Discrete Forecast Reconciliation

Bohan Zhang
Beihang University
The 43rd International Symposium on Forecasting

June 28, 2023

Collaborate with

Anastasios Panagiotelis
University of Sydney

Yanfei Kang
Beihang University

Feng Li
Central University of
Finance and Economics

CONTENTS

- Discrete-valued time series, especially those with low counts, commonly arise in various fields.
- Research attention paid on hierarchical forecasting for **discrete-valued hierarchical time series (HTS)** is limited.
- The **optimal combination reconciliation framework** was designed for continuous-valued HTSs and **can not be directly applied to** discrete-valued HTSs.
  - Support of forecasts should match the support of the variable (Freeland & McCabe 2004).
  - Transformation from continuous forecasts to integer decision introduces additional operational costs.

To address these concerns, we propose a discrete forecast reconciliation framework which

- first produces probabilistic forecasts for each series, then obtains coherent **joint predictive distribution** for the HTS through reconciliation,
- utilises scoring rules such as Brier Score to evaluate the forecasts and train the reconciliation matrix,
- allows for the employment of forecasting methods for univariate count time series in the literature.

A series of work on forecast reconciliation for count HTSs:

- Corani, G., Azzimonti, D., Rubattu, N., & Antonucci, A. (2022). Probabilistic Reconciliation of Count Time Series (arXiv:2207.09322). arXiv.
- Zambon, L., Azzimonti, D., & Corani, G. (2022). Efficient probabilistic reconciliation of forecasts for real-valued and count time series (arXiv:2210.02286). arXiv.
- Zambon, L., Agosto, A., Giudici, P., & Corani, G. (2023). Properties of the reconciled distributions for Gaussian and count forecasts (arXiv:2303.15135). arXiv.

The proposed framework conditions base probabilistic forecasts of the most disaggregated series on base forecasts of aggregated series. However, it fails to restore the dependence structure within hierarchical time series.

Coherent and incoherent domain for discrete HTS

HTS	$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$
basis time series	$(Y_1, Y_2, \ldots, Y_m)'$
domain of *i*-th variable	$\mathcal{D}(Y_i) = \{0, 1, \ldots, D_i\}$
Complete domain of HTS	$\mathcal{D}(\mathbf{Y}) = \{0, \ldots, D_1\} \times \cdots \times \{0, \ldots, D_n\}$
Coherent domain of HTS	$\tilde{\mathcal{D}}(\mathbf{Y}) = \{0, \ldots, D_1\} \times \cdots \times \{0, \ldots, D_m\}$
Incoherent domain of HTS	$\hat{\mathcal{D}}(\mathbf{Y}) = \mathcal{D}(\mathbf{Y}) - \tilde{\mathcal{D}}(\mathbf{Y})$

- We assume domains of time series are finite.
- Coherent(Incoherent) domain is the set of all *coherent*(*incoherent*) points.

Variables

$$\mathcal{D}(Y_1) = \{0,1\}, \mathcal{D}(Y_2) = \{0,1\},$$
$$Y_3 = Y_1 + Y_2, \mathcal{D}(Y_3) = \{0,1,2\}$$

Complete domain

$$\begin{aligned}\hat{\mathcal{D}}(\mathbf{Y}) = \{&(\mathbf{0},\mathbf{0},\mathbf{0})', (0,1,0)', (1,0,0)', (1,1,0)',\\ &(0,0,1)', (\mathbf{0},\mathbf{1},\mathbf{1})', (\mathbf{1},\mathbf{0},\mathbf{1})', (1,1,1)',\\ &(0,0,2)', (0,1,2)', (1,0,2)', (\mathbf{1},\mathbf{1},\mathbf{2})'\}\ ,\end{aligned}$$

Coherent domain

$$\tilde{\mathcal{D}}(\mathbf{Y}) = \{(0,0,0)', (0,1,1)', (1,0,1)', (1,1,2)'\}\ .$$

Definition (Coherent forecast)

A probabilistic forecast is said to be *coherent* if it only assigns positive probability to coherent points.

For example,

$$\hat{\boldsymbol{\pi}} = (\hat{\pi}_{(000)}, \hat{\pi}_{(010)}, \dots, \hat{\pi}_{(112)})' = (0.01, 0.02, \dots, 0.03)'$$
$$\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_{(000)}, \tilde{\pi}_{(011)}, \tilde{\pi}_{(101)}, \tilde{\pi}_{(112)})' = (0.2, 0.3, 0.4, 0.1)'$$

- Modelling and forecasting multivariate discrete-valued time series directly can be challenging.
- We construct the incoherent base forecasts by assuming the independence of the univariate forecasts.

$$\hat{\pi}_{(001)} = \hat{Pr}(Y_1 = 0) \times \hat{Pr}(Y_2 = 0) \times \hat{Pr}(Y_3 = 1)$$

The framework reconciles the base forecasts by proportionally **assigning the probability of each point in incoherent domain to points in the coherent domain**:

$$\tilde{\boldsymbol{\pi}} = \mathbf{A}\hat{\boldsymbol{\pi}},$$

where

- $A = [a_{ij}], i = 1, \ldots, r, j = 1, \ldots, q$ is an r $\times$ q reconciliation matrix with following constraints:

$$0 \leq a_{ij} \leq 1, \forall i, j$$
$$\sum_{i=1}^{r} a_{ij} = 1, \forall j$$

Movement restriction

- Probability of an incoherent point is proportionally assigned to **the cloest coherent points**, in spirit similar with the projection idea in the optimal combination reconciliation framework.
- We choose the L1 norm as the distance measure.

$$d((0,0,0),(0,0,1)) = |(0,0,0) - (0,0,1)|_1 = 1$$

Example

	000	010	100	110	001	011	101	111	002	012	102	112
000	1	0.4	0.3	0.25	0.4	0	0	0	0.25	0	0	0
011	0	0.6	0	0.25	0.3	1	0	0.3	0.25	0.35	0	0
101	0	0	0.7	0.25	0.3	0	1	0.3	0.25	0	0.4	0
112	0	0	0	0.25	0	0	0	0.4	0.25	0.65	0.6	1

We use Brier Score to evaluate the reconciled forecasts. Given reconciled forecasts $\tilde{\boldsymbol{\pi}}$ and real observation $\mathbf{Y}$, the Brier Score of the forecasts is

$$\mathrm{BS} = \sum_{k=1}^{r} (\tilde{\pi}_i - z_i)^2,$$

where $z_i = 1$ if $\mathbf{Y}$ takes the i-th coherent point, otherwise $z_i = 0$.

- We employ the rolling origin strategy to construct forecast-observation pairs, which are used to train the reconciliation matrix **A**.
- The objective function is

$$\min_{A} \frac{1}{\tau} \sum_{t=1}^{\tau} (\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)'(\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)$$

$$= \min_{a_{ij}} \frac{1}{\tau} \sum_{t=1}^{\tau} \left[\sum_{i=1}^{r} \left(\sum_{j=1}^{q} a_{ij}\hat{\pi}_{jt} - z_i^t \right)^2 \right]$$

$$s.t. \sum_{i=1}^{r} a_{ij} = 1, 0 \leq a_{ij} \leq 1$$

- This is a standard quadratic programming problem.

The DFR algorithm

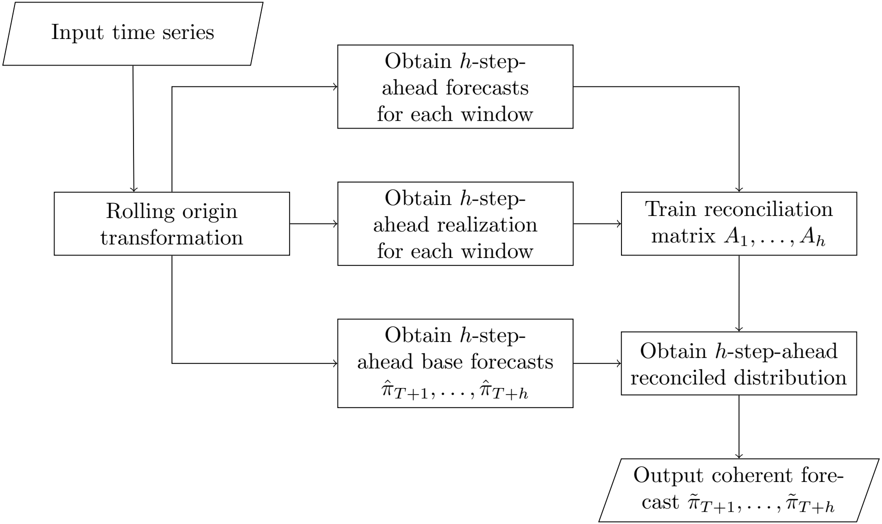
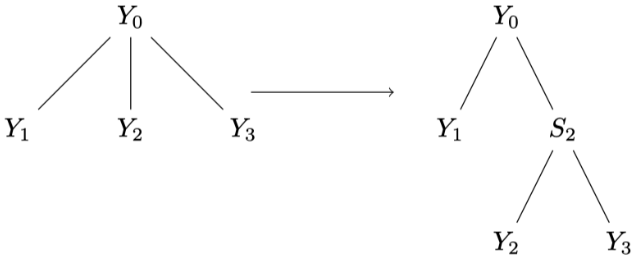


Figure: Flowchart of the DFR algorithm

Challenge of the DFR algorithm

The number of unknown parameters grows exponentially as the number of series in the hierarchy and the domain of bottom-level series grow.

Stepwise Discrete Forecast Reconciliaiton (SDFR)



1. Decompose the big hierarchy into multiple small sub-hierarchies.
2. Train the reconciliation model for each sub-hierarchy.
3. Combine the reconciled forecasts together under assumptions.

Stepwise Discrete Forecast Reconciliaiton (SDFR)

Algorithm 1: **S**tepwise **D**iscrete **F**orecast **R**econciliation (**SDFR**)

Input : $\hat{\pi}_0, \ldots, \hat{\pi}_k$

for $i = 1, \ldots, k-1$ **do**

  $\hat{\pi}_{\mathbf{S}_{k-i}} \leftarrow \texttt{BottomUp}\ (\hat{\pi}_{i+1}, \ldots, \hat{\pi}_k)$;

  if $i = \mathit{1}$ **then**

    $\hat{\pi}_{\mathbf{S}_{k-i+1}} \leftarrow \hat{\pi}_0$;

  else

    $\hat{\pi}_{\mathbf{S}_{k-i+1}} \leftarrow \sum_{\mathbf{S}_{k-i+2}, y_{i-1}} \tilde{\boldsymbol{\pi}}(\mathbf{S}_{k-i+2}, y_{i-1}, \mathbf{S}_{k-i+1})$;

  end

  $\tilde{\pi}(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i}) \leftarrow \texttt{DFR}_i\,(\hat{\pi}_{\mathbf{S}_{k-i+1}}, \hat{\pi}_i, \hat{\pi}_{\mathbf{S}_{k-i}})$

end

for $i = 2, \ldots, k-1$ **do**

  $\tilde{\pi}^1_{\mathbf{S}_{k-i+1}} \leftarrow \sum_{\mathbf{Y}_{i-1}} \tilde{\boldsymbol{\pi}}(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1})$;

  $\tilde{\pi}^2_{\mathbf{S}_{k-i+1}} \leftarrow \sum_{y_i, \mathbf{S}_{k-1}} \tilde{\boldsymbol{\pi}}(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i})$;

  $\tilde{\pi}'_{\mathbf{S}_{k-1+1}} \leftarrow \frac{1}{2}(\tilde{\pi}^1_{\mathbf{S}_{k-i+1}} + \tilde{\pi}^2_{\mathbf{S}_{k-i+1}})$;

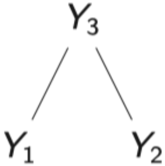
  $\tilde{\boldsymbol{\pi}}'(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1}) \leftarrow \texttt{Adjust}\ (\tilde{\boldsymbol{\pi}}(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1}), \tilde{\pi}'_{\mathbf{S}_{k-i+1}})$;

  $\tilde{\boldsymbol{\pi}}'(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i}) \leftarrow \texttt{Adjust}\ (\tilde{\boldsymbol{\pi}}(\mathbf{S}_{k-i+1}, \mathbf{S}_{k-i+1}), y_i, \tilde{\pi}'_{\mathbf{S}_{k-i+1}})$;

  $\tilde{\boldsymbol{\pi}}(\mathbf{Y}_i, \mathbf{S}_{k-i}) \leftarrow \texttt{ConstructJointDist}\ (\tilde{\boldsymbol{\pi}}'(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1}), \tilde{\boldsymbol{\pi}}'(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i}))$;

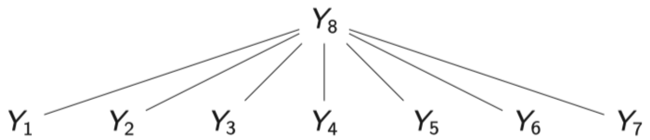
end

Output: $\tilde{\boldsymbol{\pi}}(\mathbf{Y}_k)$



	Base	DBU	DTD	DFR
Y_1	25.4	25.4	34.9	**24.4**
Y_2	27.8	27.8	34.8	**25.7**
Y_3	49.7	49.5	49.7	**42.0**
$\mathbf{Y}$	74.4	47.8	56.1	**44.0**

- $\mathcal{D}(Y_1) = \{0,1\}, \mathcal{D}(Y_2) = \{0,1\}, \mathcal{D}(Y_3) = \{0,1,2\}$
- Discrete Bottom-Up(DBU) and Discrete Top-Down(DTD) are discrete and probabilistic extensions of traditional bottom-up and top-down methods.
- Please refer to our paper for more details.



- We construct a weekly-daily temporal hierarchy in this simulation.
- $\mathcal{D}(Y_i) = \{0, 1\}, i = 1, \dots, 7.$
- SDFR is used in this simulation to handle the big hierarchy.
- Base probabilistic forecasts are produced using integer-valued GARCH models (see e.g.,Liboschik et al. 2017).

	Base	DBU	DTD	SDFR
Y_1	**40.8**	**40.8**	49.4	41.0
Y_2	**41.4**	**41.4**	49.6	41.6
Y_3	**42.1**	**42.1**	49.9	42.1
Y_4	43.0	43.0	50.0	**42.8**
Y_5	43.6	43.6	50.2	**43.1**
Y_6	44.0	44.0	50.3	**43.3**
Y_7	44.3	44.3	50.3	**43.9**
Y_8	**82.6**	83.5	**82.6**	83.1
$\mathbf{Y}$	99.5	97.8	99.4	**97.7**

Table: Summarised Brier Score($\times 10^{-2}$) of test samples in temporal setting.

- The dataset contains 231 weekly time series of offence crime numbers from 2014 to 2022; each time series corresponds to one census tracts in Washington D.C.
- We construct two-level temporal hierarchies (i.e., weekly and four-weekly) and forecast the offence numbers in the next four weeks for each time series.
- Samples whose forecast origin starts from 2022 are used for evaluation.
- Base probabilistic forecasts are produced using integer-valued GARCH models.
- DFR are used to reconcile the forecasts.

Forecasting crime number in Washington D.C.

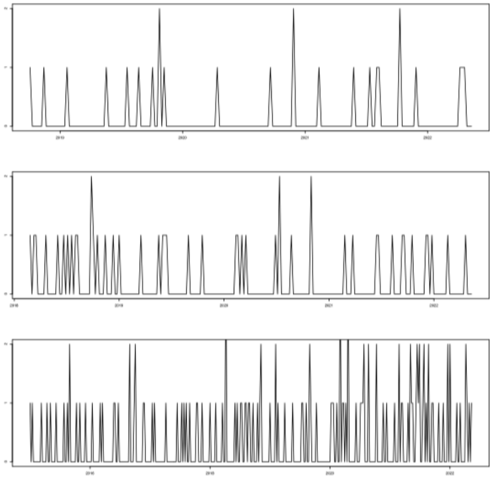


Figure: Example time series.

	Mean				Median			
	Base	DBU	DTD	DFR	Base	DBU	DTD	DFR
Total	58.47	**58.07**	58.47	58.12	66.64	65.28	66.64	**64.75**
Bottom	34.41	34.41	34.80	**34.30**	13.73	13.73	13.28	**10.82**
Hierarchy	73.87	**67.87**	68.33	67.97	97.66	92.70	93.08	**92.42**

Table: Summarised Brier Score($\times 10^{-2}$) of test samples in crime forecasting application.

Forecasting crime number in Washington D.C.

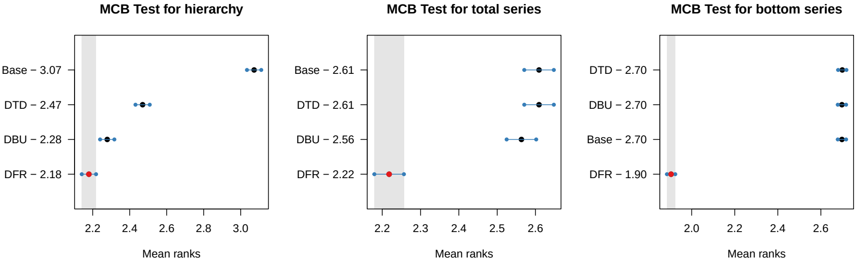


Figure: MCB test results

- We develop a novel forecast reconciliation framework for count hierarchical time series, which involves assigning probabilities from incoherent points to coherent points.
- We further propose a linear reconciliation algorithm that minimizes brier score of reconciled probabilistic forecasts.
- To address the exponential growth of the domain, we introduce a stepwise discrete reconciliation algorithm by breaking down a large hierarchy into smaller ones.
- Our DFR and SDFR algorithms produce coherent probabilistic forecasts and improve forecast accuracy, as shown in simulation and empirical studies.

Thank you!

Any questions/suggestions/comments?

Paper: `https://arxiv.org/abs/2305.18809`

Code: `https://github.com/AngelPone/DiscreteRecon`

Freeland, R. K. & McCabe, B. P. M. (2004), 'Forecasting discrete valued low count time series', *International Journal of Forecasting* **20**(3), 427–434.

Liboschik, T., Fokianos, K. & Fried, R. (2017), 'Tscount: An R Package for analysis of count time series following generalized linear models', *Journal of Statistical Software, Articles* **82**(5), 1–51.