Review Report: Forecast reconciliation: A geometric view with new insights on bias correction

November 18, 2019

Summary:

The authors provide a geometric interpretation for reconciliation of hierarchical forecasts. They show why and how reconciliation via projection is guaranteed to improve squared forecast errors. They explore a couple of different ways for dealing with biased base forecasts in an application to Australian tourism flows.

Overall the paper is well written and the geometric interpretations are an important contribution to the growing literature on forecast reconciliation. The authors do a very good job explaining the geometric aspects, which lead to new insights. That being said, the contribution of the paper in its current form is mainly theoretical, as the empirical evaluation is not much of a contribution. I do not think the paper meets the high standard set by the IJF in terms of empirical evaluation. I recommend that the authors revise their paper with a particular focus on strengthening its empirical contribution to more clearly show the practical value of the geometric insights they derive. I hope that the authors will find my comments useful for improving their paper.

Major comments:

1. **Lack of empirical contribution:** The authors apply a couple of different transformations to the tourism data before constructing base forecasts. Failing to correctly transform the base forecasts back before they are reconciled introduces a bias. That the highest accuracy is obtained when applying the correct back-transformation is a very small contribution. When the authors introduce the log-transformation and in the graphs in Figures 6 and 7, they do not clarify the interplay between the transformation and the reconciliation constraints. If $A + B = C$ then clearly $\log A + \log B \neq \log C$. In other words, a log-transformation affects the reconciliation constraints. This needs to be explained more clearly; e.g., is the reconciliation constraint imposed on the transformed or the raw data?

When the empirical evaluation is focused on bias, then it would make sense to include an error that measures bias in addition to the squared error. Moreover, simply showing the MSE without any confidence intervals or measures of significance is not sufficient for the reader to assess the results. It would also be useful to show the MSE relative to the base forecasts or the percentage improvement that is obtained. The best performing reconciliation method is MinT with shrinkage, but the authors never state the value of the shrinkage parameter or how it was chosen. Similarly, they compare with variance scaling without explaining what they mean by variance scaling.

In addition to the above mentioned shortcomings, I think the authors should reconsider their empirical evaluation. Maybe a second case study or a simulation study is needed to show the value of the geometric insights provided. We already know that MinT is better than OLS and WLS. What is the new and better reconciliation approach that has come from the geometric insights?

2. **Improvement guarantees:** The boxplot in Figure 8 shows that OLS always improves MSE, while this is not the case for the other reconciliation approaches. To gain a better understanding of the implications of Theorem 3.2, it would be useful to show that the other approaches always improve accuracy in their transformed spaces. What is the interpretation of the transformed spaces and can the authors make the connection between these spaces and the choice of reconciliation approach and error measure more clear? For example, Hyndman et al. (2011); van Erven and Cugliari (2015) argued for selecting OLS to increase the importance of forecasting the aggregate. What is the argument for WLS or MinT and what is the corresponding consistent error measure?

Minor comments:

1. In the first half of the paper it feels like every other sentence includes a *however*. I suggest reducing the use of however.

2. P. 2, l. 11: In several places the authors talk about adjusting forecasts *ex post*. Although I understand what is meant, it gives the impression that forecasts are adjusted after observing the realized values.

3. P. 2, l. 12: The authors discuss the regression formulation of forecast reconciliation. It would be useful to also make the connection to the optimization formulation considered by, e.g., van Erven and Cugliari (2015); Nystrup et al. (2020). This could also be useful for clarifying the connection between reconciliation approaches and error measures.

4. P. 4, l. 22: forf

5. P. 10, l. 12: the comma should not be there.

6. P. 11, Figure 3: usually a small square is drawn in the corner of the triangle to show orthogonality.
7. P. 17, l. 18: i,e.
8. P. 26, l. 14: the authors mention that the full results are available upon request. I suggest including them in an online supplementary appendix.
9. P. 27, Conclusions: The authors should comment on the implications of the non-uniqueness of the $\boldsymbol{S}$ matrix for future work on cross-temporal reconciliation (Kourentzes and Athanasopoulos, 2019).

References

R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9):2579–2589, 2011.

N. Kourentzes and G. Athanasopoulos. Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, 75:393–409, 2019.

P. Nystrup, E. Lindström, P. Pinson, and H. Madsen. Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, 280(3):876–888, 2020.

T. van Erven and J. Cugliari. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In A. Antoniadis, J.-M. Poggi, and X. Brossat, editors, *Modeling and Stochastic Learning for Forecasting in High Dimensions*, volume 217 of *Lecture Notes in Statistics*, pages 297–317. Springer: Cham, 2015.