Discussion* of “Cross-Temporal Probabilistic Forecast Reconciliation”

Presenter: Casey Lichtendahl (lichtendahlc@google.com)
Presented at: 2023 IIF Workshop on Forecast Reconciliation (Prato, Italy)
Date/time: 2023-09-08 (1:30-2:15pm)
Paper link: https://arxiv.org/abs/2303.17277

*The views expressed here are my own personal views and not those of my employer.

Agenda

Overview
(What does the paper do?)

Contributions
(Why should we care?)

Suggestions
(How can we improve the paper?)

Future directions
(What comes next?)

Overview

Problem

- “Cross-temporal probabilistic forecast reconciliation”
  - Beyond either “cross-sectional” or “temporal” reconciliation.
  - Beyond point forecast reconciliation.

Theory

- To sample from the coherent distribution, do the following:
  - First sample from the incoherent distribution.
  - Then reconcile the incoherent sample as usual by applying point forecast reconciliation.

Practical implementation

- Draw incoherent samples using base forecasts and bootstrap of concurrent one-step residuals.
- Use multi-step residuals to estimate covariance matrix of forecast errors.

Empirical studies

- Australian Quarterly National Accounts (QNA)
  - 95 base forecasts, 91 rolling origins, 4-steps-ahead forecasts.
  - Optimal cross-temporal approach wins on geometric mean of relative CRPS.
- Australian Tourism Demand
  - 525 base forecasts, 85 rolling origins, 12-steps-ahead forecasts.
  - Optimal cross-temporal approach wins on geometric mean of relative CRPS.

Contributions

Theory

- To sample from the coherent distribution, do the following:
  - First sample from the incoherent distribution.
  - Then reconcile the incoherent sample as usual by applying point forecast reconciliation.
  - **C1. Extension of previous cross-sectional work to the cross-temporal problem.**
  - **C2. Conveniently, there is no need to derive the reconciled distribution and to sample from it.**

Practical implementation

- Draw incoherent samples using base forecasts and bootstrap of concurrent one-step residuals.
  - **C3. Concurrency respects inherent dependency both cross-sectionally and temporally.**
- Use multi-step residuals to estimate covariance matrix of forecast errors.
  - **C4. Makes sense for the temporally dependencies at the higher frequencies.**

Empirical studies

- Australian Quarterly National Accounts (QNA) and Australian Tourism Demand
  - **C5. Optimal cross-temporal approach is best among many approaches on real data.**

Suggestions

Problem statement

- “Cross-temporal probabilistic forecast reconciliation”
  - **S1. Consider adding “sectional” to the name.**
    - Prior work gave it the name “cross-temporal”, but this reader finds that name a bit confusing.

    There are only two hard things in Computer Science: cache invalidation and naming things.
    -- Phil Karlton

Practical implementation

- Draw incoherent samples using base forecasts and bootstrap of concurrent one-step residuals.
  - **S2. Use multi-step residuals for the prediction intervals. (See also S10 on next slide.)**
- Use multi-step residuals to estimate covariance matrix of forecast errors.
  - **S3. Consider reporting on this approach in the Appendix only.**
    - Except on Tourism/ES, this approach does not perform as well.
    - Also, there may be too many approaches for the typical reader of the main text to comprehend
      - There are 60 approaches on Tourism.

Suggestions, cont.

$$\overline{\mathrm{RelCRPS}}_{j,s}^{[k]} = \left(\prod_{i=1}^{n} \frac{CRPS_{i,j,s}^{[k]}}{CRPS_{i,0,0}^{[k]}} \right)^{\frac{1}{n}}$$

Table 1. Example of mean relative performance

	Model 1	Model 2	Difference
Forecast 1	0.8000	0.8000	
Forecast 2	0.5000	0.6000	
Geometric mean	0.6325	0.6928	0.0604
Arithmetic mean	0.6500	0.7000	0.0500

Empirical studies

- Australian Quarterly National Accounts (QNA) and Australian Tourism Demand
  - **S4. Consider reporting arithmetic mean scores, in addition to geometric mean scores.**
    - The geometric mean accentuates differences. (See Table 1 above).
  - **S5. Consider using a scale-dependent score.**
    - RelCRPS (above) is scale-free and may underweight the economic consequences of top-level series relative to the bottom levels.
  - **S6. Report the pinball loss of an upper quantile, in addition to CRPS of all quantiles.**
    - Pinball loss is more aligned with economic consequences via the newsvendor problem (Jose and Winkler 2009).
  - **S7. Add a note about the mean rank test's confidence intervals.**
    - Rank confidence intervals may be too narrow because of positive serial correlation. (See S7 below.)
  - **S8. Add a Diebold-Mariano test on arithmetic mean differences in pinball loss.**
    - Accounts for serial correlation (Gneiting and Katzfuss 2014) and tests a more economically relevant hypothesis.
  - **S9. Consider reporting on some ensemble of approaches.**
  - **S10. Examine the coverage of multi-step-ahead forecasts.**
    - Prediction intervals may be too narrow because you sample one-step errors, rather than multi-step errors.
  - **S11. Visualize the cones from different approaches.**
  - **S12. Comment on aggregation in terms of averaging (vs. summing) in the temporal dimension.**

Future directions

Other loss functions

- Leading theory on reconciliation is based on minimizing mean squared error (MSE) and quadratic programming.
    - MSE encourages a good mean forecast.
    - Convenient because it leads to linear solution.
- Empirical studies, however, typically use metrics such as CRPS (a collection of pinball losses).
    - The mean pinball loss of the 0.50-quantile is MAE/2, which encourages a good median forecast.
    - Does this create some misalignment between the theory and empirics?
- **F1. Explore theory of reconciliation based on minimizing other loss functions.**
    - Could try something like the following formulation (similar to the now-standard quadratic program). It has an analytical (non-linear) solution (Owadally 2012, JCAM).

$$\min \quad \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \quad \text{s.t.} \quad \mathbf{B}\mathbf{x} = \mathbf{c}$$

Non-linear solution: $$\mathbf{x}^{*} = \boldsymbol{\rho} + \sqrt{\frac{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}}{\lambda^2 - \boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}}\,\boldsymbol{\tau}$$

where (i) $\boldsymbol{\rho} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c}$, *(ii)* $\boldsymbol{\tau} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu} - \mathbf{A}^{-1}\boldsymbol{\mu}$, *and (iii)* $\mathbf{U} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}$.

Future directions

Non-separable-error models

- The practice of reconciliation is built up around separable-error models: additive-error or multiplicative-error models. That is, either $\mathbf{y} = \mu + \varepsilon$ or $\mathbf{y} = \mu\,(1 + \varepsilon)$.
- Because of Poisson arrivals (e.g., at a call center), demand often follows something more like a normal variance-mean mixture. That is, $\mathbf{y} = \mu + \mu^{1/2}\,\varepsilon$.
- **F2. Explore reconciliation using base forecast models that do not have separable errors.**
  - Might use Bayesian inference or MAP on a multi-source-of-error state-space model.
    - Not clear how to work with non-separable-error models using exponential smoothing when there is a single source of error in the model.

Larger empirical studies

- Australian Tourism has only 525 series, and results based on these data may not generalize well.
- **F3. Examine larger hierarchical datasets (e.g., M5's 42,840 series).**

Final thoughts

Really nice paper

- Is on an important topic
- Has excellent notation
- Contains clear exposition
- Is well-grounded by the theory that it introduces
- Offers practically useful results

I recommend the following:

- Read this paper
- Use its results
- Cite it (because, among other things, there is a very promising young researcher on it)