Hierarchical forecasting for aggregated curves with an application to day-ahead electricity price auctions

Paul Ghelasi, Florian Ziel

Universität Duisburg-Essen

INREC 2022, Sep 27-28

Overview

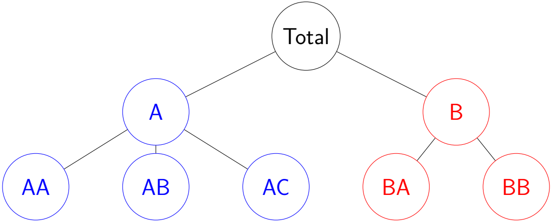
Hierarchical time series

Definition

A hierarchical time series is a set of time series that naturally relate to each other, i.e. sum up or break down, according to a certain logic.

Example

Sales broken down by region, where the sales in larger areas such as a country (Total) must equal the sum of sales in smaller areas such as regions (A,B) within that country, which in turn can be broken down into zones (AA, ..., BB).



Terminology

- Lower levels add up to form the higher levels resulting in a hierarchy between time series, e.g:

$$y_t = y_{A,t} + y_{B,t}$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t}$$

- The hierarchical relationship always holds for historical values; this property is referred to as **coherency**.
- Coherency is generally *not* given for individual forecasts, called **base forecasts**, e.g:

$$\hat{y}_{T+1} \neq \hat{y}_{A,T+1} + \hat{y}_{B,T+1}$$

- **Reconciliation** methods are transformations of the individual forecasts by which they are made coherent, i.e. it then holds that:

$$\tilde{y}_{T+1} = \tilde{y}_{A,T+1} + \tilde{y}_{B,T+1}$$

- Advantages of coherent time series: 1) aligning forecasts across organization for decision-making, 2) reconciliation methods can increase forecasting accuracy.

Notation

- Bottom level values are collected in $\mathbf{b}_t$ $(m \times 1)$
- All values of the hierarchy are collected in $\mathbf{y}_t$ $(n \times 1)$
- The summing matrix $\mathbf{S}$ $(n \times m)$ is multiplied by $\mathbf{b}_t$ and dictates how the bottom-level series are aggregated to from each element of $\mathbf{y}_t$
- Thus, any (linear) hierarchy can be summarized as:

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

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

Reconciliation - classical approaches

Bottom-up approach: uses the forecasts of the bottom-level values $\hat{\mathbf{b}}_{T+1}$:

$$\tilde{\mathbf{y}}_{T+1} = \mathbf{S}\hat{\mathbf{b}}_{T+1}$$

Top-down approach: uses the forecast of the top-level value $\hat{y}_{T+1}$ from which the bottom-level values $\hat{\mathbf{b}}_{T+1}$ are generated by multiplying $\hat{y}_{T+1}$ with proportions:

$$\hat{\mathbf{b}}_{T+1} = \mathbf{p}\hat{y}_{T+1}$$

where $\mathbf{p} = \begin{bmatrix} p_{AA} & \dots & p_{BB} \end{bmatrix}'$ $(m \times 1)$.

- There are many ways to calculate proportions, e.g: the historical average of the top-level to bottom-level ratio

$$p_{AA} = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{AA,t}}{y_t}$$

- Usually the best results are achieved by calculating proportions using the forecasted $\hat{\mathbf{y}}_{T+1}$ instead of historical values. However, the formula can become quite complex, e.g:

$$p_{AA} = \left(\frac{\hat{y}_{AA,T+1}}{\hat{y}_{AA,T+1} + \hat{y}_{AB,T+1} + \hat{y}_{AC,T+1}}\right)\left(\frac{\hat{y}_{A,T+1}}{\hat{y}_{A,T+1} + \hat{y}_{B,T+1}}\right)$$

Reconciliation - classical approaches

- The results for the top-down approach using using forecasted values are generally better because one make use of *all* the information at *every* level of the hierarchy.
- This key aspect of hierarchical forecasting is generalized via the **mapping matrix P** $(m \times n)$, which constructs bottom-level values from the full vector as $\mathbf{b}_t = \mathbf{P}\mathbf{y}_t$. Thus, every linear reconciliation method can be described as:

$$\tilde{\mathbf{y}}_t = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_t$$

Reconciliation - classical approaches

- For the top-down approach

$$\mathbf{P}_{TD} = \begin{bmatrix} \mathbf{p} & \mathbf{O}_{m \times (n-1)} \end{bmatrix},$$

where **O** is a zero matrix. Hence, in our example:

$$\underbrace{\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \end{bmatrix}}_{\tilde{\mathbf{y}}_t} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} p_{AA} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{AB} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{AC} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{BA} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{BB} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{P}_{TD}} \underbrace{\begin{bmatrix} \hat{y}_t \\ \hat{y}_{A,t} \\ \hat{y}_{B,t} \\ \hat{y}_{AA,t} \\ \hat{y}_{AB,t} \\ \hat{y}_{AC,t} \\ \hat{y}_{BA,t} \\ \hat{y}_{BB,t} \end{bmatrix}}_{\hat{\mathbf{y}}_t}$$

Reconciliation - classical approaches

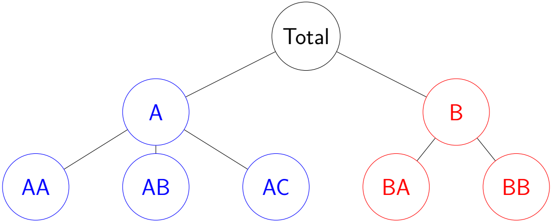
- For the bottom-up approach $\mathbf{P}_{BU}$ simply returns the bottom-level values:

$$\mathbf{P}_{BU} = \begin{bmatrix} \mathbf{O}_{m\times(n-m)} & \mathbf{I}_m \end{bmatrix}$$

Hence, in our example:

$$\underbrace{\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{AC,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \end{bmatrix}}_{\tilde{\mathbf{y}}_t} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{P}_{BU}} \underbrace{\begin{bmatrix} \hat{y}_t \\ \hat{y}_{A,t} \\ \hat{y}_{B,t} \\ \hat{y}_{AA,t} \\ \hat{y}_{AB,t} \\ \hat{y}_{AC,t} \\ \hat{y}_{BA,t} \\ \hat{y}_{BB,t} \end{bmatrix}}_{\hat{\mathbf{y}}_t}$$

Reconciliation - classical approaches



Reconciliation - classical approaches

Optimal (linear) reconciliation approach (minimum trace reconciliation): it is possible to choose $\mathbf{P}$ such that the coherent forecasts $\tilde{\mathbf{y}}_{T+1}$ are unbiased and minimize the variances of the coherent forecast errors $tr(\mathbf{V}) = tr(Var[\mathbf{y}_{T+1} - \tilde{\mathbf{y}}_{T+1}])$:

Theorem

The minimum trace mapping matrix is:

$$\boldsymbol{P}_{OP} = (\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$$

where $\boldsymbol{W} = Var[\boldsymbol{y}_{T+1} - \hat{\boldsymbol{y}}_{T+1}]$ is the variance-covariance matrix of the base forecast errors.

Different approaches that vary in complexity exist for how to estimate $\mathbf{W}$, for example $\mathbf{W} = I$ or $\mathbf{W} = diag(\hat{\mathbf{W}})$, where $\hat{\mathbf{W}}$ is the estimated base error variance-covariance matrix.

Novelties

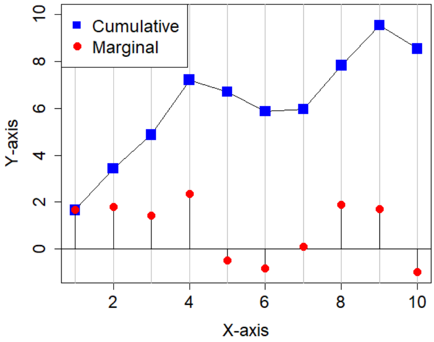
Our paper brings four novelties to the field of hierarchical forecasting:

1. We show that aggregated curves have an implicit hierarchical structure. We show various methods of constructing and deconstructing the curves.
2. We introduce a new reconciliation approach tailored to aggregated curves entitled **aggregated-down**, similar in complexity to top-down, which we recommend to use as benchmark method alongside bottom-up and top-down.
3. We study minimum trace optimal reconciliation approaches for aggregated curves in detail. This includes a result which states that under some assumptions the reconciliation approach is independent of the representations of the curve.
4. We apply all of these in an empirical setting to forecast the supply and demand curves of day-ahead electricity price auctions. We conclude that forecast accuracy can be improved through hierarchical reconciliation.

Overview

Canonical representation

Every curve whose increments are well-defined can easily be aggregated or disaggregated into marginal and cumulative values:

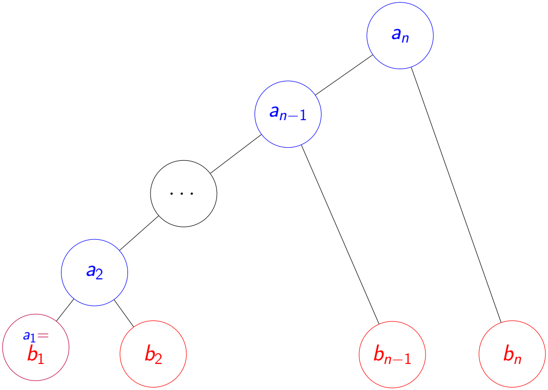


Note that the whole curve is a single observation in time, not a time series.

Canonical representation

The curve can result from an aggregation procedure of the marginals, or we can start with the curve and disaggregate to receive the bottom values.

There is a natural way of representation, which we refer to as the **canonical** hierarchical structure of an aggregated curve:



Canonical representation - notation

Let $\boldsymbol{b} = (b_1, \ldots, b_n)'$ be the bottom-level values and their aggregation $\boldsymbol{a} = (a_1, \ldots, a_n)'$, where it holds that

$$a_i = \sum_{j=1}^{i} b_j.$$

and also the recursive relationship:

$$a_i = a_{i-1} + b_i \text{ for } 1 \leq i < n \text{ and } a_1 = b_1.$$

We could also introduce the canonical representation starting from the aggregated values $\boldsymbol{a}$. Then, we can receive the same bottom values $\boldsymbol{b}$ by differencing:

$$b_i = a_i - a_{i-1}.$$

$\boldsymbol{b}$ can be expressed as $\boldsymbol{b} = \boldsymbol{D}_n \boldsymbol{a}$ where $\boldsymbol{D}_n$ is an invertible n-dimensional quadratic matrix with $\boldsymbol{a} = \boldsymbol{D}_n^{-1} \boldsymbol{b}$.

Canonical representation - notation

We can compactly write all values of the hieararchy as the $(2n-1) \times 1$ vector

$$\boldsymbol{y} = \begin{bmatrix} a_n & \dots & a_2 & \boldsymbol{b}' \end{bmatrix}'$$

which contains the values of $\boldsymbol{a}$ in the inverse order except for its first value $a_1 = b_1$.

For the considered canonical representation the summation matrix is:

$$\boldsymbol{S} = \left[\begin{array}{c:c} \boldsymbol{1}_{n-1} & \boldsymbol{U}_{n-1} \\ \hdashline \multicolumn{2}{c}{\boldsymbol{I}_n} \end{array}\right] = \left[\begin{array}{c:cccc} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 0 & 0 \\ \hdashline 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{array}\right].$$

Alternative representations

Representations other than the canonical one for aggregated curves are possible. We consider a situation where the aggregated values $\boldsymbol{a}$ are given and the scope is to define a disaggregation rule to receive the bottom values $\boldsymbol{b}$.

- Let k be the element of $\boldsymbol{a}$ where from we start the disaggregation rule.
- Let $\boldsymbol{b}_{[k]}$ be an alternative bottom-values vector resulting from starting the disaggregation from the kth element.
- For $k = 1$, we disaggregate from the start, thus getting the canonical representation, hence $\boldsymbol{b} = \boldsymbol{b}_{[1]}$. It holds that

$$b_{[1],1} = b_1 = a_1,\ b_{[1],i} = b_i = a_i - a_{i-1} \text{ for } i > k.$$

- For $k = n$, we disaggregate from the end. It holds that $b_{[n],1} = a_n$, $b_{[n],2} = a_{n-1} - a_n$, and in general

$$b_{[n],i} = a_{n-i+1} - a_{n-i+2}.$$

  - This approach can be embedded in the canonical representation which we receive by defining $b_i = b_{[n],n-i+1}$.

Alternative representations

We can further consider the disaggregation starting at k with $1 < k < n$ with the corresponding value $b_{[k],k} = a_k$. We then require two directions of aggregation, one for bottom values larger than k, and the other one for smaller ones resulting in:

$$b_{[k],i} = \begin{cases} a_i & \text{, if } i = k \\ a_i - a_{i-1} & \text{, if } i > k \\ a_i - a_{i+1} & \text{, if } i < k \end{cases}$$

The special cases $\boldsymbol{b}_{[1]}$ (the canonical representation) and $\boldsymbol{b}_{[n]}$ can be defined using this definition as well.

Example

An $n = 6$-dimensional example for $k = 1, 3, 6$:

	i	1	2	3	4	5	6
$\boldsymbol{a}$	a_i	1	4	6	7	10	15
$\boldsymbol{b}_{[1]}$	$b_{[1],i}$	1	3	2	1	3	5
$\boldsymbol{b}_{[3]}$	$b_{[3],i}$	-3	-2	6	1	3	5
$\boldsymbol{b}_{[6]}$	$b_{[6],i}$	-3	-2	-1	-3	-5	15

Alternative representations

The previous notations also generalize to include the canonical case as a special case with:

- The mapping matrix:

$$\boldsymbol{S}_{[k]} = \left[\begin{array}{c:c:c} \mathbf{O}_{n-k,k-1} & \mathbf{1}_{n-k} & \boldsymbol{U}_{n-k} \\ \hdashline \boldsymbol{L}_{k-1} & \mathbf{1}_{k-1} & \mathbf{O}_{k-1,n-k} \\ \hdashline \multicolumn{3}{c}{\boldsymbol{I}_n} \end{array}\right]$$

  where $\boldsymbol{L}_{k-1}$ is a $(k-1)$-dimensional matrix which contains 1 on the lower anti-diagonal.
- The vector $\boldsymbol{y}_{[k]}$ in the corresponding hierarchy which satisfies $\boldsymbol{y}_{[k]} = \boldsymbol{S}_{[k]}\boldsymbol{b}$ is

$$\boldsymbol{y}_{[k]} = \begin{bmatrix} \boldsymbol{a}_{[-k]} \\ \boldsymbol{b}_{[k]} \end{bmatrix}$$

  where we define $\boldsymbol{a}_{[-k]}$ as the reversed vector $\boldsymbol{a}$ without the kth element, i.e. $\boldsymbol{a}_{[-k]} = (a_n, \ldots, a_{k+1}, a_{k-1}, \ldots, a_1)'$. For $k = 1$ we have $\boldsymbol{y} = \boldsymbol{y}_{[k]}$.

Alternative representations

However, the different representations of the hierarchical structure are only formal representations and do not automatically provide different reconciled forecasts by themselves.

To see this we introduce $\boldsymbol{b}_{[k]} = \boldsymbol{A}_{[k]}\boldsymbol{a}$ with matrix $\boldsymbol{A}_{[k]}$ which yields

$$\boldsymbol{y}_{[k]} = \boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{a}.$$

For $k = 1$ we have $\boldsymbol{A}_{[k]} = \boldsymbol{D}_n$ from Slide 13.

In addition, with definition (20) there exists a matrix $\boldsymbol{B}_{[k]}$ which satisfies

$$\boldsymbol{y} = \boldsymbol{B}_{[k]}\boldsymbol{y}_{[k]}.$$

It is easy to check that $\boldsymbol{B}_{[k]}$ is orthogonal, i.e. it holds $\boldsymbol{B}_{[k]}^{-1} = \boldsymbol{B}'_{[k]}$. $\boldsymbol{B}_{[k]}$ is a generalized permutation matrix which contains permutation and reflection components.

Finally, we receive with $\boldsymbol{y} = \boldsymbol{B}_{[k]}\boldsymbol{y}_{[k]}$, $\boldsymbol{y}_{[k]} = \boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{a}$ and $\boldsymbol{a} = \boldsymbol{D}_n^{-1}\boldsymbol{b}$ that $\boldsymbol{y} = \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}\boldsymbol{b}$. Thus, it holds that

$$\boldsymbol{S} = \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}.$$

Alternative representations - structure invariance

Recall that for optimal (minimum trace) reconciliation we have a mapping matrix given a k:

$$\widetilde{\boldsymbol{P}}_{[k]} = (\boldsymbol{S}'_{[k]}\boldsymbol{W}_{[k]}^{-1}\boldsymbol{S}_{[k]})^{-1}\boldsymbol{S}'_{[k]}\boldsymbol{W}_{[k]}^{-1}.$$

We can show that under mild assumption of the forecast method and the reconciling matrix $\boldsymbol{W}_{[k]}$ that the reconciliation approach preserves the hierarchical structure, i.e. **the result does not depend on the choice of** k:

Theorem

If $\widehat{\boldsymbol{y}} = \boldsymbol{B}_{[k]}\widehat{\boldsymbol{y}}_{[k]}$ *and* $\boldsymbol{W}_{[k]}^{-1} = \boldsymbol{B}'_{[k]}\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}$ *then it holds that*

$$\widetilde{\boldsymbol{y}} = \boldsymbol{B}_{[k]}\widetilde{\boldsymbol{y}}_{[k]}.$$

$\boldsymbol{B}_{[k]}$ is orthogonal, thus the assumption $\widehat{\boldsymbol{y}} = \boldsymbol{B}_{[k]}\widehat{\boldsymbol{y}}_{[k]}$ is satisfied if the forecasting algorithm which provides $\widehat{\boldsymbol{y}}$ is invariant to orthogonal transformations. This holds for instance for linear regressions. Also $\boldsymbol{W}_{[k]}^{-1} = \boldsymbol{B}'_{[k]}\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}$ is trivially satisfied if $\boldsymbol{W} = \boldsymbol{I}_{2n-1}$ due to orthogonality of $\boldsymbol{B}_{[k]}$.

Thus, we will only be considering the canonical representation.

Alternative representations - structure invariance proof

Proof.

Consider $\widetilde{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{P}\widehat{\boldsymbol{y}}$. With definition of $\boldsymbol{P}$ and $\boldsymbol{S} = \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}$ it holds with the assumptions of the Theorem that:

$$
\begin{aligned}
\widetilde{\boldsymbol{y}} &= \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\widehat{\boldsymbol{y}} \\
&= \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}\left((\boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1})'\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}\right)^{-1} \\
&\quad (\boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1})'\boldsymbol{W}^{-1}\widehat{\boldsymbol{y}}(\boldsymbol{D}_n^{-1})'\boldsymbol{A}_{[k]}'\boldsymbol{S}_{[k]}'\boldsymbol{B}_{[k]}'\boldsymbol{W}^{-1} \\
&= \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}\boldsymbol{A}_{[k]}\boldsymbol{D}_n^{-1}\left(\boldsymbol{D}_n\boldsymbol{A}_{[k]}^{-1}(\boldsymbol{S}_{[k]}'\boldsymbol{B}_{[k]}'\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]})^{-1}(\boldsymbol{A}_{[k]}^{-1})'\boldsymbol{D}_n'\right) \\
&\quad (\boldsymbol{D}_n^{-1})'\boldsymbol{A}_{[k]}'\boldsymbol{S}_{[k]}'\boldsymbol{B}_{[k]}'\boldsymbol{W}^{-1}\widehat{\boldsymbol{y}} \\
&= \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}(\boldsymbol{S}_{[k]}'\boldsymbol{B}_{[k]}'\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]})^{-1}\boldsymbol{S}_{[k]}'\boldsymbol{B}_{[k]}'\boldsymbol{W}^{-1}\boldsymbol{B}_{[k]}\widehat{\boldsymbol{y}}_{[k]} \\
&= \boldsymbol{B}_{[k]}\boldsymbol{S}_{[k]}(\boldsymbol{S}_{[k]}'\boldsymbol{W}_{[k]}^{-1}\boldsymbol{S}_{[k]})^{-1}\boldsymbol{S}_{[k]}'\boldsymbol{W}_{[k]}^{-1}\widehat{\boldsymbol{y}}_{[k]} \\
&= \boldsymbol{B}_{[k]}\widetilde{\boldsymbol{y}}_{[k]}.
\end{aligned}
$$

□

Overview

Classical approaches

With the canonical hierarchical structure of aggregated curves, the formulas for the classical approaches are as follows:

- Bottom-up approach: $\boldsymbol{P}_{BU} = \begin{bmatrix} \mathbf{O}_{n\times(n-1)} & \boldsymbol{I}_n \end{bmatrix}$
- Top-down approach:

$$\boldsymbol{P}_{\text{TD}} = \begin{bmatrix} \boldsymbol{p} & \mathbf{O}_{n\times(2n-2)} \end{bmatrix}$$
$$\boldsymbol{p} = (p_1, \dots, p_n)'$$

Average ratio: $\widehat{p}_{\text{ar},j} = \frac{1}{T}\sum_{t=1}^{T} \frac{b_{j,t}}{a_{n,t}}$

Ratio of averages: $\widehat{p}_{\text{ra},j} = \frac{\frac{1}{T}\sum_{t=1}^{T} b_{j,t}}{\frac{1}{T}\sum_{t=1}^{T} a_{n,t}}$

Forecasted values:

$$\widehat{p}_{\text{fo},j} = \begin{cases} \frac{\widehat{b_n}}{\widehat{a_{n-1}} + \widehat{b_n}}, & \text{for } j = n \\ \frac{\widehat{b_j}}{\widehat{a_{j-1}} + \widehat{b_j}} \prod_{i=j}^{n-1} \left(\frac{\widehat{a_i}}{\widehat{a_i + b_{i+1}}} \right) & \text{for } 1 < j < n, \\ \prod_{i=1}^{n-1} \left(\frac{\widehat{a_i}}{\widehat{a_i + b_{i+1}}} \right) & \text{for } j = 1. \end{cases}$$

Classical approaches

- Optimal reconciliation (minimum trace) approach:
  - Recall that the optimal mapping matrix that returns the best, unbiased coherent forecasts is given by

    $$\boldsymbol{P}_{\mathrm{op}} = (\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$$

    where $\boldsymbol{W} = \mathrm{Var}[\boldsymbol{y} - \widehat{\boldsymbol{y}}]$ is the variance-covariance matrix of the base forecast errors. The equation is the result of minimizing the variance of the coherent forecasts.
  - $\boldsymbol{W}$ is not known and has to be estimated, for which we consider multiple estimators:

1. $\boldsymbol{W}_{\text{opols}} = \boldsymbol{I}_{2n-1}$, where $\boldsymbol{I}_{2n-1}$ is the identity matrix.
2. $\boldsymbol{W}_{\text{oplambda}} = \Lambda$, $\Lambda = \mathrm{Diag}(\boldsymbol{S}\mathbf{1}_n)$, $\boldsymbol{S}$ is the summation matrix and $\mathbf{1}_n$ is a unit vector of the same dimension as the number of bottom-level time series. In out aggregated curves setting it holds that $\mathrm{Diag}(\Lambda) = \boldsymbol{S}\mathbf{1}_n = (n, n-1, \ldots, 2, 1, 1, \ldots, 1)'$.

Classical approaches

3. $\boldsymbol{W}_{\text{opwls}} = \widehat{\boldsymbol{W}}_{\text{dcov}}$, where $\widehat{\boldsymbol{W}}_{\text{dcov}}$ is an estimator for $\boldsymbol{W}_{\text{dcov}} = \text{Diag}(\boldsymbol{W}_{\text{cov}})$ with $\boldsymbol{W}_{\text{cov}}$ as the covariance matrix of the errors associated to $\boldsymbol{y}$. $\boldsymbol{W}_{\text{cov}}$ can be estimated by the sample covariance $\widehat{\boldsymbol{W}}_{\text{cov}}$, i.e. $\widehat{\boldsymbol{W}}_{\text{cov}} = \frac{1}{n}\mathbf{E}'\mathbf{E} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{e}_i \boldsymbol{e}_i'$ and $\boldsymbol{E}$ is the matrix of residuals generated by and arranged in the same order as the base forecasts. The $\boldsymbol{W}_{\text{opwls}}$ approach may be regarded as a generalization of the $\boldsymbol{W}_{\text{oplambda}}$ approach. The design corresponds to a setting where the individual forecast errors have different variances but are uncorrelated.
4. $\boldsymbol{W}_{\text{opcov}} = \widehat{\boldsymbol{W}}_{\text{cov}}$, where $\boldsymbol{W}_{\text{cov}}$ is the full sample covariance matrix of the error terms. The underlying setting corresponds to a situation where the forecasts errors have varying variances and exhibit linear dependence.

Classical approaches

5. $\boldsymbol{W}_{\text{opshrink}} = \lambda\widehat{\boldsymbol{W}}_{\text{dcov}} + (1-\lambda)\widehat{\boldsymbol{W}}_{\text{cov}}$, where λ is the shrinkage intensity parameter. propose to set

$$\lambda = \frac{\sum_{i \neq j} \widehat{\text{Var}}(\widehat{r}_{ij})}{\sum_{i \neq j} \widehat{r}_{ij}^2},$$

where $\widehat{r}_{ij}$ is the ijth element of the 1-step-ahead sample correlation matrix. The authors implemented the formulas in the `corpcor` R package;

6. Ledoit-Wolf covariance matrix estimator with shrinkage towards constant correlation: $\boldsymbol{W}_{\text{opledoitwolf}} = \delta\mathbf{F} + (1-\delta)\widehat{\boldsymbol{W}}_{\text{cov}}$, where $\widehat{\boldsymbol{W}}_{\text{cov}}$ is the sample covariance matrix, $\mathbf{F}$ is the shrinkage target with constant correlation defined with element $f_{ij} = \bar{r}\sqrt{\widehat{w}_{ii}\widehat{w}_{jj}}$ on the ith row and jth column, $\widehat{w}_{ij}$ is the corresponding element of $\widehat{\boldsymbol{W}}_{\text{cov}}$, and

$$\bar{r} = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}, \quad r_{ij} = \frac{\widehat{w}_{ij}}{\sqrt{\widehat{w}_{ii}, \widehat{w}_{jj}}}$$

.

Classical approaches

7. Covariance matrix estimation using graphical lasso:

$$\boldsymbol{W}_{\text{opglasso}} = \begin{bmatrix} \boldsymbol{W}_{11} & \boldsymbol{w}_{12} \\ \boldsymbol{w}_{12}' & w_{22} \end{bmatrix},$$

where $\boldsymbol{W}_{\text{opglasso}}$ is partitioned as shown and estimated by using coordinated descent to solve

$$\beta_{\text{opglasso}} = argmin_{\beta} \left\{ \frac{1}{2} \| \boldsymbol{W}_{11}^{1/2} \beta - \boldsymbol{b} \|^2 + \rho \| \beta \|_1 \right\},$$

where $\boldsymbol{b} = \boldsymbol{W}_{11}^{-1/2} \widehat{\boldsymbol{w}}_{12}$. The optimal β is used to generate the optimal $\boldsymbol{w}_{12} = \boldsymbol{W}_{11} \beta_{\text{opglasso}}$. Initially, $\boldsymbol{W}_{\text{opglasso}}$ is set to $\boldsymbol{W} = \widehat{\boldsymbol{W}}_{\text{cov}} + \rho \boldsymbol{I}_n$. We used the algorithm as implemented in the `glasso` R package.

Aggregated-down approach

We introduce a new reconciliation method called **aggregated-down** similar in complexity to bottom-up and top down.

It is essentially a localized top-down approach, where the disaggregating proportions are calculated based on the node above and not based on the top-most aggregated level.

Motivation: proportions calculated using values that are closer in the hierarchical structure should be more accurate than when using values that are further away.

We denote the corresponding disaggregating proportions by q_j with $\boldsymbol{q} = (q_1, \dots, q_n)'$ being the vector of proportions. The mapping matrix is defined as

$$\boldsymbol{P}_{\mathrm{AD}} = \begin{bmatrix} \boldsymbol{Q} & \mathbf{O}_{n \times (n-1)} \end{bmatrix},$$

where $\boldsymbol{Q} = \text{Antidiag}(\boldsymbol{q})$ is a $n \times n$ anti-diagonal matrix with the elements on the anti-diagonal, starting from top to bottom, i.e.

$$\boldsymbol{Q} = \begin{bmatrix} 0 & \dots & 0 & q_1 \\ 0 & \dots & q_2 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ q_n & 0 & \dots & 0 \end{bmatrix}.$$

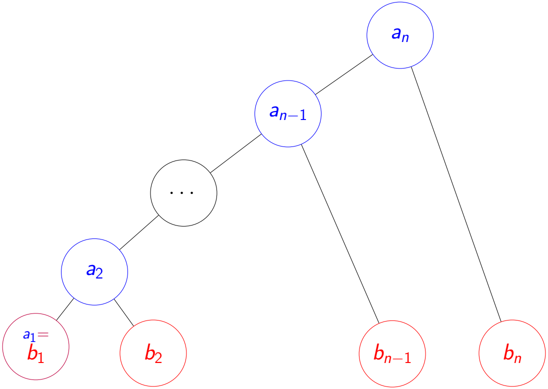
Aggregated-down approach

Hence the bottom-level values are obtained as:

$$\boldsymbol{P}_{\mathrm{AD}}\hat{\boldsymbol{y}} = \begin{bmatrix} 0 & \dots & 0 & q_1 & 0 & \dots & 0 \\ 0 & \dots & q_2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ q_n & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_2 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

We require that $q_1 = 1$ every time which corresponds to $\tilde{b}_1 = \hat{b}_1$.

Aggregated-down approach



Aggregated-down approach

Analogous to the top-down approach the proportions for aggregated-down can be estimated in various ways:

$$\boldsymbol{P}_{\mathrm{AD}} = \begin{bmatrix} \boldsymbol{Q}_{n \times n} & \mathbf{O}_{n \times (n-1)} \end{bmatrix}$$

$$\boldsymbol{Q}_{n \times n} = \mathrm{Antidiag}(\mathbf{q}) = \mathrm{Antidiag}((q_1, \dots q_n)')$$

$$q_1 = 1$$

Average ratio: $\widehat{q}_{\mathrm{ar},j} = \frac{1}{T} \sum_{t=1}^{T} \frac{a_{j,t} - a_{j-1,t}}{a_{j,t}} = \frac{1}{T} \sum_{t=1}^{T} \frac{b_{j,t}}{a_{j,t}}$

Ratio of averages: $\widehat{q}_{\mathrm{ra},j} = \frac{\frac{1}{T} \sum_{t=1}^{T} a_{j,t} - a_{j-1,t}}{\frac{1}{T} \sum_{t=1}^{T} a_{j,t}} = \frac{\frac{1}{T} \sum_{t=1}^{T} b_{j,t}}{\frac{1}{T} \sum_{t=1}^{T} a_{j,t}}$

Forecasted values: $\widehat{q}_{\mathrm{fo},j} = \frac{\widehat{a_j} - \widehat{a_{j-1}}}{\widehat{a_j}}$ for $j > 1$,

Aggregated-down approach - simulation

We provide a simulation study to support the idea of closeness, i.e. proportions calculated using values that are closer in the hierarchical structure should be more accurate due to avoidance of error aggregation.

Simulation study design:

We simulated VAR(1) processes for the bottom-level values. We replicated the simulation 1000 times for each combination of the parameters:

Number of historical observations (N)	$N \in \{16, 64, 256\}$
Number of bottom-level values (n)	$n \in \{4, 16, 64\}$
Coefficient matrix ($\mathbf{\Phi}$)	$\mathbf{\Phi} = a\boldsymbol{I}_n$ for $a \in \{0.2, 0.5, 0.7, 0.95\}$
Error variance-covariance matrix (A)	$A = \boldsymbol{I}_n$

We also considered the case of correlated errors with the variance-covariance matrix $A = 0.3\boldsymbol{I}_n + 0.7\mathbf{1}\mathbf{1}'$ in combination with the coefficient matrix $\mathbf{\Phi} = 0.7\boldsymbol{I}_n$.

We generated the aggregated values from the bottom-level simulations and fitted an AR(1) process without intercept for each level of the hierarchy. We calculated the forecast accuracy from 1-step-ahead forecasts for each level via RMSE.

Aggregated-down approach - simulation

Simulation study results:

n	4			16			64		
N	**16**	**64**	**256**	**16**	**64**	**256**	**16**	**64**	**256**
BASE	1.403	1.388	1.322	2.241	2.268	2.239	4.335	4.126	4.152
BU	1.403	1.394	1.320	2.252	2.260	2.238	4.321	4.117	4.148
TDFO	18.039	8.070	1.615	55.658	99.499	74.052	24155.015	1606.731	605.168
ADFO	1.418	1.390	1.323	2.260	2.271	2.240	4.347	4.129	4.153
OPOLS	1.390	1.386	1.321	2.228	2.265	2.239	4.326	4.125	4.152
OPWLS	1.387	1.386	1.321	2.215	2.261	2.238	4.295	4.118	4.150
OPLAMBDA	1.386	1.387	1.321	2.213	2.261	2.238	4.293	4.118	4.150
OPSHRINK	1.394	1.388	1.321	2.230	2.266	2.239	4.349	4.132	4.155

- We left out the results for historical proportion calculations since they were always vastly inferior to the base case.
- The aggregated-down approach using forecasted values is very similar to the other methods, even optimal ones.
- The top-down approach using forecasted values is markedly inferior in the considered setups.
- The accuracy decreases as more bottom-levels (n) are added, as expected. However, with the top-down approach this is disproportionally the case. We believe this to be due to the effect of error aggregation.

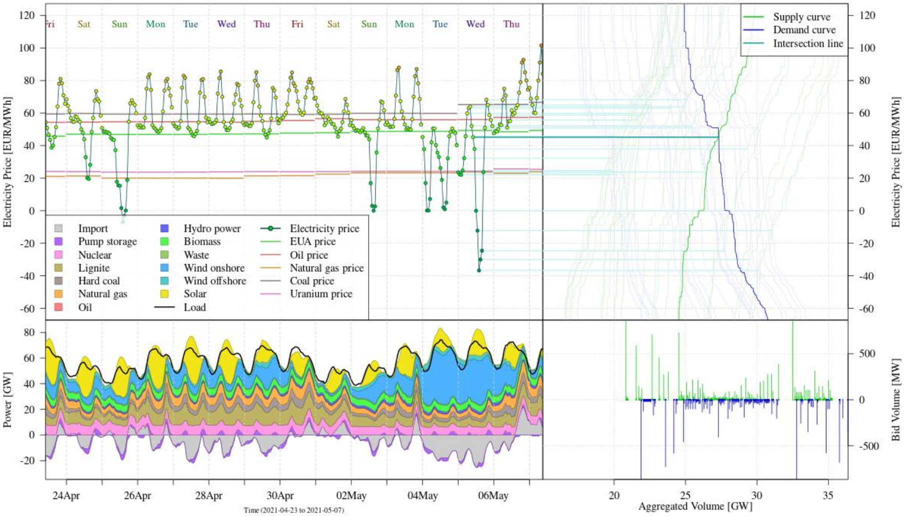
Overview

German day-ahead electricity market

- The European day-ahead electricity is a daily, blind auction market where electricity prices for each hour of the next day are determined.
- The auction closes at 12:00 CET of each day, up until when the market participants can submit buy (bid) and sell (ask) orders for each hour of the next day.
- These bids are the volumes that buyers are willing to buy and that producers are willing to sell at certain prices.
- For the considered time range of data, bid volumes can be specified for prices ranging from -500 EUR/MWh up to 3000 EUR/MWh with an increment of 0.1 EUR/MWh (since 2022-05-11 the upper price limit was increased to 4000 EUR/MWh).
- After gate closure, the volumes bid across all participating countries are aggregated and unique prices and volumes are generated for each separate market area.
- The market clearing results along with 24 supply and demand curves are published shortly after gat closure at 12:00 CET.

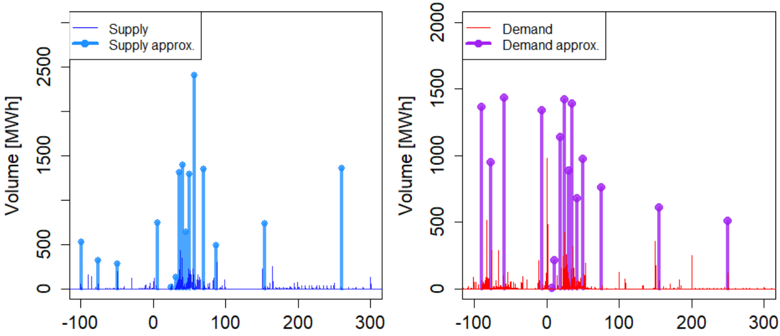
German day-ahead electricity market

- All major relevant aspects are illustrated in the figure below: Hourly price time series with supply and demand curves, volumes and power production.



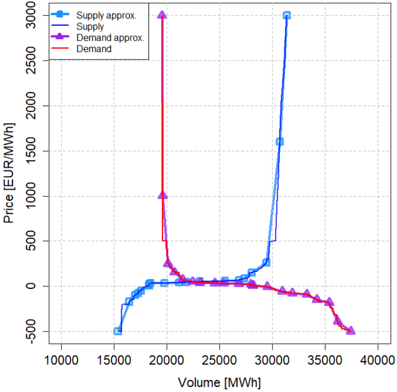
The X-Model

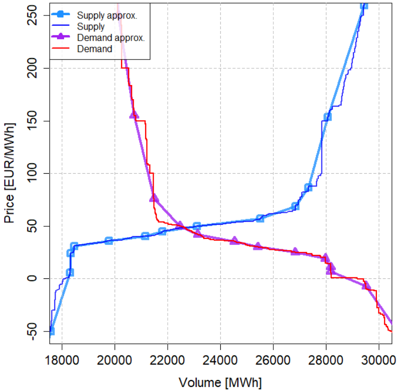
- The X-Model introduced by *F. Ziel and R. Steinert* forecasts day-ahead electricity prices as the intersection of supply and demand curves.
- Prices are grouped into price classes, thus reducing the dimensionality of the curves, e.g. from 350001 prices to ca. 40 price classes (ca. 20 per curve). The prices are split into price classes by inverting the supply and demand curves at a pre-specified grid of equidistant volumes.
- Thus the 40 marginal bids represent the sum of bids within their price range:



The X-Model

- The hourly bids are aggregated to produce the day-ahead supply and demand curves. The price classes which approximate the curves by ca. 20 points each are also shown.
- The characteristic step-like appearance of the curve is the result of the many zero-volume prices.





The X-Model - Model setup

Model Setup:

- The response variable is the sum of the volume within a price class, modeled separately for every hour and class. For 40 price classes, a total of 40×24 regressions have to be done to forecast curves for all hours of the next day.
- In the original X-Model paper,each price class volume was modeled marginally, i.e. the supply and demand curves were generated by cumulatively summing up the forecasted values. This is equivalent to using bottom-up reconciliation approach.
- In our paper, we extended this and modeled both the marginal and cumulative responses and use them to compare and contrast the different reconciliation approaches.

The X-Model - Model setup

- To model the responses, we used a combination of autoregressive and external regressors.
- Let then $X^{(c)}_{S,d,h}$ and $X^{(c)}_{D,d,h}$ be the supply and demand volumes at day d and hour h of price class c of the price classes generated for the supply and demand curves. These will constitute the bulk of the regressor matrix as each price class volume will depend on its lags according to a specific lag structure.
- Let the external regressors be denoted by $X^{(1)}_{X,d,h}, \dots, X^{(M_X)}_{X,d,h}$ for a total of M_X external regressors, and consist of prices for coal, gas, oil and CO_2 emissions (EUAs), the day-ahead prices and volumes of the previous day, as well as the day-ahead forecasts for the country-wide load, solar, onshore wind, and offshore wind production. The day-ahead data was taken from *www.epexspot.com* and the forecasts from *www.entsoe.eu*.

The X-Model - Model setup

- Let M_S and M_D be the number of price classes for the supply and demand curve respectively.
- All regressors can be compactly written as

$$\mathbf{X}_{d,h} = (X_{1,d,h}, \dots, X_{M,d,h})' = \left(\left(X_{S,d,h}^{(c \in C_S)} \right), \left(X_{D,d,h}^{(c \in C_D)} \right), \left(X_{X,d,h}^{(c \in C_X)} \right) \right)',$$

  where C_S and C_D is the set of price classes for the supply and demand side respectively, C_X is the set of external regressors $C_X = \left\{ X_{X,d,h}^{(1)} 1, \dots, X_{X,d,h}^{(M_X)} \right\}$, and $M = M_S + M_D + M_X$.
- To capture the weekly seasonality, we also included dummy regressors for every day of the week. Let these be denoted by a function $W_k(d)$ that returns the day of the week of day d.

The X-Model - Model setup

- The full model can be written out as

$$X_{m,d,h} = \sum_{k=1}^{M} \sum_{j=1}^{24} \sum_{k \in \mathcal{I}_{m,h}(l,j)} \phi_{m,h,l,j,k} X_{l,d-k,j} + \sum_{k=2}^{7} \psi_{m,h,k} W_k(d) + \varepsilon_{m,d,h},$$

  for $m \in \{1, \dots, M_S + M_D\}$ and $\mathcal{I}_{m,h}(l,j)$ represents the sets of possible lags, which we defined as

$$\mathcal{I}_{m,h}(l,j) = \begin{cases} \{1, \dots, 30\}, \text{for } m = l \text{ and } h = j \\ \{1, \dots, 8\}, \text{ for } (m = l \text{ and } h \neq j) \text{ or } (m \neq l \text{ and } h = j) \\ \{1\}, \text{ for } m \neq l \text{ and } h \neq j \end{cases} .$$

- We fitted the models using LASSO as implemented in the **glmnet** R package.

The X-Model - Data & results

Data:

- We conducted a day-ahead rolling window forecasting study for each day between 2019-01-01 and 2019-06-30.
- We used a rolling window length of $730 = 2 \times 365$ days to forecast each day.
- The price classes were generated once for the first day-ahead forecast using the 2017 and 2018 data, and were kept constant throughout the study.
- All data were hourly except for coal, gas, oil, and EUA prices which were daily.
- We used marginal values as regressors to forecast the marginal values $\widehat{b}_i$ and cumulative values as regressors to forecast the cumulative values $\widehat{a}_i$.

The X-Model - Data & results

- To measure the forecasting accuracy we used the mean absolute errors (MAEs)

$$MAE_m^{test} = \frac{1}{24 \cdot \#(\mathcal{D})} \sum_{d \in \mathcal{D}} \sum_{h=0}^{23} |X_{m,d,h} - \widehat{X}_{m,d,h}|$$

and the root mean square errors (RMSEs)

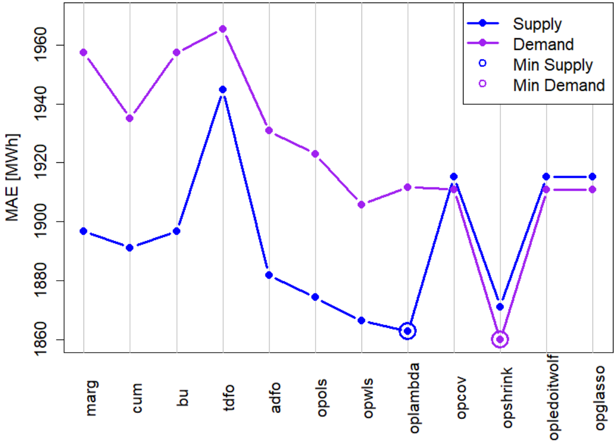
$$RMSE_m^{test} = \frac{1}{\#(\mathcal{D})} \sum_{d \in \mathcal{D}} \sqrt{\frac{1}{24} \sum_{h=0}^{23} (X_{m,d,h} - \widehat{X}_{m,d,h})^2}$$

where $\mathcal{D}$ is a set containing all 181 forecasted days, $\#(\cdot)$ is a function that returns the number of elements in a set, and $\widehat{X}_{m,d,h}$ represents the respective forecasted value.

The X-Model - Data & results

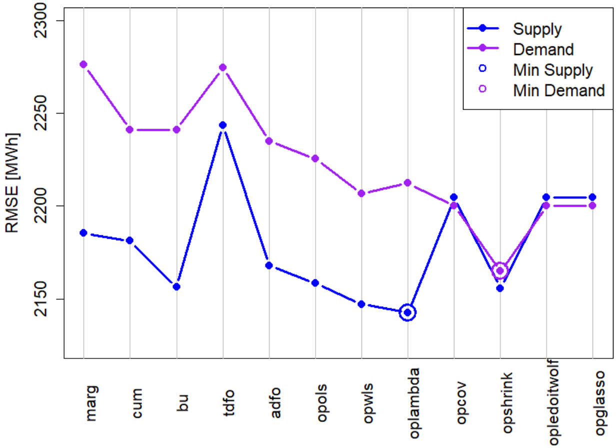
Results:

- Results for the top-down and aggregated-down approaches using historical proportions were not included due to very high inaccuracy compared to the base case. MAE:



The X-Model - Data & results

- Results for the top-down and aggregated-down approaches using historical proportions were not included due to very high inaccuracy compared to the base case. RMSE:



The X-Model - Data & results

- Aggregated-down using forecasted values yields better results than the bottom-up case on average
- Aggregated-down is superior to bottom-up and top-down when using forecasted values for the proportions.
- For the Supply curve: no approach consistently yielding the lowest errors. The optimal WLS reconciliation approach yielded the lowest errors for the largest number of classes, followed by the shrinkage and lambda approaches.
- For the Demand curve: the results were consistent for all classes with the lowest errors achieved by the optimal shrinkage reconciliation approach.

Overview

Conclusion

- We discussed the hierarchical structure of aggregated curves and different representations.
- We presented a number of reconciliation methods including established bottom-up, top-down, and minium-trace optimal reconciliation approaches in an aggregated curves setting
- We introduced a new *aggregated-down* approach which has a methodological complexity comparable to the bottom-up and top-down approaches.
- We provided theoretical insights that under mild assumptions on the forecasting and reconciling method, the reconciling result is independent of the representation of the curve.
- The approaches were then applied in a simulation study and for forecasting supply and demand curves of the German day-ahead electricity market.
- We showed how the considered reconciling approaches for aggregated curves can improve the forecasting accuracy compared to standard approaches.

Conclusion

- The results show that there is not a single reconciliation method to outperform the others every time.
- The optimal reconciliation approaches, more specifically the shrinkage, WLS, and lambda approaches, showed to have the best results most of the time, yielding considerable improvements compared to the bottom-up base case. Which reconciliation method mostly improves a forecast will likely be specific to the data.
- We see potential in considering multiple methods since even simple approaches such as aggregated-down with forecasted values yield improvements at certain points of the curve compared to the bottom-up approach, which is current state-of-the-art.
- We conclude that it is worth using the aggregated-down approach as a simple benchmark method superior to top-down.
- We conclude that it is important to have access to all base forecasts in order to calculate the proportions since these lead to substantial improvements in forecasts in contrast to using only historical values.

Conclusion

- Our study could further be extended by including more recent approaches, such as machine-learning-based and conditional coherency reconciliation methods.
- Averaging or usage of different reconciliation approaches for each point of the aggregated curve could also be considered, especially if coherency is not necessary.

Overview

References I

Aitchison, John, and James AC Brown. 1954. "A Synthesis of Engel Curve Theory." *The Review of Economic Studies* 22 (1): 35–46.

Athanasopoulos, George, Rob J. Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. 2017. "Forecasting with Temporal Hierarchies." *European Journal of Operational Research* 262 (1): 60–74. https://doi.org/https://doi.org/10.1016/j.ejor.2017.02.046.

Banks, James, Richard Blundell, and Arthur Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *Review of Economics and Statistics* 79 (4): 527–39.

Brégère, Margaux, and Malo Huard. 2022. "Online Hierarchical Forecasting for Power Consumption Data." *International Journal of Forecasting* 38 (1): 339–51.

Di Fonzo, Tommaso, and Daniele Girolimetto. 2021. "Forecast Combination Based Forecast Reconciliation: Insights and Extensions." *arXiv Preprint arXiv:2106.05653*.

References II

Eiteman, Wilford J, and Glenn E Guthrie. 1952. "The Shape of the Average Cost Curve." *The American Economic Review* 42 (5): 832–38.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. "Sparse Inverse Covariance Estimation with the Graphical Lasso." *Biostatistics* 9 (3): 432–41.

Gross, Charles W, and Jeffrey E Sohl. 1990. "Disaggregation Methods to Expedite Product Line Forecasting." *Journal of Forecasting* 9 (3): 233–54.

Gürkaynak, Refet S, Brian Sack, and Jonathan H Wright. 2007. "The US Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics* 54 (8): 2291–2304.

Haben, Stephen, Julien Caudron, and Jake Verma. 2021b. "Probabilistic Day-Ahead Wholesale Price Forecast: A Case Study in Great Britain." *Forecasting* 3 (3): 596–632.

———. 2021a. "Probabilistic Day-Ahead Wholesale Price Forecast: A Case Study in Great Britain." *Forecasting* 3 (3): 596–632. https://doi.org/10.3390/forecast3030038.

References III

Hollyman, Ross, Fotios Petropoulos, and Michael E Tipping. 2021. "Understanding Forecast Reconciliation." *European Journal of Operational Research* 294 (1): 149–60.

Huard, Malo, Rémy Garnier, and Gilles Stoltz. 2020. "Hierarchical Robust Aggregation of Sales Forecasts at Aggregated Levels in e-Commerce, Based on Exponential Smoothing and Holt's Linear Trend Method." *arXiv Preprint arXiv:2006.03373*.

Hull, John C. 2003. *Options Futures and Other Derivatives*. Pearson Education India.

Hyndman, Rob J, and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. OTexts. https://otexts.com/fpp2/.

Hyndman, Rob J, and Md Shahid Ullah. 2007. "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics & Data Analysis* 51 (10): 4942–56.

Kulakov, Sergei. 2020. "X-Model: Further Development and Possible Modifications." *Forecasting* 2 (1): 20–35.

References IV

Ledoit, Olivier, and Michael Wolf. 2004a. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis* 88 (2): 365–411.

———. 2004b. "Honey, i Shrunk the Sample Covariance Matrix." *The Journal of Portfolio Management* 30 (4): 110–19.

Mankiw, N Gregory. 2014. *Principles of Economics*. Cengage Learning.

Marshall, Alfred. 2009. *Principles of Economics: Unabridged Eighth Edition*. Cosimo, Inc.

Mestre, Guillermo, José Portela, Antonio Muñoz San Roque, and Estrella Alonso. 2020. "Forecasting Hourly Supply Curves in the Italian Day-Ahead Electricity Market with a Double-Seasonal SARMAHX Model." *International Journal of Electrical Power & Energy Systems* 121: 106083.

Narajewski, Michał, and Florian Ziel. 2022. "Optimal Bidding in Hourly and Quarter-Hourly Electricity Price Auctions: Trading Large Volumes of Power with Market Impact and Transaction Costs." *Energy Economics* 110: 105974.

References V

Park, Mijung, and Marcel Nassar. 2014. "Variational Bayesian Inference for Forecasting Hierarchical Time Series." In *International Conference on Machine Learning (ICML), Workshop on Divergence Methods for Probabilistic Inference, Beijing*. Citeseer.

Petropoulos, Fotios, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, et al. 2022. "Forecasting: Theory and Practice." *International Journal of Forecasting*.

Phillips, A. W. 1958. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957." *Economica* 25 (100): 283–99. http://www.jstor.org/stable/2550759.

Pindyck, Robert S, Daniel L Rubinfeld, and Prem L Mehta. 1995. *Microeconomics*. Vol. 4. Prentice Hall Englewood Cliffs, NJ.

Schäfer, Juliane, and Korbinian Strimmer. 2005. "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4 (1).

References VI

Shang, Han Lin, and Rob J Hyndman. 2011. "Nonparametric Time Series Forecasting with Dynamic Updating." *Mathematics and Computers in Simulation* 81 (7): 1310–24.

———. 2017. "Grouped Functional Time Series Forecasting: An Application to Age-Specific Mortality Rates." *Journal of Computational and Graphical Statistics* 26 (2): 330–43.

Soloviova, Mariia. n.d. "Modeling of Supply and Demand Curves for Day-Ahead Electricity Market." *Seminario Dottorato 2019/20*, 113.

Soloviova, Mariia, and Tiziano Vargiolu. 2021. "Efficient Representation of Supply and Demand Curves on Day-Ahead Electricity Markets." *Journal of Energy Markets* 14 (1).

Spiliotis, Evangelos, Mahdi Abolghasemi, Rob J Hyndman, Fotios Petropoulos, and Vassilios Assimakopoulos. 2021. "Hierarchical Forecast Reconciliation with Machine Learning." *Applied Soft Computing*, 107756.

References VII

Van Erven, Tim, and Jairo Cugliari. 2015. "Game-Theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts." In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, 297–317. Springer.

Wickramasuriya, Shanika L, George Athanasopoulos, and Rob J Hyndman. 2019. "Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization." *Journal of the American Statistical Association* 114 (526): 804–19. https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825.

Ziel, Florian, and Rick Steinert. 2016. "Electricity Price Forecasting Using Sale and Purchase Curves: The x-Model." *Energy Economics* 59: 435–54. https://doi.org/https://doi.org/10.1016/j.eneco.2016.08.008.

———. 2018. "Probabilistic Mid-and Long-Term Electricity Price Forecasting." *Renewable and Sustainable Energy Reviews* 94: 251–66.