

# Probabilistic Forecast Reconciliation

Puwasala Gamakumara, Anastasios Panagiotelis, George Athanasopoulos and Rob Hyndman

June 20, 2018

# Hierarchical and Grouped Time Series

- Collections of Time Series often characterised by aggregation constraints
  - Cross-Sectionally
  - Temporally
- **Coherent** forecasts respect such constraints.
- Independently produced forecasts are generally incoherent.

picture of hierarchy

# Forecast Reconciliation

- Forecast reconciliation involves
  - ① Producing incoherent **base** forecasts for all series in an  $n \times 1$  vector  $\hat{\mathbf{y}}$
  - ② Adjusting base forecasts to obtain **reconciled** forecasts in an  $n \times 1$  vector  $\tilde{\mathbf{y}}$
- Why do we care?
  - ① Aligned decision making.
  - ② Improved forecast accuracy

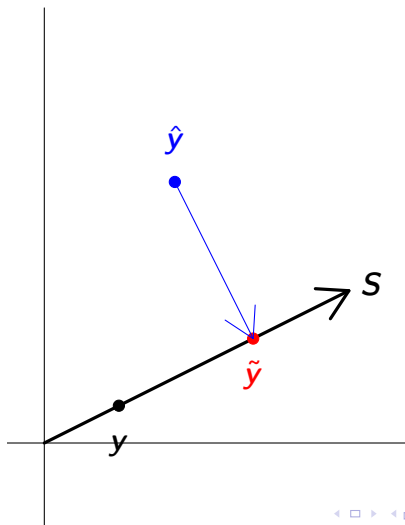
# Reconciliation in two steps

- Many reconciliation methods involve two steps
  - ① Pre-multiply  $\hat{\mathbf{y}}$  by a  $m \times n$  matrix  $\mathbf{G}$  to obtain **bottom** level series  $\mathbf{b} = \mathbf{G}\hat{\mathbf{y}}$
  - ② Pre-multiply  $\mathbf{b}$  by a  $n \times m$  matrix  $\mathbf{S}$  to obtain  $\tilde{\mathbf{y}}$ , i.e.  $\tilde{\mathbf{y}} = \mathbf{S}\mathbf{b}$
- The matrix  $\mathbf{S}$  defines the aggregation constraints, e.g.

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

- Choices of  $\mathbf{G}$  defines reconciliation method, e.g. OLS:  $\mathbf{G} = (\mathbf{S}'\mathbf{S})^{-1} \mathbf{S}'$  and Bottom Up:  $\mathbf{G} = (\mathbf{0}_{m \times n-m} \mathbf{I}_{m \times m})$

# Geometry



# Coherent Subspace

## Definition

The **coherent subspace** is the linear subspace spanned by the columns of  $\mathbf{S}$ , i.e.  $\mathfrak{s} = \text{sp}(\mathbf{S})$

This definition is not unique, it could be defined as the span of:

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

# Coherent Probabilistic Forecast

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

## Definition

The probability measure  $\nu_{\mathfrak{s}}$  is coherent if

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

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

# Reconciled Probabilistic Forecast

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

## Definition

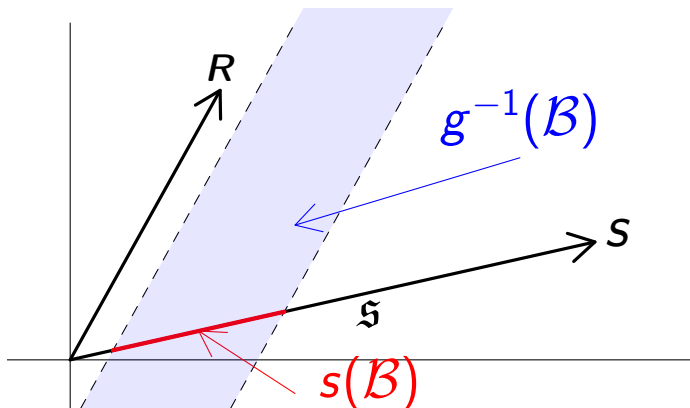
The probability triple  $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \tilde{\nu})$  is the reconciles the probability triple  $(\mathbb{R}^n, \mathcal{F}_{\mathbb{R}^n}, \hat{\nu})$  with with respect to  $g$  iff

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

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



# Geometry



# Analytically

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

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

# Elliptical distributions

Let the unreconciled density be elliptical with location  $\hat{\mu}$  and scale  $\hat{\Sigma}$  and let the true predictive density be elliptical with location  $\mu$  and scale  $\mathbf{S}\Omega\mathbf{S}'$ .

## Theorem

*The true predictive distribution can be recovered via linear reconciliation. The optimal (but infeasible) mapping is  $g(\check{\mathbf{y}}) = \mathbf{G}_{opt}\check{\mathbf{y}} + \mathbf{d}_{opt}$  where  $\mathbf{G}_{opt} = \Omega^{1/2}\hat{\Sigma}^{-1/2}$  and  $\mathbf{d}_{opt} = \mathbf{S}\mathbf{G}_{opt}(\mu - \hat{\mu})$ .*

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

# With a sample

- Often densities are unavailable but we can simulate a sample from the predictive distribution.
- Suppose  $\hat{\mathbf{y}}^{[1]}, \dots, \hat{\mathbf{y}}^{[J]}$  is a sample from the unreconciled probabilistic forecast.
- Then setting  $\tilde{\mathbf{y}}^{[j]} = s \circ g(\hat{\mathbf{y}}^{[j]}) = \mathbf{SG}\hat{\mathbf{y}}^{[j]}$  produces a sample from the reconciled distribution with respect to  $g$ .

# Univariate v Multivariate Scores

Scoring rules can be used to evaluate probabilistic forecasts

- Univariate
  - Log Score
  - Continuous Rank Probability Score
- Multivariate
  - Log Score
  - Energy Score

These may be computed using densities or a sample.

# Approaches

- 1 Use a summary of all univariate scores.
- 2 Make comparisons on the joint distribution of bottom level series only.
- 3 Make comparisons using the full joint distribution

There are pitfalls to the third approach.

# Reconciled v Unreconciled

When comparing reconciled and unreconciled probabilistic forecasts on the basis of log score

## Theorem

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

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

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

# Reconciled v Reconciled

- For two reconciled probabilistic forecasts log score can be used.
- Comparisons can be made on the basis of bottom level series (or any basis series).
- By the definition of coherence  $\log(f(\mathbf{b})) = \log(f(\mathbf{Sb})\mathbf{J})$
- The Jacobian does not affect the ordering of log score.



## Energy score

- More care must be taken using the energy score.
- Energy score is invariant to orthogonal transformation but not affine transformations.
- Since  $\mathbf{S}$  is not a rotation the ranking of different methods based on the full hierarchy may differ from the ranking based on bottom level series only.

# Simulations

- If you want to see tables with numbers look at the paper.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.
  - Bottom up does not do well.
  - MinT (an oblique projection) does well.

# Looking ahead

- The optimal feasible reconciliation method remains an open question even for elliptical distributions.
  - It is likely to depend on the specific score used.
- Are non-linear reconciliation methods worthwhile?
- How should probabilistic reconciliation work for non-elliptical distributions.
- Further development of multivariate scoring rules.