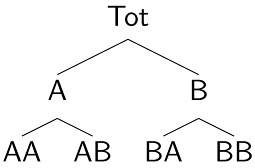
Probabilistic Forecast Reconciliation

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.



Forecast Reconciliation

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

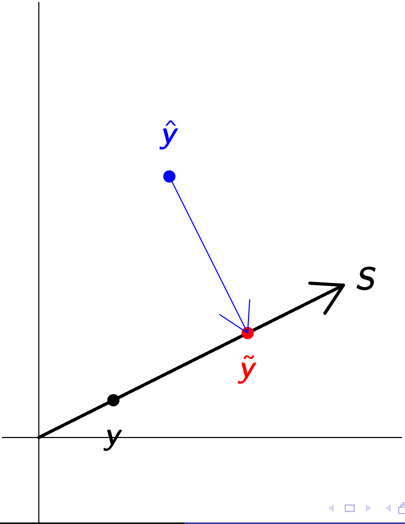
Reconciliation in two steps

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

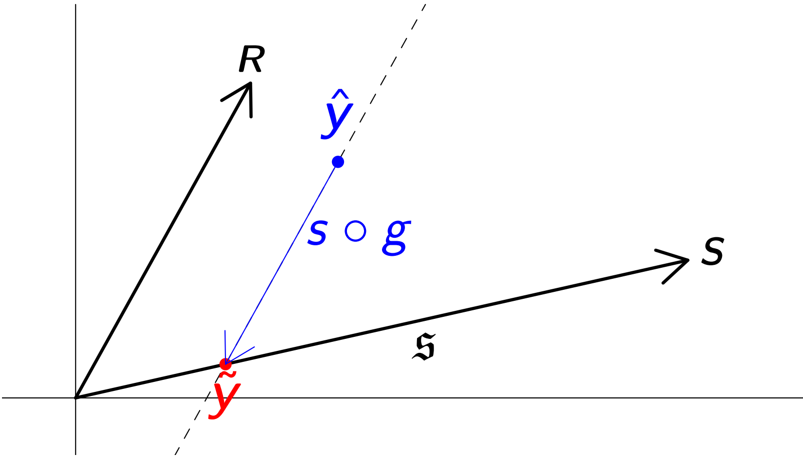
$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ & \boldsymbol{I}_{4\times 4} & & \end{pmatrix}$$

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

Geometry



Geometry: Oblique Projection



Coherent Subspace

Definition

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

Instead of using bottom-level series a different combination of m **basis series** could be used (e.g. top and $m-1$ bottom). Although $\boldsymbol{S}$ would be different $\mathfrak{s}$ would be the same.

Coherent Probabilistic Forecast

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

Definition

The probability measure ν is coherent if

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

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

Reconciled Probabilistic Forecast

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

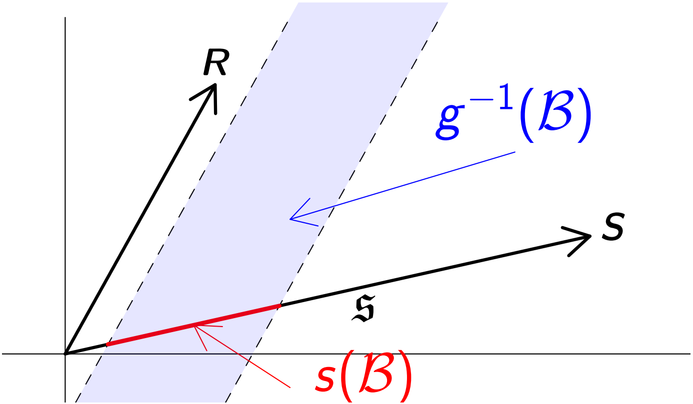
Definition

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

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

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.

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

Elliptical distributions

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

Theorem

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

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 $\hat{\boldsymbol{y}}^{[1]},\ldots,\hat{\boldsymbol{y}}^{[J]}$ is a sample from the unreconciled probabilistic forecast.
- Then setting $\tilde{\boldsymbol{y}}^{[j]}=s\circ g(\hat{\boldsymbol{y}}^{[j]})=\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}^{[j]}$ 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 $f(\boldsymbol{y})$ be the true predictive density (on $\mathfrak{s}$) and LS be the (negatively-oriented) log score. Then there exists an unreconciled density $\hat{f}(\boldsymbol{y})$ on $\mathbb{R}^n$ such that

$$E_{\boldsymbol{y}}\left[S(\hat{f},\boldsymbol{y})\right] < E_{\boldsymbol{y}}\left[S(f,\boldsymbol{y})\right]$$

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 $\log(f(\boldsymbol{b})) = \log(f(\boldsymbol{Sb})\boldsymbol{J})$
- 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 $\boldsymbol{S}$ 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 see the paper.
- The main takeaway messages are:
  - Reconciliation is better than no reconciliation.
  - Bottom up does not do well.
  - OLS (an orthogonal projection) does poorly.
  - MinT (an oblique projection) does best.

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.

The paper

- A paper will be available soon (end of July).
- Google *EBS Monash Working Paper*.
- Or email me :)