Cross-Temporal Hierarchies for Advanced Wind Energy Forecasting

Mahdi Abolghasemi
In-Collaboration with Daniele Girolimetto, and Tommaso Di Fonzo

Senior Lecturer in Statistical Data Science
School of Mathematical Sciences
Queensland Univresity of Technology
Energy Finance Workshop 2024, Sydney, Australia.

December 11, 2024

What is a hierarchical time series?

- A hierarchical time series is a collection of time series that are linked together in a hierarchical structure.
- The hierarchy has a certain structure and there is a logical relationship between series.

Time series Hierarchies

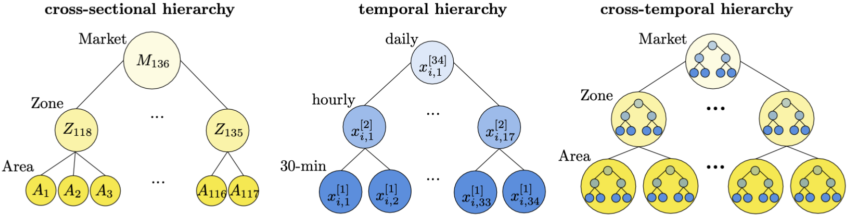


Figure from Cross-Temporal Forecast Reconciliation at Digital Platforms with Machine Learning, Jeroen Romboutsa, Marie Ternesb, Ines Wilmsb

Why hierarchical forecasting?

- Many real-world problems have hierarchical form, e.g., sales for different products, electricity consumption across states , tourism across cities in a country, etc.
- Different levels help us for different decisions.
  Example: Top level forecasts are often used for Long term, strategic decision making; Bottom level forecasts are used for operational and short term planning.
- Improving the forecast accuracy.
- Saving time and resources by generating less forecasts.

What is unique about hierarchical series?

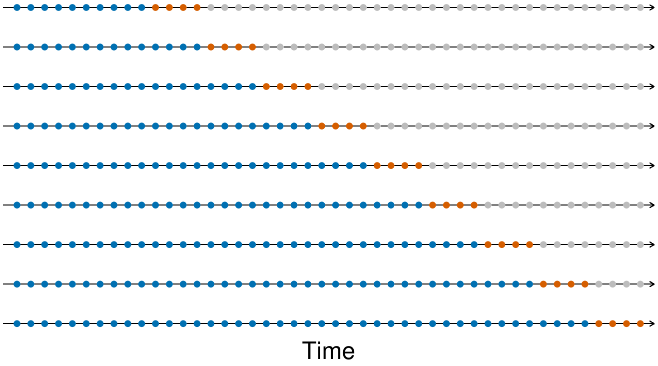
- Forecasting each node individually may generate inconsistent forecasts, so we must find a way to reconcile forecasts to generate coherent forecasts.
- Hierarchical Forecasting methods
  - Bottom-Up (BU)
  - Top-Down(TD)
  - Linear combination
  - Non-linear reconciliation with ML

What is unique about hierarchical series?

- Forecasting each node individually may generate inconsistent forecasts, so we must find a way to reconcile forecasts to generate coherent forecasts.
- Hierarchical Forecasting methods
  - Bottom-Up (BU)
  - Top-Down(TD)
  - Linear combination
  - Non-linear reconciliation with ML
- Still many open research problems, e.g., how to improve across all series/levels, big hierarchies, decision aspect, etc.

Research Gap in Hierarchcail Forecasting

- We use validation errors, rather than traditional in-sample errors, for covariance matrix estimation and forecast reconciliation.



Research Gap in Hierarchy and Decision Making

- We propose decision-based hierarchies and investigate the performance of them versus statistical-based hierarchies. We use a statistical-based strategy aggregating up to 8 hours, $\mathcal{K}_{SB} = \{48, 24, 16, 12, 8, 6, 4, 3, 2, 1\}$, and a decision-based strategy aggregating up to 1 hour, $\mathcal{K}_{DB} = \{6, 3, 2, 1\}$.
- We move beyond conventional metrics of forecast accuracy to incorporate decision costs into the evaluation of hierarchical forecasting models. We assess models not only on their ability to minimize errors but also on their potential to reduce penalties for forecast errors and inefficiencies in grid management.

Hierarchical Forecasting notations

- Consider two levels, where the top variable X represents the sum of two lower-level series, W and Z.
- The cross-sectional hierarchy is simply $X = W + Z$
- The superscript $[k]$ denotes the temporal aggregation order for each level of granularity, with $k = 1$ for 10-minute intervals until $k = 6$ for hourly.
- This holds across any temporal aggregation order $k \in \mathcal{K} = \{6, 3, 2, 1\}$, i.e., $x_\tau^{[k]} = w_\tau^{[k]} + z_\tau^{[k]}$, for $\tau = 1, \ldots, m/k$ and $m = \max(\mathcal{K})$
- Let $\boldsymbol{y}_\tau^{[k]} = \left[x_\tau^{[k]} \quad w_\tau^{[k]} \quad z_\tau^{[k]} \right]'$ represent the (3×1) vector of observations at temporal granularity $k \in \mathcal{K}$ for time $\tau = 1, \ldots, m/k$.

Cross-sectional notations

The cross-sectional aggregation relationships are expressed as follows:

$$x_{\tau}^{[k]} = \boldsymbol{A}_{cs} \begin{bmatrix} w_{\tau}^{[k]} \\ z_{\tau}^{[k]} \end{bmatrix}, \quad \boldsymbol{y}_{\tau}^{[k]} = \boldsymbol{S}_{cs} \begin{bmatrix} w_{\tau}^{[k]} \\ z_{\tau}^{[k]} \end{bmatrix}, \quad \boldsymbol{C}_{cs} \boldsymbol{y}_{\tau}^{[k]} = 0$$

where $\boldsymbol{A}_{cs}$ is the cross-sectional aggregation matrix, $\boldsymbol{S}_{cs}$ is the cross-sectional summing matrix, and $\boldsymbol{C}_{cs}$ represents cross-sectional constraints in homogeneous form, given by:

$$\boldsymbol{A}_{cs} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \boldsymbol{S}_{cs} = \begin{bmatrix} \boldsymbol{A}_{cs} \\ \boldsymbol{I}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{C}_{cs} = \begin{bmatrix} \boldsymbol{I}_1 & -\boldsymbol{A}_{cs} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix}$$

where $\boldsymbol{I}_l$ represents the identity matrix of order l.

Temporal

To capture the full temporal aggregation relationships for a single variable, we use matrices:

$$\boldsymbol{A}_{te} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{S}_{te} = \begin{bmatrix} \boldsymbol{A}_{te} \\ \boldsymbol{I}_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ & & & \boldsymbol{I}_6 & & \end{bmatrix}$$

$$\boldsymbol{C}_{te} = \begin{bmatrix} \boldsymbol{I}_6 & -\boldsymbol{A}_{te} \end{bmatrix},$$

$$\boldsymbol{v} = \boldsymbol{S}_{cs}\boldsymbol{v}^{[1]}, \quad \boldsymbol{C}_{cs}\boldsymbol{v} = \boldsymbol{0}_{6\times 1}$$

where $\boldsymbol{v} = \begin{bmatrix} v_1^{[4]} & v_1^{[2]} & v_2^{[2]} & v_1^{[1]} & v_2^{[1]} & v_3^{[1]} & v_4^{[1]} \end{bmatrix}'$ and $\boldsymbol{v}^{[1]} = \begin{bmatrix} v_1^{[1]} & v_2^{[1]} & v_3^{[1]} & v^{[1]}4 \end{bmatrix}'$ for $v \in \{x, w, z\}$.

Cross-temporal

- All nodes of the cross-temporal hierarchy can be represented based on the hourly series $w_\tau^{[1]}$ and $z_\tau^{[1]}$, $\tau = 1, \ldots, 6$
- $\boldsymbol{y} = \boldsymbol{S}_{ct}\boldsymbol{b}^{[1]} \quad \text{and} \quad \boldsymbol{C}_{ct},$
- $\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}' & \boldsymbol{w}' & \boldsymbol{z}'\end{bmatrix}'$ contains data for all variables at each granularity, $\boldsymbol{b}^{[1]} = \begin{bmatrix}\boldsymbol{w}^{[1]\prime} & \boldsymbol{z}^{[1]\prime}\end{bmatrix}'$ is the high-frequency bottom series vector, $\boldsymbol{C}_{ct}$ is the cross-temporal constraints matrix and $\boldsymbol{S}_{ct} = \boldsymbol{S}_{te} \otimes \boldsymbol{S}_{cs}$ is the summing matrix mapping $\boldsymbol{b}^{[1]}$ to $\boldsymbol{y}$ with $\otimes$ indicates the Kronecker product.
- The dimension of $\boldsymbol{S}_{ct}$ is influenced by the temporal granularities that we want to consider, i.e., stat-based vs decision-based.

Reconciliation

- Consider forecasting an n-dimensional high-frequency hierarchical time series $\left\{\boldsymbol{y}_t^{[1]}\right\}_{t=1}^{T}$
- Forecast horizon equal to the temporal aggregation order m.
- For a factor k of m, we can define several temporally aggregated versions of each component in $\boldsymbol{y}_t^{[1]}$, based on non-overlapping sums of k successive values, each having a seasonal period of $M_k = m/k$.
- Define $\mathcal{K}$ as the set of p factors of m in descending order, $\mathcal{K} = \{k_p, k_{p-1}, \ldots, k_2, k_1\}$, with $k_p = m$ and $k_1 = 1$, and let $m^* = \sum_{j=1}^{p} \frac{m}{k_j}$ and $k^* = m^* - m$.

Cross-temporal Reconciliation cntd.

Define the matrix of base forecasts $\widehat{\boldsymbol{Y}}$ as:

$$\widehat{\boldsymbol{Y}} = \left[\widehat{\boldsymbol{Y}}^{[m]} \; \widehat{\boldsymbol{Y}}^{[k_{p-1}]} \dots \widehat{\boldsymbol{Y}}^{[k_2]} \; \widehat{\boldsymbol{Y}}^{[1]}\right] = \begin{bmatrix} \widehat{\boldsymbol{A}}^{[m]} & \widehat{\boldsymbol{A}}^{[k_{p-1}]} & \dots & \widehat{\boldsymbol{A}}^{[k_2]} & \widehat{\boldsymbol{A}}^{[1]} \\ \widehat{\boldsymbol{B}}^{[m]} & \widehat{\boldsymbol{B}}^{[k_{p-1}]} & \dots & \widehat{\boldsymbol{B}}^{[k_2]} & \widehat{\boldsymbol{B}}^{[1]} \end{bmatrix}.$$

base forecasts are generally cross-sectionally and/or temporally inconsistent:

$$\boldsymbol{C}_{cs}\widehat{\boldsymbol{Y}} \neq \boldsymbol{0}_{n_a \times m^*} \quad \text{and/or} \quad \boldsymbol{C}_{te}\widehat{\boldsymbol{Y}}' \neq \boldsymbol{0}_{k^* \times n},$$

While target forecasts are expected to be consistent across time and space,

$$\boldsymbol{C}_{cs}\boldsymbol{Y} = \boldsymbol{0}_{n_a \times m^*} \quad \text{and} \quad \boldsymbol{C}_{te}\boldsymbol{Y}' = \boldsymbol{0}_{k^* \times n},$$

where $\boldsymbol{C}_{cs}$ and $\boldsymbol{C}_{te}$ are the constraint matrices corresponding to the cross-sectional and temporal frameworks, respectively.

Cross-temporal Reconciliation cntd.

The optimal cross-temporal forecast reconciliation solution using the projection approach is given by

$$\widetilde{\boldsymbol{y}} = \left[\boldsymbol{I}_{nm^*} - \boldsymbol{\Omega}_{ct}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{\Omega}_{ct}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}\right]\widehat{\boldsymbol{y}} = \boldsymbol{M}_{ct}\widehat{\boldsymbol{y}},$$

where $\boldsymbol{\Omega}_{ct}$ is a positive definite (covariance) matrix and $\boldsymbol{M}_{ct} = \boldsymbol{I}_{nm^*} - \boldsymbol{\Omega}_{ct}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{\Omega}_{ct}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}$. Alternatively, the cross-temporally reconciled forecasts can be derived based on the structural approach for cross-sectional reconciliation

$$\widetilde{\boldsymbol{y}} = \boldsymbol{S}_{ct}\left(\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}\boldsymbol{S}_{ct}\right)^{-1}\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}\widehat{\boldsymbol{y}} = \boldsymbol{S}_{ct}\boldsymbol{G}_{ct}\widehat{\boldsymbol{y}},$$

where $\boldsymbol{G}_{ct} = \left(\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}\boldsymbol{S}_{ct}\right)^{-1}\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}$ and $\boldsymbol{M}_{ct} = \boldsymbol{S}_{ct}\boldsymbol{G}_{ct}$.

Estimating the Covariance Matrix

Since $\mathbf{\Omega}_{ct}$ is generally unknown, we consider the following approximations for the cross-temporal covariance matrix:

- $[ct - ols]$ identity matrix, $\mathbf{\Omega}_{ct} = \boldsymbol{I}_{nm^*}$.
- $[ct - str]$ structural matrix, $\mathbf{\Omega}_{ct} = \mathrm{diag}(\boldsymbol{S}_{ct}\mathbf{1}_{mn_b})$.
- $[ct - wlsv]$ variance scaling matrix, an extension of the variance scaling approach.
- $[ct - acov]$ auto-covariance scaling matrix, based on the auto-covariance extension.
- $[ct - bdshr]$ block-diagonal shrunk cross-covariance matrix where we assume incorrelation along the temporal dimension and shrunk cross-sectional cross-covariance matrix for each temporal aggregation level k.

Partially Bottom-Up

When dealing with cross-temporal hierarchies, the bottom-up %(ct(bu)) approach is represented as

$$\widetilde{\boldsymbol{y}} = \boldsymbol{S}_{ct}\boldsymbol{b}^{[1]} = \boldsymbol{S}_{ct}\boldsymbol{G}_{\mathrm{ct}(bu)}\widehat{\boldsymbol{y}},$$

where $\boldsymbol{S}_{ct} = (\boldsymbol{S} \otimes \boldsymbol{R})$ is the cross-temporal summing matrix, $\widehat{\boldsymbol{b}}^{[1]} = \mathrm{vec}\left(\widehat{\boldsymbol{B}}^{[1]\prime}\right)$ is the vector of base forecasts for the high-frequency bottom time series, $\boldsymbol{G}_{\mathrm{ct}(bu)} = \boldsymbol{G}_{\mathrm{cs}(bu)} \otimes \boldsymbol{G}_{\mathrm{te}(bu)}$ with $\otimes$ denoting the Kronecker product,

$$\boldsymbol{G}_{\mathrm{cs}(bu)} = \begin{bmatrix}\boldsymbol{0}_{n_b \times n_a} & \boldsymbol{1}_{n_b}\end{bmatrix} \quad \text{and} \quad \boldsymbol{G}_{\mathrm{te}(bu)} = \begin{bmatrix}\boldsymbol{0}_{m \times k^*} & \boldsymbol{1}_{m}\end{bmatrix}.$$

Partially Bottom-Up Cntd

The cross-temporal bottom-up reconciliation can be viewed as a two-step sequential reconciliation: cross-sectional (temporal) reconciliation of the high-frequency base forecasts $\widehat{\boldsymbol{Y}}^{[1]}$ followed by temporal (cross-sectiona) bottom up.

Let $\boldsymbol{G}_{cs} = \left(\boldsymbol{S}_{cs}' \boldsymbol{W}^{-1} \boldsymbol{S}_{cs}\right)^{-1} \boldsymbol{S}_{cs}' \boldsymbol{W}^{-1}$ the weigthed matrix that cross-sectionally reconcile $\widehat{\boldsymbol{Y}}^{[1]}$ with $\boldsymbol{W}$ a p.d. matrix, then the partly bottom-up reconciled forecasts are obtained as

$$\widetilde{\boldsymbol{y}} = \boldsymbol{S}_{ct} \left(\boldsymbol{G}_{cs} \otimes \boldsymbol{G}_{\text{te}(bu)}\right) \widehat{\boldsymbol{y}}.$$

We refer to these approaches as cs(rec)+te(bu), where 'rec' denotes a generic reconciliation approach in the cross-sectional framework.

Partially Bottom-Up- Iterative Method

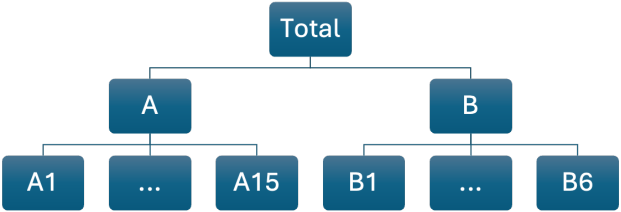
Iterative method alternates reconciliation across one dimension (cross-sectional and temporal). For iteration $j \geqslant 1$, the approach follows these steps:

Step 1 Compute the temporally reconciled forecasts, denoted $\widetilde{\boldsymbol{Y}}_{te}^{(j)}$, for each variable $i \in 1, \ldots, n$ using the cross-sectionally reconciled forecasts from the previous iteration, $\widetilde{\boldsymbol{Y}}_{cs}^{(j-1)}$;

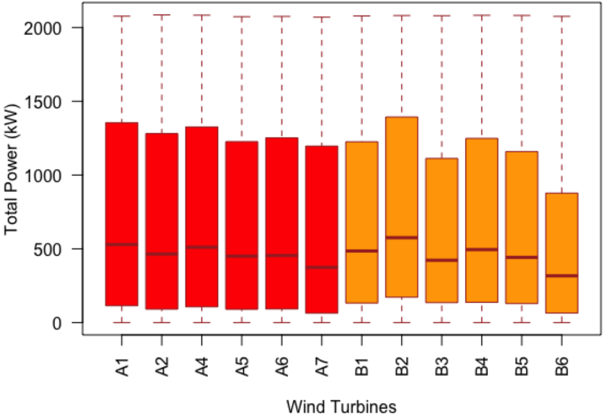
Step 2 Compute the cross-sectionally reconciled forecasts, $\widetilde{\boldsymbol{Y}}_{cs}^{(j)}$, for each temporal aggregation level based on $\widetilde{\boldsymbol{Y}}_{te}^{(j)}$.

The initial values ($j = 0$) for the iterations are set to $\widetilde{\boldsymbol{Y}}_{cs}^{(0)} = \widehat{\boldsymbol{Y}}$. The iterative process ends once the entries in the matrix $\boldsymbol{D}_{te} = \boldsymbol{Z}'\widetilde{\boldsymbol{Y}}_{cs}^{(j)\prime}$, which holds all temporal discrepancies, become sufficiently small according to the convergence criterion.

Case study: Wind Farm Hierarchy



Turbines Time Series Boxplot



Experiment Design

- We used 10-minuetly data from 2020-01-01 to 2020-10-01 for training and data from 2020-10-01 to 2020-12-31 for evaluation.
- The testset spans the period from 01/01/2021 to 31/03/2021.
- We trained one model per turbine and temporal frequency, resulting in 23 models for each level of temporal granularity
- Forecasts via Linear regression and Light Gradient Boosting Machine(LGBM)
- Features: Lags of wind speed, Lags of wind power, Wind speed moving average , Wind speed moving standard deviation, Power moving average, Power standard deviation, Calendar feastures such as time, season, month.
- Bayesian Hyperparameter optimisation for (*num_leaves*), maximum depth of a tree (*max_depth*), learning rate (*learning_rate*), and minimum number of data points in a leaf (*min_data_in_leaf*).

Experiment Design (ctd.)

- Accuracy metric:

$$AvgRelMSE_j^{[k]} = \left(\prod_{i=1}^{23} \frac{MSE_{i,j}^{[k]}}{MSE_{i,b}^{[k]}} \right)^{\frac{1}{23}}$$

where

$$MSE_{i,j}^{[k]} = \frac{1}{QH_k} \sum_{q=1}^{Q} \sum_{h=1}^{H_k} \left(\bar{y}_{q,h,i,j}^{[k]} - y_{q,h,i}^{[k]} \right)^2$$

where $\bar{y}$ represents either the base forecast ($\widehat{y}$) or the reconciled forecast ($\widetilde{y}$), $q = 1, \ldots, Q$ denotes the index of the forecast origins in the test set, $h = 1, \ldots, H_k$ corresponds to the forecast horizon ($H_k = m/k$), $i = 1, \ldots, 23$ indexes the individual time series, j indicates forecast approach, and b refers to the benchmark model (naive).

Statistical-based Hierarchy-Validation Erroe

Approach	Temporal aggregation orders										
	10	20	30	40	60	80	120	160	240	480	All
LR base models											
base	0.968	0.961	0.954	0.951	0.942	0.936	0.924	0.897	0.874	0.881	0.928
pbu	0.948	0.966	0.964	0.958	0.946	0.938	0.919	0.902	0.850	0.691	0.904
ct(*str*)	0.900	0.916	0.913	0.906	0.894	0.885	0.864	0.846	0.794	0.635	0.851
ct(*wlsv*)	0.927	0.944	0.942	0.935	0.923	0.915	0.895	0.877	0.825	0.666	0.881
ct(*bdshr*)	0.916	0.932	0.930	0.923	0.911	0.903	0.882	0.865	0.813	0.654	0.869
ct(*acov*)	0.897	0.912	0.909	0.902	0.890	0.881	0.861	0.843	0.792	0.637	0.848
ite	*0.890*	*0.905*	*0.902*	*0.895*	*0.882*	*0.874*	*0.854*	*0.835*	*0.784*	*0.629*	*0.841*
LGBM base models											
base	0.962	0.952	0.947	0.940	0.932	0.925	0.917	0.908	0.896	0.945	0.932
pbu	0.937	0.955	0.953	0.946	0.934	0.926	0.907	0.889	0.838	0.679	0.892
ct(*str*)	0.888	0.903	0.900	0.893	0.880	0.871	0.850	0.832	0.779	0.621	0.837
ct(*wlsv*)	0.916	0.933	0.930	0.923	0.911	0.903	0.883	0.865	0.813	0.655	0.869
ct(*bdshr*)	0.904	0.920	0.917	0.910	0.898	0.889	0.869	0.851	0.799	0.641	0.855
ct(*acov*)	0.882	0.897	0.894	0.887	0.874	0.865	0.845	0.827	0.776	0.622	0.832
ite	**0.875**	**0.890**	**0.886**	**0.879**	**0.866**	**0.857**	**0.837**	**0.819**	**0.767**	**0.613**	**0.825**

Decision-based Hierarchy-Insample Error

	Temporal aggregation orders										
Approach	10	20	30	40	60	80	120	160	240	480	All
LR base models											
base	0.968	0.961	0.954	0.951	0.942	0.936	0.924	0.897	0.874	0.881	0.928
pbu	0.948	0.966	0.964	0.958	0.946	0.938	0.919	0.902	0.850	0.691	0.904
ct(*str*)	0.937	0.955	0.953	0.946	0.934	0.926	0.906	0.889	0.837	0.678	0.892
ct(*wlsv*)	0.944	0.962	0.960	0.953	0.942	0.934	0.914	0.897	0.845	0.686	0.900
ct(*bdshr*)	0.931	0.948	0.945	0.939	0.927	0.919	0.899	0.881	0.830	0.671	0.885
ct(*acov*)	0.935	0.952	0.950	0.943	0.931	0.923	0.903	0.886	0.834	0.675	0.889
ite	*0.923*	*0.940*	*0.938*	*0.931*	*0.919*	*0.910*	*0.891*	*0.873*	*0.821*	*0.662*	*0.876*
LGBM base models											
base	0.962	0.952	0.947	0.940	0.932	0.925	0.917	0.908	0.896	0.945	0.932
pbu	0.937	0.955	0.953	0.946	0.934	0.926	0.907	0.889	0.838	0.679	0.892
ct(*str*)	0.928	0.945	0.943	0.936	0.924	0.916	0.896	0.879	0.827	0.668	0.882
ct(*wlsv*)	0.935	0.953	0.951	0.944	0.932	0.924	0.904	0.887	0.835	0.677	0.890
ct(*bdshr*)	0.920	0.936	0.934	0.927	0.915	0.907	0.887	0.869	0.817	0.659	0.873
ct(*acov*)	0.925	0.942	0.939	0.932	0.920	0.912	0.892	0.875	0.823	0.664	0.878
ite	**0.914**	**0.931**	**0.928**	**0.921**	**0.909**	**0.900**	**0.880**	**0.862**	**0.810**	**0.652**	**0.867**

Decision-based Hierarchy-Validation Error

Approach	Temporal aggregation orders										
	10	20	30	40	60	80	120	160	240	480	All
LR base models											
base	0.968	0.961	0.954	0.951	0.942	0.936	0.924	0.897	0.874	0.881	0.928
pbu	0.948	0.966	0.964	0.958	0.946	0.938	0.919	0.902	0.850	0.691	0.904
ct(*str*)	0.937	0.955	0.953	0.946	0.934	0.926	0.906	0.889	0.837	0.678	0.892
ct(*wlsv*)	0.944	0.962	0.960	0.953	0.942	0.934	0.914	0.897	0.845	0.686	0.900
ct(*bdshr*)	0.931	0.948	0.945	0.939	0.927	0.919	0.899	0.881	0.830	0.671	0.885
ct(*acov*)	0.935	0.952	0.950	0.943	0.931	0.923	0.903	0.886	0.834	0.675	0.889
ite	*0.923*	*0.940*	*0.938*	*0.931*	*0.919*	*0.910*	*0.891*	*0.873*	*0.821*	*0.662*	*0.876*
LGBM base models											
base	0.962	0.952	0.947	0.940	0.932	0.925	0.917	0.908	0.896	0.945	0.932
pbu	0.937	0.955	0.953	0.946	0.934	0.926	0.907	0.889	0.838	0.679	0.892
ct(*str*)	0.928	0.945	0.943	0.936	0.924	0.916	0.896	0.879	0.827	0.668	0.882
ct(*wlsv*)	0.935	0.953	0.951	0.944	0.932	0.924	0.904	0.887	0.835	0.677	0.890
ct(*bdshr*)	0.920	0.936	0.934	0.927	0.915	0.907	0.887	0.869	0.817	0.659	0.873
ct(*acov*)	0.925	0.942	0.939	0.932	0.920	0.912	0.892	0.875	0.823	0.664	0.878
ite	**0.914**	**0.931**	**0.928**	**0.921**	**0.909**	**0.900**	**0.880**	**0.862**	**0.810**	**0.652**	**0.867**

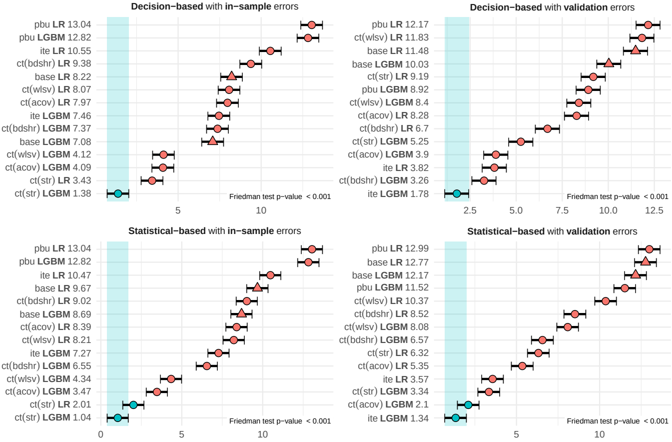
Statistical-based Hierarchy-Insample Error

	Temporal aggregation orders										
Approach	10	20	30	40	60	80	120	160	240	480	All
LR base models											
base	0.968	0.961	0.954	0.951	0.942	0.936	0.924	0.897	0.874	0.881	0.928
pbu	0.979	0.998	0.997	0.991	0.979	0.972	0.953	0.937	0.886	0.726	0.938
ct(*str*)	*0.900*	*0.916*	*0.913*	*0.906*	*0.894*	*0.885*	*0.864*	*0.846*	*0.794*	*0.635*	*0.851*
ct(*wlsv*)	0.941	0.959	0.957	0.950	0.939	0.931	0.911	0.894	0.842	0.683	0.897
ct(*bdshr*)	0.944	0.962	0.960	0.953	0.942	0.934	0.914	0.897	0.846	0.686	0.900
ct(*acov*)	0.942	0.959	0.957	0.950	0.939	0.931	0.911	0.894	0.842	0.683	0.897
ite	0.949	0.967	0.965	0.959	0.947	0.939	0.920	0.902	0.851	0.692	0.905
LGBM base models											
base	0.962	0.952	0.947	0.940	0.932	0.925	0.917	0.908	0.896	0.945	0.932
pbu	0.977	0.997	0.995	0.989	0.978	0.971	0.952	0.935	0.884	0.725	0.936
ct(*str*)	**0.888**	**0.903**	**0.900**	**0.893**	**0.880**	**0.871**	**0.850**	**0.832**	**0.779**	**0.621**	**0.837**
ct(*wlsv*)	0.930	0.947	0.945	0.938	0.926	0.918	0.899	0.881	0.829	0.671	0.884
ct(*bdshr*)	0.937	0.955	0.953	0.946	0.934	0.926	0.906	0.889	0.838	0.679	0.892
ct(*acov*)	0.930	0.946	0.944	0.937	0.926	0.917	0.898	0.880	0.828	0.670	0.883
ite	0.939	0.956	0.954	0.947	0.935	0.927	0.908	0.890	0.839	0.680	0.893

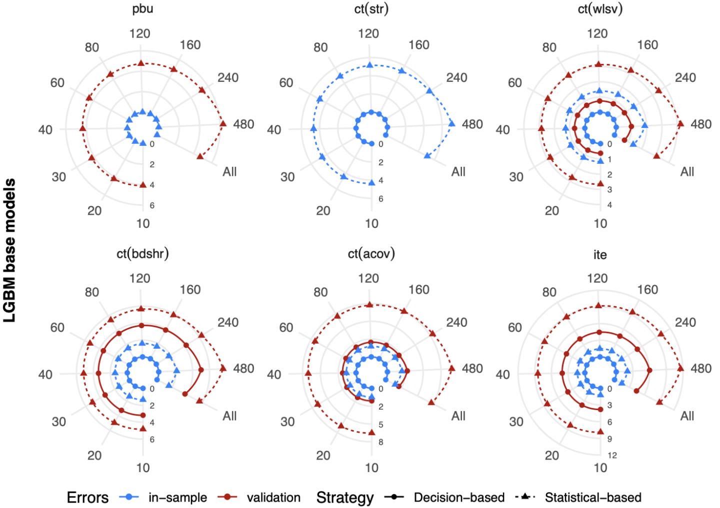
Statistical Significance Test

- Multiple Comparison with Best (MCB)
  $r_{\alpha,K,N} = q_a \times \sqrt{\frac{K*(K+1)}{12N}}$,
  N is the number of the time series
  K is the number of the examined methods
  q_a is the quantile of the confidence interval.

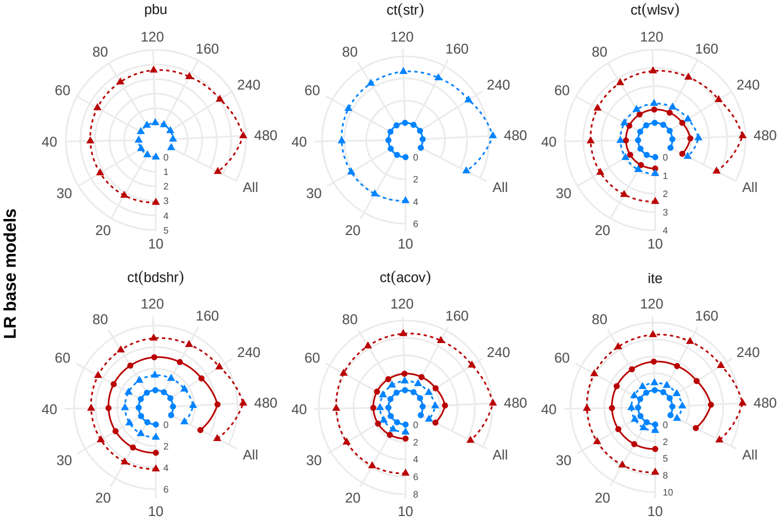
Models Comparison- MCB Test



Relative Models Improvement to Worst-LGBM



Relative Models Improvement to Worst-LR



Decision Cost

- Fines and penalties, δ^{-}: The proportion of total profit for an eight-hour production period that is lost due to fines and regulatory penalties when actual energy production falls below the forecast amount. It measures the percentage of expected profit consumed by fines and penalties when actual production does not meet the forecast levels.
- Revenue loss, δ^{+} : It measure the percentage of additional income that could have been earned. When a company forecasts an output of $\bar{y}$, but the actual production exceeds this amount with $y > \bar{y}$, excess production $(y - \bar{y})$ cannot typically be put on the market due to the regulatory and consumption limits, resulting in unrealized profit. Thus, δ^{+} quantifies the economic loss associated with overproduction.

Decision Cost-Cntd

The indices δ^- and δ^+ are calculated via

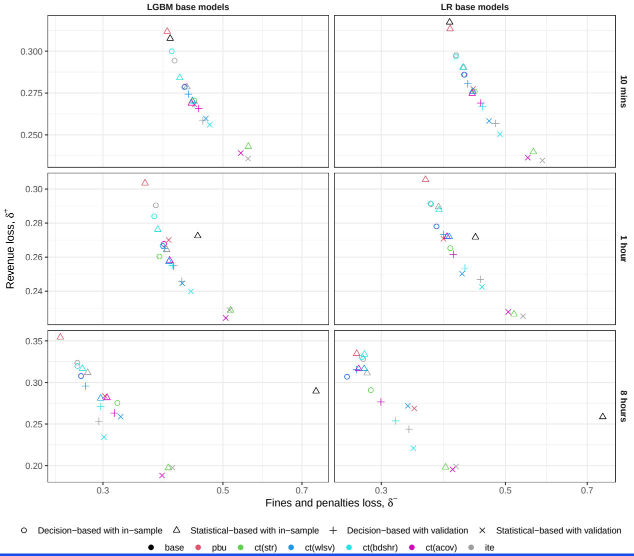
$$\begin{cases} e^-_{q,j,k} = \dfrac{\bar{y}^{[k]}_{q,1,j} - y^{[k]}_{q,1}}{y^{[k]}_{q,1}} & \text{if } y^{[k]}_{q,1} < (1-\Delta)\bar{y}^{[k]}_{q,1,j} \\ e^+_{q,j,k} = \dfrac{y^{[k]}_{q,1} - \bar{y}^{[k]}_{q,1,j}}{\bar{y}^{[k]}_{q,1,j}} & \text{if } y^{[k]}_{q,1} > (1+\Delta)\bar{y}^{[k]}_{q,1,j} \end{cases}$$

where Δ represents a threshold set to 1%. Then averaged across all the forecast origins:

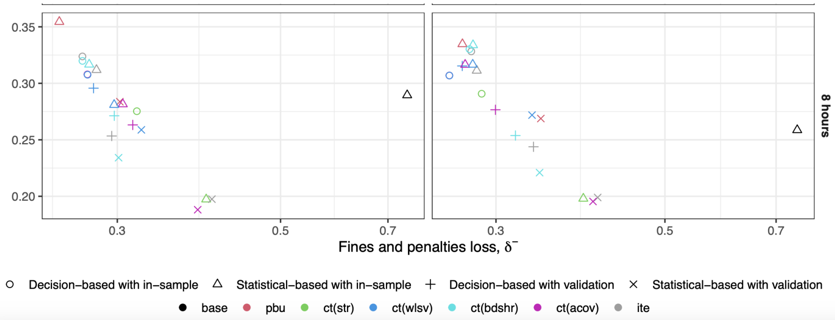
$$\delta^-_{j,k} = \left(\prod_{q \in \mathcal{Q}^-_{j,k}} e^-_{q,j,k} \right)^{\left|\mathcal{Q}^-_{j,k}\right|^{-1}} \quad \text{and} \quad \delta^+_{j,k} = \left(\prod_{q \in \mathcal{Q}^+_{j,k}} e^+_{q,j,k} \right)^{\left|\mathcal{Q}^+_{j,k}\right|^{-1}}$$

where
$\mathcal{Q}^-_{j,k} = \{q = 1, \ldots, Q \mid y^{[k]}_{q,1} < \bar{y}^{[k]}_{q,1,j}\}$ and
$\mathcal{Q}^+_{j,k} = \{q = 1, \ldots, Q \mid y^{[k]}_{q,1} > \bar{y}^{[k]}_{q,1,j}\}$.

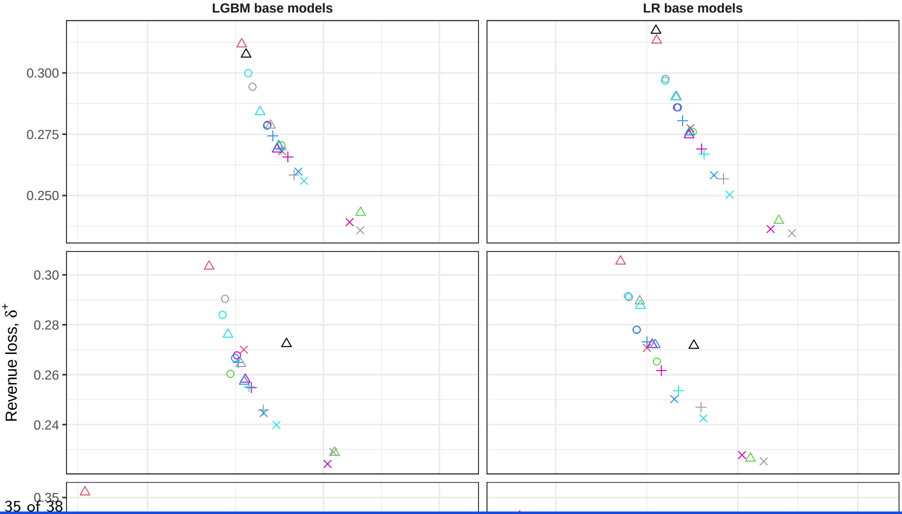
Decision Cost- Results



Decision Cost- Results cntd



Decision Cost- Results cntd



Conclusion

- We proposed a robust alternative to traditional in-sample error-based approaches in hierarchical forecasting.
- While statistical-based approaches achieve higher accuracy, decision-driven models offer a practical balance between accuracy and economic performance.
- Decision-based methods, particularly those leveraging validation error strategies such as ct(bdshr) and ct(wlsv), provide a versatile and effective framework for balancing forecast accuracy with real-world economic impacts.
- Statistical-based hierarchies tend to adopt less conservative forecasts, reducing revenue losses.
- Decision-based methods offered a more balanced compromise between accuracy and decision costs, making them particularly attractive for practical applications.

References

- Di Fonzo, T., Girolimetto, D. (2023a). Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. International Journal of Forecasting, 39(1), 39–57.
- Athanasopoulos, G., Kourentzes, N. (2023). On the evaluation of hierarchical forecasts. International Journal of Forecasting, 39(4), 1502–1511. https://doi.org/10.1016/j.ijforecast.2022.08.003
- George Athanasopoulos, Roman A Ahmed, and Rob J Hyndman. "Hierarchical forecasts for Australian domestic tourism". International Journal of Forecasting 25.1 (2009), pp. 146–166
- Cross-Temporal Forecast Reconciliation at Digital Platforms with Machine Learning, Jeroen Romboutsa, Marie Ternesb, Ines Wilmsb, EJOR, 2024

References

- R. J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, 2nd Edition, OTexts, Melbourne, Australia, 2018.URL http://OTexts.com/fpp2
- R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, H. L. Shang, Optimal combination forecasts for hierarchical time series, Computational Statistics & Data Analysis 55 (9) (2011) 2579–2589.
- S. L. Wickramasuriya, G. Athanasopoulos, R. J. Hyndman, Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization, J American Statistical Association 114 (526) (2019) 804–819.
- G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, F. Petropoulos, Forecasting with temporal hierarchies, European Journal of Operational Research 262 (1) (2017) 60–74