Probabilistic Hierarchical Reconciliation via a Non-parametric Bootstrap Approach

Puwasala Gamakumara

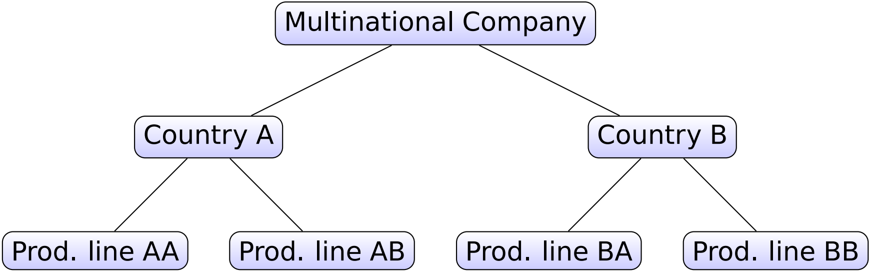
with
Anastasios Panagiotelis, George Athanasopoulos and Rob J. Hyndman

39th International Symposium on Forecasting

Introduction

Introduction

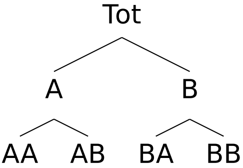
- **Example**: Forecasting the revenue of a large organisation



- **Hierarchical time series:** A collection of multiple time series that has an inherent aggregation structure
- Forecasts should add up. We call it *coherent*
- Why coherent forecasts?

Preliminaries

Notations



$$\boldsymbol{y}_t = [y_{Tot,t}, y_{A,t}, y_{B,t}, y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$\boldsymbol{b}_t = [y_{AA,t}, y_{AB,t}, y_{BA,t}, y_{BB,t}]^T$$

$$m = 4$$

$$n = 7$$

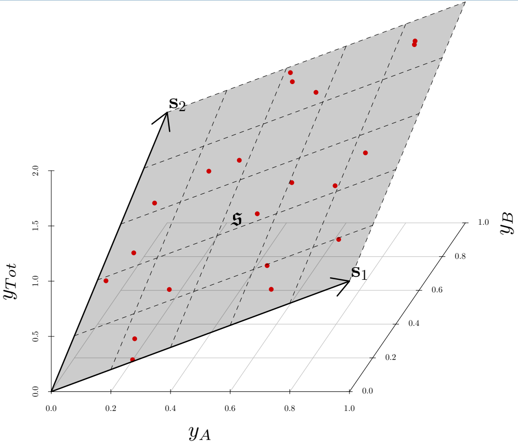
$$\boldsymbol{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Due to the aggregation nature of the hierarchy we have,

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

- *Coherent subspace:* $\mathfrak{s} = Span(\boldsymbol{S}), \quad \mathfrak{s} \subset \mathbb{R}^n$

Preliminaries: *Coherent forecasts*



- Three dimensional hierarchy, $y_{Tot} = y_A + y_B$.
- $\vec{s}_1 = (1, 1, 0)'$ and $\vec{s}_2 = (1, 0, 1)'$ form a basis for $\mathfrak{s}$.

Preliminaries: *Point forecasts reconciliation*

- Let $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ be an incoherent forecast. Then, the reconciled forecasts $\tilde{\boldsymbol{y}}_{T+h}$ are given by,

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h}$$

Method	$\boldsymbol{G}$
BU	$\left(\boldsymbol{0}_{m \times n-m} \ \boldsymbol{I}_{m \times m}\right)$
OLS	$\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'$
WLS	$\left(\boldsymbol{S}'\hat{\boldsymbol{W}}^{wls}_{T+1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}^{wls}_{T+1}$
MinT(Shrink)	$\left(\boldsymbol{S}'\hat{\boldsymbol{W}}^{shr}_{T+1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\hat{\boldsymbol{W}}^{shr}_{T+1}$

$$\begin{aligned} \hat{\boldsymbol{W}}^{shr}_{T+1} &= \tau \mathrm{Diag}(\hat{\boldsymbol{W}}^{sam}_{T+1}) + (1-\tau)\hat{\boldsymbol{W}}^{sam}_{T+1} \\ \hat{\boldsymbol{W}}^{wls}_{T+1} &= \mathrm{Diag}(\hat{\boldsymbol{W}}^{shr}_{T+1}) \end{aligned}$$

Probabilistic forecasts for hierarchical time series

Motivation

- Lack of attention in probabilistic forecasts
  - Ben Taieb et al. (2017)
  - Jeon, Panagiotelis, and Petropoulos (2018)

- Probabilistic forecasts should reflect the inherent properties of real data. In particular,
  - ⋆ Aggregation structure
  - ⋆ Correlation structure

- Extending the “reconciliation" method into probabilistic framework

Probabilistic forecast reconciliation

- Often parametric densities are unavailable but we can simulate a sample from the predictive distribution.
- Suppose $\hat{\boldsymbol{y}}_{T+h}^{[1]}, \ldots, \hat{\boldsymbol{y}}_{T+h}^{[J]}$ is a sample from the incoherent predictive distribution.
- Then setting $\tilde{\boldsymbol{y}}_{T+h}^{[j]} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_{T+h}^{[j]}$ produces a sample from the reconciled predictive distribution with respect to $\boldsymbol{G}$.

Non-parametric bootstrap approach

1. Fit univariate models at each node using data up to time T.
2. Let $\boldsymbol{\Gamma}_{(T\times n)} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_T)'$ be a matrix of in-sample residuals where $\boldsymbol{e}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$.
3. Let $\boldsymbol{\Gamma}^b_{(H\times n)} = (\boldsymbol{e}^b_1, ..., \boldsymbol{e}^b_H)'$ be a block bootstrap sample of size H from $\boldsymbol{\Gamma}$.
4. Generate h-step ahead sample paths from the fitted models incorporating $\boldsymbol{\Gamma}^b$. Denote these by $\hat{\boldsymbol{y}}^b_{T+h}$, for $h = 1, ..., H$.
5. Repeat step 3 and 4 for $b = 1, ..., B$ times. Denote these as $\hat{\boldsymbol{\Upsilon}}_{T+h} = (\hat{\boldsymbol{y}}^1_{T+h}, ..., \hat{\boldsymbol{y}}^B_{T+h})'$ for all h.
6. Setting $\tilde{\boldsymbol{\Upsilon}}_{T+h} = \boldsymbol{SG}\hat{\boldsymbol{\Upsilon}}'_{T+h}$ produces a sample from the reconciled distribution.

Optimal reconciliation of future paths

- We propose to find an optimal $\boldsymbol{G}_h$ matrix by minimizing Energy score.

$$\underset{\boldsymbol{G}_h}{\operatorname{argmin}} \quad \mathrm{E}_{\boldsymbol{Q}}[eS(\tilde{\boldsymbol{F}}, \boldsymbol{y}_{T+h})], \quad \tilde{\boldsymbol{F}} := \tilde{\Upsilon}_{T+h} = \boldsymbol{S}\boldsymbol{G}_h \hat{\Upsilon}'_{T+h}$$

where,

$$\mathrm{eS}(\tilde{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = \mathrm{E}_{\tilde{\boldsymbol{F}}} \|\tilde{\boldsymbol{Y}}_{T+h} - \boldsymbol{y}_{T+h}\|^{\alpha} - \frac{1}{2}\mathrm{E}_{\tilde{\boldsymbol{F}}} \|\tilde{\boldsymbol{Y}}_{T+h} - \tilde{\boldsymbol{Y}}^{*}_{T+h}\|^{\alpha},$$

$$\alpha \in (0, 2]$$

Optimal reconciliation of future paths

- Monte-Carlo approximation to the objective function is,

$$\underset{\boldsymbol{G}_h}{\operatorname{argmin}} \sum_{i=1}^{N} \Bigg\{ \frac{1}{B} \sum_{b=1}^{B} ||\boldsymbol{S}\boldsymbol{G}_h \hat{\boldsymbol{y}}_{T+h,i}^{b} - \boldsymbol{y}_{T+h}|| - \frac{1}{2(B-1)} \sum_{b=1}^{B-1} ||\boldsymbol{S}\boldsymbol{G}_h(\hat{\boldsymbol{y}}_{T+h,i}^{b} - \hat{\boldsymbol{y}}_{T+h,i,}^{b+1})|| \Bigg\}$$

Optimal reconciliation of future paths *Cont.*

- We impose the following structure to the $\boldsymbol{G}_h$ matrix

$$\boldsymbol{G}_h = \left(\boldsymbol{S}'\boldsymbol{W}_h\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_h \qquad (1)$$

- We propose four methods to optimise $\boldsymbol{G}_h$

Method 1: Optimising $\boldsymbol{W}_h$

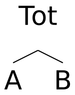
Method 2: Optimising Cholesky decomposition of $\boldsymbol{W}_h$
$\boldsymbol{W}_h = \boldsymbol{R_h'R_h}$ where $\boldsymbol{R}_h$ is an upper triangular matrix

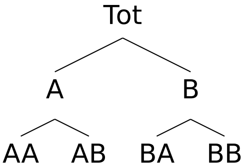
Method 3: Optimising Cholesky of $\boldsymbol{W}_h$ - restricted for scaling
$\boldsymbol{W}_h = \boldsymbol{R_h'R_h}$ s.t $\boldsymbol{i'W_hi} = 1$ where $\boldsymbol{i} = (1, 0, .., 0)'$

Method 4: Optimising $\boldsymbol{G}_h$ such that $\boldsymbol{G}_h\boldsymbol{S} = \boldsymbol{I}$

Monte-Carlo Simulation

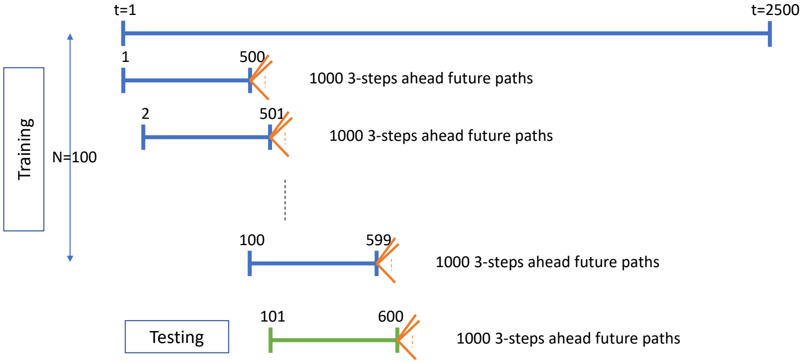
- Data generating process





- DGP was designed such that we have much noisier series in the bottom level.

Monte-Carlo Simulation



Monte-Carlo Simulation *Cont.*

Optimisation method	Hierarchy 1				Hierarchy 2			
	$h = 1$		$h = 3$		$h = 1$		$h = 3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Method 1	2.48	0.11	2.75	0.11	5.36	1.21	5.83	1.38
Method 2	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 3	2.48	0.11	2.75	0.11	5.37	1.21	5.83	1.37
Method 4	2.48	0.11	2.75	0.11	5.38	1.21	5.83	1.38

- Parameterisation does not matter

Monte-Carlo Simulation *Cont.*

- Comparison with point forecast reconciliation methods.

Reconciliation method	Hierarchy 1				Hierarchy 2			
	$h=1$		$h=3$		$h=1$		$h=3$	
	ES	VS	ES	VS	ES	VS	ES	VS
Optimal $\boldsymbol{G}$	2.48*	0.106	2.75*	0.106	5.36*	1.21*	5.83*	1.38*
MinT(Shrink)	2.47*	0.105	2.74*	0.105	5.33*	1.19*	5.77*	1.34*
WLS	2.46*	0.105	2.74*	0.105	5.43*	1.23	5.98*	1.40*
OLS	2.54*	0.105	2.80*	0.105	5.51*	1.23	5.98*	1.40*
Base	*2.67*	*0.105*	*2.94*	*0.105*	*5.71*	*1.28*	*6.27*	*1.49*

"*" indicates if the average score for a particular reconciliation method is significantly different from that of base forecasts.

- Reconciliation methods perform better than Base forecasts.

- MinT(Shrink) is at least as good as Optimal method. Thus going forward with MinT projection.

Forecasting Australian domestic tourism flows

Forecasting Australian domestic tourism flows

Geographical hierarchical structure for Australia

Level	**No.Series per level**
Total (Australia)	1
Level-1 (States)	7
Level-2 (Zones)	27
Level-3 (Regions)	75

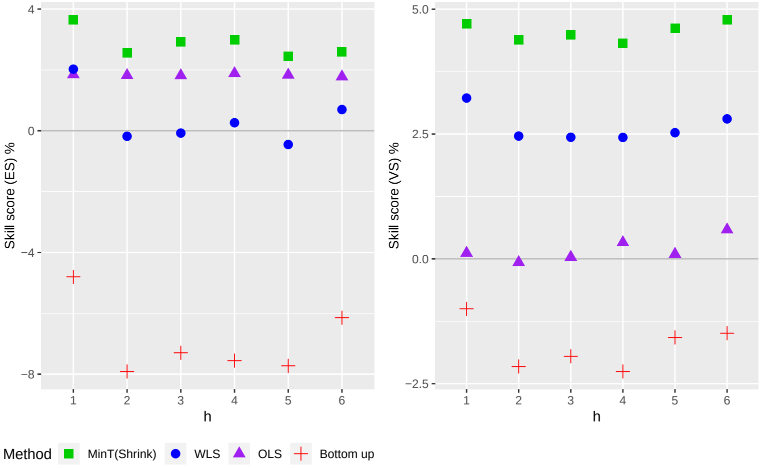
- Data: monthly “overnight trips" over the period January 1998 - December 2017
- Source of data: National Visitor Survey (NVS) [Tourism Research Australia]

Forecasting Australian domestic tourism flows

- **Analysis set up:**
  - First training sample is set from 1998:Jan to 2006:Apr
  - Univariate ARIMA models were fitted for each series in the hierarchy.
  - Reconciled probabilistic forecasts were produced for six months ahead (2006:May to 2006:Oct)
  - Then the training window is rolled by one month ahead at a time.
  - This leads to 140 1-step-ahead, 139 2-steps-ahead, through 135 6-step-ahead forecasts available for evaluation.

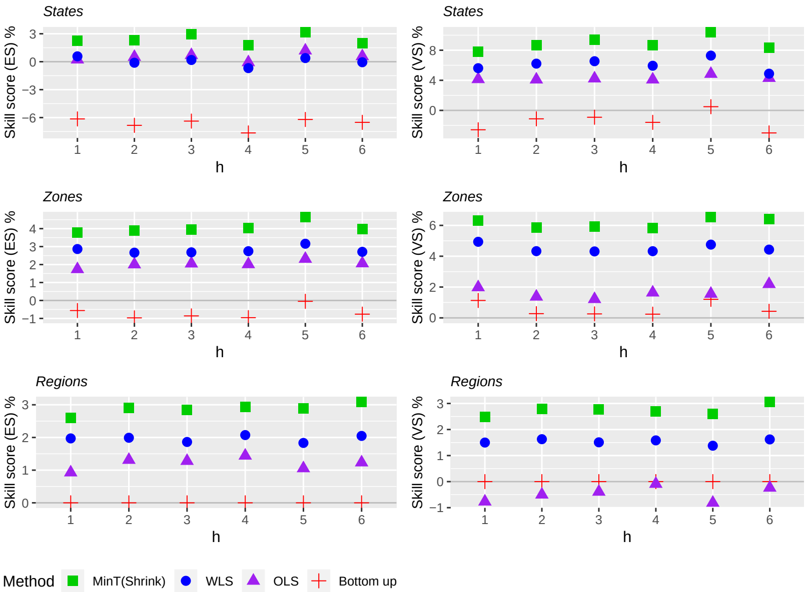
Results

Overall probabilistic forecast performance



Results

Probabilistic forecast performance for different levels



Conclusions

Conclusions

- We introduce a novel non-parametric bootstrap approach for producing reconciled probabilistic forecasts
- Simulation study evident that the optimal reconciliation with respect to energy score is equivalent to reconciling each sample path via MinT approach
- We apply this non-parametric bootstrap approach to obtain coherent probabilistic forecasts for domestic tourism flow in Australia

References

Ben Taieb, S., R. Huser, R. J. Hyndman, and M. G. Genton (2017). "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression". In: *IEEE Transactions on Smart Grid* 7.5, pp. 2448–2455.

Gneiting, T. and A. E. Raftery (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association* 102.477, pp. 359–378.

Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson (2008). "Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds". In: *Test* 17.2, pp. 211–235.

Jeon, J., A. Panagiotelis, and F. Petropoulos (2018). "Reconciliation of probabilistic forecasts with an application to wind power". In: *European Journal of Operational Research*. arXiv: `arXiv:1808.02635v1`.

References

Scheuerer, M. and T. M. Hamill (2015). “Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities *”. In: *Monthly Weather Review* 143.4, pp. 1321–1334.

THANK YOU!

Email: `puwasala.gamakumara@monash.edu`

Appendix

Appendix: *Probabilistic forecasts evaluation*

A2

Energy score (Gneiting et al., 2008)

$$\mathrm{eS}(\check{\boldsymbol{Y}}_{T+h}, \boldsymbol{y}_{T+h}) = \mathrm{E}_{\check{\boldsymbol{F}}}\|\check{\boldsymbol{Y}}_{T+h} - \boldsymbol{y}_{T+h}\|^{\alpha} - \frac{1}{2}\mathrm{E}_{\check{\boldsymbol{F}}}\|\check{\boldsymbol{Y}}_{T+h} - \check{\boldsymbol{Y}}^{*}_{T+h}\|^{\alpha}, \quad \alpha \in (0,2]$$

Log score (Gneiting and Raftery, 2007)

$$\mathrm{LS}(\check{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = -\log \check{\boldsymbol{f}}(\boldsymbol{y}_{T+h})$$

Variogram score (Scheuerer and Hamill, 2015)

$$\mathrm{VS}(\check{\boldsymbol{F}}, \boldsymbol{y}_{T+h}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\Big(|y_{T+h,i} - y_{T+h,j}|^{p} - \mathrm{E}_{\check{\boldsymbol{F}}}|\check{Y}_{T+h,i} - \check{Y}_{T+h,j}|^{p}\Big)^{2}$$

CRPS (Gneiting and Raftery, 2007)

$$\mathrm{CRPS}(\check{F}_i, y_{T+h,i}) = \mathrm{E}_{\check{F}_i}|\check{Y}_{T+h,i} - y_{T+h,i}| - \frac{1}{2}\mathrm{E}_{\check{F}_i}|\check{Y}_{T+h,i} - \check{Y}^{*}_{T+h,i}|$$

$\check{\boldsymbol{Y}}_{T+h}$ and $\check{\boldsymbol{Y}}^{*}_{T+h}$: Independent random vectors from the coherent forecast distribution $\check{\boldsymbol{F}}$.

$\boldsymbol{y}_{T+h}$: Vector of realizations.