Discrete Forecast Reconciliation

Bohan Zhang
Beihang University
IIF Workshop on Forecast Reconciliation
September 7, 2023

Collaborate with

Anastasios Panagiotelis
University of Sydney

Yanfei Kang
Beihang University

Feng Li
Central University of
Finance and Economics

CONTENTS

- Non-negative and discrete-valued time series, particularly those with low counts, commonly arise in various fields. Examples include:
  - occurrences of “black swan” events
  - intermittent demand in the retail industry
- Despite the great concern of hierarchical forecasting in these applications, limited research have been conducted.

The forecast reconciliation approach

- first produces base forecasts for each series in the hierarchy; then optimally reconciles the base forecasts through projection;
- utilises forecast combination, which improves forecast accuracy and reduces the risk of model misspecification;
- has been shown to improve forecast accuracy in various applications.

But it was designed for continuous-valued HTS and can not be directly applied to discrete-valued HTS: *projection* may produce non-integer and negative forecasts.

- While point and interval forecasts are most widely applied in practice, attention has been shifted towards full predictive distribution.
- When forecasting discrete-valued time series, it is also more natural to produce predictive distribution.

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.

To address these concerns, we

- introduce the definition of *coherent domain and coherent forecasts* in the context of multivariate discrete random variables.
- propose a discrete forecast reconciliation framework.
- develop the DFR and Stepwise DFR (SDFR) algorithms to train the reconciliation matrix.
- extend the top-down and bottom-up method to discrete probabilistic setting for comparison.
- verify the applicability of the algorithms in simulation experiments and real-world applications.

HTS	$\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$
basis (e.g., bottom-level) time series	$\mathbf{Y}_b = (Y_1, Y_2, \dots, Y_m)'$
domain of *i*-th variable	$\mathcal{D}(Y_i) = \{0, 1, \dots, D_i\}$

Complete domain of $\mathbf{Y}$ is the Cartesian product of domains of all variables.

$$\hat{\mathcal{D}}(\mathbf{Y}) = \{0, \dots, D_1\} \times \cdots \times \{0, \dots, D_n\} \quad q := |\hat{\mathcal{D}}(\mathbf{Y})|$$

Coherent domain of $\mathbf{Y}$ is a subset of $\hat{\mathcal{D}}(\mathbf{Y})$, in which every point respects the aggregation constraints.

$$\tilde{\mathcal{D}}(\mathbf{Y}) = \{\mathbf{y} | \mathbf{y} \in \hat{\mathcal{D}}(\mathbf{Y}), S\mathbf{y}_b = \mathbf{y}\} \quad r := |\tilde{\mathcal{D}}(\mathbf{Y})|$$

Incoherent domain of $\mathbf{Y}$

$$\bar{\mathcal{D}}(\mathbf{Y}) = \hat{\mathcal{D}}(\mathbf{Y}) \backslash \tilde{\mathcal{D}}(\mathbf{Y})$$

Variables

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

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 (Discrete Coherence)

A coherent discrete distribution has the property $Pr(\mathbf{Y} = \mathbf{y}) = 0, \forall y \in \bar{\mathcal{D}}(\mathbf{Y})$. Any distribution not meeting this condition is an incoherent distribution.

- We use a probability vector to represent the discrete predictive distribution.
- Denote (potentially) incoherent base forecasts by $\hat{\boldsymbol{\pi}}$ and reconciled forecasts by $\tilde{\boldsymbol{\pi}}$.

$$\hat{\boldsymbol{\pi}} := [\hat{\pi}_1, \hat{\pi}_2 \ldots, \hat{\pi}_q]' := [\hat{\pi}_{(y_1,\ldots,y_n)^{(1)}}, \ldots, \hat{\pi}_{(y_1,\ldots,y_n)^{(q)}}]$$
$$\tilde{\boldsymbol{\pi}} := [\tilde{\pi}_1, \tilde{\pi}_2 \ldots, \tilde{\pi}_r]' := [\tilde{\pi}_{(y_1,\ldots,y_n)^{(1)}}, \ldots, \tilde{\pi}_{(y_1,\ldots,y_n)^{(r)}}]$$

Base forecast

$$\hat{\boldsymbol{\pi}} = [\hat{\pi}_1, \hat{\pi}_2 \ldots, \hat{\pi}_{12}]' = [\hat{\pi}_{(001)}, \hat{\pi}_{(011)} \ldots, \hat{\pi}_{(112)}]'$$
$$= [0.01, 0.02, \ldots, 0.03]'$$

Reconciled forecast

$$\tilde{\boldsymbol{\pi}} = [\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4]' = [\tilde{\pi}_{(000)}, \tilde{\pi}_{(011)}, \tilde{\pi}_{(101)} \tilde{\pi}_{(112)}]'$$
$$= [0.2, 0.3, 0.40.1]'$$

$$\tilde{\boldsymbol{\pi}} = \psi(\hat{\boldsymbol{\pi}}) \quad \psi : [0, 1]^q \rightarrow [0, 1]^r$$

The linear reconciliation function

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

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

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

The framework reconciles the base forecasts by proportionally assigning the probability of each point in complete domain to points in the coherent domain.

Example

	000	010	100	110	001	011	101	111	002	012	102	112
000	0	0.4	0.3	0.25	0	0	0	0	0.3	0	1	0.2
011	0	0.4	0.3	0.25	0.2	0	0	1	0.3	1	0	0.3
101	0	0	0.3	0.25	0.4	0	1	0	0.3	0	0	0.5
112	1	0.2	0.1	0.25	0.4	1	0	0	0.1	0	0	0

- The framework allows the probability of a point in the complete domain assigned to any point in the coherent domain.
- For example, in an extreme case, from a coherent point (000) to another coherent point (112).

Movement restriction strategy requires that the probability is only assigned to **the closest coherent points**.

- This is similar to the projection idea in the optimal combination reconciliation framework.
- A coherent point in the complete domain moves all of its probability to the same point in the coherent domain.
- 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	0.3	0	0	0
011	0	0.6	0	0.25	0.3	1	0	0.3	0.3	0.35	0	0
101	0	0	0.7	0.25	0.3	0	1	0.3	0.3	0	0.4	0
112	0	0	0	0.25	0	0	0	0.4	0.1	0.65	0.6	1

- Scoring rules assess the quality of probabilistic forecasts, by assigning a numerical score based on the predictive distribution and on the corresponding observation.
- Brier Score can be used to evaluate the probabilistic forecasts of discrete variables.

$$\text{BS} = \sum_{k=1}^{r} (\tilde{\pi}_k - z_k)^2,$$

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

By minimising the average Brier Score of τ reconciled forecasts, we can find the optimal reconciliation matrix.

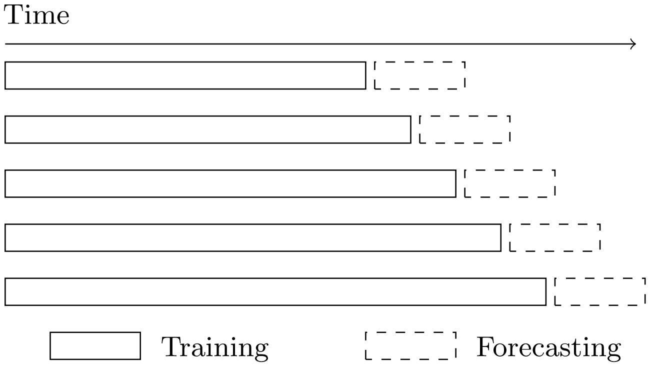
$$\min_{A} \frac{1}{\tau} \sum_{t=1}^{\tau} (\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)'(\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)$$

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

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

This is a standard quadratic programming problem.

We employ the expanding window strategy to construct the τ training samples.



We construct the incoherent base forecasts $\hat{\boldsymbol{\pi}}$ by assuming the independence of univariate base forecasts.

1. Generate predictive distributions for each time series in the hierarchy using arbitrary univariate forecasting model.
2. Construct the joint distribution by assuming independence.

Example

$\hat{\boldsymbol{\pi}}_{Y_1} = [0.4, 0.6]' \quad \hat{\boldsymbol{\pi}}_{Y_2} = [0.3, 0.7]' \quad \hat{\boldsymbol{\pi}}_{Y_3} = [0.2, 0.2, 0.6]'$
$\hat{\pi}_{(001)} = Pr(Y_1 = 0) \times Pr(Y_2 = 0) \times Pr(Y_3 = 1) = 0.024$

$$\begin{aligned}\hat{\boldsymbol{\pi}} = [&0.024, 0.056, 0.036, 0.084,\\ &0.024, 0.056, 0.036, 0.084,\\ &0.072, 0.168, 0.108, 0.252]'\end{aligned}$$

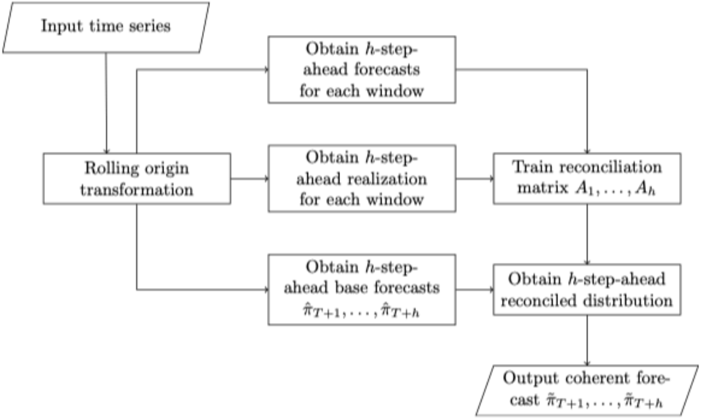
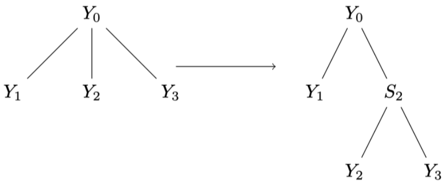
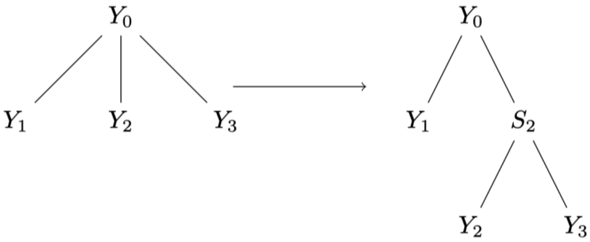


Figure: Flowchart of the DFR algorithm

- The number of unknown parameters in **A** grows exponentially as the number of time series and the cardinality of domains of bottom-level series grow.
- We propose the Stepwise DFR (SDFR) algorithm to deal with this problem.



- It reduces the number of unknown parameters from exponential level to cubic level.



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.

$$P(Y_0, Y_1, Y_2, Y_3) = P(Y_0, Y_1, S_2)P(Y_2, Y_3|S_2)$$

Stepwise DFR (SDFR): handling high-dimensional hierarchy

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**
  $\hat{\pi}_{\mathbf{S}_{k-i}} \leftarrow$ `BottomUp` $(\hat{\pi}_{i+1}, \ldots, \hat{\pi}_k)$;
  if $i = 1$ **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
  $\tilde{\pi}(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i}) \leftarrow$ `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**
  $\tilde{\pi}^1_{\mathbf{S}_{k-i+1}} \leftarrow \sum_{\mathbf{Y}_{i-1}} \tilde{\boldsymbol{\pi}}(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1})$;
  $\tilde{\pi}^2_{\mathbf{S}_{k-i+1}} \leftarrow \sum_{y_i, \mathbf{S}_{k-1}} \tilde{\boldsymbol{\pi}}(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i})$;
  $\tilde{\pi}'_{\mathbf{S}_{k-1+1}} \leftarrow \frac{1}{2}(\tilde{\pi}^1_{\mathbf{S}_{k-i+1}} + \tilde{\pi}^2_{\mathbf{S}_{k-i+1}})$;
  $\tilde{\boldsymbol{\pi}}'(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1}) \leftarrow$ `Adjust` $(\tilde{\boldsymbol{\pi}}(\mathbf{Y}_{i-1}, \mathbf{S}_{k-i+1}), \tilde{\pi}'_{\mathbf{S}_{k-i+1}})$;
  $\tilde{\boldsymbol{\pi}}'(\mathbf{S}_{k-i+1}, y_i, \mathbf{S}_{k-i}) \leftarrow$ `Adjust` $(\tilde{\boldsymbol{\pi}}(\mathbf{S}_{k-i+1}, \mathbf{S}_{k-i+1}), y_i, \tilde{\pi}'_{\mathbf{S}_{k-i+1}})$;
  $\tilde{\boldsymbol{\pi}}(\mathbf{Y}_i, \mathbf{S}_{k-i}) \leftarrow$ `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: $\tilde{\boldsymbol{\pi}}(\mathbf{Y}_k)$

- The discrete bottom-up method constructs a coherent distribution by assuming independent bottom-level forecasts.
- This method follows the same procedure as constructing base forecasts explained earlier except that the base forecasts of aggregated series are excluded.
- The mean point forecasts obtained from this coherent distribution's marginal distribution are identical to those obtained by directly aggregating mean forecasts of bottom-level series.

Example

$$\hat{\boldsymbol{\pi}}_{Y_1} = [0.4, 0.6]' \quad \hat{\boldsymbol{\pi}}_{Y_2} = [0.3, 0.7]'$$
$$\tilde{\pi}_{(000)} = Pr(Y_1 = 0) \times Pr(Y_2 = 0) = 0.12$$
$$\tilde{\boldsymbol{\pi}} = [0.12, 0.28, 0.18, 0.42]'$$

The discrete top-down method extends the traditional top-down by proportionally disaggregating the probabilities of each point of the total series into all possible coherent points, using a ratio computed from historical occurrences.

Example

40 $(1, 0, 1)$ and 60 $(0, 1, 1)$ observed in the history.

$$\hat{\boldsymbol{\pi}}_{Y_3} = [0.2, 0.3, 0.5]'$$

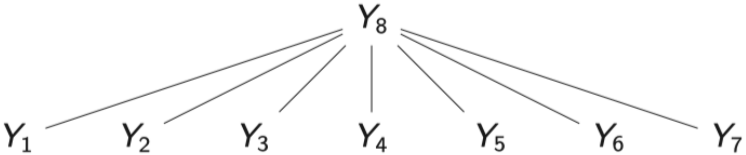
$$\tilde{\pi}_{(011)} = Pr(Y_3 = 1) \times \frac{60}{60 + 40} = 0.18$$

$$\tilde{\boldsymbol{\pi}} = [0.2, 0.18, 0.12, 0.5]'$$

- $\mathcal{D}(Y_1) = \{0, 1\}, \mathcal{D}(Y_2) = \{0, 1\}, Y_3 = Y_1 + Y_2$
- We produce and evaluate one-step-ahead forecast in this experiment.
- For each binary series, we generate 480 observations; expanding window strategy yields 330 samples for training and 30 samples for testing.
- The performances for each time series are evaluated based on the average Brier scores of test samples.
- The base probabilistic forecasts are obtained using the binomial AR(1) model.
- The procedure was repeated 1000 times.

Table: Average Brier score ($\times 10^{-2}$) of 1000 simulations for cross-sectional setting

	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**



- We construct a weekly-daily temporal hierarchy in this simulation.
- $\mathcal{D}(Y_i) = \{0, 1\}, i = 1, ..., 7.$
- SDFR is used in this simulation to handle the big hierarchy.

Table: Average Brier score ($\times 10^{-2}$) of 1000 samples for temporal setting.

	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**

- 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.

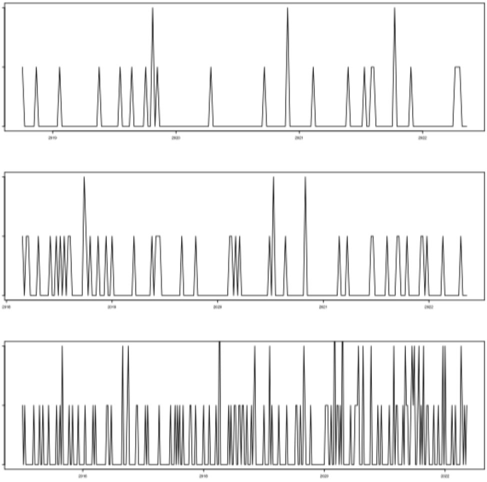


Figure: Example time series

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

	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**

Forecasting crime number in Washington D.C.

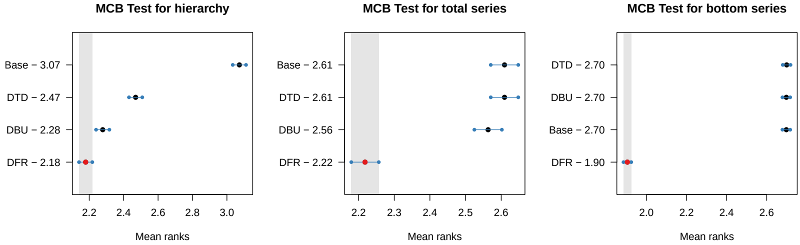


Figure: MCB test results

- 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 minimises 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`
Package: `https://github.com/AngelPone/DiscreteRecon`