Discussion: Forecast Reconciliation for Quantiles using Bilevel Optimisation

Talk Anastasios Panagiotelis

Discussion Lorenzo Zambon

IIF Workshop on Forecast Reconciliation

8th September 2023

Probabilistic forecast reconciliation

- $\widehat{\mathbf{y}}^{(1)}_{t+h|t}, \dots, \widehat{\mathbf{y}}^{(J)}_{t+h|t}$ (incoherent) samples from the base forecast distr
- Find map $\psi : \mathbb{R}^n \to \mathfrak{s}$, where $\mathfrak{s}$ is the coherent linear subspace $\widetilde{\mathbf{y}}^{(1)}, \dots, \widetilde{\mathbf{y}}^{(J)}$ are the reconciled samples, where

$$\widetilde{\mathbf{y}}^{(j)} := \psi(\widehat{\mathbf{y}}^{(j)}_{t+h|t})$$

- ψ is usually a linear mapping $\mathbf{SG}$, and $\mathbf{G}$ is optimized wrt some loss function L:

$$\mathbf{G}^{opt} = \underset{\mathbf{G}}{\operatorname{argmin}} \sum_{t \in \mathcal{T}_{train}} L\left(\mathbf{y}, \left(\widetilde{\mathbf{y}}^{(j)}\right)_j\right)$$

Panagiotelis, Gamakumara, Athanasopoulos, Hyndman (EJOR, 2023)

Quantile reconciliation

Quantiles of a coherent forecast distribution are **not** coherent! (Kolassa, IJF 2022)

Aim: coherent samples whose α-quantiles are optimal

Hence, pick the pinball loss function:

$$L_\alpha(y, q) := \alpha(y - q)I(y \geq q) + (1 - \alpha)(q - y)I(y < q)$$

The expected loss is minimized if q is the α-quantile
$\longrightarrow$ compute loss between actual value $y_{i,t}$ and sample quantile $\widetilde{q}_{i,t}$

Optimization

Problem:

$$\underset{\mathbf{G}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t \in \mathcal{T}_{train}} L_\alpha \left(y_{i,t}, \widetilde{q}_{i,t} \right)$$

where

$$\widetilde{q}_{i,t} = \underset{q}{\operatorname{argmin}} \sum_{j} L_\alpha \left(\widetilde{\mathbf{y}}_{i,t}^{(j)},\ q \right)$$

is the α-quantile of the reconciled samples

- Bi-level optimization
- Can be expressed as Mixed Integer Programming problem
- Use gradient descent to speed up computations:
  - smooth approximation of the pinball loss
  - trick to pass the gradient through the argmin

Questions

- Choice of β?
- Intuition about why base forecasts outperform OLS, WLS, MinT?
- Method is less robust for extreme quantiles?
  Regularization that depends on the chosen quantile?

Questions

$$\mathbf{G} \longrightarrow \widetilde{\mathbf{y}}_{i,t}^{(j)} \longrightarrow \widetilde{q}_{i,t} \longrightarrow L_\alpha$$

- $\widetilde{\mathbf{y}}_{i,t}^{(j)} = \mathbf{S}\mathbf{G}\widehat{\mathbf{y}}_{i,t}^{(j)}$
- $\widetilde{q} = \operatorname{argmin}_q \sum_j L_\alpha\big(\widetilde{\mathbf{y}}^{(j)},\, q\big)$
- $L_\alpha = L_\alpha\left(y_{i,t}, \widetilde{q}_{i,t}\right)$

Bi-level optimization arises from the definition of the sample quantile

Other definitions, e.g. using (a smooth version of) sorting?

Blondel et al., *Fast Differentiable Sorting and Ranking*, ICML 2020

Then: optimize using gradient descent (no bi-level optimization)

References

A. Panagiotelis, P. Gamakumara, G. Athanasopoulos, and R. J. Hyndman. Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. *European Journal of Operational Research*, 2023.

S. Kolassa. Do we want coherent hierarchical forecasts, or minimal MAPEs or MAEs? (we won't get both!). *International Journal of Forecasting*, 2022.

M. Blondel, O. Teboul, Q. Berthet, J. Djolonga. Fast Differentiable Sorting and Ranking. *Proceedings of the 37th International Conference on Machine Learning*, 2020.