdot)$, the skill score is given by

$$SS[S_B(\cdot)] = \frac{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}} - S_B(\check{\boldsymbol{Y}}, \boldsymbol{y})}{S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}} \times 100\%,$$

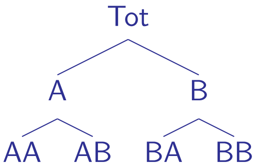
where

$S_B(\cdot)$	:	Average score over B samples
$S_B(\boldsymbol{Y}, \boldsymbol{y})^{\text{ref}}$	:	Average score of the reference forecasting methods

- $SS[S_B(\cdot)]$ gives the percentage improvement of the preferred forecasting method relative to the reference method

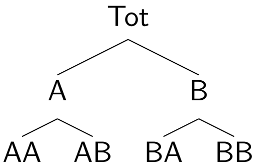
Monte-Carlo simulation

- **Data generating process** A3



Monte-Carlo simulation

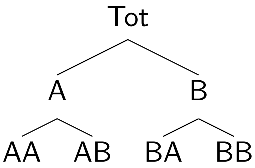
- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels

Monte-Carlo simulation

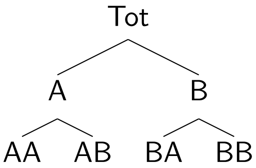
- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels
- **Simulation setup**

Monte-Carlo simulation

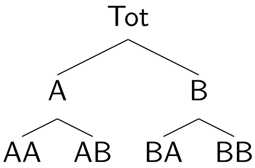
- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels
- **Simulation setup**
  - 501 observations were generated

Monte-Carlo simulation

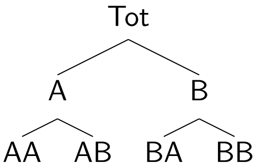
- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels
- **Simulation setup**
  - 501 observations were generated
  - Univariate ARIMA models were fitted to the first 500 observations and 1- step ahead incoherent forecasts were generated

Monte-Carlo simulation

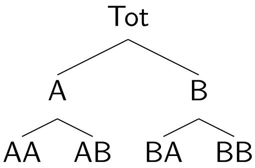
- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels
- **Simulation setup**
  - 501 observations were generated
  - Univariate ARIMA models were fitted to the first 500 observations and 1- step ahead incoherent forecasts were generated
  - Mean and variance forecasts of incoherent Gaussian densities were reconciled through different estimates of $\boldsymbol{R}'_{\perp}$

Monte-Carlo simulation

- **Data generating process** A3



- DGP was designed such that we have noisier series in the bottom levels than in the aggregate levels
- **Simulation setup**
  - 501 observations were generated
  - Univariate ARIMA models were fitted to the first 500 observations and 1- step ahead incoherent forecasts were generated
  - Mean and variance forecasts of incoherent Gaussian densities were reconciled through different estimates of $\boldsymbol{R}'_{\perp}$
  - This process was replicated using 1000 different series from the same (DGP)

Monte-Carlo simulation *Cont..*

- Comparison of reconciled vs bottom-up Gaussian forecasts

Forecasting method	Energy score		Log score		Variogram score	
	Average score	Skill score (%)	Average score	Skill score (%)	Average score	Skill score (%)
MinT(Shrink)	7.47	10.11	11.34	6.44	3.05	4.69
MinT(Sample)	7.47	10.11	11.33	6.52	3.05	4.69
MinT(WLS)	7.91	4.81	12.64	−4.29	3.23	−0.94
OLS	10.14	−22.02	135.13	−1014.93	4.60	−43.75
Bottom up	8.31		12.12		3.20	

Note: *Positive(/negative) entry in "Skill score (%)" column represent the percentage increase(/decrease) of preferred forecasting method over Bottom-up method.*

Monte-Carlo simulation *Cont..*

- Comparison of coherent vs incoherent Gaussian forecasts

Forecasting method	Total CRPS	Total LogS	Series - A CRPS	Series - A LogS	Series - B CRPS	Series - B LogS
MinT(Shrink)	1.12	0.34	10.07	2.93	5.41	1.52
MinT(Sample)	1.12	0.34	10.07	2.93	5.41	1.52
MinT(WLS)	−2.61	−2.02	5.28	−4.40	2.70	−4.24
OLS	−38.06	−698.99	−24.70	−1368.33	−24.86	−1159.09
Bottom up	−89.55	−21.83	−8.87	−2.35	−9.46	−2.73
Incoherent	*2.68*	*2.97*	*4.17*	*3.41*	*3.70*	*3.30*

Forecasting method	Series - AA CRPS	Series - AA LogS	Series - AB CRPS	Series - AB LogS	Series - BA CRPS	Series - BA LogS	Series - BB CRPS	Series - BB LogS
MinT(Shrink)	8.71	2.71	10.57	3.04	5.95	1.86	7.91	2.46
MinT(Sample)	8.71	2.71	10.57	3.04	5.95	1.86	8.19	2.46
MinT(WLS)	5.54	0.30	5.96	0.30	2.43	−0.62	5.08	0.62
OLS	−22.43	−931.63	−22.49	−886.32	−26.01	−834.67	−23.45	−812.92
Incoherent	*3.79*	*3.32*	*3.69*	*3.29*	*3.46*	*3.23*	*3.54*	*3.25*

Note: *"Incoherent" row represents the average score for incoherent forecasts.*
Each entry above this row represent the percentage skill score with reference to the Incoherent forecasts.

Probabilistic forecasts in a non-parametric framework: A bootstrap approach

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

2. Let $\Gamma_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

2. Let $\boldsymbol{\Gamma}_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

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

2. Let $\boldsymbol{\Gamma}_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$

3. Let $\boldsymbol{E}^b_{(h\times n)} = (\boldsymbol{e}_{i+1}, \boldsymbol{e}_{i+2}, \dots, \boldsymbol{e}_{i+h})'$ be a block bootstrap sample of size h from $\boldsymbol{\Gamma}$

4. Generate h-step ahead sample paths from the fitted models incorporating $\boldsymbol{E}^b$. Denote these by $\boldsymbol{y}^b_{T+h}$

5. Repeat step 3-5 for J times

Probabilistic forecast reconciliation in a non-parametric framework: A bootstrap approach

1. Fit univariate models at each node using data up to time T
2. Let $\Gamma_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$
3. Let $\boldsymbol{E}^b_{(h\times n)} = (\boldsymbol{e}_{i+1}, \boldsymbol{e}_{i+2}, \dots, \boldsymbol{e}_{i+h})'$ be a block bootstrap sample of size h from Γ
4. Generate h-step ahead sample paths from the fitted models incorporating $\boldsymbol{E}^b$. Denote these by $\boldsymbol{y}^b_{T+h}$
5. Repeat step 3-5 for J times
6. $\tilde{\boldsymbol{y}}^b_{T+h} = \boldsymbol{S}(\boldsymbol{R}'_\perp \boldsymbol{S})^{-1}\boldsymbol{R}'_\perp \boldsymbol{y}^b_{T+h}$,
   where $\tilde{\boldsymbol{y}}^b_{T+h}$ denotes h-step ahead coherent sample paths

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

- **Simulation setup**

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

- **Simulation setup**
  - 1510 observations were generated

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

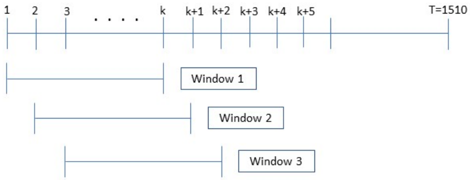
- **Simulation setup**
  - 1510 observations were generated
  - Univariate ARIMA models were fitted using a rolling window of 500 observations as the training set

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

- **Simulation setup**
  - 1510 observations were generated
  - Univariate ARIMA models were fitted using a rolling window of 500 observations as the training set
  - 5000 reconciled future sample paths were generated for forecast horizons $h = 1, ..., 5$

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

- **Simulation setup**
  - 1510 observations were generated
  - Univariate ARIMA models were fitted using a rolling window of 500 observations as the training set
  - 5000 reconciled future sample paths were generated for forecast horizons $h = 1, ..., 5$
  - The training window was rolled one observation ahead and the process repeated until we get 1000 reconciled future sample paths



Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework

- Comparison of reconciled vs bottom-up forecasts

Forecasting method	h=1 ES	h=1 VS	h=2 ES	h=2 VS	h=3 ES	h=3 VS	h=4 ES	h=4 VS	h=5 ES	h=5 VS
MinT(Shrink)	6.39	5.83	6.48	5.54	5.62	5.43	3.94	3.98	3.45	3.64
MinT(Sample)	6.39	5.83	6.35	5.19	5.29	4.89	3.47	3.09	3.04	2.73
MinT(WLS)	3.83	0.89	3.89	1.04	3.35	1.09	2.44	1.33	2.06	1.28
OLS	3.35	0	3.24	0.35	2.69	0.27	2.06	0.66	1.73	0.73
Bottom Up	*6.26*	*2.23*	*7.72*	*2.89*	*9.26*	*3.68*	*10.67*	*4.52*	*12.16*	*5.49*

Note: *"Bottom-up" row represent the average score for incoherent forecasts. Each entry above this row represent the percentage skill score with reference to the Bottom-up forecasts.*

Analyzing the predictive performance of probabilistic forecasts in a non-parametric framework *Cont..*

- Comparison of coherent vs incoherent forecasts

Forecasting method	Total	Series-A	Series-B	Series-AA	Series-AB	Series-BA	Series-BB
				h=1			
MinT(Shrink)	0.28	7.17	6.57	10.04	2.93	7.35	0.41
MinT(Sample)	0.29	7.10	6.45	10.04	3.02	7.22	0.32
MinT(WLS)	−1.78	6.46	5.80	5.17	−0.03	6.37	−0.68
OLS	−6.30	5.62	5.49	4.50	0.09	5.42	−1.59
Bottom Up	−69.68	−5.25	−4.51	0.00	0.00	0.00	0.00
Incoherent	*1.86*	*2.73*	*2.68*	*2.93*	*2.78*	*2.98*	*2.35*
				h=5			
MinT(Shrink)	0.63	6.30	0.67	7.56	0.35	1.63	−2.01
MinT(Sample)	0.56	6.39	0.58	7.15	0.02	1.29	−3.47
MinT(WLS)	1.24	5.45	0.62	5.39	−3.10	2.41	−3.36
OLS	1.16	5.07	1.20	4.64	−2.78	2.43	−5.33
Bottom Up	−9.44	−0.52	−1.42	0.00	0.00	0.00	0.00
Incoherent	*8.72*	*7.46*	*5.96*	*6.78*	*4.66*	*5.95*	*2.68*
Note:	*"Incoherent" row represent the average score for incoherent forecasts. Each entry above this row represent the percentage skill score with reference to the Incoherent forecasts.*						

Probabilistic forecasts for Australian domestic tourism flows

Probabilistic forecasts for Australian domestic tourism flows

- Domestic tourism in Australia can be disaggregated based on a geographical hierarchy

Probabilistic forecasts for Australian domestic tourism flows

- Domestic tourism in Australia can be disaggregated based on a geographical hierarchy

Probabilistic forecasts for Australian domestic tourism flows

- Domestic tourism in Australia can be disaggregated based on a geographical hierarchy

Geographical hierarchical structure for Australia

Level	**No.Series per level**
Total (Australia)	1
Level-1 (States)	7
Level-2 (Zones)	27
Level-3 (Regions)	75

Probabilistic forecasts for Australian domestic tourism flows

- Domestic tourism in Australia can be disaggregated based on a geographical hierarchy

Geographical hierarchical structure for Australia

Level	**No.Series per level**
Total (Australia)	1
Level-1 (States)	7
Level-2 (Zones)	27
Level-3 (Regions)	75

- Reconciliation of point forecasts (Athanasopoulos, Ahmed, and Hyndman, 2009, Hyndman et al., 2011, Wickramasuriya, Athanasopoulos, and Hyndman, 2018)

Probabilistic forecasts for Australian domestic tourism flows

- Domestic tourism in Australia can be disaggregated based on a geographical hierarchy

Geographical hierarchical structure for Australia

Level	**No.Series per level**
Total (Australia)	1
Level-1 (States)	7
Level-2 (Zones)	27
Level-3 (Regions)	75

- Reconciliation of point forecasts (Athanasopoulos, Ahmed, and Hyndman, 2009, Hyndman et al., 2011, Wickramasuriya, Athanasopoulos, and Hyndman, 2018)
- We use a non-parametric bootstrap approach to obtain reconciled probabilistic forecasts

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly “overnight trips” over the period January 1998 - December 2016

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016

- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016

- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly “overnight trips” over the period January 1998 - December 2016
- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]
- **Analysis:**

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016
- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]
- **Analysis:**
  - Univariate ARIMA models were fitted using a rolling window of 100 observations

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016
- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]
- **Analysis:**
  - Univariate ARIMA models were fitted using a rolling window of 100 observations
  - 5000 reconciled future sample paths were generated for forecast horizons $h = 1, ..., 5$

Probabilistic forecasts for Australian domestic tourism flow

- We consider monthly "overnight trips" over the period January 1998 - December 2016
- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]
- **Analysis:**
  - Univariate ARIMA models were fitted using a rolling window of 100 observations
  - 5000 reconciled future sample paths were generated for forecast horizons $h = 1, ..., 5$
  - The training window was rolled one observation ahead and the process repeated until we get 123 reconciled future sample paths

Probabilistic forecasts for Australian domestic tourism flow

- Comparison of reconciled vs. bottom-up probabilistic forecasts

Forecasting method	h=1 ES	h=1 VS	h=2 ES	h=2 VS	h=3 ES	h=3 VS	h=4 ES	h=4 VS	h=5 ES	h=5 VS
States										
MinT(Shrink)	19.58	27.95	14.65	23.12	12.05	21.25	7.76	14.09	9.50	15.22
MinT(WLS)	4.50	6.58	4.37	6.31	3.32	6.01	1.89	3.52	3.58	4.67
OLS	7.00	3.15	5.58	5.01	2.65	3.31	1.83	−0.66	3.58	2.01
Bottom Up	*345.63*	*271.05*	*348.68*	*281.63*	*355.98*	*290.47*	*367.41*	*300.25*	*364.81*	*299.75*
Zones										
MinT(Shrink)	13.11	19.30	10.94	19.48	9.40	18.24	6.99	15.10	7.62	15.26
MinT(WLS)	2.32	1.99	2.55	2.99	1.95	2.61	1.05	1.28	1.27	1.97
OLS	1.91	−5.82	1.58	−2.36	0.47	−1.78	0.25	−2.07	1.29	−1.57
Bottom Up	*228.70*	*2366.28*	*229.54*	*2440.32*	*234.09*	*2510.51*	*238.98*	*2556.95*	*237.98*	*2552.26*
Regions										
MinT(Shrink)	10.43	15.96	9.03	16.11	7.87	15.05	6.28	12.81	6.96	13.21
MinT(WLS)	1.54	1.43	1.75	2.16	1.30	1.83	0.68	0.95	0.97	1.46
OLS	1.37	−3.01	1.28	−0.77	0.49	−0.90	0.30	−1.08	1.14	−0.18
Bottom Up	*194.19*	*10956.41*	*193.50*	*11145.87*	*196.26*	*11355.30*	*198.25*	*11461.30*	*197.76*	*11438.31*

Note: *"Bottom-up" row represent the average score for incoherent forecasts. Each entry above this row represent the percentage skill score with reference to the Bottom-up forecasts.*

Probabilistic forecasts for Australian domestic tourism flow

- Comparison of coherent vs. incoherent probabilistic forecasts in total Australia and each states

Forecasting method	Australia	NSW	Victoria	Queensland	South Australia	Western Australia	Tasmania	Northern Territory
				h=1				
MinT(Shrink)	8.07	9.61	16.84	22.38	23.38	11.25	13.38	20.41
MinT(WLS)	−12.31	−7.54	2.64	4.95	1.12	1.31	6.93	2.55
OLS	1.39	2.96	5.14	7.68	1.58	−0.48	−13.04	−22.4
Bottom Up	−20.00	−16.22	−1.12	3.07	−4.64	−6.52	4.43	−0.96
Incoherent	*478.81*	*188.48*	*159.49*	*136.27*	*51.29*	*52.97*	*35.23*	*24.27*
				h=5				
MinT(Shrink)	1.72	6.65	11.96	6.30	14.61	3.66	16.54	24.58
MinT(WLS)	−6.83	−0.30	2.01	2.21	2.26	−1.77	1.40	6.45
OLS	−0.29	2.20	3.41	4.23	9.50	2.58	9.99	−4.52
Bottom Up	−9.07	−3.33	0.24	1.48	0.44	−7.35	−2.59	4.70
Incoherent	*552.45*	*217.09*	*170.55*	*145.01*	*54.99*	*59.59*	*38.76*	*33.87*

Note: *"Incoherent" row represent the average score for incoherent forecasts. Each entry above this row represent the percentage skill score with reference to the Incoherent forecasts.*

Probabilistic forecasts for Australian domestic tourism flow

- Comparison of coherent vs. incoherent probabilistic forecasts in Zones and Regions

Forecasting	Zones					Regions				
method	h=1	h=2	h=3	h=4	h=5	h=1	h=2	h=3	h=4	h=5
MinT(Shrink)	14.38	11.25	10.20	8.67	9.39	9.96	8.58	7.71	6.53	6.98
MinT(WLS)	2.88	2.15	1.85	1.76	1.75	1.54	1.73	1.42	0.82	0.94
OLS	1.79	0.55	−0.12	1.00	2.22	−0.11	−0.07	−0.51	−0.20	0.74
Bottom Up	0.21	−0.66	−0.43	0.35	0.26					
Incoherent	*31.75*	*31.87*	*32.71*	*33.73*	*33.49*	*14.36*	*14.45*	*14.70*	*14.87*	*14.81*

Note: *"Incoherent" row represent the average score for incoherent forecasts. Each entry above this row represent the percentage skill score with reference to the Incoherent forecasts.*

Probabilistic forecasts for Australian domestic tourism flow

- Comparison of coherent vs. incoherent mean forecasts

Forecasting	Australia					States				
method	h=1	h=2	h=3	h=4	h=5	h=1	h=2	h=3	h=4	h=5
MinT(Shrink)	11.09	5.56	4.61	2.17	1.51	16.64	11.28	12.19	6.39	8.10
MinT(WLS)	−9.90	−5.33	−2.17	−0.71	−5.48	0.08	0.26	1.31	−0.16	0.41
OLS	1.62	1.53	1.05	1.16	0.12	2.26	0.47	0.43	−0.22	3.06
Bottom Up	−18.88	−12.33	−7.45	−3.66	−8.23	−5.86	−4.64	−2.60	−2.64	−2.16
Incoherent	*669.21*	*697.51*	*744.54*	*812.17*	*764.24*	*128.71*	*132.00*	*138.92*	*142.92*	*142.07*
	Zones					Regions				
	h=1	h=2	h=3	h=4	h=5	h=1	h=2	h=3	h=4	h=5
MinT(Shrink)	14.64	11.15	10.71	8.58	9.16	10.87	8.81	8.05	6.71	6.88
MinT(WLS)	2.77	2.13	1.92	1.59	1.68	1.75	1.80	1.54	0.88	0.95
OLS	1.59	0.55	0.04	0.55	1.80	−0.07	−0.17	−0.53	−0.41	0.31
Bottom Up	−0.09	−0.90	−0.66	0.26	0.28					
Incoherent	*43.95*	*44.38*	*45.58*	*47.01*	*46.53*	*20.11*	*20.23*	*20.58*	*20.85*	*20.72*

Note: *"Incoherent" row represent the average MAE for incoherent forecasts. Each entry above this row represent the percentage improvement with reference to the Incoherent forecasts.*

Conclusions and future work

Conclusions

- We define probabilistic forecast reconciliation

Conclusions

- We define probabilistic forecast reconciliation

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation
- Linear reconciliation involves two steps

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation
- Linear reconciliation involves two steps

  Step 1: Transform coordinates of incoherent probabilistic forecasts

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation
- Linear reconciliation involves two steps

  Step 1: Transform coordinates of incoherent probabilistic forecasts

  Step 2: Marginalize over the null space of the coherent subspace

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation
- Linear reconciliation involves two steps

  Step 1: Transform coordinates of incoherent probabilistic forecasts

  Step 2: Marginalize over the null space of the coherent subspace

Conclusions

- We define probabilistic forecast reconciliation
- Our definition holds for linear and non-linear reconciliation
- Linear reconciliation involves two steps

  Step 1: Transform coordinates of incoherent probabilistic forecasts

  Step 2: Marginalize over the null space of the coherent subspace
- MinT not only produces optimal point forecasts, but also optimally reconcile Gaussian forecast densities

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts

- Results from a simulation study favored MinT for reconciling bootstrapped future paths

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts
- Results from a simulation study favored MinT for reconciling bootstrapped future paths

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts
- Results from a simulation study favored MinT for reconciling bootstrapped future paths
- Finally we apply the non-parametric approach to obtain coherent probabilistic forecasts for domestic tourism flow in Australia

Time plan

	Thesis Chapter	Task description	Time duration	Progress
2.	Literature Review	Writing the chapter.	May/2018 - July/2018	Incomplete
3.	Probabilistic forecast reconciliation for hierarchical time series	Completing the paper.	Mar/2018 - Apr/2018	90% Completed
4.	Probabilistic forecast reconciliation in non-parametric framework	Methodology. Simulation study. Providing theoretical foundation.	Nov/2016 - Feb/2017 Mar/2017 - June/2017 Apr/2018 - July/2018	Completed Completed Incomplete
5.	Application	Forecasting Australian domestic tourism flow. Forecasting Walmart sales.	Dec/2017 - Mar/2017 Oct/2017 - Mar/2019	Completed Incomplete

Athanasopoulos, G., R. A. Ahmed, and R. J. Hyndman (2009). "Hierarchical forecasts for Australian domestic tourism". In: *International Journal of Forecasting* 25.1, pp. 146–166.

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". In: *IEEE Transactions on Smart Grid* 7.5, pp. 2448–2455.

Gneiting, T. and A. E. Raftery (2007). "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American Statistical Association* 102.477, pp. 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".

References II

Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang (2011). "Optimal combination forecasts for hierarchical time series". In: *Computational Statistics and Data Analysis* 55.9, pp. 2579–2589.

Scheuerer, M. and T. M. Hamill (2015). "Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities *". In: *Monthly Weather Review* 143.4, pp. 1321–1334.

Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman (2018). "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization". In: *J American Statistical Association*. to appear.

Thank You!!

Appendix

A1

◂ A1

- By substituting the Gaussian distribution function for $\hat{\boldsymbol{f}}(\cdot)$ in $\boldsymbol{f_B}(\cdot)$ we get,

$$\boldsymbol{f_B}(\cdot) = \frac{1}{(2\pi)^{\frac{n}{2}} \left| \begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{pmatrix} \boldsymbol{W}_{t+h} \begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{pmatrix}' \right|^{\frac{1}{2}}} \exp\Bigg\{ -\frac{1}{2} \begin{pmatrix} \tilde{\boldsymbol{b}}_{t+h} - \boldsymbol{P}\hat{\boldsymbol{y}}_{t+h} \\ \boldsymbol{t}_{t+h} - \boldsymbol{Q}\hat{\boldsymbol{y}}_{t+h} \end{pmatrix}' \left[\begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{pmatrix} \boldsymbol{W}_{t+h} \begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{pmatrix}' \right]^{-1} \begin{pmatrix} \tilde{\boldsymbol{b}}_{t+h} - \boldsymbol{P}\hat{\boldsymbol{y}}_{t+h} \\ \boldsymbol{t}_{t+h} - \boldsymbol{Q}\hat{\boldsymbol{y}}_{t+h} \end{pmatrix} \Bigg\}.$$

where

$$(\boldsymbol{S} \vdots \boldsymbol{R})^{-1} = \begin{pmatrix} (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp} \\ \cdots \\ (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp} \end{pmatrix} = \begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{pmatrix},$$

and $\boldsymbol{P} = (\boldsymbol{R}'_{\perp}\boldsymbol{S})^{-1}\boldsymbol{R}'_{\perp}$ and $\boldsymbol{Q} = (\boldsymbol{S}'_{\perp}\boldsymbol{R})^{-1}\boldsymbol{S}'_{\perp}$.

Appendix

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

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

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

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

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

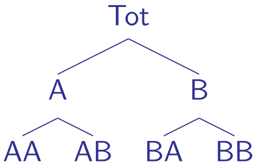
- **Data generating process** A3

Monte-Carlo simulation

- **Data generating process** A3

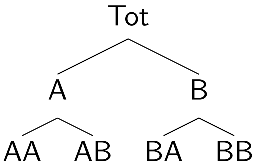
Monte-Carlo simulation

- **Data generating process** A3



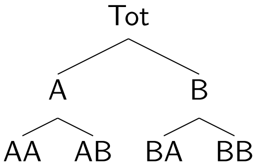
Monte-Carlo simulation

- **Data generating process** A3



Monte-Carlo simulation

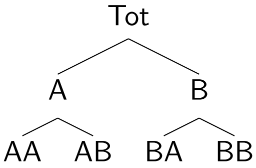
- **Data generating process** A3



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

Monte-Carlo simulation

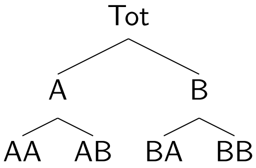
- **Data generating process** A3



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

Monte-Carlo simulation

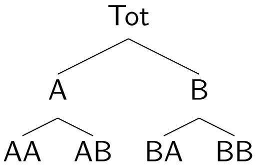
- **Data generating process** A3



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

Monte-Carlo simulation

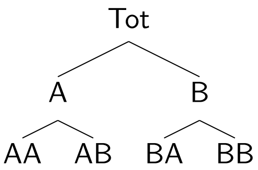
- **Data generating process** A3



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

Monte-Carlo simulation

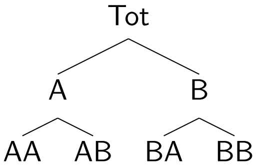
- **Data generating process** A3



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

Monte-Carlo simulation

- **Data generating process** A3

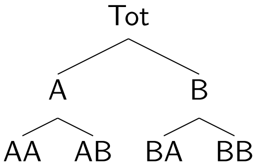


- $\{w_{AA,t}, w_{AB,t}, w_{BA,t}, w_{BB,t}\} \sim ARIMA(p, d, q)$
- $p \in \{1, 2\}$ and $d \in \{0, 1\}$
- $\{\epsilon_{AA,t}, \epsilon_{AB,t}, \epsilon_{BA,t}, \epsilon_{BB,t}\} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
- Parameters for AR and MA components were randomly and uniformly generated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

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

- $\boldsymbol{y}_t$ are then generated as follows

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

Monte-Carlo simulation *Cont..*

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

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

Monte-Carlo simulation *Cont..*

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

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

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

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

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