Série dos Seminários de Acompanhamento à Pesquisa

DEI
DEPARTAMENTO DE ENGENHARIA INDUSTRIAL

Número 23 | 09 2021

Robust reconciliation for hierarchical time series

Autor:

Maurício Franca Lila

PUC RIO

Série dos Seminários de Acompanhamento à Pesquisa

Número 23 | 09 2021

Robust reconciliation for hierarchical time series

Autor:

Maurício Franca Lila

Orientador: Fernando Cyrino

CRÉDITOS:

SISTEMA MAXWELL / LAMBDA
https://www.maxwell.vrac.puc-rio.br/

Organizadores: Fernanda Baião / Soraida Aguilar

Layout da Capa: Aline Magalhães dos Santos

Outline

Introduction

- The demand for information to support an effective decision-making process is an increasing issue. The society and politicians want to understand the effects of policies on certain groups of interest, especially for hierarchical or grouped domains of interest.

- Time series sometimes offer special features, such as the natural disaggregation of their components according to a hierarchical structure.

Introduction

- Hierarchical forecasting methods can take advantage over such structure by considering the base forecasts reconciliation, producing results which are usually unbiased and more accurate than the ones provided by standard methods.

- When forecasting series aggregated according to a hierarchical structure one should not overlook their aggregation constraints.

Introduction

- Early studies were basically focused on two main strategies: the *top-down* and the *bottom-up*.
- The former aims to provide base forecasts for the series on the most aggregate level of the hierarchy and then produce forecasts for series situated in the lower levels using weighting systems.
- The *bottom-up* approach, in turn, works in the opposite way, i.e. by forecasting at the most granular level of the hierarchy and then adding up these forecasts to the top.
- The *middle-out* is a hierarchical level dependent approach. From a given stage of the hierarchy, a model is estimated and the forecasts above are obtained through the *bottom-up* approach while lower level forecasts are obtained using *top-down*.

Introduction

- Another approach, called optimal combination (or reconciliation) that addresses to the topic of hierarchical forecasting by taking into account forecasts from all levels of the hierarchy.
- The idea is based on a regression model that combines a set of incoherent forecasts into a set of base forecasts for the lowest hierarchical level.
- The mapping from a set of incoherent forecasts to a coherent space is called reconciliation, where forecasts sum up properly (Hyndman et al., 2011).

Introduction

- Wickramasuriya et al. (2019) proposed the Minimum Trace (MinT) reconciliation approach that minimizes the sum of variances of the reconciled forecast errors under the assumption of unbiasedness.
- Recently, Machine Learning techniques were implemented to derive the combination weights for the forecasts across the various aggregation levels (Spiliotis et al., 2020).
- In addition, a probabilistic framework was enhanced in order to deal with uncertainty from an inference based approach (Taieb et. al, 2017).

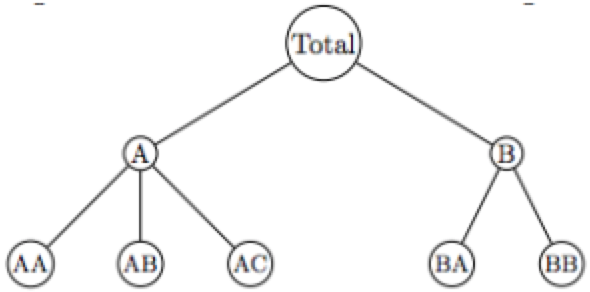
HTS: Basic strategies and state-of-the-art

In order to characterize a HTS, consider $\mathrm{y_t}$ a vector of size m having observations from all hierarchical levels at time t. It is possible to define a summing matrix S of dimension $m \times n$ such that,

$$\mathrm{y_t} = \mathrm{S}\mathrm{y_t^b} \tag{1}$$

where $\mathrm{y_t^b}$ is a n-vector containing the observations in the most disaggregated level of the hierarchy.

HTS: Basic strategies and state-of-the-art



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

Figure 1: Source: Hymdman et al.(2018)

A similar structure can be defined for forecast reconciliation of hierarchical time series. Consider $\hat{\mathrm{y}}_{t+h|t}$ a vector of h steps ahead forecasts with the same arrangement as $\mathrm{y_t}$. Thus, for a given matrix P of dimension $n \times m$, we have the following equation

$$\tilde{\mathrm{y}}_{t+h|t} = \mathrm{SP}\hat{\mathrm{y}}_{\mathrm{t+h|t}} \tag{2}$$

where $\tilde{\mathrm{y}}_{t+h|t}$ are the reconciled forecasts. The SP matrices represent the reconciliation process, which maps independent (or incoherent) forecasts into coherent ones.

HTS: Basic strategies and state-of-the-art

Depending on how P is structured, it is possible to reproduce several traditional hierarchical forecasting approaches.

By letting $\mathrm{P} = \left[0_{\mathrm{n}\times(\mathrm{m}-\mathrm{n})}|\mathrm{I}_{\mathrm{n}}\right]$ where $0_{n\times(m-n)}$ is a null matrix, one can reproduce the *bottom-up* approach.

Conversely, *top-down* forecasts can be achieved by making $\mathrm{P} = \left[\mathrm{p}|0_{\mathrm{n}\times(\mathrm{m}-1)}\right]$, where p is the set of proportions of forecasts .

HTS: Basic strategies and state-of-the-art

Common approaches to obtain these proportions are described as follows:

- Average of historical proportions or Gross-Sohl method A (TDGSA) :

$$p_j = \frac{1}{T} \sum_{t=1}^{T} \frac{y_{j,t}}{y_t} \tag{3}$$

- Proportion of historical averages or Gross-Sohl method F (TDGSF):

$$p_j = \sum_{t=1}^{T} \frac{y_{j,t}}{T} \Big/ \sum_{t=1}^{T} \frac{y_t}{T} \tag{4}$$

  for $j = 1, \ldots, m$.

- Proportion of forecasts (TDFP):

$$p_j = \prod_{\ell=0}^{K-1} \frac{\hat{y}_{j,h}^{(\ell)}}{\hat{S}_{j,h}^{(\ell+1)}} \tag{5}$$

HTS: Basic strategies and state-of-the-art

Wickramasuriya et al. (2019) proposed the Minimum *MinT* reconciliation approach.

It aims to find a matrix P that it minimizes $tr(\mathrm{S}\mathrm{G}\mathrm{W}_\mathrm{h}\boldsymbol{G}'\mathrm{S}')$ subject to $\mathrm{SPS} = \mathrm{S}$, the unbiasedness condition.

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

Regardless of the approach selected for estimating W_h, the optimal reconciled forecasts are given by

$$\tilde{\mathrm{y}}_h = \mathrm{S}(\mathrm{S}'\mathrm{W}_\mathrm{h}^{-1}\mathrm{S})^{-1}\mathrm{S}'\mathrm{W}_\mathrm{h}^{-1}\hat{\mathrm{y}}_\mathrm{h}. \tag{6}$$

This equation can be written in another fashion as follows.

$$\tilde{\mathrm{y}}_h = \mathrm{S}\hat{\beta}_\mathrm{h}. \tag{7}$$

HTS: Basic strategies and state-of-the-art

In order to use this approach in practice, it is necessary to estimate $\mathrm{W_h}$.

- OLS - $\mathrm{W_h} = \mathrm{k_h I}$ where $k_h > 0$.
- $WLS(v)$ - $\mathrm{W_h} = \mathrm{k_h diag}(\hat{\mathrm{W}}_1)$ where $k_h > 0$ and $\hat{\mathrm{W}}_1$
- $WLS(s)$ - $\mathrm{W_h} = \mathrm{k_h}\Lambda$ where $k_h > 0$, $\Lambda = \mathrm{diag}(\mathrm{S1})$ and 1 is a vector of ones
- MinT(Sample) - $\mathrm{W_h} = \mathrm{k_h W_1}$.
- MinT(Shrink) - $\mathrm{W_h} = \mathrm{k_h}\hat{\mathrm{W}}^*_{1,\mathrm{D}}$ where $k_h > 0$, and $\hat{\mathrm{W}}^*_{1,D} = \lambda\hat{\mathrm{W}}_{1,D} + (1-\lambda)\hat{\mathrm{W}}_1$

The Robust Reconciliation for HTS

- As observed, all methods described in the previously are obtained following the same guidelines of linear regression models.

- In such cases, a parameter estimation method which is less affected by data imperfections and/or contamination is desirable.

- Techniques based on M-estimators are useful in this context.

The Robust Reconciliation for HTS

Let the residuals from the reconciliation process be defined as

$$\mathrm{y_h} - \tilde{\mathrm{y}}_\mathrm{h} = \epsilon(\beta_\mathrm{h}) \tag{8}$$

Consider a function ρ having the following properties:

- Nonnegative, $\rho(Z) \geq 0$.
- $\rho(0) = 0$.
- Symmetric, $\rho(Z) = \rho(-Z)$.
- Monotone in $|Z_i|$, $\rho(Z_i) \geq \rho(Z_{i'})$ for $|Z_i| > |Z_{i'}|$

Then, the M-estimator based on the residuals from the equation (8) is given as

$$\hat{\beta}_{h,M} = \arg\min_{\beta_h} \sum_{i=1}^{n} \rho(\epsilon_i(\beta_h)). \tag{9}$$

In order to solve this minimization problem we need to find $\psi(\cdot) = \rho'(\cdot))$ which is the influence function.

The Robust Reconciliation for HTS

In this work, we consider the Huber influence function in light of its desirable properties for computational convergence.

$$\rho(z) = \begin{cases} z^2, & \text{if } |z| < c; \\ |2z|c - c^2, & \text{if } |z| \geq c \end{cases} \tag{10}$$

$$\psi(z) = \begin{cases} z, & \text{if } |z| < c; \\ c[\operatorname{sgn}(z)], & \text{if } |z| \geq c \end{cases} \tag{11}$$

for a given constant c. We also need to define a set weights $w(z)$ in order to obtain the optimal solution. In this case we have the weights given by

$$w(z) = \frac{\psi(z)}{z} \tag{12}$$

In this case, an iterative procedure called Iteratively Reweighted Least Squares (IRLS) is required.

The Robust Reconciliation for HTS

Some M-estimators are influenced by the scale of the residuals, so a scale-invariant version of the M-estimator is used:

$$\hat{\beta}_{\mathrm{M},h} = \arg\min_{\beta_h} \sum_{i=1}^{n} \rho\left(\frac{\epsilon_i(\beta_h)}{\sigma}\right), \tag{13}$$

There are two common ways of estimating σ. The first one is based on the mean absolute deviation (MAD).

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{median\{|\epsilon_i(\beta_h)|\}}{0.6745} \tag{14}$$

the second approach, also called Huber's Proposal 2 comes from the solution of:

$$\frac{1}{n-p}\sum_{i=1}^{n}\psi^2\left[\frac{\epsilon_i(\beta_h)}{\hat{\sigma}}\right] = E_Z[\psi^2(\epsilon)] \tag{15}$$

where $E_Z[\psi^2(\epsilon)]$ is the expected value of ψ^2 when ϵ has standard normal distribution and p is the number of regression parameters.
In the results section we denote these approaches by HUBER (1) and HUBER (2) respectively.

Experimental Setup - Dataset

Our study is based on the Brazilian Monthly Labor Force Survey The survey covered the metropolitan areas of Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre.

For our empirical experiments, we opted to use data from March 2002 to February 2016, the last date before the substantial changes in the way that data were collected and processed took place.

In this study we use data from the number of unemployed people in a hierarchical fashion.

Hierachy	Number of series
Overall	1
Metropolitan Areas	6
Sex	12
Total	19

Table 1: Number of time series according to hierarchical levels

Experimental Setup - Dataset

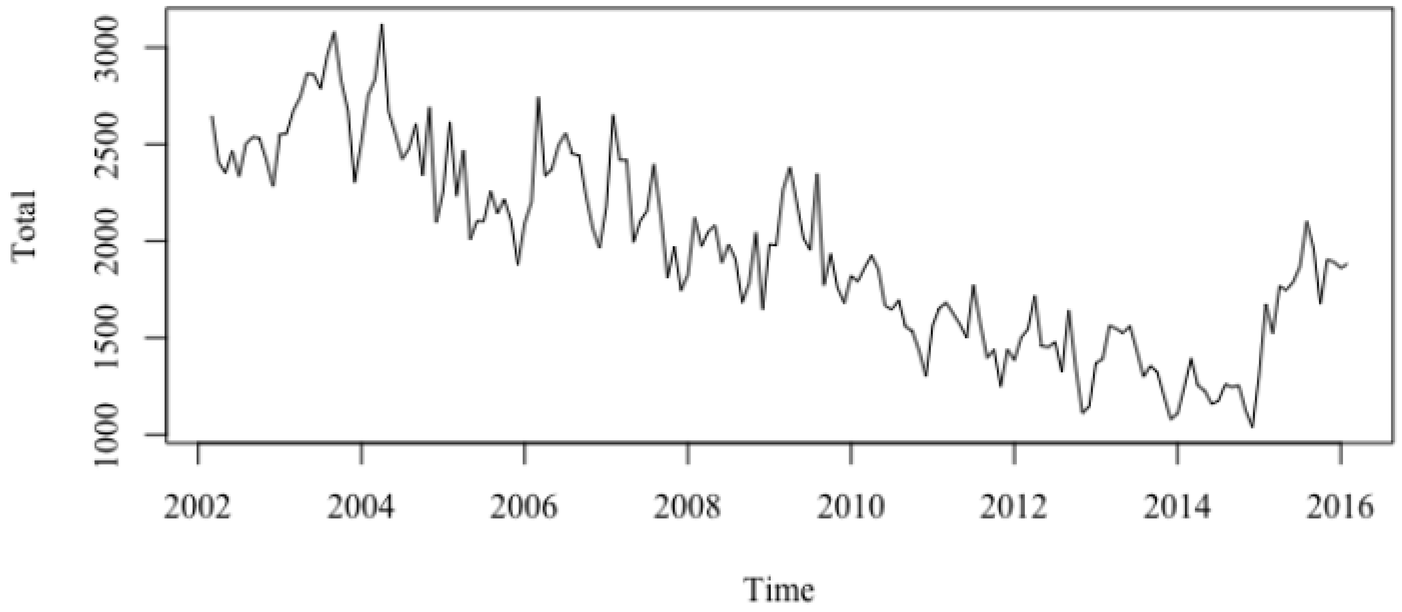


Figure 2: Overall number of unemployed people (in thousands) in six metropolitan areas

Experimental Setup - Dataset

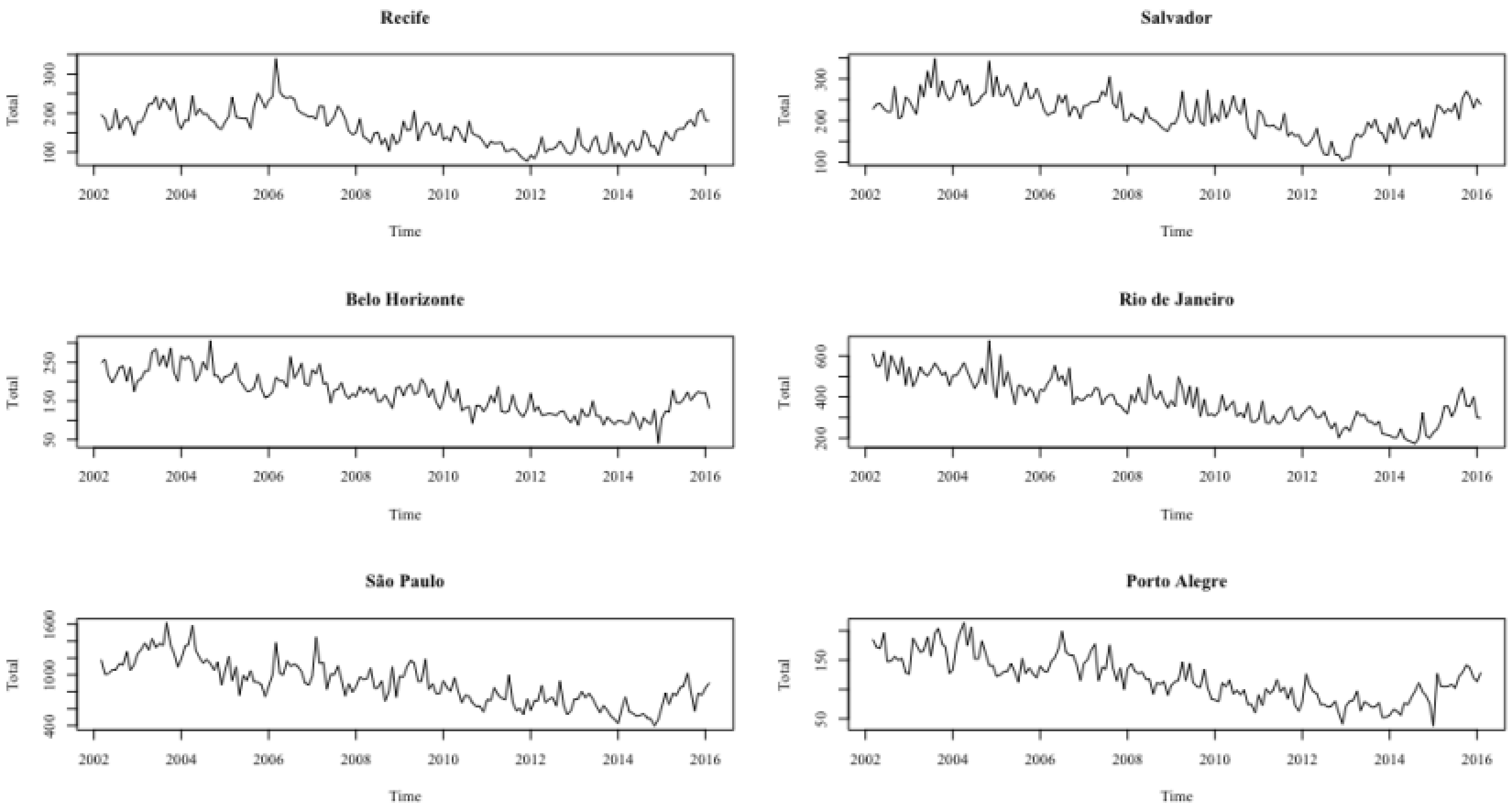


Figure 3: Number of unemployed people (in thousands) by metropolitan areas

Experimental Setup - Dataset

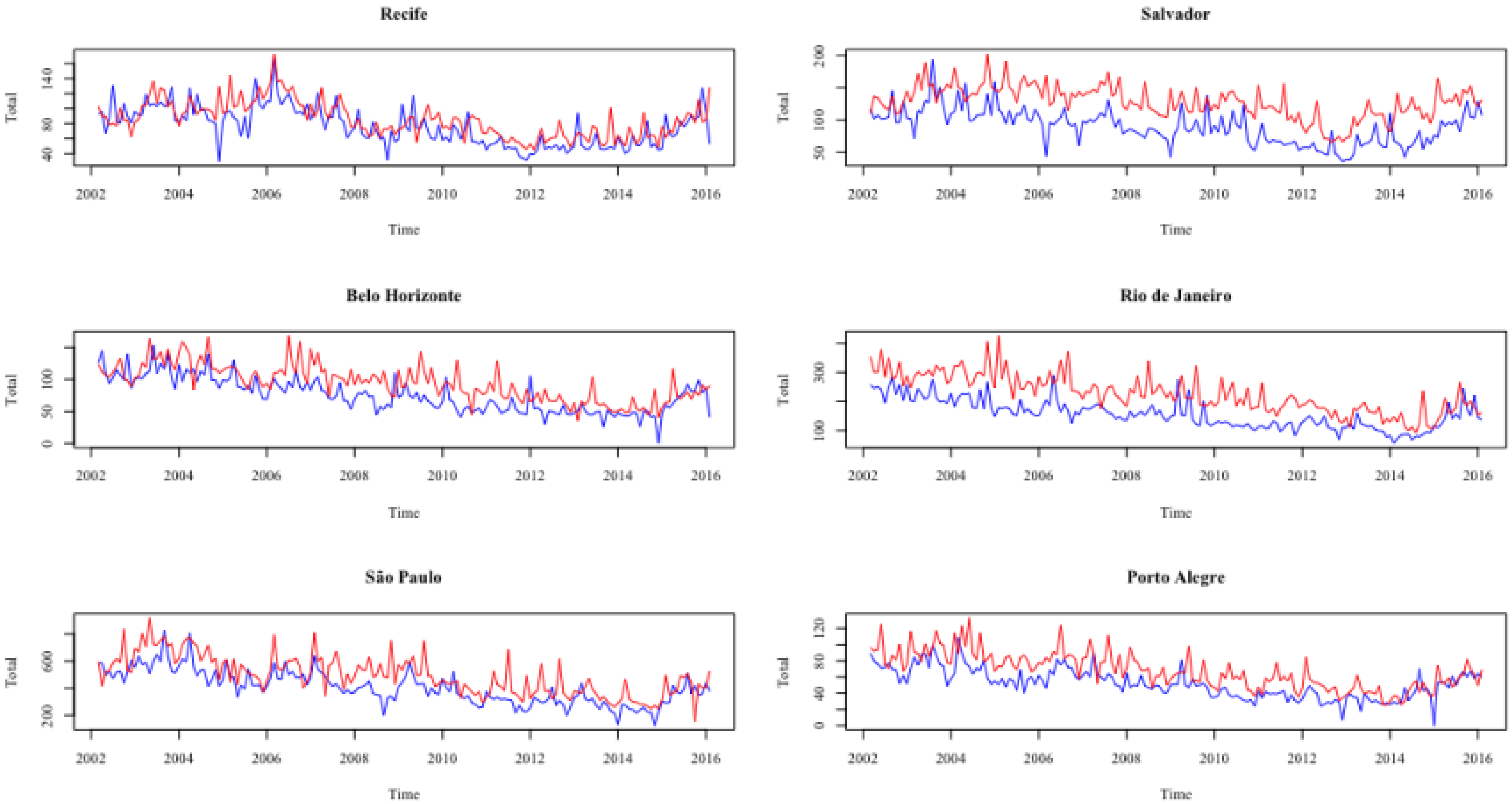


Figure 4: Number of unemployed people (in thousands) for each metropolitan areas by sex (male in blue and female in red)

Experimental Setup - Base Forecasting strategies

The reconciliation strategy relies on producing a set of base forecasts and then combining then into coherent ones.

In order to implement the strategies described previously, we choose a well-known forecasting approach: exponential smoothing models.

We rely on an automatic model selection routine commonly known as ETS - an acronym for Error, Trend and Seasonality.

ETS(**E**xponen**T**ial **S**moothing)
An automatic algorithm, implemented in the `ets()` function of the `forecast` package in the software `R` , selects the best state space formulation for each series from a set of 30 possible combinations.

Experimental Setup - Assessment Metrics

Assessment metrics are an important set of tools for model selection in the context of time series.

Mean Squared Error (MSE)

$$MSE = \frac{1}{h}\sum_{t=1}^{h}(\hat{y}_t - y_t)^2 \tag{16}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{h}\sum_{t=1}^{h}|\hat{y}_t - y_t|^2 \tag{17}$$

Experimental Setup - Assessment Metrics

Given that in this study several strategies of reconciliation will be tested to obtain predictions for the entire hierarchy , a relative measure to these metrics are obtained through independent models for each hierarchical level

$$RelMSE_{i,h} = \frac{MSE_{i,h}^{rec}}{MSE_{i,h}^{base}} \tag{18}$$

$$RelMAE_{i,h} = \frac{MAE_{i,h}^{rec}}{MAE_{i,h}^{base}} \tag{19}$$

Experimental Setup - Assessment Metrics

Since the combination of techniques and levels of the hierarchy generates a large number of results, the geometric mean within each level of the hierarchy provides a measure of improvement in terms of MSE and MAE.

$$AveRelMSE = \sqrt[\#L]{\prod_{i \in L} RelMSE_i} \tag{20}$$

$$AveRelMAE = \sqrt[\#L]{\prod_{i \in L} RelMAE_i} \tag{21}$$

where L is the corresponding level of the hierarchy

Experimental Setup - Assessment Metrics

We implemented an evaluation on a rolling forecasting origin using the last 12 months period in order to produce the cross-validation for different steps ahead.
The advantage of using these metrics is that $(1 - AveRelMSE) \times 100$ yields the percentage of improvement on MSE over the base forecasting strategy. Analogously we have the same interpretation for the MAE.

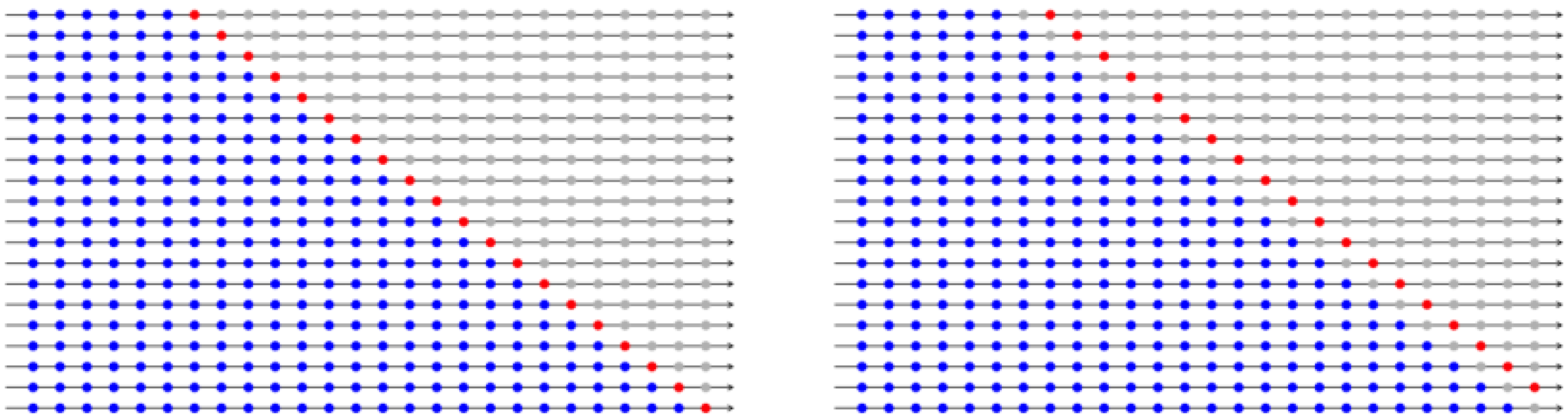


Figure 5: Evaluation on a rolling forecasting origin implemented for different reconciliation strategies

Results

Hierarchy	h					
	1	2	3	4	5	6
Overall	X	X	X	MinT-Shrink	MinT-Shrink	MinT-Shrink
	X	X	X	0.9991	X	0.9900
	X	X	X	OLS	X	OLS
	X	X	X	0.9997	X	0.9984
Metropolitan Areas	HUBER (2)	MinT-Shrink	HUBER (1)	MinT-Shrink	HUBER (1)	MinT-Shrink
	0.9606	0.9671	0.9648	0.9704	0.9838	0.9918
	OLS	HUBER (1)	MinT-Shrink	HUBER (1)	MinT-Shrink	HUBER (1)
	0.9642	0.9706	0.9651	0.9775	0.9925	0.9971
Sex	HUBER (2)	MinT-Shrink	MinT-Shrink	MinT-Shrink	HUBER (1)	MinT-Shrink
	0.9450	0.9399	0.9685	0.9743	0.9699	0.9758
	OLS	HUBER (1)	HUBER (1)	HUBER (1)	MinT-Shrink	HUBER (1)
	0.9478	0.9511	0.9705	0.9840	0.9724	0.9849

Table 2: Methods that presented the largest improvement in *AveRelMSE* and their corresponding *AveRelMSEs*. ETS as base forecasting method.

Results

Hierarchy	h					
	1	2	3	4	5	6
Overall	X	X	X	X	X	X
	X	X	X	X	X	X
	X	X	X	X	X	X
	X	X	X	X	X	X
Metropolitan Areas	MinT-Shrink	MinT-Shrink	HUBER (1)	HUBER (1)	HUBER (1)	HUBER (1)
	0.9843	0.9849	0.9725	0.9904	0.9898	0.9959
	HUBER (1)	HUBER (1)	MinT-Shrink	MinT-Shrink	HUBER (2)	MinT-Shrink
	0.9883	0.9813	0.9839	0.9906	0.9941	0.9987
Sex	MinT-Shrink	MinT-Shrink	HUBER (1)	HUBER (1)	HUBER (1)	HUBER (1)
	0.9688	0.9720	0.9819	0.9854	0.9798	0.9899
	HUBER (2)	HUBER (1)	MinT-Shrink	MinT-Shrink	MinT-Shrink	MinT-Shrink
	0.9733	0.9822	0.9842	0.9868	0.9872	0.9904

Table 3: Methods that presented the largest improvement in *AveRelMAE* and their corresponding *AveRelMAEs*. ETS as base forecasting method.

Conclusions and suggestions for future studies

- From an empirical evaluation, our results indicate that the robust reconciliation techniques provided a substantial contribution to an effective decision-making process.
- Although, these estimators were not uniformly dominant for all scenarios, they behave as good as the ones proposed in the state-of-the-art literature.
- We found considerable gains on accuracy according to the assessment metrics for the intermediate and lower levels of the hierarchy. We also provided an original contribution to the field of HTS adopting the framework of robust estimation.
- Of all robust reconciliation strategies adopted in this work, the HUBER (1) showed to be the most consistent strategy. The HUBER (2) approach, in turn, presented the best performance for short-term forecasting in terms of MSE.

Conclusions and suggestions for future studies

- In this work we used only forecasts from a single model strategy to compute the reconciled forecasts. Another option would be to consider forecast combination as base inputs so as to improve the final forecasts.

- Another extension to this work can be useful in the context of dynamic data with reduced dependencies on covariance structures in the process of reconciliation.

- It is also possible to evaluate the benefits of using alternative approaches related to Machine Learning in the processes of model fitting, such as Bagging or other ensemble methods, to reduce base forecasting errors.

- Finally, this contribution to HTS should be extended to other time series in order to provide a better understanding about this framework.

Bibliography I

George Athanasopoulos, Roman A Ahmed, and Rob J Hyndman. "Hierarchical forecasts for Australian domestic tourism". In: *International Journal of Forecasting* 25.1 (2009), pp. 146–166.

Charles W Gross and Jeffrey E Sohl. "Disaggregation methods to expedite product line forecasting". In: *Journal of Forecasting* 9.3 (1990), pp. 233–254.

Rob J Hyndman, Roman A Ahmed, et al. "Optimal combination forecasts for hierarchical time series". In: *Computational Statistics & Data Analysis* 55.9 (2011), pp. 2579–2589.

Rob J Hyndman, George Athanasopoulos, et al. "Package 'forecast'". In: *Online] https://cran. r-project. org/web/packages/forecast/forecast. pdf* (2020).

Bibliography II

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020. URL: https://www.R-project.org/.

Evangelos Spiliotis et al. "Hierarchical forecast reconciliation with machine learning". In: *ArXiv* abs/2006.02043 (2020).

Souhaib Ben Taieb, James W Taylor, and Rob J Hyndman. "Coherent probabilistic forecasts for hierarchical time series". In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 3348–3357.

Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman. "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization". In: *Journal of the American Statistical Association* 114.526 (2019), pp. 804–819.

Thank you!