Probabilistic Forecast Reconciliation

Puwasala Gamakumara, Anastasios Panagiotelis, George Athanasopoulos and Rob Hyndman

August 15, 2018

Motivating Examples

- Gross Domestic Product
  - An aggregate of consumption, investment, government spending and trade balance
  - Further breakdown, e.g. consumption of food, rent, etc.
- Retail sales
  - Aggregate of different product category sales
  - Further breakdown into individual stock keeping units.
- Potentially need forecasts of all time series.

Incoherent Forecasts

- Potential approaches
  - Univariate models
  - Multivariate models
  - Judgemental forecasts
- Forecasts do not respect aggregation structure (**Incoherent**)
- Outcome *does* respect aggregation structure (**Coherent**)
- Motivation is aggregation but can be generalised to any linear constraints.

Reconciliation

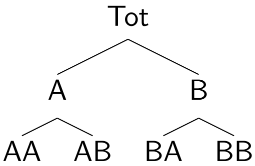
- Begin with a vector of forecasts that are incoherent.
- Adjust these **ex post** to make them coherent.
- There are good solutions for point forecasting that:
  - Guarantee coherent forecasts.
  - Improve forecast accuracy overall.
- Generalisation to probabilistic forecasts is our contribution.
- Getting there necessitates a rethink of the existing point forecasting literature.

What we do NOT do

- All information is contained in the most disaggregate series.
- In principle using the correct multivariate model for the most disaggregate series and aggregating them should work.
- Disaggregate series are:
  - Very noisy
  - High-dimensional
  - Prone to model misspecification

A simple hierarchy

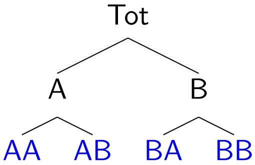
Consider a hierarchy given by



- Let n be the number of series, $\boldsymbol{y}$ be an n-vector of all series.

A simple hierarchy

Consider a hierarchy given by



- Let n be the number of series, $\boldsymbol{y}$ be an n-vector of all series.
- Let m be the bottom level series and $\boldsymbol{b}$ be an m-vector of the bottom level series.

The $\boldsymbol{S}$ matrix

The matrix $\boldsymbol{S}$ defines the aggregation constraints, e.g.

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

Coherence holds when

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

As a regression model

- Cast the problem as a regression model with base forecasts $\hat{\boldsymbol{y}}$ as the "dependent variable" and $\boldsymbol{S}$ as the "design matrix".

$$\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{\beta} + \boldsymbol{e}$$

- Initial approach (Athanasopoulos et al, 2009; Hyndman et al, 2011) was to fit by OLS yielding reconciled forecasts:

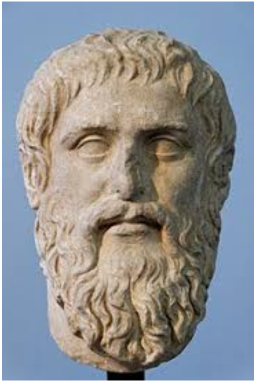
$$\tilde{\boldsymbol{y}} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}$$

Generalisation

- Wherever we can use OLS we can use GLS

$$\tilde{\boldsymbol{y}} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\hat{\boldsymbol{y}}$$

- Diagonal $\boldsymbol{W}$ considered by Athanasopoulos et al (2017)
- MinT approach (Wickremasuriya et al, 2018) use a $\boldsymbol{W}$ that is an estimate of the *in-sample* forecast error covariance matrix.



ΑΓΕΩΜΤΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ

Those without knowledge of geometry may not enter.

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 Point Forecast

Definition

A **coherent point forecast** is any forecast lying in the linear subspace $\mathfrak{s}$

Reconciled Point Forecast

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

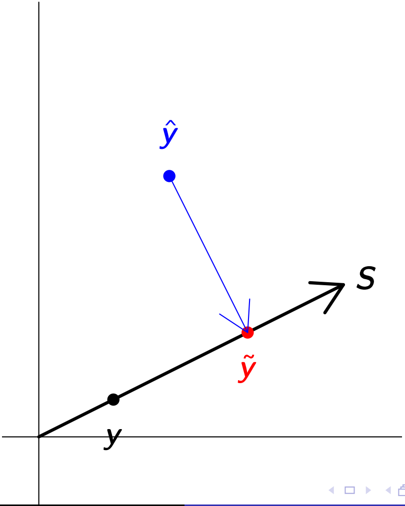
Definition

A **point forecast** $\tilde{\boldsymbol{y}}$ is reconciled with respect to $g(.)$ iff

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

when $g(.)$ is linear it is easier to write $\tilde{\boldsymbol{y}} = \boldsymbol{SG}\hat{\boldsymbol{y}}$

Geometry



Why reconciliation works

- The realised observation always lies on $\mathfrak{s}$.
- Orthogonal projections always get us 'closer' to all points in $\mathfrak{s}$ including the actual realisation.
- Ergo reconciliation reduces the error and not just in expectation.
- What about the MinT approach?

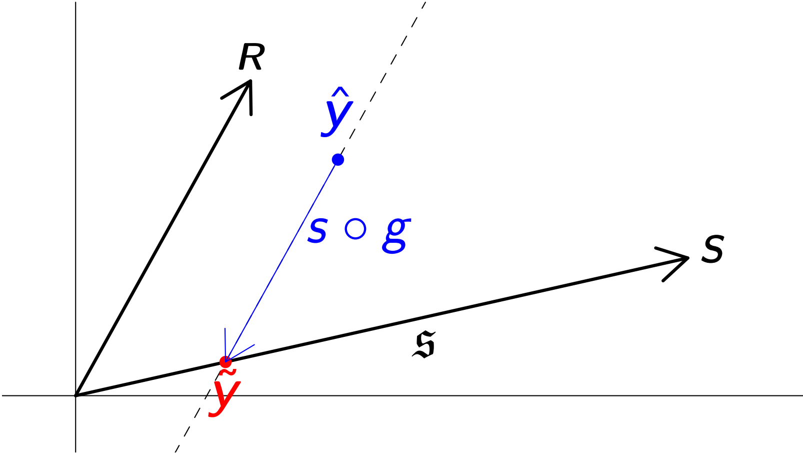
Finding a direction

- Consider the covariance matrix of $\boldsymbol{y} - \hat{\boldsymbol{y}}$.
- This can be estimated using in-sample forecast errors.
- This provides information about the likely direction of an error.
- Projecting along this direction is more likely to result in a reconciled forecast closer to the target.

Animation

content...

Geometry: Oblique Projection



Is this overkill?

- Is this geometric interpretation really necessary?
- When we generalise to probabilistic forecasts the regression interpretation does not really fit.
- Geometric ideas can easily be generalised.

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

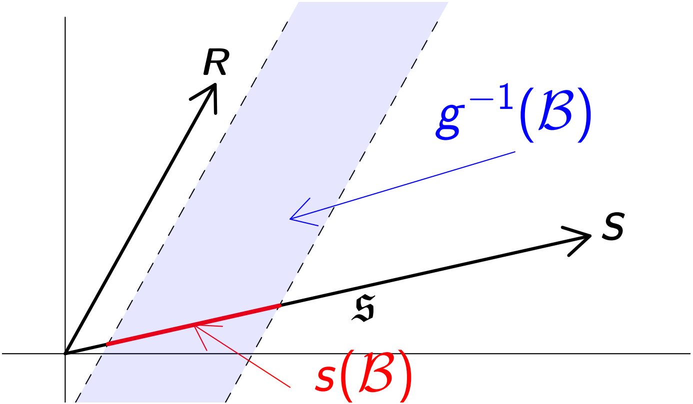
Definition

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

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

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

Geometry



Analytically

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

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

Elliptical distributions

Let the true predictive densities be elliptical.

Theorem

There exists a function $g(.)$ such that the true predictive distribution can be recovered by linear reconciliation as long as the unreconciled probabilistic forecast comes from the correct elliptical class.

This follows from the closure property of elliptical distributions under affine transformations and marginalisation. Conditions are also derived for when this $g(.)$ is also a projection.

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.

Coherent v Incoherent

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

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

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

- Using bottom level series only is a bad idea for 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

- 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.