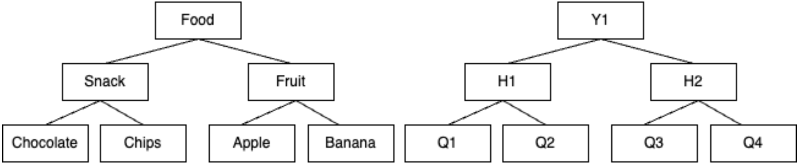
Forecast reconciliation: applications in complex scenarios

Bohan Zhang
Beihang University
NUMBATs Seminar
August 31, 2023

Hierarchical time series refers to multivariate time series that adhere to specific linear constraints, with aggregation constraints being the most common. For example,



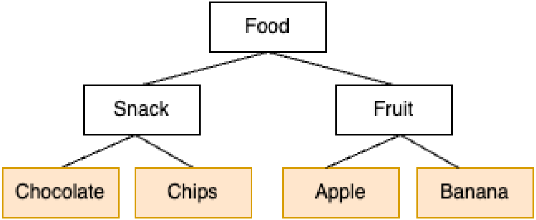
- In a retail network, product time series can be grouped by categories, geographical regions, etc. (Cross-sectional hierarachy, left figure)
- A quartly time series can be aggregated into half-yearly and yearly time series. (Temporal hierarchy, right figure)

The objective of hierarchical forecasting is to produce *coherent* forecasts for hierarchical time series.

- *Coherence* means that multivariate forecasts should respect the linear constraints.
- Coherent forecasts support aligned decisions across different levels and across different forecast horizons in a large forecasting system.

Single-level methods first produce forecasts at one level, then obtain forecasts at other levels through aggregation or/and disaggregation.

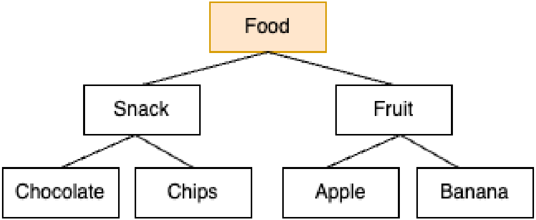
- Bottom-up
- Top-down
- Middle-out



The bottom-level series can be highly volatile and challenging to forecast, leading to poorly performing aggregated forecasts.

Single-level methods first produce forecasts at one level, then obtain forecasts at other levels through aggregation or/and disaggregation.

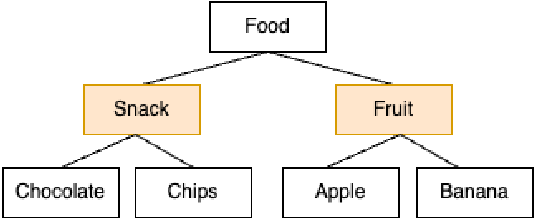
- Bottom-up
- Top-down
- Middle-out



- Top-down method ignores useful information from other levels.
- Top-down method does not produce unbiased coherent forecasts even if the base forecasts are unbiased.

Single-level methods first produce forecasts at one level, then obtain forecasts at other levels through aggregation or/and disaggregation.

- Bottom-up
- Top-down
- Middle-out



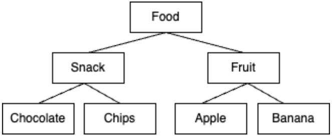
The performance highly depends on the selection of level.

The optimal combination approach, aka *forecast reconciliation*, first produces *base forecasts* for each time series in the hierarchy, then optimally reconciles them to be coherent.

- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis, 55(9), 2579–2589.
- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. European Journal of Operational Research, 262(1), 60–74.
- Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. Journal of the American Statistical Association, 114(526), 804–819.
- Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. International Journal of Forecasting, 37(1), 343–359.

Suppose there are n time series in the hierarachy. m of them are *basis*(bottom-level) series.

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{I}_m \end{bmatrix} \boldsymbol{b} \Longleftrightarrow \boldsymbol{U}'\boldsymbol{y} = \begin{bmatrix} \boldsymbol{I}_{n-m} & -\boldsymbol{C} \end{bmatrix} \boldsymbol{y} = \boldsymbol{0}$$



$$\begin{bmatrix} y_{\text{Food}} \\ y_{\text{Snack}} \\ y_{\text{Fruit}} \\ y_{\text{Chocolate}} \\ y_{\text{Chips}} \\ y_{\text{Apple}} \\ y_{\text{Banana}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{\text{Chocolate}} \\ y_{\text{Chips}} \\ y_{\text{Apple}} \\ y_{\text{Banana}} \end{bmatrix}$$

See Section 3 of Di Fonzo & Girolimetto (2021) for details of *zero constraints kernel representation*.

The S matrix is not unique.

$$\begin{bmatrix} y_{\text{Food}} \\ y_{\text{Snack}} \\ y_{\text{Fruit}} \\ y_{\text{Chocolate}} \\ y_{\text{Chips}} \\ y_{\text{Apple}} \\ y_{\text{Banana}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{\text{Food}} \\ y_{\text{Chips}} \\ y_{\text{Apple}} \\ y_{\text{Banana}} \end{bmatrix}$$

Reconciliation approach finds $\tilde{\boldsymbol{y}}$ such that

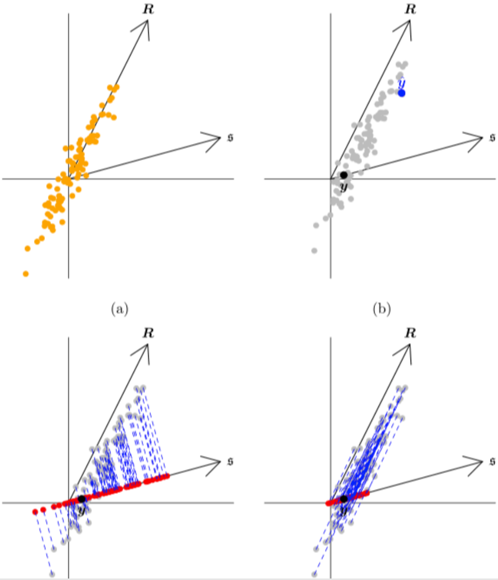
$$\min_{\tilde{\boldsymbol{y}}} \frac{1}{2}(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})' \boldsymbol{W}^{-1}(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$$
$$\text{s.t.} \boldsymbol{U}' \tilde{\boldsymbol{y}} = 0$$

It leads to

$$\tilde{\boldsymbol{y}} = [\boldsymbol{I} - \boldsymbol{W}\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}\boldsymbol{U})^{-1}\boldsymbol{U}']\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}},$$

where $\boldsymbol{G} = \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}\boldsymbol{U})^{-1}\boldsymbol{U}'$, $\boldsymbol{J} = [\boldsymbol{0}_{m\times(n-m)} \quad \boldsymbol{I}_m]$.
Equivalently (Wickramasuriya et al. 2019),

$$\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$$


R
s
y
(a)
(b)

- $\boldsymbol{W}$ is the variance-covariance matrix of the base forecast error (Wickramasuriya et al. 2019, Wickramasuriya 2021).
- Usually it is estimated using the in-sample residuals.
- Several common alternatives include:

OLS	$\boldsymbol{I}_n$
Structural scaling (WLSs)	$\boldsymbol{S}\mathbf{1}$
Variance scaling (WLSv)	$\text{diag}(\boldsymbol{R})$
MinT(Shrinkage)	$\lambda\text{diag}(\boldsymbol{R}) + (1-\lambda)\boldsymbol{R}$

where $\boldsymbol{R} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}_t'$, λ is the shrinkage intensity parameter.

- The intuition behind $\boldsymbol{W}$: bigger residual variance, bigger adjustments.

$$(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})'\boldsymbol{W}^{-1}(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$$

- Forecast reconciliation combines information from all levels.
- It utilises forecast combination, which reduces the risk of model misspecification, mitigates forecast uncertainty and improves forecast accuracy.

But ...

- It generally modifies forecasts of every time series in the hierarachy, which might be unexpected in practice.
- The performances depend on the choices of W.
  - The simple alternatives such as OLS and structural scaling performs very poorly in a large diverse hierarachy.
  - The estimation of complex $\boldsymbol{W}$ can be infeasible, e.g., unable to estimate a covariance matrix and MinT(Shrinkage) when length of time series are unequal.

The sales time series hierarchy from a Chinese online retailer:

- top level: category food
- middle level: 40 subcategories
- bottom level: 1905 items

Our objective:

- Forecast sales in 8 promotion periods (e.g., 11.11).

The challenges:

- 80% of the bottom-level series are intermittent time series.
- Bottom-level series have unequal historical lengths. Thus, we cannot use the MinT(Shrinkage) estimator.
- Sales pattern during the promotion period differ significantly from regular periods. Thus, we use the residuals during promotion periods to estimate variance.

The results using the reconciliation method:

Level	Base	OLS		WLS_s		WLS_v	
		U	U+NN	U	U+NN	U	U+NN
Top	2.94	2.93	2.92	2.72	2.72	2.75	2.77
Middle	2.66	272.83	48.84	16.09	6.50	2.39	2.40
Bottom-1	2.04	3.98	2.70	2.96	2.32	1.86	1.83
Bottom-2	0.11	42.66	15.43	26.99	8.34	1.52	1.52
Bottom-3	1.08	1.64	1.48	1.36	1.25	1.58	1.19

We propose a variant to the reconciliation framework that keeps forecasts of a pre-specified subset of variables unchanged or "immutable".

It can be formulated into following objective:

$$\min_{\tilde{\boldsymbol{y}}} \frac{1}{2}(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})' \boldsymbol{W}^{-1}(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$$

$$\text{s.t.} \quad \check{\boldsymbol{U}}' \tilde{\boldsymbol{y}} = \boldsymbol{d}$$

where $\check{\boldsymbol{U}}' = \begin{bmatrix} \boldsymbol{I}_n^{[\mathcal{A}]} \\ \boldsymbol{U}' \end{bmatrix}$ and $\boldsymbol{d} = \begin{bmatrix} \hat{\boldsymbol{y}}^{[\mathcal{A}]} \\ \boldsymbol{0} \end{bmatrix}$. $\mathcal{A} \subset \{1, 2, \ldots, n\}$ is the set of indices of immutable series. It leads to

$$\tilde{\boldsymbol{y}} = \hat{\boldsymbol{y}} - \boldsymbol{W}\check{\boldsymbol{U}}(\check{\boldsymbol{U}}\boldsymbol{W}\check{\boldsymbol{U}})^{-1}\left[\check{\boldsymbol{U}}\hat{\boldsymbol{y}} - \boldsymbol{d}\right]$$

Sometimes it is necessary to make sure the reconciled forecasts to be nonnegative. The following formulation combines immutablility and nonnegativity.

$$\min_{\tilde{\boldsymbol{y}}} \frac{1}{2}(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})' \boldsymbol{W}^{-1} (\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$$

$$\text{s.t.} \quad \check{\boldsymbol{U}}' \tilde{\boldsymbol{y}} = \boldsymbol{d}$$

$$\tilde{\boldsymbol{y}} \geq 0$$

- We keep base forecasts of the top level, bottom-2 and bottom-3 level immutable.

Level	Base	OLS				WLS_s				WLS_v			
		C	C+NN	U	U+NN	C	C+NN	U	U+NN	C	C+NN	U	U+NN
Top	2.94	2.94	2.94	2.93	2.92	2.94	2.94	2.72	2.72	2.94	2.94	2.75	2.77
Middle	2.66	9.31	4.94	272.83	48.84	6.41	4.83	16.09	6.50	2.43	2.47	2.39	2.40
Bottom-1	2.04	8.98	4.31	3.98	2.70	7.19	3.71	2.96	2.32	1.97	1.88	1.86	1.83
Bottom-2	0.11	0.11	0.11	42.66	15.43	0.11	0.11	26.99	8.34	0.11	0.11	1.52	1.52
Bottom-3	1.08	1.08	1.08	1.64	1.48	1.08	1.08	1.36	1.25	1.08	1.08	1.58	1.19

- For OLS and WLS_s estimators, our method mitigates the accuracy deterioration.
- For the WLS_v estimator, our method improves forecast accuracy for mutable series compared to base forecasts, without an associated decrease in accuracy for immutable series.

- Paper: Zhang, B., Kang, Y., Panagiotelis, A., & Li, F. (2022). Optimal reconciliation with immutable forecasts. European Journal of Operational Research, 308(2), 650–660.
- Code: `https://github.com/AngelPone/chf/tree/master`

- 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 non-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.

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 expriments 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

$$\tilde{\boldsymbol{\pi}} = [\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4]' = [\tilde{\pi}_{(000)}, \tilde{\pi}_{(011)}, \tilde{\pi}_{(101)}\tilde{\pi}_{(112)}]'$$
$$= [0.2, 0.3, 0.4, 0.1]'$$

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

The linear reconciliation function

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

where, $\boldsymbol{A} = [a_{ij}], i = 1, \dots, r, j = 1, \dots, q$ 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

$$
\begin{array}{c}
 \\ 000 \\ 011 \\ 101 \\ 112
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccc}
000 & 010 & 100 & 110 & 001 & 011 & 101 & 111 & 002 & 012 & 102 & 112
\end{array} \\
\begin{bmatrix}
0 & 0.4 & 0.3 & 0.25 & 0 & 0 & 0 & 0 & 0.3 & 0 & 1 & 0.2 \\
0 & 0.4 & 0.3 & 0.25 & 0.2 & 0 & 0 & 1 & 0.3 & 1 & 0 & 0.3 \\
0 & 0 & 0.3 & 0.25 & 0.4 & 0 & 1 & 0 & 0.3 & 0 & 0 & 0.5 \\
1 & 0.2 & 0.1 & 0.25 & 0.4 & 1 & 0 & 0 & 0.1 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

- 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 assgined to **the cloest 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

$$\begin{array}{c c}
 & \begin{array}{cccccccccccc} 000 & 010 & 100 & 110 & 001 & 011 & 101 & 111 & 002 & 012 & 102 & 112 \end{array} \\
\begin{array}{c} 000 \\ 011 \\ 101 \\ 112 \end{array} & \begin{bmatrix} 1 & 0.4 & 0.3 & 0.25 & 0.4 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.25 & 0.3 & 1 & 0 & 0.3 & 0.3 & 0.35 & 0 & 0 \\ 0 & 0 & 0.7 & 0.25 & 0.3 & 0 & 1 & 0.3 & 0.3 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0.4 & 0.1 & 0.65 & 0.6 & 1 \end{bmatrix}
\end{array}$$

- 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.

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

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

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

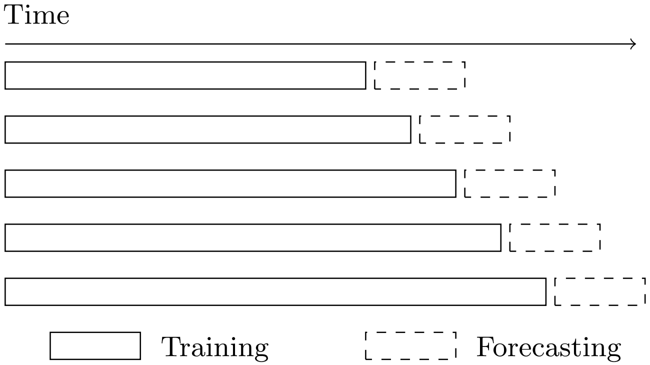
$$\min_{A} \frac{1}{\tau} \sum_{t=1}^{\tau} (\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)'(\mathbf{A}\hat{\boldsymbol{\pi}}_t - \mathbf{z}_t)$$

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

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

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 hierarachy using arbitrary univariate forecasting model.
2. Construct the joint distribution by assuming independence.

Example

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

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

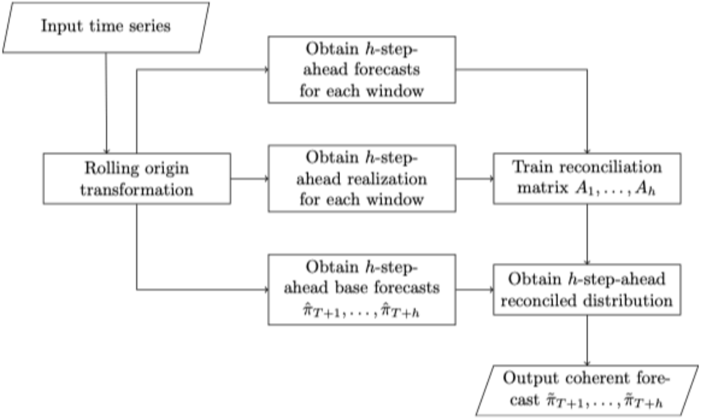
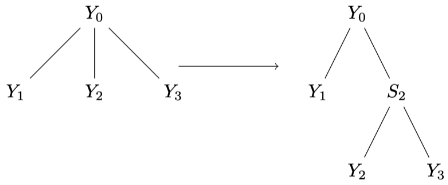
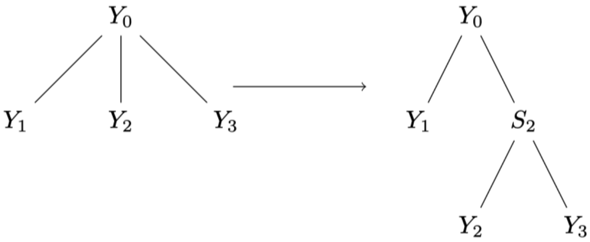


Figure: Flowchart of the DFR algorithm

- 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 fo 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.

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

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.

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

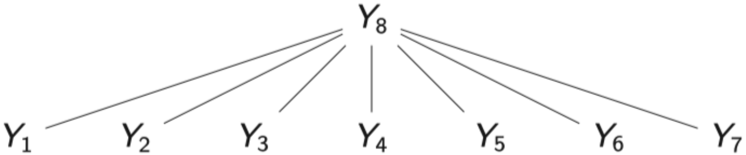
$$\tilde{\pi}_{(011)} = Pr(Y_3 = 1) \times \frac{60}{60 + 40} = 0.18$$

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

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

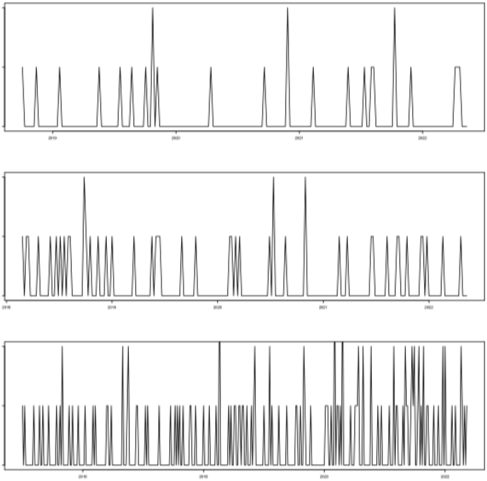


Figure: Example time series

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**

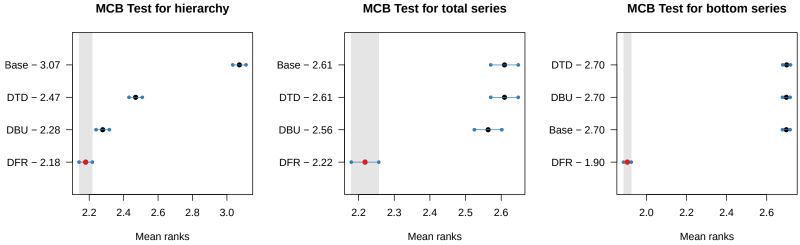


Figure: MCB test results

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

Di Fonzo, T. & Girolimetto, D. (2021), 'Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives', *International Journal of Forecasting* .

Wickramasuriya, S. L. (2021), 'Properties of point forecast reconciliation approaches', *arXiv:2103.11129 [stat]* .

Wickramasuriya, S. L., Athanasopoulos, G. & Hyndman, R. J. (2019), 'Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization', *Journal of the American Statistical Association* **114**(526), 804–819.