Pontifícia Universidade Católica
do Rio de Janeiro

Maurício Franca Lila

Essays on Hierarchical Time Series Forecasting

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Engenharia de Produção, do Departamento de Engenharia Industrial da PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção.

Advisor : Prof. Fernando Luiz Cyrino Oliveira
Co-advisor: Dr. Erick Meira de Oliveira

Rio de Janeiro
April 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Maurício Franca Lila

Essays on Hierarchical Time Series Forecasting

Thesis presented to the Programa de Pós–graduação em Engenharia de Produção da PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção. Approved by the Examination Committee:

Prof. Fernando Luiz Cyrino Oliveira
Advisor
Departamento de Engenharia Industrial – PUC-Rio

Dr. Erick Meira de Oliveira
FINEP

Prof. Reinaldo Castro Souza
Departamento de Engenharia Industrial – PUC-Rio

Prof. Hélio Côrtes Vieira Lopes
Departamento de Informática – PUC-Rio

Prof. Lilian Manoel de Menezes Willenbockel
UL

Prof. Lupercio França Bessegato
UFJF

Prof. Paulo Jorge Canas Rodrigues
UFBA

Rio de Janeiro, April 28th, 2023

Maurício Franca Lila

Maurício Lila is a senior Statistician at the Brazilian Institute of Geography and Statistics (IBGE) and a Lecturer at the National School of Statistical Sciences. He graduated from the Rio de Janeiro State University of (UERJ) with a BSc. in Statistics. He received his MSc degree in Statistics from Iowa State University and is currently a PhD candidate at the Industrial Engineering Department, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil.

Bibliographic data

Franca Lila, Maurício

Essays on Hierarchical Time Series Forecasting / Maurício Franca Lila; advisor: Fernando Luiz Cyrino Oliveira; co-advisor: Erick Meira de Oliveira. – 2023.

132 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2023.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Modelos de previsão. 3. Séries temporais hierarquicas. 4. Reconciliação resistente. 5. Reconciliação robusta. 6. Demanda energética. 7. Força de trabalho. I. Cyrino Oliveira, Fernando Luiz. II. Meira de Oliveira, Erick. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

To my beloved wife and children.

Acknowledgments

I would like to express my gratitude to my advisors, Professor Fernando Cyrino and Dr. Erick Meira, who guided me throughout this project.

I would like to express my sincere gratitude to PUC-Rio for providing me with the opportunity of a tuition waiver. This support has allowed me to pursue my academic goals and advance my education, and I am truly thankful for this assistance.

I would also like to thank my friends and family who supported over the years through this journey.

I wish to acknowledge the help provided by the technical and support staff at the Department of Industrial Engineering at PUC-Rio. Particular thanks in this regard goes to Claudia Teti.

I also dedicate this thesis to the loving memory of my mother and father, who sacrificed so much to provide me with the education and opportunities that have brought me to this point. Their memory will always be a source of inspiration and motivation for me.

I cannot forget to mention my colleagues from the Brazilian Institute of Geography and Statistics (IBGE) for their support during this process.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Franca Lila, Maurício; Cyrino Oliveira, Fernando Luiz (Advisor); Meira de Oliveira, Erick (Co-Advisor). **Essays on Hierarchical Time Series Forecasting**. Rio de Janeiro, 2023. 132p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

This study presents a set of methodological proposals aimed at improving forecast reconciliation in the context of Hierarchical Time Series. The main objective is to present original solutions to the theme, seeking to obtain more accurate forecasts than those obtained by independent models for the different levels of the hierarchy. The studies were conducted using real data, showing the potentiality of application of the methods developed in different scenarios, in which the time series are structured in a hierarchical fashion. This thesis is composed of a set of essays that explore forecast reconciliation from the perspective of a regression model, which gives foundations to optimal reconciliation. The first contribution addresses the problem of forecast reconciliation from the perspective of robust estimators. The proposal presents an original contribution applied to data from labor force surveys in Brazil, presenting a set of solutions that can drive efficient public policies. In this case, the reconciled forecasts obtained through robust estimators provided consistent gains in terms of accuracy when compared to methods that represent the state-of-the-art on forecast reconciliation in hierarchical time series. The second contribution deals with the problem of optimal reconciliation applied to energy consumption time series in Brazil. We present an alternative proposal, less sensitive to outlying forecasts at the reconciliation stage. The results obtained in this second study show considerable improvements in standard evaluation metrics with regard to the new forecasts. The third proposal seeks to offer robust covariance structures for forecasting errors, which expands the set of strategies presented in the literature. The main contribution is to incorporate robust covariance estimates into the MinT (Minimum Trace) reconciliation approach, which minimizes reconciliation errors, offering an estimator with minimum variance.

Keywords

Forecasting; Hierarchical Time Series; Resistant Reconciliation; Robust Reconciliation; Electricity Demand; Labor Force Survey.

Resumo

Franca Lila, Maurício; Cyrino Oliveira, Fernando Luiz; Meira de Oliveira, Erick. **Ensaios sobre Previsão de Séries Temporais Hierárquicas**. Rio de Janeiro, 2023. 132p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

O presente estudo, apresenta um conjunto de propostas metodológicas relacionadas a reconciliação de previsões em Séries Temporais Hierárquicas. O principal objetivo é apresentar soluções originais ao tema, buscando obter previsões mais acuradas do que as obtidas por modelos independentes para os diferentes níveis da hierarquia. Os estudos foram realizados considerando dados reais, mostrando a potencialidade de aplicação dos métodos desenvolvidos em diferentes cenários, onde as series temporais são estruturadas de forma hierárquica. Esta tese é composta por um conjunto de ensaios que exploram a reconciliação de previsão sob a ótica de um modelo de regressão, que dá origem a reconciliação ótima. A primeira contribuição trata do problema da reconciliação de previsões na perspectiva de estimadores robustos. A proposta apresenta uma contribuição original aplicada a dados dos de pesquisas de força de trabalho no Brasil, apresentando um conjunto de soluções que podem direcionar políticas públicas eficientes. Neste caso, as previsões reconciliadas obtidas através de estimadores robustos possibilitaram um maior ganho em termos acurácia e uma performance equivalente aos métodos que representam o estado da arte sobre reconciliação de previsões em séries temporais hierárquicas. A segunda contribuição trata do problema da reconciliação ótima em séries de consumo de energia no Brasil, apresentado uma proposta alternativa, menos sensível a valores estremos. Os resultados obtidos neste segundo trabalho apresentam melhoramentos consideráveis em métricas de avaliação padrão no que diz respeito as novas previsões. Uma terceira proposta busca oferecer uma estrutura alternativa de covariância dos erros de previsão, que irá ampliar o conjunto de propostas apresentadas na literatura para o método de reconciliação denominado por MinT (do inglês, Minimum Trace) , que minimiza os erros de reconciliação, oferecendo um estimador de variância mínima.

Palavras-chave

Modelos de previsão; Séries temporais hierarquicas; Reconciliação resistente; Reconciliação robusta; Demanda energética; Força de trabalho.

Table of contents

List of figures

List of tables

List of abbreviations

HTS – Hierarchical Time Series

PME – Brazilian Monthly Labor Force Survey

PNADC – Brazilian Continuous National Household Sample Survey

BU – Bottom-Up

TD – Top-Down

OLS – Ordinary Least Squares

TDGSA – Top-Down Gross-Sohl method A

TDGSF – Top-Down Gross-Sohl method F

TDFP – Top-Down based on forecast proportions

WLS(v) – Weighted Leatst Squares

WLS(s) – Weighted least squares with structural scaling

MinT-Sample – Minimum trace reconciliation using the full covariance matrix

MinT-Shrink – Minimum trace reconciliation using the shrinkage factor

HUBER (1) – Robust Reconciliation using M-Estimator

HUBER (2) – Robust Reconciliation M-Estimator proposal 2

MSE – Mean Squared Error

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

RelMSE – Relative Mean Squared Error

RelMAE – Relative Mean Absolute Error

RelMAPE – Relative Mean Absolute Percentage Error

AveRelMSE – Average Relative Mean Squared Error

AveRelMAE – Average Relative Mean Absolute Error

AveRelMAPE – Average Relative Mean Percentage Error

ARIMA – Autoregressive, Integrated, Moving Average

ETS– Error, Trend and Seasonality

EPE – Brazilian Energy Research Company

MME – Brazilian Ministry of Mines and Energy

PDE – Energy Expansion Plan

PNE – National Energy Planning

SIN – Brazilian National Interlinked System

MW – Midwest

NE – Northeast

NO – North

SE – Southeast

SO – South

SARIMA – Seasonal Autoregressive, Integrated, Moving Average

MLE – Maximum Likelihood Estimation

MCD – Minimum Covariance Determinant

In God we trust. All others must bring data

William Edwards Deming .

1
Introduction

In recent years, the interest for hierarchical levels in forecasting has increased due to, among other things, their growing use in policy-making, allocation of government funds and regional planning. Thus, Hierarchical Time Series (HTS) has been a promising area of research.

The HTS framework, as presented in the literature (ATHANASOPOULOS et al., 2009), stands for a set of time series that can be aggregated at different levels, according to a well-defined hierarchical structure. In general, these hierarchies can be classified into *cross-sectional*, *temporal* or a combination of both, called *cross-temporal* hierarchies. In *cross-sectional* hierarchies (HYNDMAN et al., 2011; HYNDMAN; LEE; WANG, 2016), a variable of interest can be structured in different partitions, such as geographical subdivisions, population groups of interest or different industrial sectors. *Temporal* hierarchies, in turn, data are organized in partitions of the time frequency , for instance, annual information on energy demand is broken down into quarterly or monthly results (ATHANASOPOULOS et al., 2017). In the case of *cross-temporal* hierarchies, there is a double constraint added to this framework (KOURENTZES; ATHANASOPOULOS, 2019), for example, different geographical regions must be aligned to aggregation restrictions over time frequencies.

Hierarchical settings involving time series can be quite challenging. In some cases, by adding an extra level in a given hierarchy, the number of involved time series increases exponentially. Hierarchies can be unbalanced in some fashion, adding more complexity to their arrangement.

The importance of cross-sectional and temporal hierarchical forecasts is evident from the recent literature (WICKRAMASURIYA et al. 2019,WICKRAMASURIYA et al. 2020, ATHANASOPOULOS et al. 2017). However, the literature of *cross-temporal* hierarchies (KOURENTZES; ATHANASOPOULOS, 2019) is still incipient.

Regardless of the approach related to a hierarchical structure, an internal consistency is expected because of the additive property of the data, i.e. , the upper levels correspond to the sum of those at lower levels. When using independent forecasting methods for each time series in a given hierarchy, the original additive property is lost/not respected, mostly due to model specifications, which are defined and estimated independently at each node. In such cases, the internal consistency is lost due to independent and incoherent forecasts.

The strategy of using hierarchical forecasting methods connects with the idea of taking advantage of the hierarchical structure of the data through base forecast reconciliation, thus generating results that are usually unbiased and more accurate than those provided by benchmark methods (NYSTRUP et al., 2021; PANAGIOTELIS et al., 2021). The reconciliation process returns the additive property of the data to the independent forecasts.

The current research topic focuses mainly on the development of alternative reconciliation methods to obtain accurate forecasts for a wide range of time series, which are structured in a hierarchical fashion.

The thesis is divided into a series of essays. The first contribution refers to a published article LILA et al. (2022), reproduced in full in Chapter 3. It explores the regression-based perspective of forecast reconciliation. The main contribution addresses the fact that when combining base forecasts through regression-based reconciliation strategies, some forecasts may behave like outliers, causing distortions to the reconciliation process. This work introduces the concept of robust estimation applied to hierarchical forecasting, when reconciliation methods take place, by formalizing two different robust-based approaches. This research was applied to unemployment data from multiple labor force surveys in Brazil. In doing so, we address a significant gap in the modelling and forecasting of unemployment, taking into account the hierarchical structure of the data. Besides the methodological contribution to the field of HTS, the extension to Labor Statistics plays an important are role in the modern societies. Accurate forecasts under this framework, allows policymakers and regulators to strike disparities in under-utilization of the labour supply.

To demonstrate the potential and validity of the proposed approaches, we compared their performance with those from traditional and state-of-the-art methods. Overall, the robust reconciliation approaches show promising forecasting results under multiple settings and through the lens of different evaluation metrics. Furthermore, the methodology developed is flexible, in the sense that it can be readily applied to other time series and deliver equally reliable results.

The second contribution, benefits from the framework established in Chapter 3. This methodological contribution refers to another published article MEIRA et al. (2023), reproduced in full in Chapter 4. When considering dependency on error structures during the reconciliation process, the reconciled forecasts will also have minimum variance amongst all possible combinations of forecasts. This strategy is adequate when there are no outliers in the reconciliation process. We introduce the concept of hierarchical forecast reconciliation based resistant regression and formalize a modified resistant-based strategy applied to electricity consumption time series in Brazil. By considering resistant statistics in the forecast reconciliation process, we provide another valuable contribution in the research topic of HTS: the substantial reduction of contamination in the involved reconciled forecasts due to departures on observations. The new methodology was considered in several experiments and contrasted with traditional and innovative benchmarks. The proposed approach showed superior forecasting accuracy under different experimental setups. Consequently, the new approach is shown to be suitable to support decision making in the energy and related sectors.

The third contribution is an ongoing research topic that aims to improve the idea of optimally combining forecasts subject to linear constraints, providing an extension to MinT reconciliation with robust covariance structures. In this case, we contextualize in Chapter 5 the problem and possible research directions.

This document is structured as follows. In Chapter 2 we present some previous work relevant to our problem. In Chapter 3 we transcribe in full a published paper, and introduce the robust reconciliation method applied to Labor Statistics in Brazil, in Chapter 4 we show the idea of resistant reconciliation applied to the Brazilian energy sector . In Chapter 5 we present the overall idea of the third contribution. Finally, in Chapter 6 we present our concluding remarks and future work.

2
Literature Overview

HTS has become a promising field of study, with many forecasting applications to the industry, government planning, trading and other topics (KARMY; MALDONADO, 2019; ATHANASOPOULOS et al., 2020). When forecasting time series are aggregated according to a hierarchical structure, one should not overlook their aggregation constraints. Ideally, the forecasts from all levels should result in coherent quantities, i.e., the sum of the forecasts for the time series located in a particular hierarchical level should match the series results at the immediate upper level. Without such constraints, an intuitive strategy would require only independent forecasts for each stage of the hierarchy, at the cost of producing incoherent forecasts. Early studies on forecast aggregation were basically focused on three main streams: the *top-down*, the *bottom-up* and a combination of both called *middle-out*. While the *top-down* disagregates the information on forecasts from the top level of the hierarchy from a weighting system, the *bottom-up* borrows this information from the most granular levels of the hierarchy. The *middle-out* is a hierarchical level dependent approach that applies both techniques in a given hierarchical level. In the work of SYNTETOS et al. (2016) a rich discussion on different approaches and their implications is provided. Some authors compared the performance of these hierarchical forecasting methods as in (FLIEDNER, 1999; GRUNFELD; GRILICHES, 1960, BARNEA; LAKONISHOK, 1980; FOGARTY et.al, 1991; NARASIMHAN et.al, 1995), which advocate in favor of *top-down*, whilst others advocated in favor of the *bottom-up* strategy, (KINNEY, 1971; ORCUTT et al. 1968; TOBIAS; ZELLNER, 2000 ; SILVA et al. 2019). More recently, a modified *top-down* approach was presented in the work of ATHANASOPOULOS et al. (2009) using data from the tourism sector. Another top-down strategy was also recently explored in the work of MANCUSO et al. (2020) applying Machine Learning techniques to compute the disaggregation weights.

Finally, some authors argued that there were no substantial differences between the two approaches in terms of superiority (FLIEDNER; MABERT, 1992; NENOVA; MAY, 2016; TORRINI et al., 2016).

The idea of optimally combine forecasts, generating coherent forecasts from a regression-based perspective appears in the work of HYNDMAN et al. (2011). By adopting this strategy, the authors showed that, if the error covariance matrix is known, a generalized least squares regression solution would provide the optimal combination of forecasts. However, since this covariance matrix is often challenging to estimate in practice, the authors adopted a simplifying assumption of additivity on the forecasts errors.

The idea of optimally combining forecasts has gained considerable methodological contributions. As the complexity of the hierarchy arises, the computational efforts also increase. In the work of HYNDMAN et al. (2016), the authors show how to handled efficiently and propose a solution when there are millions of time series at the most disaggregated level. A Game-Theoretically Optimal (GTOP) reconciliation technique was developed (ERVEN; CUGLIARI, 2015), mapping any given set of independent forecasts into a new aggregate consistent forecasts that are guaranteed to be at least as good as the independent ones.

The concept of hierarchical time series was also extended to the context of temporal hierarchies, i.e., time series with frequency dependencies through their time span (KOURENTZES; ATHANASOPOULOS, 2019). Based on this concept, some authors proposed the use of different approaches to represent the covariance structure of the time series: hierarchy variance scaling; series variance scaling; and structural scaling (ATHANASOPOULOS et al., 2017). When reconciliation merges with temporal aggregation, it aims to obtain important features about a time series at different time frequencies. In the work of NYSTRUP et al. (2021), the authors emphasize the importance of the correlation structure of the forecast errors in order to improve the accuracy of the reconciled forecasts. In the work of KOURENTZES & ATHANASOPOULOS (2019), the authors presented a solution to cross-temporal aggregations, providing a framework based on the idea of optimally combining forecasts with a minimum variance, initially proposed in the work of WICKRAMASURIYA et al. (2019) for cross-sectional aggregation.

In a framework dealing with the presence of missing values STRATIGAKOS et al. (2022) show that a class of machine learning models directly provides both point and probabilistic coherent hierarchical forecasts according to their findings. Recently, Machine Learning techniques were implemented to derive the combination weights for the forecasts across the various aggregation levels (SPILIOTIS et al., 2020). From an distributional point of view, TAIEB et al. (2017) built a coherent probabilistic forecast framework throughout a bottom-up manner in which the dependency between nodes at each level is obtained by reordering quantile forecasts. In the work of JEON et al. (2019), the authors proposed new approaches for reconciling probabilistic forecasts ensuring coherence, when combining information from density forecasts at all hierarchical levels. PRITULARGA et al. (2021) proposed the idea of defining coherency as stochastic allowing to better understanding some overlooked uncertainties in the forecast reconciliation processes, which come from producing base forecasts and the estimation of the reconciliation matrix.

Despite the relevant growth in the hierarchical time series forecasting literature, to date, the only methodological contribution which considered the use of robust estimators to produce reconciled forecasts comes from this thesis presented in LILA et al. (2022). From a regression-based perspective, this gap should not be overlooked, considering the possibility of outliers occurring in several time series of a given hierarchy due to multiple reasons, ranging from false or misleading information provided, measurement errors, incorrect data processing, among others. In the second contribution deriving from this original research, we propose an outlier-resistant method for hierarchical time series and applying them to multiple time series hierarchies of electricity demand in Brazil at different levels. The motivation behind our approach is because forecast reconciliation based on resistant regression is less affected by outliers or influential points. In the third contribution we propose an extension to MinT reconciliation using two subset-based covariance estimators, which are less sensitive to outliers. In this case, we use the Minimum Volume Ellipsoid (MVE) and Minimum Covariance Determinant (MCD) method (ROUSSEEUW, 1985; HUBERT et al. 2018; AELST; ROUSSEEUW, 2009), which is are highly robust estimators of multivariate location and scatter.

3
First contribution:Forecasting unemployment in Brazil: a robust reconciliation approach using hierarchical data

This Chapter introduces the concept of robust estimation for hierarchical forecast reconciliation methods. This study was published in the Socio-Economic Planning Sciences journal (ISSN: 0038-0121), (LILA et al., 2022).

3.1
Introduction

Official statistics are the fundamental building blocks of our society, providing essential insights for policy implementation and assessment. Labor statistics, in particular, is of paramount importance to understand the effects of policies on certain groups of interest. Understanding labor market functioning for a given sex, age or educational group, allows policymakers and regulators to strike disparities that still happen in modern societies (VEEN; EVERS, 1983; BARNICHON; GARDA, 2016; BAGCHI; PAUL, 2018).

National Statistical Offices worldwide regularly produce a set of indicators about their populations based on specific surveys. In Brazil, for instance, the Monthly Labor Force Survey (PME[1]) produced a set of indicators on six metropolitan areas' workforce from 2002 to 2016. This survey was followed by the Brazilian Continuous National Household Sample Survey (PNADC[2]), designed to produce a set of quarterly indicators about the characteristics of the Brazilian Labor market and supplementary topics, serving as a benchmark tool for monitoring the labor force in Brazil.

Labor time series can sometimes be presented as of Hierarchical Time Series (HTS). These stand for a set of time series that can be aggregated at different levels, according to a well-defined hierarchical structure. For instance, a macroeconomic variable for a given country can be stratified first into states, then into cities and finally, into sex or age groups, if data is available. HTS has become a promising field of study, with many forecasting applications to the industry, government planning, trading and other topics (KARMY; MALDONADO, 2019; ATHANASOPOULOS et al., 2020). When forecasting time series aggregated according to a hierarchical structure, one should not overlook their aggregation constraints. Ideally, the forecasts from all levels should result in coherent quantities, i.e., the sum of the forecasts for the time

[1]PME stands for Pesquisa Mensal de Emprego.
[2]PNADC stands for Pesquisa Nacional por Amostra de Domicílios Contínua.

series located in a particular hierarchical level should match the series results on the immediate upper level. Without such constraints, an intuitive strategy would require only independent forecasts for each stage of the hierarchy, at the cost of producing incoherent forecasts.

Early studies on HTS were basically focused on two main strategies: the *top-down* and the *bottom-up* (BU). The former aims to provide base forecasts for the series at 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 works oppositely, i.e., by forecasting at the most granular the hierarchy level 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*. A considerable number of articles focused on comparing the performance of hierarchical forecasting methods. Some authors presented favourable results to the *top-down* approach (FLIEDNER, 1999; GRUNFELD; GRILICHES, 1960; FOGARTY et al., 1991; NARASIMHAN et al., 1995), whilst others advocated in favor of the *bottom-up* strategy (KINNEY, 1971; ORCUTT et al., 1968; TOBIAS; ZELLNER, 2000; SILVA et al., 2019). Finally, some authors argued that there were no substantial differences between the two approaches in terms of superiority (FLIEDNER; MABERT, 1992; NENOVA; MAY, 2016; TORRINI et al., 2016).

Recent studies demonstrated that it is possible to improve upon the original hierarchical forecasting strategies. Using data from the tourism sector, a new strategy based on the *top-down* approach was presented by ATHANASOPOULOS et al. (2009). The strategy consisted of using the forecasted proportions of the lower-level series to compute the disaggregation weights, rather than using the historical proportions of the data. Another *top-down* strategy was recently presented by MANCUSO et al. (2021), who advocated using of Machine Learning techniques to compute the disaggregation weights.

A recently proposed set of approaches for hierarchical data are the so-called optimal combination (or reconciliation) methods. In short, these are based on a regression model that maps a set of incoherent forecasts to a coherent space using reconciliation, i.e., a process that aims at generating forecasts that are unbiased and add up correctly across the hierarchy (HYNDMAN et al., 2011). By adopting this strategy, the authors showed that, if the error covariance matrix is known, a generalized least squares regression solution would provide the optimal combination of forecasts. However, since this covariance matrix is often challenging to estimate in practice, the authors adopted a simplifying assumption of additivity on the forecasts errors, suggesting the use of Ordinary Least Squares (OLS). The idea of optimally combining forecasts has gained considerable methodological contributions, with some authors recommending the use of Weighted Least Squares (WLS) to ignore the covariance terms in reconciliation (HYNDMAN et al. 2016).

The concept of hierarchical time series was also extended to the context of temporal hierarchies, i.e., time series with frequency dependencies through their time span. Based on this concept, some authors proposed the use of different approaches to represent the covariance structure of the time series: hierarchy variance scaling; series variance scaling; and structural scaling (ATHANASOPOULOS et al., 2017). Recently, Machine Learning techniques were implemented to derive the combination weights for the forecasts across the various aggregation levels (SPILIOTIS et al., 2020). A probabilistic framework was proposed to deal with the uncertainty arising from inference-based approaches (TAIEB et al. 2017).

Despite the relevant growth in the hierarchical time series forecasting literature, to date, no work has considered the use of robust estimators to produce reconciled forecasts from a regression-based perspective. This gap should not be overlooked, considering the possibility of outliers occuring in several time series of a given hierarchy due to multiple reasons, ranging from false or misleading information provided, measurement errors, incorrect data processing, among others. In this work, we fill this gap by proposing two robust methods for hierarchical time series and applying them to multiple time series hierarchies of unemployed people in Brazil at different frequencies (monthly and quarterly). The motivation behind our approach is because robust regression is less affected by outliers or influential point forecasts. In this case, when using an M-estimator based regression, it is possible control undesirable effects of high leverage data points.

3.1.1 How robust reconciliation is conducted

In a hierarchical forecasting framework, all series are first forecasted using a pre-determined independent forecasting approach, giving birth to 'base forecasts'. Then, an optimal reconciliation (combination) approach is sought, aiming to deliver 'coherent', final forecasts. This term stands for forecasts which are both unbiased, have minimum variance amongst all combinations and add up correctly across the hierarchy. The idea of reconciliation consists of revising the forecasts so that, theoretically, better estimates of a given process can be found than using only independent forecasting methods. The process of base forecasts generation and forecast reconciliation in HTS methods can be represented as follows in the flowchart described in Figure 3.1.

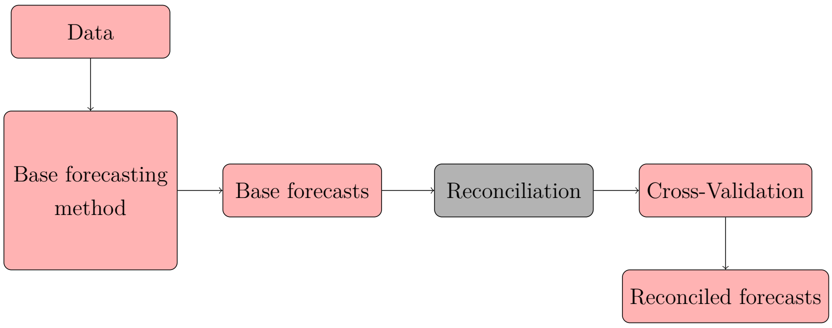

Figure 3.1: Base forecasts generation and forecast reconciliation in HTS methods.

Robust settings can be applied prior to or in conjunction with the base forecasting approaches to account for outliers or other influential points. This would lead to the flowchart illustrated in Figure 3.2. In this case, the primary concern is to address measurement issues occurring in the original time series. Several articles address the topic of attenuating the effects of outliers and its detection in time series in the modeling stage (BARROW et al., 2020; ROUSSEEUW et al., 2019; CROUX et al., 2010). Although this is a relevant issue related to time series forecasting, this process is usually infeasible in most HTS settings, mainly when the hierarchy comprises a large number of time series. This is because a tailored forecasting approach would have to be proposed to each time series in the hierarchy, as not all time series would require the same treatment. Since multiple forecasting routines would be applied at the base forecasts generation stage, estimating the error covariance matrix of the base forecasts would become impossible in practice, thereby invalidating the use of most forecast reconciliation approaches.

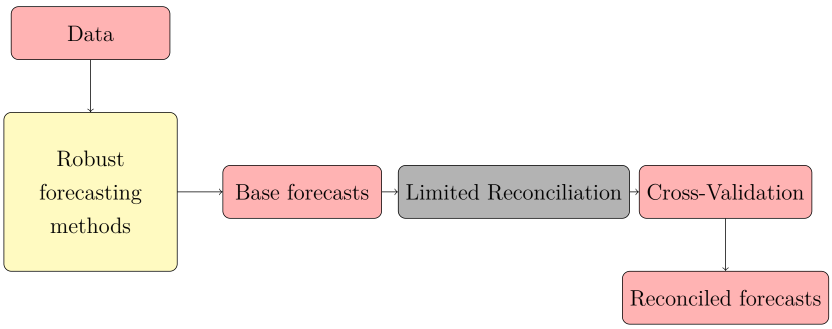

Figure 3.2: Robust base forecasts generation and forecast reconciliation.

This paper focuses on addressing outliers and influential points at the reconciliation stage. That way, not only we allow for the estimation of the base forecasts error covariance structure, but we can also address disparities and other potentially unwanted effects that may occur in established regression-based reconciliation methods. To the best of our knowledge, no forecasting approach, be it statistical, AI, or hybrid-based, has considered the use of robust estimators for hierarchical reconciliation, which represents our main methodological contribution in this paper. Our idea of robust reconciliation is depicted in the flowchart of Figure 3.3.

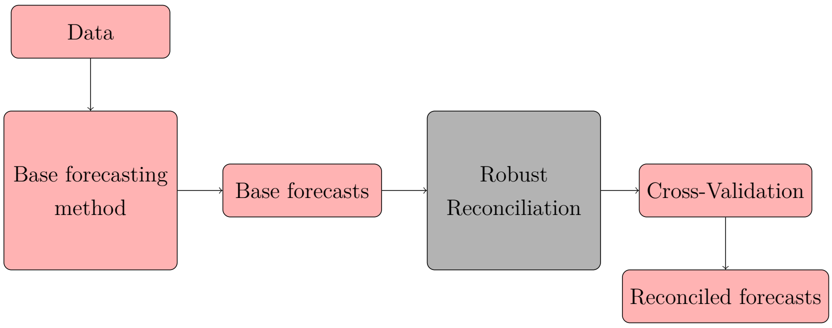

Figure 3.3: Our approach to robust forecast reconciliation.

To demonstrate the benefits of robust reconciliation, we use monthly data of total unemployed persons in Brazil, collected from the Brazilian Monthly Labor Force Survey (PME) and the same statistics from the quarterly results from the Continuous Brazilian Household Survey (PNADC). The first survey covered six metropolitan areas in Brazil and produced several key indicators for the Brazilian economy. The second is part of the Integrated System of Household Surveys covering the entire Brazilian territory. Nowadays, the leading labor statistics produced by the Brazilian Statistical Office come from this survey. Our experiments show encouraging results in favor of the proposed robust reconciliation strategies, which ranked among the first or second best methods in almost every case considered. In brief terms, we take advantage of the information provided at all hierarchical levels to improve the forecasts' quality in every case.

The rest of the paper unfolds as follows: Section 3.2 outlines the basic ideas of HTS and details the most up-to-date framework. Section 3.3 introduces the concept of robust reconciliation in hierarchical forecasting. Section 3.4 describes the experimental setup. The results are presented in Section 3.5. Finally, Section 3.6 provides a set of conclusions based on our findings.

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

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

$$\mathbf{y_t} = \mathbf{S}\mathbf{y_t^b} \tag{3-1}$$

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

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

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SP}\hat{\mathbf{y}}_{\mathbf{t+h|t}} \tag{3-2}$$

where $\tilde{\mathbf{y}}_{t+h|t}$ are the reconciled forecasts. The $\mathbf{SP}$ matrices represent the reconciliation process, which maps independent (or incoherent) forecasts into coherent ones. Depending on how $\mathbf{P}$ is structured, it is possible to reproduce several traditional hierarchical forecasting approaches. For instance, by letting $\mathbf{P} = \left[\mathbf{0_{n\times(m-n)}}|\mathbf{I_n}\right]$ where $0_{n\times(m-n)}$ is a null matrix, one can reproduce the *bottom-up* (BU) approach. In this case, the $\mathbf{P}$ matrix keeps only forecasts for the most granular level, then $\mathbf{S}$ completes the processes by generating reconciled forecasts at every level of the hierarchy. On the other hand, *top-down* forecasts can be achieved by making $\mathbf{P} = \left[\mathbf{p}|\mathbf{0_{n\times(m-1)}}\right]$, where $\mathbf{p}$ is the set of proportions of forecasts (ATHANASOPOULOS et al. 2009). Common approaches to obtain these proportions are described as follows:

– Average of historical proportions or Gross-Sohl method A (TDGSA) (GROSS; SOHL, 1990):

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

for $j = 1, \ldots, m$, where $y_{j,t}$ are the bottom-level series over the period $t = 1, \ldots, T$ and y_t is the total aggregate. Under this approach, the weights for the forecasts at the top of the hierarchy takes into account the relative importance of each value of lower hierarchical levels of the time series for a given time instant. The approach then averages the weights out for the entire time series.

– Proportion of historical averages or Gross-Sohl method F (TDGSF)(GROSS; SOHL, 1990):

$$p_j = \sum_{t=1}^{T}\frac{y_{j,t}}{T} \Bigg/ \sum_{t=1}^{T}\frac{y_t}{T} \tag{3-4}$$

for $j = 1, \ldots, m$. In this case, the weighting system takes the relative importance of the entire series from the past at lower levels with respect to the time series at the top level.

- Forecast proportions (TDFP)(ATHANASOPOULOS et al. 2009):

$$p_j = \prod_{\ell=0}^{K-1} \frac{\hat{y}_{j,h}^{(\ell)}}{\hat{S}_{j,h}^{(\ell+1)}} \tag{3-5}$$

where K stands for the number of levels of the hierarchy, $\hat{y}_{j,h}^{(\ell)}$ are the h-steps ahead forecasts of the series that corresponds to the node that is ℓ levels above node j and $\hat{S}_{j,h}^{(\ell)}$ is the sum of h-steps ahead forecasts below the node which is ℓ levels above node j, where $j = 1, \ldots, m$.

Optimal combination (or reconciliation) approaches, in turn, can be can be expressed according to the following regression model:

$$\hat{\mathbf{y}}_{t+h|t} = S\beta_{\mathbf{t+h|t}} + \epsilon_{\mathbf{t+h|t}} \tag{3-6}$$

where $\beta_{\mathbf{t+h|t}} = \mathbf{E}\left[\mathbf{y^b_{t+h}}|\mathcal{I}_\mathbf{t}\right]$, $\mathcal{I}_t = y_1, y_2, \ldots, y_t$ and $V\left(\epsilon_{t+h|t}|\mathcal{I}_t\right) = \Sigma_h$.

Later, authors advocated the use of the Weighted Least Squares (WLS) estimator to obtain an estimate of the variance-covariance matrix, while ignoring the elements outside the diagonal (HYNDMAN et al.2016). Following this idea, new studies (WICKRAMASURIYA et al., 2019; WICKRAMASURIYA et al., 2020; KOURENTZES; ATHANASOPOULOS, 2019) proposed alternative methods to obtain Σ_h, whose calculation is dependent on the forecasting models. For instance, the Minimum Trace (*MinT*) reconciliation approach, introduced in WICKRAMASURIYA et al. (2019), aims to find a matrix $\mathbf{P}$ that it minimizes $tr(\mathbf{SPW_hP'S'})$ subject to $\mathbf{SPS} = \mathbf{S}$, the unbiasedness condition.

In order to use *MinT* reconciliation, it is necessary to estimate $\mathbf{W_h}$, the variance-covariance matrix of the h-step-ahead base forecast errors. Given the difficulty in estimating this matrix, several approximation approaches were proposed, giving birth to variants of the optimal combination approach depicted in eq. (3-6). Some of them are listed as follows:

- Ordinary least squares estimator (OLS): $\mathbf{W_h} = \mathbf{k_h I}$ where $k_h > 0$. This is the simplest hypothesis and means assuming that the $\mathbf{P}$ matrix is independent of the data, providing an advantage in computational terms. However, by ignoring the dependency structures between residuals, relevant information is lost (ATHANASOPOULOS et al., 2009; HYNDMAN et al., 2011).

- Weighted least squares estimator with variance scaling: $\mathbf{W_h} = \mathbf{k_h}\text{diag}(\hat{\mathbf{W}}_\mathbf{1})$ where $k_h > 0$ and $\hat{\mathbf{W}}_1$ is the unbiased sample covariance estimator of the in-sample one-step-ahead base forecast errors, represented as follows:

$$\hat{\mathbf{W}}_1 = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e_t}(\mathbf{1})\mathbf{e_t}(\mathbf{1})' \tag{3-7}$$

This specification scales the base forecasts using the variance of the one step-ahead residuals. In this situation, the elements outside the main diagonal of the covariance matrix are set to zero (HYNDMAN et al., 2016). This approach was considered as a benchmark method in the results, and is represented as $WLS(v)$.

– Weighted least squares estimator with structural scaling: $\mathbf{W_h} = \mathbf{k_h}\mathbf{\Lambda}$ where $k_h > 0$, $\mathbf{\Lambda} = \text{diag}(\mathbf{S1})$ and $\mathbf{1}$ is a $n \times 1$ vector of ones. This specification assumes that the errors of the bottom-level base forecasts have variance k_h and are not correlated between the different nodes. This estimator depends only on the number of series in each node of the hierarchy. By adopting this weighting structure, the reconciliation is conducted by solving a weighted least squares problem (KOURENTZES; ATHANASOPOULOS, 2019). This approach will be represented by $WLS(s)$.

– Estimator by minimizing the matrix trace using the full covariance matrix – *MinT-Sample*: $\mathbf{W_h} = \mathbf{k_h}\hat{\mathbf{W}}_\mathbf{1}$. In this case, the only assumption is that the error covariance matrices are proportional to each other and the full one-step ahead covariance matrix is directly estimated (WICKRAMASURIYA et al., 2019). This approach presents implementation problems when the covariance matrix is not is positive definite. Therefore, it will not be used for comparison purposes in this paper.

– The shrinkage estimator – *MinT-Shrink*: $\mathbf{W_h} = \mathbf{k_h}\hat{\mathbf{W}}^*_{\mathbf{1,D}}$ where $k_h > 0$, and $\hat{\mathbf{W}}^*_{1,D} = \lambda\hat{\mathbf{W}}_{1,D} + (1-\lambda)\hat{\mathbf{W}}_1$ is an estimator of the covariance matrix and aims to reduce the importance of elements outside the main diagonal of $\hat{\mathbf{W}}_1$. The shrinkage parameter λ is a function of the *in-sample* correlations (KOURENTZES; ATHANASOPOULOS, 2019) and is estimated as follows.

$$\hat{\lambda} = \frac{\sum_{i\neq j}\hat{V}(\hat{r}_{ij})}{\sum_{i\neq j}\hat{r}^2_{ij}} \tag{3-8}$$

where $\hat{r}_{ij}$ corresponds to ij-element of $\hat{\mathbf{R}}_1$, the one-step-ahead *sample* correlation matrix. It is worth noticing that the shrinkage estimator, despite reducing the importance of covariance between series at different levels of the hierarchy, still takes into account some measure of relationship between the series.

Regardless of the approach selected for estimating $\mathbf{W_h}$, the optimal reconciled forecasts are given by

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{W_h^{-1}}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W_h^{-1}}\hat{\mathbf{y}}_\mathbf{h}. \tag{3-9}$$

This equation can be written in another fashion as follows.

$$\tilde{\mathbf{y}}_h = \mathbf{S}\hat{\beta}_\mathbf{h}. \tag{3-10}$$

3.3 The Robust Reconciliation for HTS

As observed, all methods described in the previous section are obtained following the same guidelines of linear regression models. However, the use of least squares may not be appropriate when solving problems containing outliers or influential points. In such cases, a parameter estimation method that is less affected by data imperfections and/or contamination is desirable. In this regard, an alternative set of tools related to robust estimation can be applied to provide reliable outcomes. Some techniques like M-estimation are helpful in this context (HUBER, 1964; ROUSSEEUW; YOHAI, 1984; SMUCLER; YOHAI, 2017). Let the residuals from the reconciliation process be defined as

$$\mathbf{y_h} - \tilde{\mathbf{y}}_\mathbf{h} = \epsilon(\beta_\mathbf{h}) \tag{3-11}$$

where $\mathbf{y_h}$ are the actual (true) values of time series and $\tilde{\mathbf{y}}_h$ are the reconciled forecasts. Consider a function ρ having the following properties:

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

Then, the M-estimator based on the residuals from the equation (5-14) is given as

$$\hat{\beta}_{M,h} = \arg\min_{\beta_h} \sum_{i=1}^{n} \rho(\epsilon_i(\beta_h)). \tag{3-12}$$

In order to solve this minimization problem we need to find $\psi(\cdot) = \rho'(\cdot)$, which is the influence function. In this work, we consider the Huber influence function (HUBER, 1964) in light of its desirable properties for computational convergence.

$$\rho(z) = \begin{cases} z^2, & \text{if } |z| < c; \\ |2z|c - c^2, & \text{if } |z| \geq c \end{cases} \tag{3-13}$$

$$\psi(z) = \begin{cases} z, & \text{if } |z| < c; \\ c[\text{sgn}(z)], & \text{if } |z| \geq c \end{cases} \tag{3-14}$$

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

$$w(z) = \frac{\psi(z)}{z} \tag{3-15}$$

These weights are a function of residuals. However, the residuals depend on the estimated coefficients, which depend on the weights. In this case, an iterative procedure called Iteratively Reweighted Least Squares (IRLS) is required.

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

$$\hat{\beta}_{\text{M},h} = \arg\min_{\beta_h} \sum_{i=1}^{n} \rho\left(\frac{\epsilon_i(\beta_h)}{\sigma}\right), \tag{3-16}$$

There are two common ways of estimating σ. The first is based on the Mean Absolute Deviation (MAD) and is represented below:

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{median\{|\epsilon_i(\beta_h)|\}}{0.6745} \tag{3-17}$$

The second approach, also known as Huber's Proposal 2, comes from the solution of:

$$\frac{1}{n-p}\sum_{i=1}^{n} \psi^2\left[\frac{\epsilon_i(\beta_h)}{\hat{\sigma}}\right] = E_Z[\psi^2(\epsilon)] \tag{3-18}$$

where $E_Z[\psi^2(\epsilon)]$ is the expected value of ψ^2 when ϵ has standard normal distribution, n is the number of observations and p is the number of regression parameters.

In the results, we denote the above approaches by HUBER (1) and HUBER (2), respectively.

3.4 Experimental Setup

Two original data sets of unemployment time series are used in the experiments. The first refers to the Brazilian Monthly Labor Force Survey (PME) and the second to the Brazilian Continuous National Household Sample Survey (PNADC). Both surveys are conducted by the Brazilian National Statistical

Office[3]. These surveys considered different periods, comprised different geographical regions and collected time series of different frequencies (monthly, in the case of PME, and quarterly, in the case of PNADC), making them interesting options to assess the robustness of the methods proposed in Section 4.2.4.

3.4.1 The Brazilian Monthly Labor Force Survey

The Brazilian Monthly Labor Force Survey (PME) produced monthly indicators of the labor market in distinct geographical regions in Brazil. The survey covered the metropolitan areas of Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre. This survey had to adapt its questionnaire to capture the changes in the labor market. For our experiments, we used 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. The selected data represent the number of unemployed people and are organized in a hierarchical fashion. The top level of the hierarchy corresponds to the aggregation of the metropolitan areas, the intermediate level refers to each metropolitan area and the bottom level splits the areas by sex. Table 3.1 depicts the number of time series per hierarchical level.

Hierarchical level	Number of time series
Overall	1
Metropolitan Areas	6
Sex	12
Total	19

Table 3.1: PME – Number of time series according to the hierarchical levels.

Given that we are interested in applying different strategies to forecast the number of unemployed people across the hierarchy, it is reasonable to show how this variable behaves through time. Figure 3.4 illustrates the top level time series, i.e., the series representing the total of unemployed people across the six metropolitan areas.

As can be noted, the total of unemployed people showed a downward trend between early 2004 and late 2014. This number then increased considerably from 2015 onwards, reaching similar levels to those observed at the end of the subprime financial crisis. To understand the regional effects, we plot the

[3]Also known as the Brazilian Institute of Geography and Statistics (IBGE), the agency responsible for the official collection of statistical, geographic, cartographic, geodetic and environmental information in Brazil.

time series for each metropolitan area. Seasonal patterns are readily apparent in all series. Although these series show slightly decreasing trends, by the end of 2014 we can spot an inflection point, showing an increase in the number of unemployed people, as depicted in Figure 3.5.

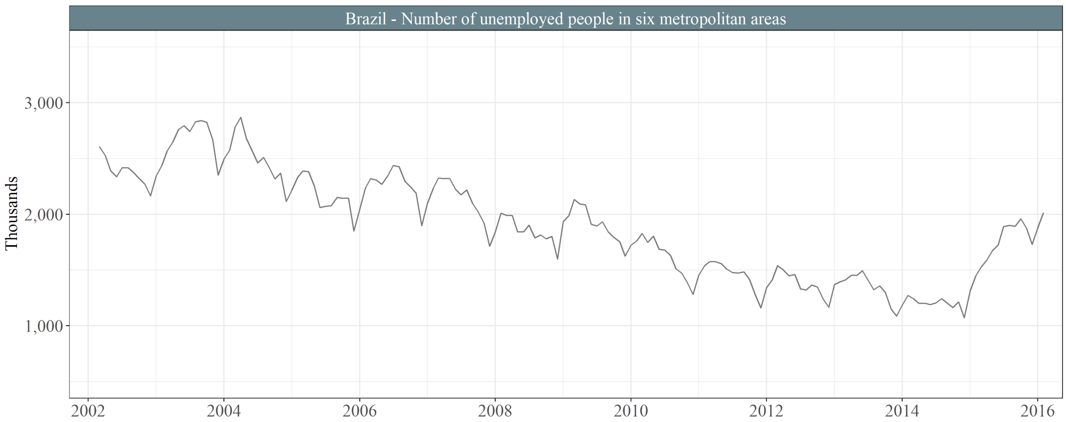

Figure 3.4: PME – Total of unemployed people (in thousands) across six metropolitan areas in Brazil.

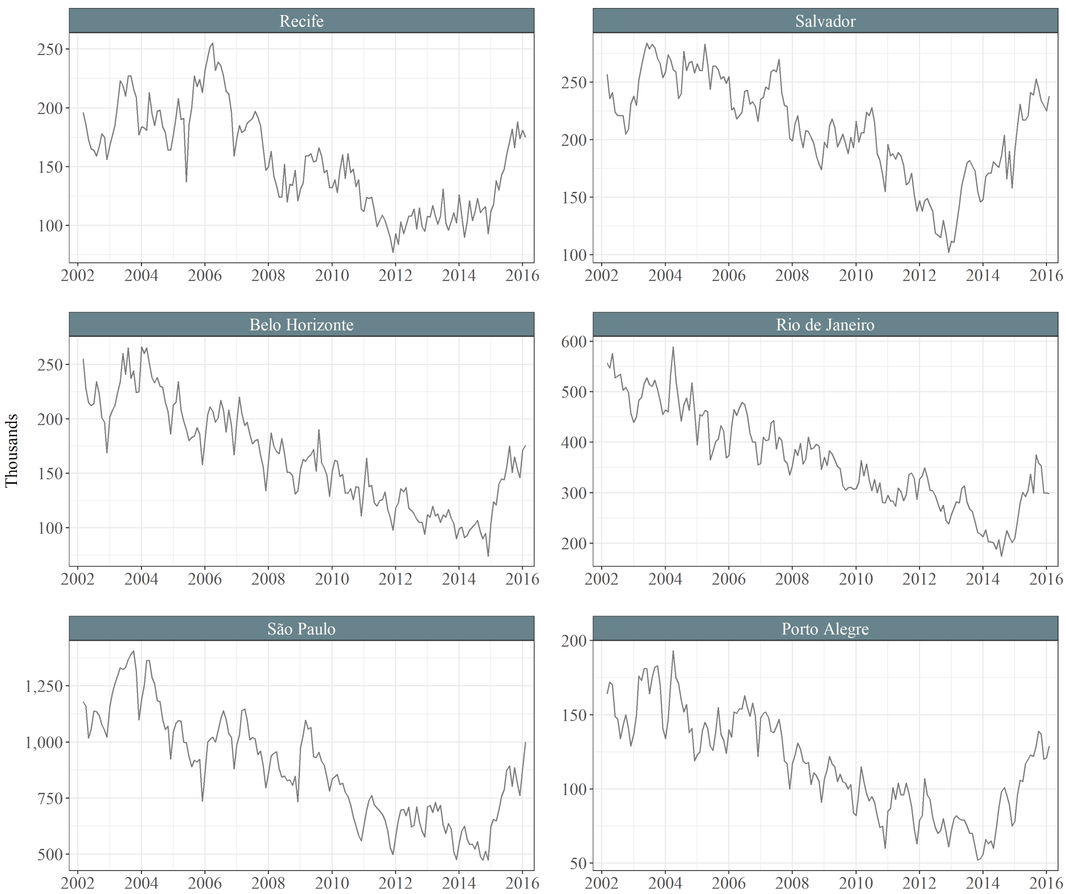

Figure 3.5: PME – Number of unemployed people (in thousands) by metropolitan areas.

The most granular level of this hierarchy divides the population into two categories. In Figure 3.6 the series for women (in red) shows a positive shift in level during most of the analyzed period. This highlights that, besides the geographical aspects, there are still informative differences to be explored in coherent hierarchical forecasting.

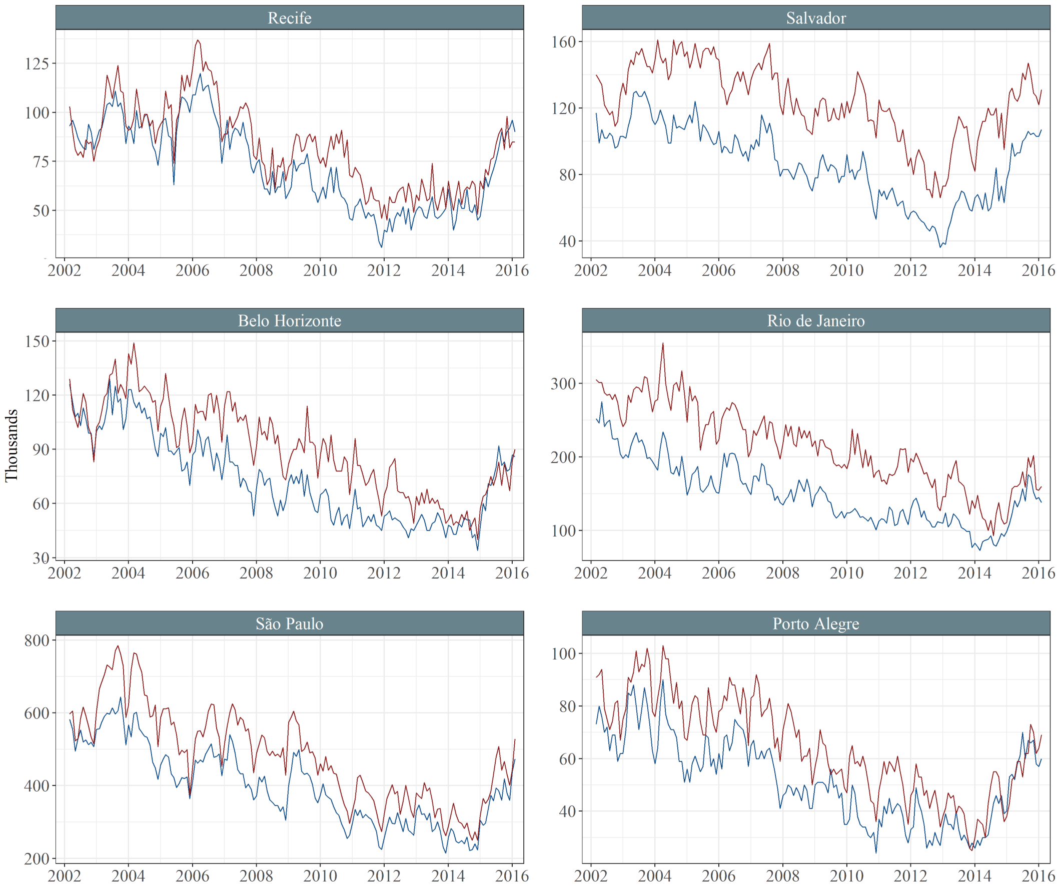

Figure 3.6: PME – Number of unemployed people (in thousands) for each metropolitan areas by sex (male in black and female in red).

3.4.2 The Brazilian Continuous National Household Sample Survey

The Brazilian Continuous National Household Sample Survey (PNADC) produces a set of critical indicators for a better understanding of Brazilian socioeconomic development. It was initially planned to produce a set of quarterly indicators about characteristics of the Brazilian labor market and other supplementary topics. Today, it constitutes one of the main tools for monitoring the labor force in Brazil. The survey covers the entire Brazilian territory and releases information for different geographical levels. In this study, we considered the geographical hierarchy described in Table 3.2. Due to the SARS-COV-2 outbreak, the household surveys conducted by the Brazilian

Statistical Office from 2020 onwards presented non-response rates higher than the minimum expected rates. Hence, to avoid spurious interpretations and provide a fair forecasting experiment to all benchmarks and reconciliation techniques herein involved, we considered only the pre-pandemic period from 2012 to 2019.

Hierachical level	Number of time series
Brazil	1
Great Regions	5
Federative Units	27
Total	33

Table 3.2: PNADC – Number of time series according to the hierarchical levels.

Figure 3.7 illustrates the behavior of the total of unemployed people in Brazil over the selected time span. As also observed in the case of PME, a clear upward trend is noted from late 2014 to late 2016. This period was marked by a political turmoil that resulted in the impeachment of President Dilma Rousseff in Brazil and widespread dissatisfaction with the political system. The time series also contains a well defined seasonal component, with considerably lower levels of unemployed people observed at the end of each fiscal year, given the year-end hiring surges which generally occur, particularly in the manufacturing and service sectors.

Figure 3.7: PNADC – Total of unemployed people (in thousands) in Brazil.

When the data disaggregates to the Great Regions, as depicted in Figure 3.8, the behavior of all series looks very similar, with a noticeable shift in level for each time series. The most populated regions present the highest values of total unemployed people. It is also possible to identify the same patterns observed at the top level of the hierarchy regarding trend and seasonal components.

Figure 3.8: PNADC – Number of unemployed people (in thousands) per Brazilian Great Region.

The most granular level of the selected hierarchical structure for PNADC corresponds to the geographical partitioning of Great Regions into Federative Units. Figure 3.9 illustrates the time series behavior for each Federative Unit within the five Brazilian Great Regions. The differences between series levels are highlighted in this case. This fact may produce forecasts that behave as outliers in regression-based reconciliation approaches, thus highlighting the importance of alternative, robust settings. Finally, we note that all time series contain trend and seasonal components. However, these facts are not readily apparent for some Federative Units in this sort of data visualization, given the considerable differences in the population numbers compared to larger and more populated Units.

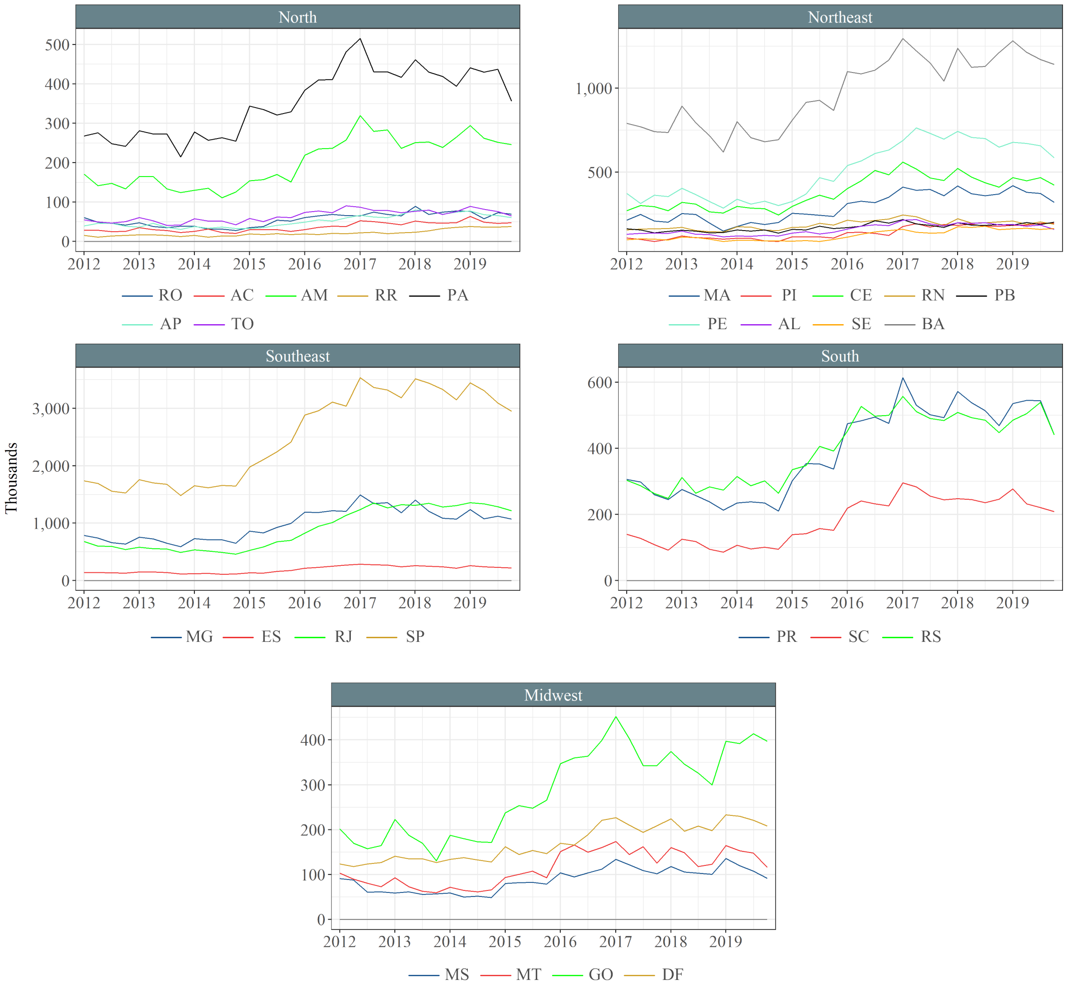

Figure 3.9: PNADC – Number of unemployed people (in thousands) according to each Great Region and Federative Unit.

3.4.3
Base Forecasting strategies

The reconciliation strategy relies on producing a set of base forecasts and then combining them into coherent ones. In some cases, the choice of a strategy depends on how data behave over time, or whether explanatory variables are considered to improve the predictive power of a particular model. In order to implement the strategies described previously, we consider here two widely applied sets of forecasting models as base forecasting strategies: the ETS class of exponential smoothing models and the Autoregressive, Integrated, Moving Average (ARIMA) formulations. These are briefly described in the following paragraphs.

In exponential smoothing models, the forecasts correspond to weighted averages of past observations. These weights decay exponentially, given the time difference between observations. These were first presented in seminal works (BROWN, 1959; COX, 1961; HOLT, 1957; WINTERS, 1960).

In this work, we rely on an automatic model selection routine commonly known as ETS – an acronym for Error, Trend and Seasonality, the three components that vary across exponential smoothing formulations. The ETS approach was addressed and placed in the form of state space equations by HYNDMAN et al. (2002). Each model consists of an observation equation, which describes the data, and one or more state equations, which describe the components of the level (ℓ_t), trend (b_t) and seasonality (s_t).

In practice, an automatic algorithm, implemented in the `ets()` function from the `forecast` (HYNDMAN et al., 2020) package in the software `R` (R Core Team, 2020), selects the best state space formulation for each series from a set of 30 possible combinations.

The Autoregressive Integrated Moving Average (ARIMA) formulations, in turn, explain a univariate time series as a combination of autoregressive and moving average components, which represent the existing autocorrelation patterns within the time series (BOX; JENKINS, 1970a). In addition, the integration order depends on the number of consecutive times that the series needs to be differenced to achieve stationarity.

In practice, an automatic model selection used in this work is provided by the argument `auto.arima` (also specified as a function) from the `forecast` package in `R`. It implements a multi-step algorithm that is a variation of the Hyndman-Khandakar algorithm (HYNDMAN; KHANDAKAR, 2008) and is currently the benchmark for automatic ARIMA selection in textbooks and recent empirical studies (HYNDMAN; ATHANASOPOULOS, 2021; BERGMEIR et al., 2018).

3.4.4 Assessment Metrics

Assessment metrics are an important set of tools for model selection in the context of time series. It is not uncommon to observe forecasting models performing better for a given time window when compared to others that usually have better predictive powers. In such contexts, using different metrics can provide an alternative point of view to select the best set of forecasts. Another important procedure is to consider rolling forecasting origins coupled with time series cross-validation to provide additional support when selecting the most accurate forecasting approach from a range of available methods. In this work, we focus on a set of related metrics used to evaluate the predictive power and gains in performance. The first set of metrics is the Mean Squared

Error (MSE) and the Mean Absolute Error (MAE). These measures are scale dependent and should be applied to models using the same data set. They are defined as follows:

$$MSE = \frac{1}{h}\sum_{t=1}^{h}(\hat{y}_t - y_t)^2 \tag{3-19}$$

$$MAE = \frac{1}{h}\sum_{t=1}^{h}|\hat{y}_t - y_t| \tag{3-20}$$

where h is the number of forecasting steps (forecast horizon), y_t are the actual (true) values of the time series and $\hat{y}_t$ are the corresponding forecasts.

Given that multiple reconciliation strategies are considered to generate forecasts for the entire hierarchies, relative measures of the above metrics are obtained by dividing the absolute values of the metrics for the reconciled strategies by the absolute values of the same metrics for the base forecasts, i.e., before reconciliation takes place. This allows us to indicate which technique presents a gain in terms of accuracy when compared to the base forecasts, as shown in the following equations:

$$RelMSE_{i,h} = \frac{MSE_{i,h}^{rec}}{MSE_{i,h}^{base}} \tag{3-21}$$

$$RelMAE_{i,h} = \frac{MAE_{i,h}^{rec}}{MAE_{i,h}^{base}} \tag{3-22}$$

where $MSE_{i,h}^{base}$ (denominator) corresponds to the MSE of the base (independent) forecast method, for the series i of the hierarchy at the forecast horizon h, and $MSE_{i,h}^{rec}$ (numerator) is the MSE obtained for the same series and time horizon after reconciliation. The same notation and concepts apply to the MAE.

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 summarized measure of improvement in terms of MSE and MAE. These quantities can be obtained through the following equations:

$$AveRelMSE = \sqrt[\#L]{\prod_{i \in L} RelMSE_i} \tag{3-23}$$

$$AveRelMAE = \sqrt[\#L]{\prod_{i \in L} RelMAE_i} \tag{3-24}$$

where L is the corresponding level of the hierarchy.

We implemented a rolling forecasting origin evaluation using the last 24 months period for the Monthly Labor Force Survey (PME) and the last 12 quarters for the Brazilian Continuous National Household Sample Survey (PNADC) in order to produce the cross-validation results for different steps ahead. In this case, the advantage of using the metrics presented in eqs. (3-23) and (3-24) is that $(1 - AveRelMSE) \times 100\%$ yields the percentage of improvement in MSE over the base forecasting strategy. The same interpretation holds for the MAE. The procedure is illustrated in Figure 3.10 for the forecasting steps $h = 1, 2, 3$ (HYNDMAN et al., 2020). In this work, we consider cross-validation for steps $h = 1$ to $h = 6$, in the case of PME, and $h = 1$ to $h = 4$, in the case of PNADC. Conducting cross-validation for accuracy evaluation has several advantages over selecting a single period, as cross-validation exposes the reconciliation approaches to different characteristics of our data. We want to find the reconciliation technique that best performs for a set of forecasts h steps ahead. The suitability of cross-validation for forecast accuracy assessment is discussed in further details in BERGMEIR et al. (2018).

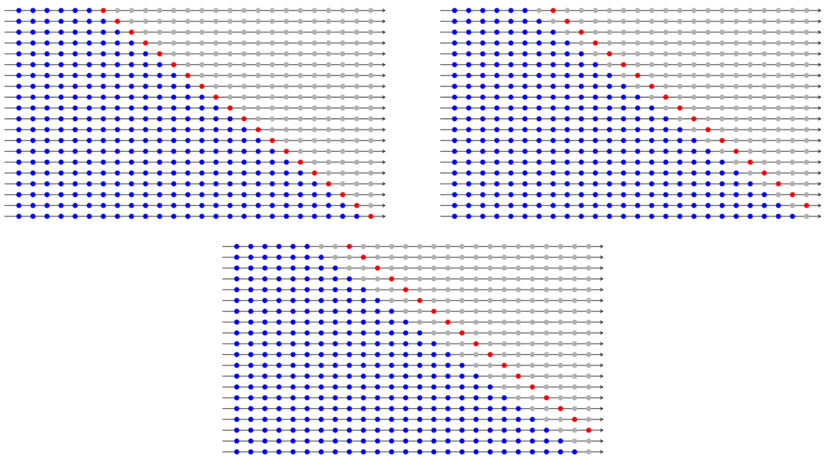

Figure 3.10: How accuracy evaluation using rolling forecasting origins is implemented for the different reconciliation strategies – illustrations for the first three forecasting steps.

In addition to the cross-validation forecast evaluation, we also implemented the non-parametric Friedman and post-hoc Nemenyi tests (DEMŠAR, 2006). The Friedman test indicates whether at least one of the reconciliation strategies is significantly different from the others. The post-hoc Nemenyi test aims to identify groups of reconciliation approaches for which there is no evidence of statistically significant differences.

The Friedman test works in the following fashion: first, the accuracy measures associated with each reconciliation technique are ranked according to their performance from the smallest to the largest throughout the forecasting horizons. Then, sums of the ranks across horizons are obtained for each reconciliation technique. Finally, the forecasting accuracy for a given approach is considered to be different than other reconciliation techniques if there is a significant difference in the sum of the ranks of at least one approach. The Friedman test statistic can be written as:

$$f_{(\alpha,h,k)} = \frac{12}{hk(k+1)} \sum_{j=1}^{h} r_j^2 - 3h(k+1), \qquad (3\text{-}25)$$

where r_j is the rank assigned to a reconciliation technique based on an accuracy measure, k is the number of competing techniques and h is the number of forecasting horizons. The test statistics has approximately a χ^2_{k-1} distribution.

Once we find evidence towards statistical differences among reconciliation techniques, we perform the Nemenyi post-hoc pairwise test to find which reconciliation techniques differ from others. The forecasting accuracies between the two approaches are significantly different if their average ranks differ by a critical distance. The critical distance for the post-hoc pairwise test is defined as:

$$c_{(\alpha,h,k)} = q_\alpha \sqrt{\frac{k(k+1)}{6h}} \qquad (3\text{-}26)$$

where q_α can be found in (DEMŠAR, 2006). We implemented these non-parametric statistical tests using the `nemenyi()` function from the `tsutils` (KOURENTZES, 2019) package in the software `R`.

3.5 Results

3.5.1 Forecasting monthly unemployment in Brazil

We consider first forecasting evaluation using monthly data from the Brazilian Labour Force Survey. The average accuracy results of the several reconciliation strategies when ETS is used as the base forecasting method are depicted in Table 3.3. The table shows the values of the average relative metrics ($AveRelMSE$ and $AveRelMAE$), computed across all hierarchical levels. Numbers highlighted in **bold** indicate the best forecasting performance in each forecasting horizon (number of forecast steps ahead), while numbers in *italics* represent the second best method. Values larger than one indicate that the reconciled forecasts show no improvement on the original accuracy measures.

Reconciliation	Forecast horizon					
Approach	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Average Relative MSE across all hierarchical levels						
HUBER (1)	*0.963*	*0.960*	*0.971*	*0.983*	**0.976**	*0.990*
HUBER (2)	**0.954**	0.985	1.000	0.997	0.990	0.998
BU	1.023	1.014	1.003	1.000	1.008	1.005
OLS	0.956	0.989	1.007	1.002	0.993	1.000
WLS(v)	0.992	0.986	0.992	0.994	0.995	0.996
MinT(Shrink)	0.967	**0.953**	**0.970**	**0.974**	*0.980*	**0.982**
WLS(s)	0.974	0.982	0.991	0.991	0.990	0.994
TDFP	0.975	0.999	1.025	1.018	1.005	1.013
TDGSA	2.031	1.583	1.368	1.235	1.201	1.196
TDGSF	2.036	1.575	1.355	1.222	1.189	1.186
Average Relative MAE across all hierarchical levels						
HUBER (1)	*0.980*	*0.980*	**0.981**	**0.989**	**0.984**	**0.993**
HUBER (2)	0.984	0.989	0.991	0.996	*0.991*	0.999
BU	1.008	1.010	1.003	1.004	1.007	1.005
OLS	0.985	0.991	0.996	0.999	0.993	1.000
WLS(v)	0.993	0.996	0.996	0.998	0.999	0.999
MinT(Shrink)	**0.977**	**0.979**	*0.986*	*0.990*	0.994	*0.994*
WLS(s)	0.987	0.993	0.994	0.996	0.995	0.998
TDFP	0.993	0.997	1.006	1.006	0.997	1.007
TDGSA	1.471	1.287	1.200	1.150	1.124	1.122
TDGSF	1.472	1.286	1.193	1.143	1.119	1.117

Table 3.3: *AveRelMSE* and *AveRelMAE* across all hierarchical levels for different forecast horizons. ETS as base forecasting method.

The table shows promising results for the robust reconciliation based on the HUBER (1) estimator. Considering Average Relative MSE and Average Relative MAE, this approach ranked first or second best in every forecast horizon considered. The improvements in terms of MSE varied from 4% to 4.6% for the two robust reconciliation approaches considered, while the improvements in terms of MAE ranged between 2% and 2.9%.

Considering once again ETS as base forecasting method, the Average Relative MSE and Average Relative MAE results for each hierarchical level are depicted in Tables 3.4 and 3.5, respectively. Figure 3.11, in turn, depicts the average ranks of each reconciliation technique and the results of the Friedman and post-hoc Nemenyi tests. They both show the potential of the robust reconciliation approach for the intermediate and the bottom levels of the hierarchy. The results suggest that the magnitude of the different forecasts to be reconciled produced an undesirable effect that was attenuated by the proposed robust reconciliation techniques. In most forecast horizons considered, the HUBER (1) and the Mint-Shrink estimators ranked among the first and second best reconciliation strategies. When accounting for long-term performance (forecasting lead times of $h = 3$ and over), the HUBER (1) reconciliation approach consistently outperformed the others in terms of MAE reduction. This behavior is coherent with the proposed accuracy measures, since the HUBER (1) works in favor of variance reduction. The HUBER (2) approach, in turn, presented the best performance for short-term forecasting in terms of MSE.

On average, the HUBER (1) presented consistent accuracy gains across all forecasting horizons for the intermediate and lower hierarchy levels. The gains in terms of MSE varied from 0.29% to 3.52% when forecasting unemployment for Metropolitan Areas and 1.51% to 4.89% when analysing the results at the bottom level (disaggregation by sex). The improvements on MAE stayed within the range of 0.41% to 2.75% for Metropolitan Areas and 1.01% to 2.57% for sex.

We note that improvements on forecast accuracy are more evident when the base forecasts are less accurate. When bottom-up or top-down strategies are dominant at a given forecast horizon, to the detriment of more sophisticated reconciliation techniques, it indicates that independent base forecasts could better capture the behavior of the time series at that level. Although in some cases no improvement is noted for the reconciliation strategies, for instance, at the top level of the hierarchy in Tables 3.4 and 3.5, it is important to note that AveRelMSE and AveRelMAE compares the relative performance of only a unique vector of forecasts produced at the top level. When the relatives for all hierarchical levels are considered in a combined metric (Table 3.3), the reconciliation approaches that use limited information, such as bottom-up or top-down, tend to worsen the results. These deficiencies in forecast accuracy are considerable at the top levels of aggregation.

Reconciliation	Forecast horizon					
Approach	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Total						
HUBER (1)	1.037	1.010	1.011	1.007	1.000	1.001
HUBER (2)	1.017	1.007	1.004	1.001	1.001	0.999
BU	1.225	1.116	1.069	1.034	1.026	1.013
OLS	1.014	1.006	1.003	1.000	1.000	0.998
WLS(v)	1.114	1.056	1.035	1.013	1.010	1.001
MinT(Shrink)	1.070	1.025	1.018	**0.999**	**0.999**	*0.990*
WLS(s)	1.081	1.040	1.025	1.008	1.006	1.000
TDFP	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
Metropolitan Areas						
HUBER (1)	0.973	*0.971*	**0.965**	*0.978*	**0.984**	*0.997*
HUBER (2)	**0.961**	1.002	0.999	0.995	0.998	1.004
BU	1.039	1.028	0.997	0.994	1.020	1.014
OLS	*0.964*	1.008	1.008	1.001	1.002	1.007
WLS(v)	1.004	1.000	0.988	0.990	1.007	1.005
MinT(Shrink)	0.980	**0.967**	*0.965*	**0.970**	*0.993*	**0.992**
WLS(s)	0.983	0.995	0.987	0.987	1.000	1.002
TDFP	0.987	1.023	1.031	1.019	1.012	1.019
TDGSA	1.988	1.545	1.295	1.233	1.235	1.252
TDGSF	1.994	1.528	1.274	1.217	1.222	1.241
Sex						
HUBER (1)	0.952	*0.951*	*0.971*	*0.984*	**0.970**	*0.985*
HUBER (2)	**0.945**	0.974	1.000	0.997	0.984	0.994
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	*0.948*	0.979	1.007	1.003	0.988	0.997
WLS(v)	0.976	0.973	0.991	0.994	0.988	0.991
MinT(Shrink)	0.953	**0.940**	**0.969**	**0.974**	*0.972*	**0.976**
WLS(s)	0.961	0.971	0.990	0.992	0.984	0.990
TDFP	0.967	0.987	1.025	1.020	1.001	1.012
TDGSA	2.178	1.666	1.443	1.257	1.202	1.187
TDGSF	2.183	1.661	1.433	1.246	1.191	1.176

Table 3.4: *AveRelMSE* per hierarchical level for different forecast horizons. ETS as base forecasting method.

Reconciliation Approach	Forecast horizon $h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
Total						
HUBER (1)	1.008	1.017	1.014	1.016	1.008	1.010
HUBER (2)	1.006	1.010	1.007	1.009	1.006	1.005
BU	1.080	1.097	1.060	1.060	1.034	1.028
OLS	1.005	1.010	1.007	1.008	1.005	1.004
WLS(v)	1.043	1.056	1.035	1.037	1.024	1.020
MinT(Shrink)	1.026	1.030	1.026	1.031	1.021	1.016
WLS(s)	1.029	1.040	1.028	1.029	1.019	1.016
TDFP	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
Metropolitan Areas						
HUBER (1)	*0.988*	*0.981*	**0.973**	**0.990**	**0.990**	**0.996**
HUBER (2)	1.002	0.999	0.987	1.002	*0.994*	1.003
BU	1.013	1.015	0.999	1.003	1.017	1.010
OLS	1.003	1.001	0.994	1.005	0.996	1.004
WLS(v)	0.998	0.997	0.990	0.999	1.007	1.003
MinT(Shrink)	**0.984**	**0.985**	*0.984*	*0.991*	1.002	*0.999*
WLS(s)	0.996	0.996	0.988	0.998	1.001	1.001
TDFP	1.010	1.012	1.004	1.011	1.001	1.007
TDGSA	1.486	1.317	1.179	1.177	1.174	1.171
TDGSF	1.483	1.318	1.169	1.170	1.167	1.166
Sex						
HUBER (1)	0.974	*0.976*	**0.982**	**0.985**	**0.980**	**0.990**
HUBER (2)	*0.973*	0.982	0.992	0.992	0.989	0.997
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	0.974	0.984	0.996	0.995	0.990	0.998
WLS(v)	0.987	0.990	0.995	0.994	0.993	0.996
MinT(Shrink)	**0.969**	**0.972**	*0.984*	*0.987*	*0.987*	*0.991*
WLS(s)	0.979	0.988	0.994	0.992	0.991	0.996
TDFP	0.984	0.990	1.007	1.004	0.996	1.007
TDGSA	1.512	1.300	1.229	1.150	1.110	1.109
TDGSF	1.515	1.298	1.222	1.143	1.106	1.104

Table 3.5: *AveRelMAE* per hierarchical level for different forecast horizons. ETS as base forecasting method.

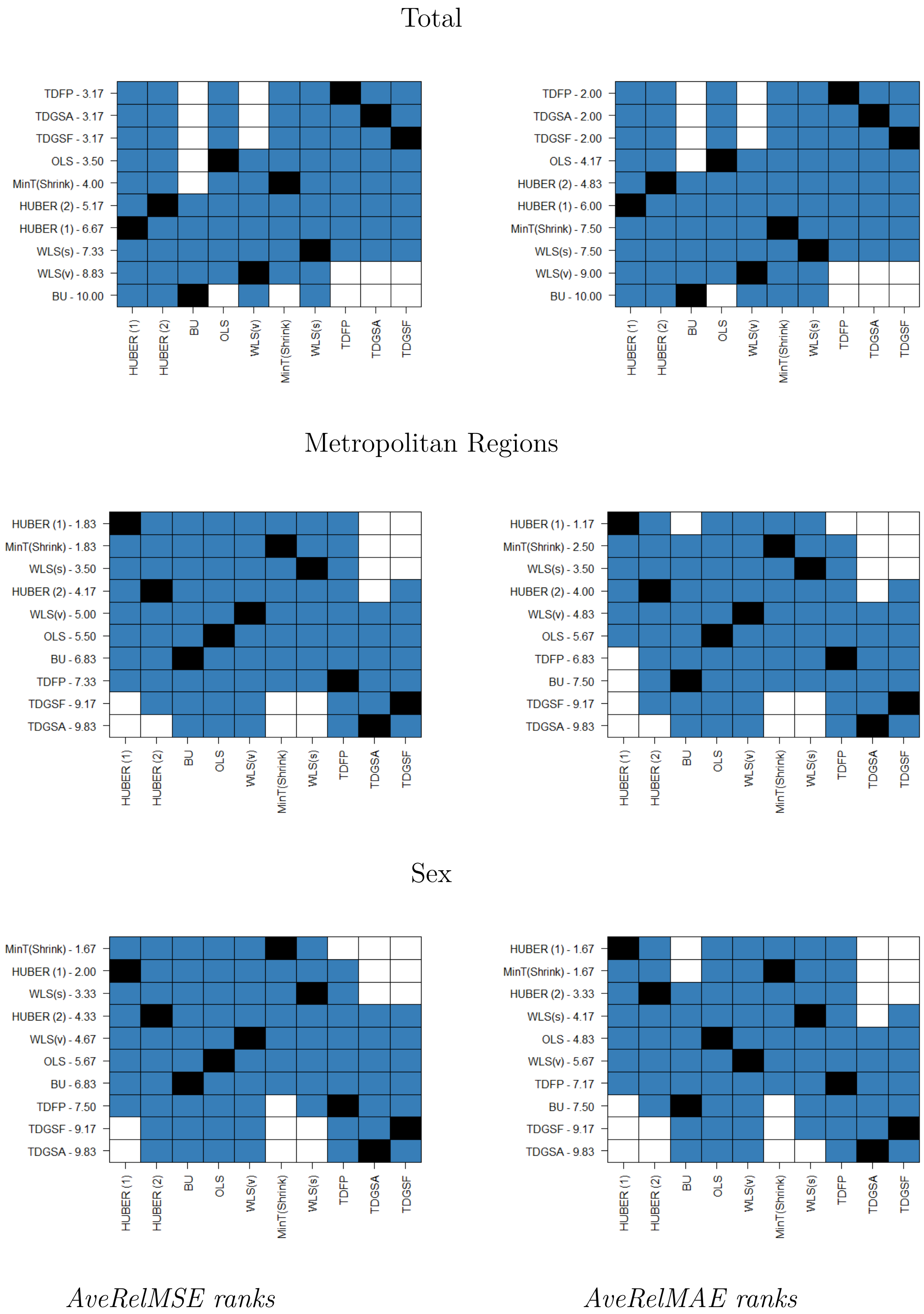

Figure 3.11: Average rank and Nemenyi post-hoc test after the Friedman test for different hierarchical levels based on *AveRelMSE* and *AveRelMAE* - 5% significance level

The vertical axis in every chart of Figure 3.11 shows the average ranks of the reconciliation strategies according to the selected evaluation metric (AveRelMSE on the left, AveRelMAE on the right). Each method is compared against the others in the columns. black cells indicate no statistical difference between the compared approaches, i.e., they are suggested to have similar accuracies. White cells, in turn, indicate that there is a difference between the performances of the compared approaches. At the top level of the hierarchy, no differences can be told concerning the performance of regression-based reconciliation strategies and the performance of the top-down approaches. At the intermediate level, however, the test results indicate that HUBER (1), *MinT-Shrink* and $WLS(s)$ belong to a different performance cluster compared with the top-down strategies in terms of MSE. When comparing the MAE, the HUBER (1) is significantly different from all top-down approaches. The scenario is very similar for the bottom hierarchical level.

As a second experiment using monthly data from PME, we investigate the performance of the reconciliation approaches when base forecasts are generated via ARIMA formulations. In this case, the accuracy gains in favor of the robust settings are noted across all hierarchical levels for short-term horizons, as shown in Table 3.6. These results corroborate the findings of Tables 3.4 and 3.5, where the improvements in terms of MSE and MAE are noted for each hierarchical level in the short-term.

The average relative metrics results per hierarchical level are shown in Tables 3.7 and 3.8. The most relevant gains in this case occur at the top level, with an approximately 9% reduction in terms of MSE and 3.4% reduction for the MAE when forecasting up to two steps ahead. Even though the BU method provides more accurate results for longer periods at the top and intermediate level of the hierarchy, by construction this approach is not able to offer reconciliation improvements at the bottom level. This fact is due to its additive construction from the bottom, as described in Section 3.2. According to Figure 3.12, it is possible to identify a cluster of methods of superior performance relative to the others when comparing the *AveRelMSE* at the top of the hierarchy. On the other hand, based on *AveRelMAE* ranks, the critical distance is not enough to divide the reconciliation strategies into groups of specific forecasting performance.

Reconciliation	Forecast horizon					
Approach	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Average Relative MSE across all hierarchical levels						
HUBER (1)	*0.950*	**0.946**	0.991	1.027	1.063	1.102
HUBER (2)	**0.948**	0.981	1.039	1.125	1.205	1.272
BU	1.012	1.015	0.997	**0.987**	**0.979**	**0.978**
OLS	**0.948**	0.981	1.039	1.125	1.205	1.273
WLS(v)	0.983	0.980	**0.987**	*0.999*	1.010	1.020
MinT(Shrink)	0.954	0.987	1.030	1.087	1.134	1.178
WLS(s)	0.960	*0.966*	*0.991*	1.033	1.068	1.096
TDFP	0.960	0.960	1.014	1.091	1.167	1.225
TDGSA	1.833	1.540	1.373	1.403	1.471	1.554
TDGSF	1.832	1.529	1.361	1.392	1.462	1.546
Average Relative MAE across all hierarchical levels						
HUBER (1)	**0.976**	**0.984**	1.001	1.013	1.036	1.059
HUBER (2)	*0.984*	1.006	1.032	1.067	1.118	1.152
BU	1.009	1.012	1.000	**0.989**	**0.984**	**0.981**
OLS	*0.984*	1.006	1.032	1.067	1.118	1.153
WLS(v)	0.991	0.995	**0.995**	*0.996*	1.007	1.011
MinT(Shrink)	0.983	1.008	1.026	1.048	1.078	1.101
WLS(s)	0.985	*0.992*	0.999	1.016	1.041	1.056
TDFP	0.989	0.995	1.007	1.047	1.100	1.132
TDGSA	1.406	1.304	1.221	1.224	1.277	1.297
TDGSF	1.406	1.299	1.215	1.219	1.272	1.293

Table 3.6: *AveRelMSE* and *AveRelMAE* across all hierarchical levels for different forecast horizons. ARIMA as base forecasting method.

Reconciliation	Forecast horizon					
Approach	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Total						
HUBER (1)	**0.964**	**0.910**	0.898	0.882	0.867	0.865
HUBER (2)	*0.996*	0.973	0.966	0.963	0.957	0.952
BU	1.140	0.949	**0.863**	**0.816**	**0.769**	**0.741**
OLS	*0.996*	0.973	0.966	0.963	0.957	0.953
WLS(v)	1.050	*0.936*	*0.891*	*0.869*	*0.836*	*0.818*
MinT(Shrink)	1.040	0.962	0.934	0.910	0.885	0.867
WLS(s)	1.028	0.934	0.900	0.884	0.858	0.843
TDFP	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
Metropolitan Areas						
HUBER (1)	*0.947*	**0.975**	1.006	1.030	1.061	1.106
HUBER (2)	**0.944**	1.011	1.054	1.135	1.212	1.287
BU	1.016	1.057	1.016	**0.993**	**0.976**	**0.979**
OLS	**0.944**	1.011	1.054	1.135	1.212	1.288
WLS(v)	0.985	1.015	1.003	1.003	1.007	1.020
MinT(Shrink)	0.953	1.022	1.049	1.099	1.140	1.186
WLS(s)	0.957	0.997	1.005	1.037	1.067	1.101
TDFP	0.956	0.993	1.023	1.095	1.167	1.231
TDGSA	1.701	1.493	1.343	1.427	1.510	1.614
TDGSF	1.699	1.475	1.327	1.415	1.501	1.607
Sex						
HUBER (1)	0.951	**0.935**	*0.992*	1.038	1.083	1.122
HUBER (2)	**0.946**	0.967	1.037	1.134	1.225	1.296
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	**0.946**	0.967	1.037	1.134	1.225	1.297
WLS(v)	0.976	0.967	**0.988**	1.009	1.027	1.039
MinT(Shrink)	*0.947*	0.971	1.029	1.097	1.155	1.204
WLS(s)	0.957	0.954	*0.992*	1.044	1.087	1.117
TDFP	0.958	*0.953*	1.010	1.098	1.183	1.242
TDGSA	2.002	1.622	1.426	1.431	1.500	1.582
TDGSF	2.001	1.613	1.414	1.419	1.490	1.573

Table 3.7: *AveRelMSE* per hierarchical level for different forecast horizons. ARIMA as base forecasting method.

Reconciliation	Forecast horizon					
Approach	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Total						
HUBER (1)	**0.979**	**0.967**	0.936	0.919	0.914	0.913
HUBER (2)	1.002	0.995	0.976	0.974	0.972	0.972
BU	1.119	0.979	*0.899*	**0.865**	**0.829**	**0.822**
OLS	1.002	0.995	0.976	0.974	0.972	0.9723
WLS(v)	1.039	0.986	**0.920**	*0.903*	*0.891*	*0.885*
MinT(Shrink)	1.024	0.989	0.950	0.938	0.931	0.927
WLS(s)	1.021	*0.984*	0.926	0.917	0.906	0.902
TDFP	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
Metropolitan Areas						
HUBER (1)	**0.979**	1.005	1.013	1.016	1.029	1.055
HUBER (2)	0.998	1.029	1.043	1.072	1.117	1.148
BU	1.009	1.041	1.019	**0.989**	**0.979**	**0.972**
OLS	0.998	1.029	1.043	1.072	1.116	1.147
WLS(v)	0.994	1.018	1.008	*0.997*	*0.999*	1.004
MinT(Shrink)	0.994	1.031	1.041	1.053	1.072	1.098
WLS(s)	*0.989*	1.011	1.008	1.018	1.034	1.050
TDFP	1.007	1.011	1.003	1.046	1.095	1.128
TDGSA	1.340	1.346	1.213	1.243	1.297	1.318
TDGSF	1.387	1.342	1.209	1.240	1.291	1.314
Sex						
HUBER (1)	*0.974*	**0.975**	1.001	1.020	1.051	1.075
HUBER (2)	0.975	0.996	1.031	1.072	1.132	1.171
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	0.975	0.996	1.031	1.072	1.131	1.172
WLS(v)	0.986	*0.985*	**0.995**	1.004	1.021	1.026
MinT(Shrink)	**0.974**	0.998	1.025	1.055	1.094	1.118
WLS(s)	0.980	0.983	1.001	1.023	1.056	1.073
TDFP	0.980	0.986	1.011	1.051	1.111	1.147
TDGSA	1.454	1.312	1.246	1.235	1.293	1.315
TDGSF	1.457	1.306	1.237	1.229	1.288	1.310

Table 3.8: *AveRelMAE* per hierarchical level for different forecast horizons. ARIMA as base forecasting method.

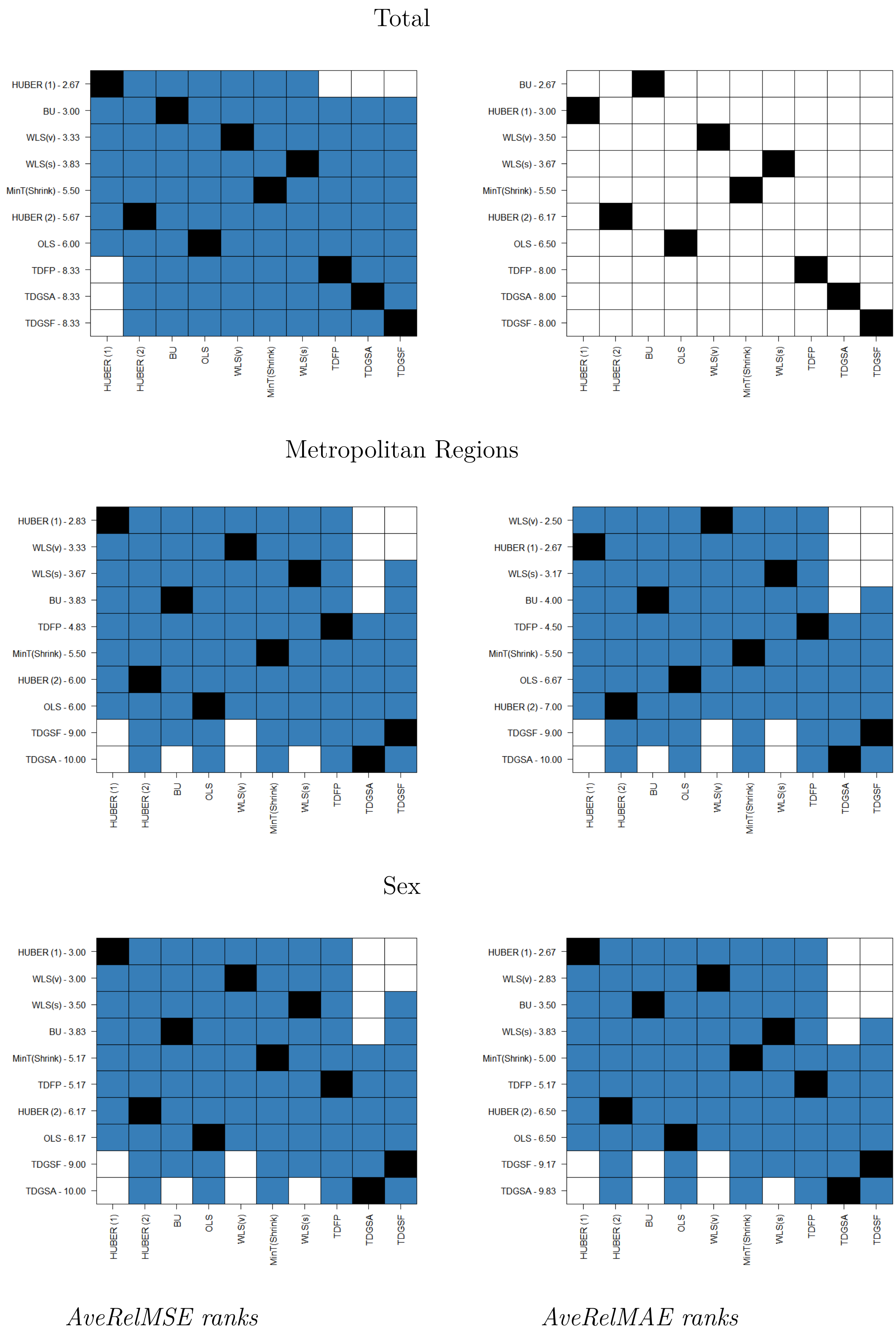

Figure 3.12: Average rank and Nemenyi post-hoc test after the Friedman test for different hierarchical levels based on *AveRelMSE* and *AveRelMAE* - 5 % significance level - ARIMA as base forecasting method

3.5.2 Forecasting quarterly unemployment in Brazil

The second data set concerns quarterly data from the Brazilian Continuous National Household Sample Survey (PNADC) on Total Unemployment. The values of the average relative (MSE and MAE) metrics across all hierarchical levels are shown in Table 3.9 for reconciliation techniques applied to ETS base forecasts. The HUBER (1) approach presented the most accurate results for long-term horizons, i.e., three and four quarters ahead, while the MinT(Shrink) provided the best results $h = 1$ and $h = 2$, followed by HUBER (1) when $h = 2$.

Reconciliation	*AveRelMSE*				*AveRelMAE*			
Approach	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
All hierarchical levels								
HUBER (1)	0.928	*0.961*	**0.925**	**0.941**	0.970	*0.964*	**0.949**	**0.949**
HUBER (2)	1.000	1.222	1.315	1.434	0.992	1.081	1.121	1.153
BU	1.018	1.003	0.999	1.008	1.019	1.008	1.015	1.011
OLS	1.000	1.222	1.315	1.434	0.992	1.081	1.121	1.153
WLS(v)	0.945	0.971	*0.967*	*0.945*	0.963	0.991	*0.983*	*0.954*
MinT(Shrink)	**0.806**	**0.937**	1.008	1.034	**0.880**	**0.952**	0.989	0.986
WLS(s)	0.915	1.004	1.001	0.981	0.956	1.006	0.992	0.975
TDFP	*0.876*	1.025	1.039	1.066	*0.935*	1.012	1.010	1.001
TDGSA	2.951	2.246	2.036	1.925	1.830	1.581	1.527	1.404
TDGSF	2.948	2.245	2.036	1.925	1.829	1.580	1.527	1.404

Table 3.9: *AveRelMSE* and *AveRelMAE* across all hierarchical levels for different forecast horizons. ETS as base forecasting method.

Table 3.10 shows the results for the *AveRelMSE* and *AveRelMAE* computed for each level of the PNADC hierarchy – top (Brazil), intermediate (the five Great Regions) and bottom (27 Federative Units, distributed across the five Great Regions). Accuracy gains for the HUBER (1) estimator are observed in all levels of the hierarchy for most forecast horizons. In this case, it is fair to say that the HUBER (1) estimator has the most dominant performance compared with the state-of-the-art approaches. The HUBER (2) estimator also present competitive results in the top level of the hierarchy, particularly in very short-term horizons. Finally, the results presented in Figure 3.13 indicate that the HUBER (1) robust reconciliation technique differs significantly, from the statistical viewpoint, from the TDGSA and TDGSF Top-down approaches for the intermediate and bottom levels. The results for the top of the hierarchy suggest that all strategies differ significantly from each other, i.e., no clusters of methods of similar performance can be identified.

Reconciliation Approach	*AveRelMSE* h = 1	h = 2	h = 3	h = 4	*AveRelMAE* h = 1	h = 2	h = 3	h = 4
Total (Brazil)								
HUBER (1)	1.084	**0.787**	**0.702**	*0.698*	0.978	**0.850**	**0.833**	0.873
HUBER (2)	*0.975*	0.928	0.893	0.835	0.949	0.975	0.932	0.902
BU	1.266	0.884	0.863	0.923	1.188	0.947	1.002	1.050
OLS	*0.975*	0.928	0.893	0.835	0.949	0.975	0.932	0.902
WLS(v)	1.049	0.816	*0.757*	0.717	0.990	0.919	0.907	0.868
MinT(Shrink)	**0.792**	*0.795*	0.792	0.728	**0.798**	*0.918*	0.901	*0.826*
WLS(s)	1.020	0.827	0.759	**0.683**	*0.931*	0.930	*0.893*	**0.786**
TDFP	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Great Regions								
HUBER (1)	0.913	**0.938**	**0.889**	**0.922**	0.972	**0.973**	**0.943**	**0.965**
HUBER (2)	0.885	1.197	1.236	1.367	0.957	1.093	1.140	1.135
BU	1.075	1.044	1.025	1.069	1.092	1.068	1.099	1.067
OLS	0.885	1.197	1.236	1.367	0.957	1.093	1.140	1.135
WLS(v)	0.928	0.977	*0.947*	0.937	0.970	1.025	1.007	0.972
MinT(Shrink)	**0.747**	0.976	1.055	1.087	**0.847**	1.008	1.031	1.009
WLS(s)	*0.884*	1.010	0.966	*0.930*	*0.943*	1.048	1.018	*0.955*
TDFP	0.907	1.151	1.177	1.258	0.983	1.102	1.110	1.112
TDGSA	3.447	2.883	2.791	2.904	2.007	1.796	1.889	1.734
TDGSF	3.445	2.883	2.792	2.905	2.006	1.796	1.889	1.735
Federative Units								
HUBER (1)	0.925	*0.973*	**0.941**	**0.954**	0.969	*0.967*	**0.955**	**0.949**
HUBER (2)	1.024	1.240	1.349	1.476	1.000	1.083	1.126	1.167
BU	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
OLS	1.024	1.240	1.349	1.476	1.000	1.083	1.126	1.167
WLS(v)	0.944	0.976	*0.979*	*0.957*	0.961	0.987	*0.981*	*0.954*
MinT(Shrink)	**0.818**	**0.936**	1.009	1.038	**0.889**	**0.943**	0.984	0.988
WLS(s)	0.917	1.010	1.018	1.004	0.959	1.002	0.991	0.986
TDFP	*0.866*	1.004	1.017	1.036	*0.924*	0.997	0.993	0.981
TDGSA	2.985	2.210	1.972	1.828	1.840	1.571	1.492	1.367
TDGSF	2.981	2.209	1.971	1.828	1.839	1.570	1.491	1.367

Table 3.10: *AveRelMSE* and *AveRelMAE* per hierarchical level for different horizons. ETS as base forecasting method.

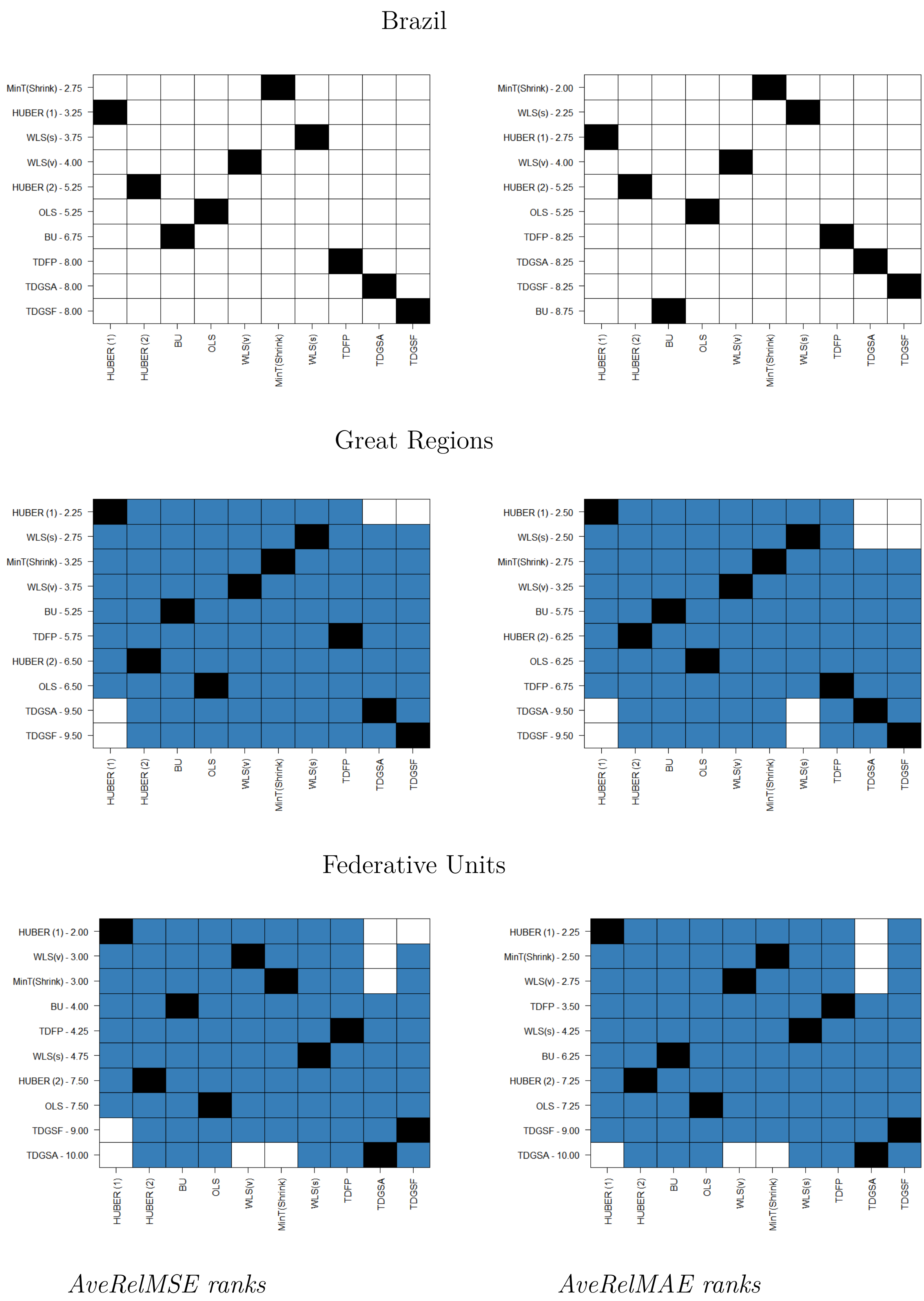

Figure 3.13: Average rank and Nemenyi post-hoc test after the Friedman test for different hierarchical levels based on *AveRelMSE* and *AveRelMAE* – Tests conducted at the 5% significance level.

3.5.3
Discussions and benefits from reconciliation

The results outlined in Sections 3.5.1 and 3.5.2 endorse the strength of our proposed reconciliation approaches. The experiments demonstrate the value of combining univariate methods with reconciliation approaches. The first set of methods usually provide reasonable estimates when the interest lies in multi-step ahead forecasting time series that present a range of stylized facts, such as unemployment data (as in our case) and selected sectoral time series, such as energy consumption (OLIVEIRA; OLIVEIRA, 2018) and tourism demand data (ATHANASOPOULOS et al., 2011). Reconciliation techniques, in turn, benefit from the hierarchical structure of the involved time series. Indeed, from a theoretical perspective, recent reconciliation techniques were proposed to guarantee that reconciled forecasts are at least as good the base forecasts. This is because the resulting revised forecasts will be unbiased and will add up appropriately across the hierarchy (HYNDMAN et al., 2011), as opposed to independently forecasting each time series. In addition, when the representation of the variance-covariance matrix of the reconciled forecast errors satisfies the $SPS = P$ condition – which is usually the case in most recent reconciliation techniques –, the reconciled forecasts will also have minimum variance amongst all possible combinations of forecasts (WICKRAMASURIYA et al., 2020). In PANAGIOTELIS et al. (2021), the authors focus on the geometric interpretation and representation of the reconciliation based on a projection into coherent spaces and how it can be addressed to two desirable properties. First, reconciliation should produce minimal adjustment on base forecasts and, secondly, reconciliation techniques should improve forecast accuracy. The article HOLLYMAN et al. (2021) adds to this discussion by pointing that forecasts generated via reconciliation techniques inherit the benefits of combined forecasts, which have been validated in a considerable range of empirical settings and have often been shown to outperform the best *ex-ante* forecasting method.

Our results corroborate the above line of reasoning, as not only the proposed approaches, but also several established reconciliation techniques delivered reliable estimates for the time series in the two data sets considered. Furthermore, our proposed methods also highlight the benefits of properly addressing the outliers in the reconciliation stage, further enhancing the quality of the revised forecasts.

As a brief note, we acknowledge that modelling and forecasting independently each time series may make more sense if the interest lies in a single or a limited set of time series belonging to the same hierarchical level. However, when the interest lies in the comprehensive set of time series under a defined hierarchical structure, as in our case, finding the best of suite approach for each time series will almost surely result in forecasts that do not add up across the hierarchy, losing interpretation of the overall results. Hence, hierarchical forecasting methods that generate coherent forecasts, as those herein proposed, should be considered to allow appropriate decision-making at the different levels (PETROPOULOS et al., 2022).

3.6
Conclusions and future studies

Providing data to support decisions involving multiple hierarchical levels typically requires consolidating estimates. In this sense, our empirical evaluation shows encouraging results in favor of the proposed robust reconciliation approaches. Even though these estimators are not uniformly dominant for all scenarios, they rank among the first or second best approaches in almost every case, being competitive with state-of-the-art techniques in hierarchical forecasting reconciliation. We find considerable gains in accuracy according to the assessment metrics for different hierarchical levels. Overall, the HUBER (1) is the most consistent strategy, delivering competitive results in both short and long-run horizons and outperforming most methods in terms of MSE and MAE.

Our results provide relevant insights to a wide range of stakeholders, such as: (i) policymakers and public administrators at the ground zero of decision-making; (ii) academics and practitioners interested in sharpening/improving the quality of their forecasting models by taking into consideration the hierarchical structure of the data; and (iii) business managers concerned with unemployment levels in the region(s) in which their operations are located.

A natural development of this work is to evaluate the unemployment expressed as a percentage of the total labor force, which translates into the unemployment rate, at the constraint of forecasting coherently for a given hierarchy. Our research focused on the numerator of the indicator since the hierarchical framework works for totals. However, the same approach can be extended to non-linear quantities that can be expressed as a function of totals.

Another point worth noting is that, in this paper, we focused on a particular class of M-estimators when proposing the robust reconciliation strategies. However, other influence functions can be implemented with different tuning parameters. In addition, other classes of robust estimators such as S-estimators and MM estimators can also be explored.

We should also highlight that we used only base forecasts from a single model strategy to compute the reconciled forecasts. A straightforward alternative would be to consider combination techniques to improve the final forecasts. Moreover, further investigation can be done to assess whether the benefits from the forecast combination in this framework come from the base forecasts or the combined reconciled ones.

Finally, even though we focused in this paper in data sets of total unemployment across several regions in Brazil, we note that the methods herein presented are flexible in the sense that they are not scale dependent and can be applied to other sets of hierarchical time series and provide competitive results.

4
Second contribution: A novel reconciliation approach for hierarchical electricity consumption forecasting based on resistant regression

This Chapter proposes a hierarchical forecast reconciliation approach aimed at delivering accurate forecasts of energy demand across all divisions in a power system. It relies on the use of resistant-based estimators to aid in the process of forecast reconciliation. This study was published at Energy (ISSN 0360-5442), (MEIRA et al., 2023).

4.1
Introduction

Power demand forecasting has become a critical task in several application systems. Reliable energy consumption forecasts, for instance, are not only required for the optimal control and scheduling of power systems, but they also form the basis of the electrical energy trade and spot price calculation in financial markets (OLIVEIRA; OLIVEIRA, 2018).

The reliance and accuracy of future electricity demand and supply forecasting have received special attention in emerging countries with vulnerable supply systems (VELASQUEZ et al., 2022). In Brazil, for instance, the National Interlinked System (SIN) has shown to be vulnerable to electricity shortages and has demanded significant overhaul in order to address its challenges (TORRINI et al., 2016). This has become clearer in the last years, when the country barely escaped forced electricity supply shortages due to severe and prolonged droughts.

If on the one hand the limited diversification of the Brazilian electric matrix, for which hydropower generation corresponds to almost two-thirds of the national installed capacity, leaves the Brazilian power system sensitive to weather conditions (OLIVEIRA et al., 2015), on the other hand, the country's energy consumption has risen rapidly in recent decades, lifting its world ranking to the sixth largest energy consumer in 2021 (ENERDATA, 2022). Furthermore, energy consumption levels in Brazil are expected to rise even further in the next decades. For instance, the Brazilian Energy Research Company (EPE) is in charge of publishing two official reports, at the request of the Brazilian Ministry of Mines and Energy: The 10-year Energy Expansion Plan and the National Energy Planning 2050. Both documents are adamant in their position, claiming that the demand for electric energy will continue to grow in the coming years. For instance, according to PDE 2031, Brazil foresees an increase in electricity consumption by around 3.5% per year, from 562.6 TWh at the end of 2021 to 791.9 TWh in 2031 (EPE, 2022a).

The official electricity demand forecasts provided by EPE relate to very long-term horizons, i.e., they provide estimates of future electric energy consumption for lead times of ten or more years. Forecasts for such horizons can aid in setting long-term energy planning goals but are of limited importance in most operations planning applications. Short to mid-term demand forecasts of energy consumption, i.e., forecasts for horizons ranging from the first few hours to several months in advance, in turn, are crucial for several reasons. First, short-term forecasts not only form the basis of the electrical energy trade and spot price calculation (CASTELLI et al., 2015), but they are also required for the control and scheduling of power systems (TAYLOR; MCSHARRY, 2007). Mid-term electricity load forecasting, in turn, is particularly important in the scenario of smart grid development (TANG et al., 2022). KABOLI et al. (2017) add that forecasting within this time frame is especially interesting for companies operating in deregulated environments, as accurate forecasts provide valuable information about the market need of energy, the need for unit maintenance, the fuel supplies, and the balance of imports and exports of energy.

Another point worth noting is that there is no official forecasting data available for consumption at the sector (Industrial, Residential, Commercial and Others) or sub-sector levels (data for the different geographic regions within sectors). Estimates of consumption across such disaggregated levels are also of paramount importance, as the lack of sufficient energy load in parts of the power system may cause the whole system to collapse.

This paper aims to fill the above-discussed gaps by estimating and subsequently forecasting Brazilian monthly electric energy demand at both total and disaggregated levels. To that end, a novel forecast reconciliation approach is proposed combining resistant regression techniques and hierarchical time series methods. By considering resistant statistics in the forecast reconciliation process, we address a significant gap in the modelling and forecasting of hierarchical time series: the substantial reduction of contamination in the involved series due to outlying observations.

4.1.1 Brief overview on forecast reconciliation

Forecast Reconciliation stands for a process by which independently generated forecasts of a collection of linearly related time series are reconciled via the introduction of accounting aggregations that naturally apply to the data (HOLLYMAN et al., 2021). For instance, since national and sub-national data of energy consumption in a power system are linearly related, as there is a natural hierarchy among these series (sub-national data aggregate into the total consumption), reconciliation methods can take advantage of the underlying hierarchical structure to slightly adjust, i.e., combine using proper weighting matrices, the original forecasts, so that the reconciled forecasts are usually more accurate than the original ones.

Hierarchical forecast reconciliation has gained considerable attention in the time series forecasting literature, particularly after the work of HYNDMAN et al. (2011), who put forth the idea of optimally combining forecasts using a regression-based perspective. In short, the authors demonstrated that if the error covariance matrix is known, a generalized least squares regression solution provides the optimal combination of forecasts. However, since this covariance matrix is often challenging to estimate in practice, the authors adopted a simplifying assumption of additivity on the forecast errors. HYNDMAN et al. (2016), in turn, showed how to efficiently handle very complex hierarchical structures, particularly when there are millions of time series at the most disaggregated level. The authors also proposed the use of the Weighted Least Squares (WLS) estimator to obtain an estimate of the variance-covariance matrix of the reconciled forecast errors.

A probabilistic hierarchical forecast framework was built, in which the dependency between the nodes at each hierarchical level is obtained by reordering quantile forecasts (TAIEB et al., 2017). JEON et al. (2019), in turn, proposed new approaches for reconciling probabilistic forecasts by combining information from density forecasts at all hierarchical levels. WICKRAMASURIYA et al. (2019) introduced the Minimum Trace ($MinT$) reconciliation approach, putting forth an alternative to estimate the variance-covariance matrix of the reconciled forecast errors. More recently, PRITULARGA et al. (2021) proposed the idea of defining coherency, i.e., the property that ensures that forecasts add up properly throughout the hierarchy, as stochastic. According to the authors, this allows practitioners to better understand overlooked uncertainties in the forecast reconciliation process.

Recently was introduced the idea of using robust estimators during forecast reconciliation to dampen the influence of outliers that may occur in selected forecasts of the hierarchy (LILA et al., 2022). However, whilst robust reconciliation methods may provide accurate reconciled forecasts on several occasions, they are only capable of partially addressing the influence of outliers. When the involved forecasts contain multiple outliers and these occur in both extremes of the distribution, robust regressions may find it hard to withstand the proportion of contamination and still provide reliable outcomes.

4.1.2
Proposal and relevance to energy issues

Our work constitutes a pioneer effort in considering the use of resistant estimators to the context of hierarchical forecast reconciliation. Drawing from the fields of statistics, optimization, and time series, we propose a resistant-based reconciliation approach that is theoretically constructed to not be influenced by outliers and other influential points. The contributions of this research can be summarized as follows:

1. Evaluate several robust and resistant parameter estimation methods and understand how these methods can be effectively adapted to be used in the context of hierarchical forecast reconciliation;

2. Propose a methodology for adapting the Least Absolute Deviation (LAD) resistant-based estimator to the context of hierarchical forecast reconciliation, allowing outliers and influential points that may be present in the base forecasts to have minimal or even null effects on reconciliation weights. This is particularly important in the context of energy demand time series, as the different stylized facts that may be present in these time series, such as nonlinearities, stochastic components (trend, seasonality, residuals), heteroscedasticity, presence of structural breaks, among others, may interfere in the production of reconciled forecasts;

3. Consider an application of the developed approach to reconcile base forecasts from a particular set of hierarchical time series, i.e., the time series that represent the monthly electric energy demand across all divisions of the Brazilian power system;

4. Conduct several forecasting experiments comparing the forecast performance of the proposed resistant reconciliation approach with that from several traditional and state-of-the-art hierarchical forecasting methods. In addition, consider a broad range of settings such as different forecast lead times, cross-validation with rolling forecast origins, among others, as well as multiple forecast evaluation metrics to attest the robustness of the proposed approach in delivering accurate forecasts across multiple levels of the hierarchy; and

5. Make the proposed methodology flexible so that it can be readily applied to other sets of hierarchical time series and deliver equally reliable results.

In terms of its relevance, we emphasize that the proposed approach not only addresses a significant gap in the modelling and forecasting of hierarchical time series, i.e., the reduction of contamination due to outlying observations, but it also delivers accurate energy consumption forecasts across distinct aggregation levels in a power system. Given the ongoing complexity of power systems throughout the years, accurate estimates of energy demand in distinct regions and classes of a power system are of paramount importance, as the lack of sufficient energy load in parts of the power system may cause the whole system to collapse.

The rest of the paper unfolds as follows: Section 4.2 outlines the basic strategies in Hierarchical Time Series (HTS) forecasting and details the most up-to-date framework. It also introduces the concept of resistant reconciliation in hierarchical forecasting. Section 4.3 describes the experimental setup. Section 4.4 summarizes the results and assesses their implications. Finally, Section 4.5 concludes and suggests directions for future research.

4.2
HTS: from benchmarks to the state-of-the-art

4.2.1
Basic concepts

Some sets of time series can be presented in the form of Hierarchical Time Series (HTS). This term stands for a collection of time series that follow a defined hierarchical aggregation structure. HTS can be found in several industries, and their analysis usually give valuable insights in terms of operations and management. For instance, total electric energy consumption in a given country can be disaggregated into Classes (or Sectors of Consumption) and these can be further divided into Geographic Regions.

In order to characterize HTS, consider a three-level, balanced hierarchy with $m = 10$ time series and information at time t. This hierarchy can be illustrated in the following fashion:

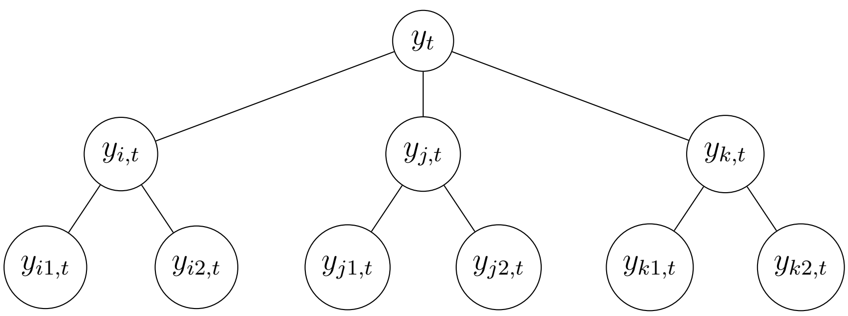

Figure 4.1: Hierarchical structure for a three levels hierarchy.

In a hierarchical structure, internal consistency is expected. In this case, by the additive property of the data, we have the following construction:

$$y_t = y_{i,t} + y_{j,t} + y_{k,t} \tag{4-1}$$

and

$$y_t = \sum_{l=1}^{2} y_{il,t} + \sum_{l=1}^{2} y_{jl,t} + \sum_{l=1}^{2} y_{kl,t} \tag{4-2}$$

A set of hierarchical time series can be represented in matrix notation. Let $\mathbf{y_t}$ be a vector of size m, as defined earlier, comprising observations from all hierarchical levels at time t. It is possible to define an appropriate matrix $\mathbf{S}$ of dimension $m \times n$ such that,

$$\mathbf{y_t} = \mathbf{S}\mathbf{y_t^b} \tag{4-3}$$

where $\mathbf{y_t^b}$ is a n-vector containing the observations at the most disaggregated level of the hierarchy, as illustrated in equation 4-4.

$$\underbrace{\begin{bmatrix} y_t \\ y_{i,t} \\ y_{j,t} \\ y_{k,t} \\ y_{i1,t} \\ y_{i2,t} \\ y_{j1,t} \\ y_{j2,t} \\ y_{k1,t} \\ y_{k2,t} \end{bmatrix}}_{\mathbf{y_t}} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} y_{i1,t} \\ y_{i2,t} \\ y_{j1,t} \\ y_{j2,t} \\ y_{k1,t} \\ y_{k2,t} \end{bmatrix}}_{\mathbf{y_t^b}} \tag{4-4}$$

4.2.2
Forecasting HTS - traditional methods

In a hierarchical forecasting framework, all time series are first forecasted using independent methods, giving birth to "base forecasts". Then, a reconciliation approach is sought for, aiming at delivering "coherent", final forecasts. By coherent, one means forecasts that are unbiased, have minimum variance amongst all combinations and add up properly across the hierarchy. In other words, the values of the forecasts should add in a manner that is consistent with the underlying aggregation structure. Turning once again to the example on power systems, the forecasts across all classes of consumption should add to the country level. The reconciliation process can be represented in the flowchart depicted in Figure 4.2.

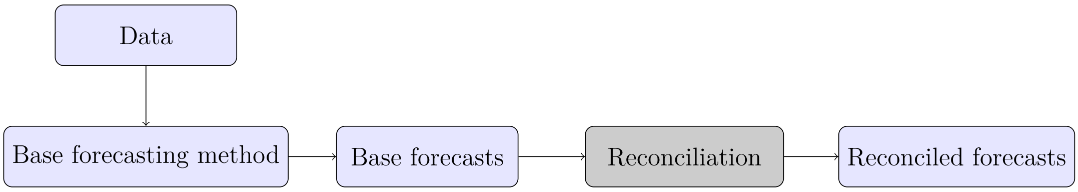

Figure 4.2: Forecast reconciliation flowchart.

To aid in the visualisation of the coming explanations, we will refer to the three-level hierarchical structure considered for the Brazilian power system. This structure first divides total (national) electricity consumption (top-level of the hierarchy) by Classes (intermediate level) and then by Geographic Regions (bottom-level). There are four Classes of Consumption in the Brazilian power system: Commercial, Industrial, Residential and Others (mainly rural, public service and public lighting). Geographic Regions, in turn, are divided into five groups: North, Northeast, Midwest, Southeast and South.

In mathematical terms, the process of forecast reconciliation can be represented as follows. First, consider $\hat{\mathbf{y}}_{t+h|t}$ a vector of h steps ahead base forecasts, generated using independent methods, with the same arrangement as $\mathbf{y_t}$. Thus, for a given matrix $\mathbf{P}$ of dimension $n \times m$, we have the following equation

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SP}\hat{\mathbf{y}}_{\mathbf{t+h|t}} \tag{4-5}$$

where $\tilde{\mathbf{y}}_{t+h|t}$ are the reconciled forecasts. The $\mathbf{SP}$ matrices represent the reconciliation process, which maps independent (or incoherent) forecasts into coherent ones. Depending on how $\mathbf{P}$ is structured, it is possible to reproduce several traditional hierarchical forecasting approaches. For instance, by letting $\mathbf{P} = \left[\mathbf{0_{n\times(m-n)}}|\mathbf{I_n}\right]$, where $0_{n\times(m-n)}$ is a null matrix, one can reproduce the *Bottom-Up* (BU) approach. In this case, the $\mathbf{P}$ matrix keeps only forecasts for the most granular level, then $\mathbf{S}$ completes the processes by generating reconciled forecasts at every other level of the hierarchy. Under this approach, and turning to the example of power system divisions, information from Brazilian geographic regions within classes of consumption (the most granular level) will provide the forecasts for the entire hierarchy of energy consumption. It is worth noting that bottom-up forecasts, although unbiased by construction, usually capture more volatility than other reconciliation approaches.

Top-Down forecasts, in turn, can be achieved by making $\mathbf{P} = \left[\mathbf{p}|\mathbf{0_{n\times(m-1)}}\right]$, where $\mathbf{p}$ is the set of proportions, or weights, applied to forecasts at the top level. Common approaches to obtain these proportions are the Top-Down Gross-Sohl methods A and (TDGSA and TDGSF), described in GROSS & SOHL (1990). Under TDGSA, the weights applied to forecasts at the top of the hierarchy take into account the average of the relative importance of each value of lower hierarchical levels of the time series. The TDGSF approach, in turn, considers a weighting system that takes the relative importance of the entire series from the past at lower levels with respect to the time series at the top level. In practice, both approaches consider the overall Brazilian energy consumption profile to produce forecasts for the classes of consumption (intermediate level) and for the Brazilian Regions within classes (bottom-level). These approaches basically assume that the seasonal behavior of the time series at the intermediate and bottom-levels is induced by the most aggregated information (the national energy consumption).

4.2.3 Optimal reconciliation

The idea of optimally reconciling/combining forecasts, generating coherent estimates from a regression-based perspective, first appeared in the work of HYNDMAN et AL. (2011). Following the same matrix notation from the previous sections, optimal reconciliation approaches can be expressed according to the following regression model:

$$\hat{\mathbf{y}}_{t+h|t} = S\beta_{\mathbf{t+h|t}} + \epsilon_{\mathbf{t+h|t}} \tag{4-6}$$

where $\beta_{\mathbf{t+h|t}} = \mathbf{E}\left[\mathbf{y}^{\mathbf{b}}_{\mathbf{t+h}}|\mathcal{I}_{\mathbf{t}}\right]$, $\mathcal{I}_t = y_1, y_2, \ldots, y_t$ and $V\left(\epsilon_{t+h|t}|\mathcal{I}_t\right) = \Sigma_h$. In this case, forecasts from all hierarchical levels of the power system, i.e., from the overall consumption to regions within classes, are optimally combined into new unbiased forecasts that add up properly according to the underlying hierarchical structure.

Later, HYNDMAN et al. (2016) proposed the use of the Weighted Least Squares (WLS) estimator to obtain an estimate of the variance-covariance matrix of the h-step-ahead final (reconciled) forecast errors (Σ_h), while ignoring the elements outside the diagonal. Recent studies proposed alternative methods to obtain this matrix. WICKRAMASURIYA et al. (2019) introduced the Minimum Trace (MinT) reconciliation approach, which aims to find a matrix $\mathbf{P}$ that minimizes $tr(\mathbf{SPW_hP'S'})$ subject to $\mathbf{SPS} = \mathbf{S}$, the unbiasedness condition. WICKRAMASURIYA et al. (2020), in turn, reconsidered the least squares minimization problem with non-negativity constraints to ensure that the coherent forecasts are strictly non-negative.

In order to use $MinT$ reconciliation, it is necessary to estimate $\mathbf{W_h}$, the variance-covariance matrix of the h-step-ahead base forecast errors. Several approaches were proposed, giving birth to variants of the optimal combination approach depicted in eq. (4-6). In this work we considered two approaches as benchmarks: the simplest version, based on Ordinary Least Squares (OLS) and the shrinkage estimator (MinT-S).

The OLS approach takes $\mathbf{W_h} = \mathbf{k_h I}$, for a given constant $k_h > 0$. This is the simplest hypothesis and means assuming that the $\mathbf{P}$ matrix is independent of the data, providing an advantage in computational terms. However, by ignoring the dependency structures between residuals, relevant information is lost (HYNDMAN et al., 2011). For instance, in the Brazilian power system, forecasts for the regions within classes of consumption, which usually contain more volatility than forecasts for the intermediate and top-levels, will have the same importance, in terms of reconciliation weights, as the ones provided at the most aggregated levels, which tend to be smoother/less volatile by construction.

The shrinkage estimator (MinT-S) takes $\mathbf{W_h} = \mathbf{k_h}\lambda\hat{\mathbf{W}}_{\mathbf{1,D}} + (\mathbf{1} - \lambda)\hat{\mathbf{W}}_{\mathbf{1}}$ where $k_h > 0$, $\hat{\mathbf{W}}_1$ is the unbiased sample covariance estimator of the in-sample one-step-ahead base forecast errors and $\hat{\mathbf{W}}_{1,D} = diag(\hat{\mathbf{W}}_1)$. In this case, the estimator of the covariance matrix aims to reduce the importance of elements outside the main diagonal of $\hat{\mathbf{W}}_1$. The shrinkage parameter λ is a function of the in-sample correlations and is estimated as follows.

$$\hat{\lambda} = \frac{\sum_{i \neq j} \hat{V}(\hat{r}_{ij})}{\sum_{i \neq j} \hat{r}_{ij}^2} \tag{4-7}$$

where $\hat{r}_{ij}$ corresponds to ij-element of $\hat{\mathbf{R}}_1$, the one-step-ahead in-sample correlation matrix. It is worth noting that the shrinkage estimator, despite reducing the importance of covariance between series at different levels of the hierarchy, still takes into account some measure of relationship between the series.

Regardless of the approach selected for estimating $\mathbf{W_h}$, the optimal reconciled forecasts are given by

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{W_h^{-1}S})^{-1}\mathbf{S}'\mathbf{W_h^{-1}}\hat{\mathbf{y}}_{\mathbf{h}}. \tag{4-8}$$

This equation can be written in another fashion as follows.

$$\tilde{\mathbf{y}}_h = \mathbf{S}\tilde{\beta}_{\mathbf{h}}. \tag{4-9}$$

4.2.4 Robust Reconciliation for HTS

As observed, the optimal reconciliation approaches described in the previous section are obtained following the same guidelines as in linear regression techniques. However, the use of Least Squares may not appropriate when dealing with time series forecasts presenting outlying observations and/or influential points. An influential point is a point that has a large impact on any part of a regression analysis, such as the predicted responses, the estimated slope coefficients, or the hypothesis test results. Outliers, in turn, are data points whose responses do not follow the general trend of the rest of the data. Regardless of the case, a reconciliation method that is less affected by forecast imperfections is desirable. The work of LILA et al. (2022) can be viewed as a seminal effort in this regard. In brief, the authors considered the use of M-Estimators in the context of hierarchical forecast reconciliation and proposed two different robust-based approaches applied to unemployment data from multiple labor force surveys.

The approach can be summarized as follows: let ρ be a function having the following properties: nonnegative, i.e., $\rho(z) \geq 0$; $\rho(0) = 0$; symmetric, $\rho(z) = \rho(-z)$ and monotone in $|Z_i|$, $\rho(z_i) \geq \rho(z_{i'})$ for $|z_i| > |z_{i'}|$. Then, the robust M-estimator based on Equation (5-13) is given as:

$$\tilde{\beta}_{M,h} = \arg\min_{\tilde{\beta}_h} \sum_{i=1}^{n} \rho(\eta_i(\tilde{\beta}_h)). \tag{4-10}$$

The solution to this minimization problem requires finding $\psi(\cdot) = \rho'(\cdot)$, the influence function. In light of its desirable properties for computational convergence, LILA et al. (2022) considered the Huber influence function (HUBER, 1964), as follows:

$$\rho(z) = \begin{cases} z^2, & \text{if } |z| < c; \\ |2z|c - c^2, & \text{if } |z| \geq c \end{cases} \tag{4-11}$$

for a given constant c. Since some robust estimators are influenced by the scale of the residuals, a scale-invariant version of the M-estimator was used in LILA et al. (2022):

$$\tilde{\beta}_{\mathrm{M},h} = \arg\min_{\tilde{\beta}_h} \sum_{i=1}^{m} \rho\left(\frac{\eta_i(\tilde{\beta}_h)}{\sigma}\right), \tag{4-12}$$

There are two common ways of estimating σ. The first is based on the Mean Absolute Deviation (MAD) and is represented below:

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{median\{|\eta_i(\tilde{\beta}_h)|\}}{0.6745} \tag{4-13}$$

The second approach, also known as Huber's Proposal 2, comes from the solution of:

$$\frac{1}{m-n} \sum_{i=1}^{m} \psi^2 \left[\frac{\eta_i(\tilde{\beta}_h)}{\hat{\sigma}}\right] = E_Z[\psi^2(\eta)] \tag{4-14}$$

where $E_Z[\psi^2(\eta)]$ is the expected value of ψ^2 when η has standard normal distribution, m is the number of observations, which are the forecasts for all levels of the hierarchy and n is the number of reconciled forecasts at the most granular level of the hierarchy when convergence is met. LILA et al. (2022) denoted the above approaches by HUBER (1) and HUBER (2), respectively.

4.2.5 Resistant Reconciliation for HTS

Whilst robust reconciliation may provide accurate reconciled forecasts on several occasions, robust regression methods are only capable of dampening the influence of outliers. In other words, these approaches do not drop outliers, but instead reduce their effects during regression.

When the involved forecasts contain multiple outliers and these occur in both extremes of the distribution, robust regressions may find it hard to withstand the proportion of contamination (due to outlying observations) and still provide reliable outcomes. On such occasions, the use of estimates that are not influenced by any outliers, regardless of the nature of the data, is desired. The use of resistant statistics provides a promising solution in this regard as these metrics are, by definition, measures of the data that are not influenced by outliers. In the context of hierarchical forecast reconciliation, some resistant techniques like the Least Absolute Deviations (LAD) can be used and provide reliable outcomes. Let the residuals from the reconciliation process be defined as:

$$\mathbf{y_h} - \tilde{\mathbf{y}}_\mathbf{h} = \epsilon(\tilde{\beta}_\mathbf{h}) \tag{4-15}$$

where $\mathbf{y_h}$ are the actual (true) values of time series and $\tilde{\mathbf{y}}_h$ are the reconciled forecasts. Considering a L_1 norm, the LAD-estimator based on the residuals from the equation (4-15) is given as

$$\tilde{\beta}_{LAD,h} = \arg\min_{\tilde{\beta}_h} \sum_{i=1}^{m} |\epsilon_i(\tilde{\beta}_h)|. \tag{4-16}$$

One of the problems that we find when reconciling forecasts is that we do not observe the reconciled residuals at the estimation stage. In this case, the distance between the reconciled forecasts and the independent ones can be used:

$$\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h = \eta(\tilde{\beta}_h) \tag{4-17}$$

Hence, the LAD-estimator based on the quantities from the equation (4-17) is given as

$$\tilde{\beta}_{LAD,h} = \arg\min_{\beta_h} \sum_{i=1}^{m} |\eta_i(\tilde{\beta}_h)|. \tag{4-18}$$

The solution of this minimization problem invokes a variant of the BARRODALE & ROBERTS (1974) simplex algorithm described in KOENKER & D'OREY (1987). The promising aspects concerning LAD estimators is that they not only offer desired robust properties in linear regression models (WANG, 2013), but they also tend to present reliable results when heavy-tailed errors are present. On a short note, we clarify that, in practice, resistant-based methods are not completely immune to unusual observations. However, the cases in which such influences occur are rare in practice as they require the presence of clusters of extreme cases of outliers and a relatively small data set. Even in such cases, the influence of an unusual observation on a resistant metric, such as the median, is bounded/limited.

4.2.6
Scientific Hypothesis and Deductive Reasoning

To summarize the rationale behind the selection of the proposed, resistant-based approach for forecast reconciliation, we present in Figure 4.3 the main steps of our deductive reasoning process. Our scientific hypothesis was built after investigation of the electric energy demand time series that comprise the Brazilian power system, presented in detail in the next section. We observed that the independent forecasts of some time series contained outliers and influential points that caused disturbances in the reconciliation processes of several forecast reconciliation methods, including not only traditional benchmarks but also state-of-the-art methodologies. To circumvent this problem, and at the same time provide accurate forecast results to most of the time series involved in the Brazilian power system hierarchy, we proposed a resistant-based reconciliation approach that is theoretically constructed to not be influenced by any source of data contamination. As previously outlined, resistant estimators, such as the Least Absolute Deviation (LAD) described in Section 4.2.5, present desired robust properties in linear regression models and usually provide reliable results when heavy-tailed errors are present.

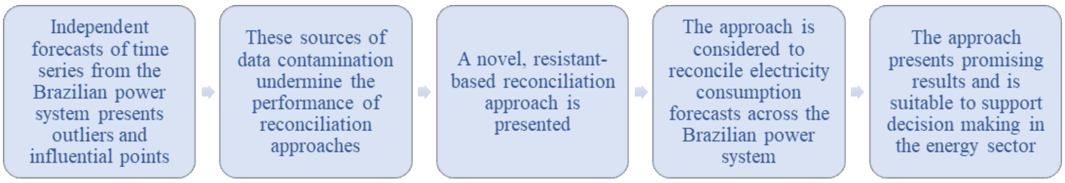

Figure 4.3: Deductive reasoning for the proposal of resistant reconciliation as the hierarchical reconciliation method for electricity demand forecasts across the Brazilian power system.

4.3 Forecasting energy consumption across the Brazilian power system

The experiments consider hierarchical monthly data of electric energy consumption across the Brazilian National Interlinked System (SIN), a set of 25 time series that follow a three-level hierarchical structure. The data are officially compiled in megawatt-hours (MWh) by the Brazilian Energy Research Company (EPE, 2022b) and spans from January 2004 to November 2021, the last official date available at the time of collection.

We consider a hierarchical setting that first divides total electricity consumption (level 0 of the hierarchy, or top-level) by Classes (also referred to as Sectors of Consumption, level 1) and then by Geographic Regions (2^{nd} and last/bottom-level). There are four classes of Consumption in the Brazilian power system: Commercial, Industrial, Residential and Others (mainly rural, public service and public lighting). Geographic Regions, in turn, are divided into five groups: North (NO), Northeast (NE), Midwest (MW), Southeast (SE) and South (SO). Figure 4.4 illustrates the hierarchical structure considered to represent the Brazilian power system whilst Figure 4.5 provides a comprehensive view of the geographic regions distributed over the Brazilian territory.

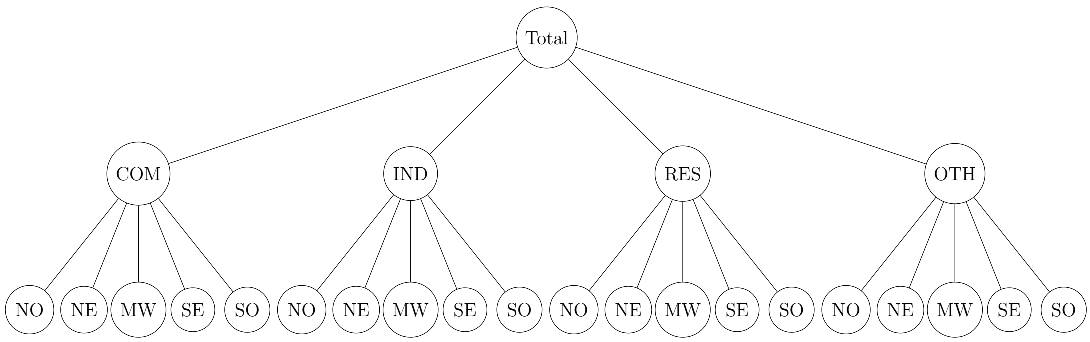

Figure 4.4: Hierarchical structure of a three-level hierarchy representing the Brazilian power system.

Figure 4.5: Brazilian geographic regions.

Figure 4.6 illustrates how total energy consumption in Brazil has varied throughout the years. Some stylized facts are readily apparent, such as a clear upward trend and a multiplicative seasonal component. Energy demand across the different classes of consumption, in turn, is illustrated in Figure 4.7. Except for the industrial demand, the behavior in every other class of consumption is similar to that observed in the most aggregated level of the hierarchy, with time series presenting an increasing trend and a well-defined seasonal component.

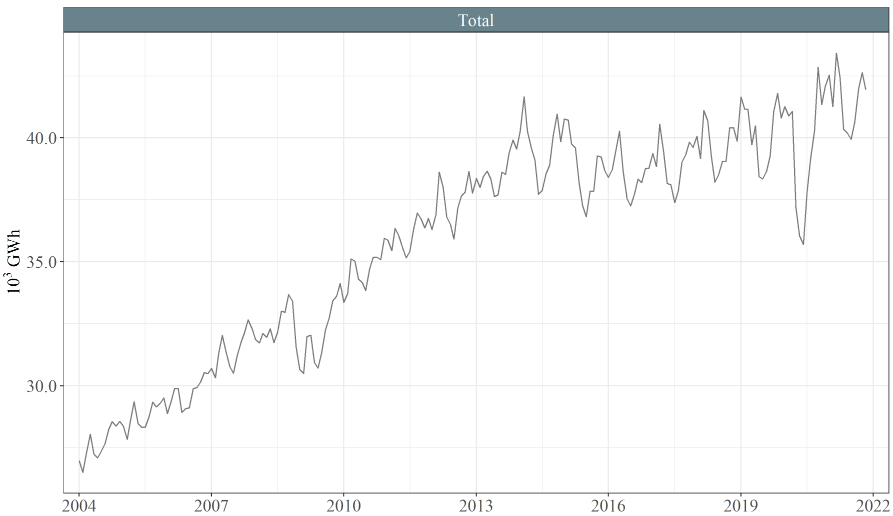

Figure 4.6: Brazil – Total electric energy consumption in the Brazilian National Interlinked System (SIN).

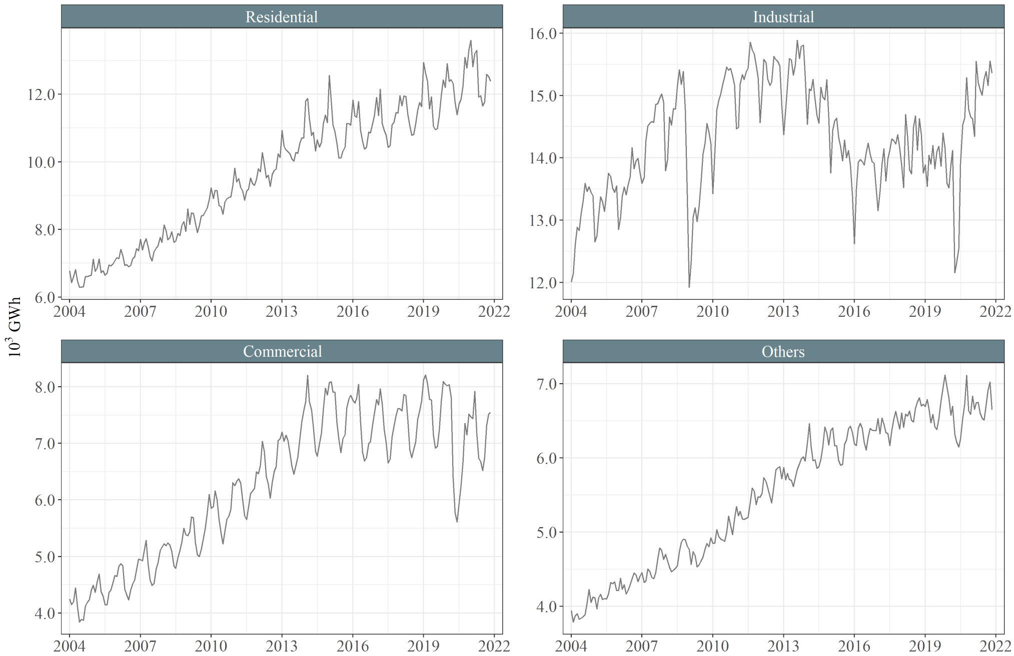

Figure 4.7: Total electric energy consumption in the SIN by classes of consumption.

Finally, Figure 4.8 illustrates the behavior of the time series at the most granular level of the hierarchy, i.e., geographic regions within classes of consumption. In most cases, the trend and seasonality components are also observed, sharing similar behaviors with the more aggregated levels of the hierarchy. One can also observe a substantial difference in the levels of the time series referring to the southeastern Region, whose historical records are considerably higher than those observed in other geographic regions.

Figure 4.8: Total electric energy consumption in the SIN, by classes of consumption and regions within classes.

4.3.1 Base forecasting methods

As previously outlined, reconciliation strategies consist of reconciling, i.e., combining, a set of base forecasts into coherent ones by finding a solution which minimizes the aggregate reconciliation error. This involves finding an orthogonal or oblique projection of the (incoherent) base forecasts onto a coherent subspace, then aggregating the information according to a hierarchical structure, such that the upper-level forecasts are obtained in an additive fashion from the most granular ones. However, the selection of the base forecasting method also plays an important role, as reconciliation methods are only able to improve the quality of the base forecasts up to a certain level.

The choice of a base forecasting strategy depends on how data behave over time, or on whether explanatory variables are useful to improve the predictive power of a particular model (OLIVEIRA et al., 2017).Given their competitive performance when forecasting energy demand and supply time series on several occasions, such as electricity demand (OLIVEIRA; OLIVEIRA, 2018), electricity supply (MEIRA et al., 2021) and natural gas consumption (MEIRA et al., 2022), and to allow straightforward comparisons with previously published papers in hierarchical forecasting reconciliation, we considered in this work two well-known forecasting approaches to generate the base forecasts: Exponential Smoothing models and Seasonal Autoregressive Integrated Moving Average (SARIMA) formulations.

To conserve space, we provide below a brief explanation on how the ETS and SARIMA family of models work. The interested reader is referred to Appendix A for detailed descriptions on these models work and how they can be properly specified in R (related packages, functions, and choice of arguments).

The ETS stands for a finite set state space based exponential smoothing formulations, which can be obtained by considering variations in the combination of the error, trend, and seasonal components of a time series . Exponential smoothing, in turn, consists of procedures that attribute exponentially decreasing weights for past data, i.e., recent observations are given relatively more weight in forecasting than older ones. Exponential smoothing formulations were first presented in the seminal works of HOLT (1957), BROWN (1959) and WINTERS (1960).

The SARIMA formulations, in turn, are an integral part of the so-called BOX & JENKINS (BOX; JENKINS, 1970b) family of models for estimating and forecasting univariate time series data. Devised as alternative approaches to exponential smoothing methods, SARIMA models are similar to the latter in the sense that they are adaptive, can model trends and seasonal patterns, and can be automated. Conversely, they are based on autocorrelations (patterns in time) rather than a structural view of level, trend and seasonality (as in ETS formulations). In practical terms, it can be argued that SARIMA formulations tend to succeed better than exponential smoothing methods for longer, more stable data sets and not as well for noisier, more volatile data (LITTERMAN et al., 1986).

4.3.2
Assessment metrics

To gauge the overall accuracy of the reconciled forecasts, we summarized the results according to a set of metrics specified in Table 4.1.

Metric	Formula	Unit of measurement
Mean Absolute Error (MAE)	$\frac{1}{h}\sum_{t=1}^{h}\lvert y_t - \hat{y}_t\rvert$	Same as the original series
Root Mean Squared Error (RMSE)	$\sqrt{\frac{\sum_{t=1}^{h}(y_t-\hat{y}_t)^2}{h}}$	Same as the original series
Mean Absolute Percentage Error (MAPE)	$\frac{100}{h}\sum_{t=1}^{h}\frac{\lvert y_t-\hat{y}_t\rvert}{\lvert y_t\rvert}$	Percentage points (%)

Table 4.1: Evaluation metrics. Notes: y_t e $\hat{y}_t$ are the real (actual) and forecasted values of the underlying series, respectively; h is the forecasting horizon (number of forecasting steps ahead).

Given that multiple reconciliation strategies are considered to generate forecasts for the whole hierarchy, relative measures from reconciled forecasts are obtained with respect to independent ones, as shown in the following equations:

$$RelMAE_{i,h} = \frac{MAE_{i,h}^{rec}}{MAE_{i,h}^{base}} \tag{4-19}$$

$$RelRMSE_{i,h} = \frac{RMSE_{i,h}^{rec}}{RMSE_{i,h}^{base}} \tag{4-20}$$

$$RelRMSE_{i,h} = \frac{MAPE_{i,h}^{rec}}{MAPE_{i,h}^{base}} \tag{4-21}$$

where the denominator corresponds to the metric associated to the independent forecasts, also called base forecasts, for the series i of the hierarchy at the forecast horizon h, and the numerator is the metric obtained for the same time series and time horizon after the reconciliation process.

The geometric mean within each level of the hierarchy (or for the entire hierarchy) provides a summarized measure of improvement in terms of MAE, RMSE and MAPE. These quantities can be obtained through the following equations:

$$AveRelMAE = \sqrt[\#L]{\prod_{i \in L} RelMAE_i} \tag{4-22}$$

$$AveRelRMSE = \sqrt[\#L]{\prod_{i \in L} RelRMSE_i} \tag{4-23}$$

$$AveRelMAPE = \sqrt[\#L]{\prod_{i \in L} RelMAPE_i} \tag{4-24}$$

where L is the corresponding level of the hierarchy.

4.3.3 The Experimental Setup

4.3.3.1 First set of experiments

Our first set of experiments explores the forecasting accuracy of the reconciliation approaches in two fixed monthly forecasting horizons. These correspond to the last three and four months of official data available in the test set, i.e., from September 2021 to November 2021 ($h = 1 - 3$) and from August 2021 to November 2021 ($h = 1 - 4$). Figure 4.9 illustrates how results based on the assessment metrics are obtained for this first exercise, using the last three and four months.

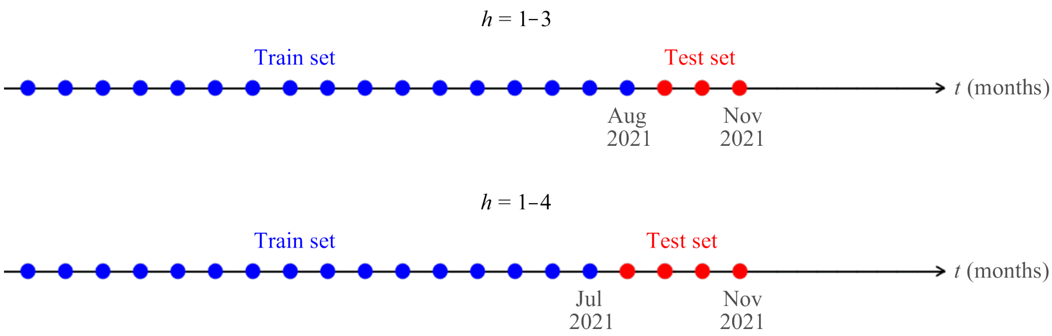

Figure 4.9: Defining the train and test sets in the first set of forecasting experiments.

Forecasting accuracy is first examined across all hierarchical levels – in this case, we consider the grand mean, i.e., the average of the geometric means obtained using Equations 4-22 to 4-24 – and then for each level of hierarchy, so as to better understand where the reconciliation processes provide the most gains/benefits to the base forecasts.

To allow for straightforward reproduction of the results, as well as to facilitate the understanding of the main steps required to generate the reconciled forecasts, we provide the database and the R code used to generate the results from the first set of experiments. The content is available in the supplementary material and can also be accessed online via GitHub. The interested reader is referred to MEIRA et al. (2023).

4.3.3.2
Second set of experiments

The second set of experiments considers rolling forecast origins throughout the last semester of observations. In brief, we consider different forecasting exercises in which the origin of the forecast horizon slides over the last six months of observations. For instance, considering a forecast horizon of three steps ahead, the experiments are conducted by first considering the end of the train set in April 2021 and the test set comprising the month of July 2021; then, the train set increases by one observation (up to May 2021) and the test set now comprises the month of August 2021. The procedure ends when the test set considers the month of November 2021, i.e., the last official data available. Then, the average of the forecast evaluation metrics obtained in each forecasting experiment is computed. In this work, we consider four different forecast horizons that vary throughout the last semester of observations: 1-step-ahead, i.e., $h = 1$ (origins varying from May 2021 to November 2021); 2-steps-ahead, i.e., $h = 2$ (origins varying from May 2021 to October 2021); 3-steps-ahead, i.e., $h = 3$ (origins varying from May 2021 to September 2021); and 4-steps-ahead, i.e., $h = 4$ (origins varying from May 2021 to August 2021).

The cross-validation experiments are designed to provide robustness to the findings obtained in the first set of experiments and is illustrated in Figure 4.10 for the four different forecast horizons considered ($h = 1, 2, 3, 4$). Conducting cross-validation for accuracy evaluation has several advantages over selecting a single period, as cross-validation exposes the reconciliation approaches to different characteristics of the data. The suitability of cross-validation for forecast accuracy assessment is discussed in further details in BERGMEIR et al. (2018).

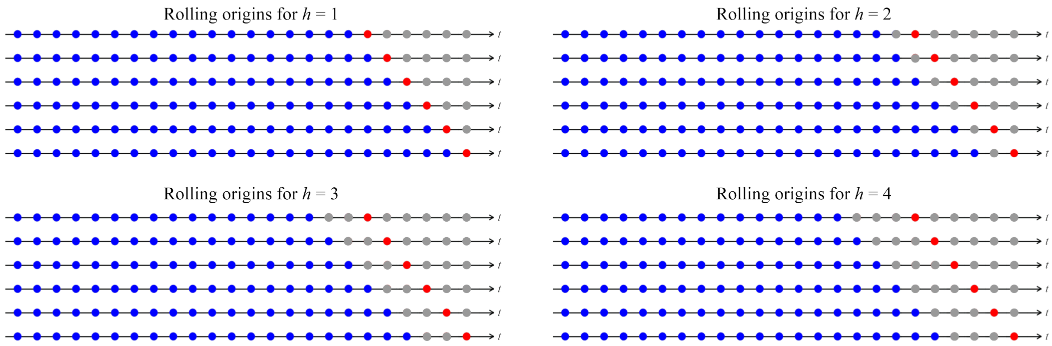

Figure 4.10: Rolling forecast origin of fixed length/horizon (fixed number of steps ahead).

4.4 Results and discussion

4.4.1 Results for the first set of experiments

4.4.1.1 Average relative metrics

Table 4.2 shows the values of the average relative metrics ($AveRelMAE$, $AveRelRMSE$ and $AveRelMAPE$), computed across all hierarchical levels, for several reconciliation methods when reconciling ETS base forecasts in two different monthly time horizons, i.e., $h = 1 - 3$ (months from September 2021 to November 2021) and $h = 1 - 4$ (months from August 2021 to November 2021). Numbers highlighted in **bold** indicate the methods that provide the best forecasting performance in each forecasting horizon, while numbers in *italics* represent the methods that rank as the second best reconciliation strategies. When the relative metrics present values larger than one, they indicate that the reconciliation strategy leads to worse (less accurate) forecasts than the original, base forecasts. The results suggest the potential of the LAD resistant regression technique for forecast reconciliation. The best outcomes in terms of all three assessment metrics are observed for the LAD estimator when forecasting up to three steps-ahead ($h = 1 - 3$), when improvements of 6.8% in both MAE and MAPE and 5.8% on RMSE are observed. The LAD also performed best when reconciling forecasts up to four steps ahead, providing improvements of 4.3% in terms of MAE and MAPE, and 4% on RMSE. It is worth noting that LAD estimators provided the best reconciliation results in all scenarios, outperforming the BU and MinT-S, which are important benchmarks from the HTS literature.

Reconciliation approach	Forecast horizon $h = 1-3$			Forecast horizon $h = 1-4$		
	AveRelMAE	*AveRelRMSE*	*AveRelMAPE*	*AveRelMAE*	*AveRelRMSE*	*AveRelMAPE*
Robust						
HUBER (1)	1.029	1.022	1.029	0.992	0.994	0.990
HUBER (2)	1.067	1.055	1.066	1.048	1.041	1.046
Resistant						
LAD	**0.932**	**0.942**	**0.932**	**0.957**	**0.960**	**0.957**
Benchmarks						
BU	*0.991*	0.990	*0.991*	1.002	1.001	1.002
OLS	1.067	1.055	1.066	1.048	1.041	1.046
MinT-S	0.992	*0.987*	0.992	*0.987*	*0.987*	*0.987*
TDGSF	3.269	3.046	3.261	3.701	3.452	3.699
TDGSA	3.395	3.159	3.385	3.866	3.628	3.860

Table 4.2: *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* across all hierarchical levels for different forecast horizons. ETS as base forecasting method.

Table 4.3 presents the results for the same evaluation metrics and forecast horizons when the forecast reconciliation strategies are applied to base SARIMA forecasts. The results are, to a greater extent, in line with those from Table 4.2. In the shorter-term, i.e., $h = 1 - 3$, LAD reconciled forecasts presented the largest gains in accuracy, varying from 3.3% to 4.5% according to the evaluation metric. When analysing the outcomes for $h = 1-4$, the LAD estimator ranked second-best, following the MinT-S reconciliation.

Reconciliation approach	Forecast horizon $h = 1-3$			Forecast horizon $h = 1-4$		
	AveRelMAE	*AveRelRMSE*	*AveRelMAPE*	*AveRelMAE*	*AveRelRMSE*	*AveRelMAPE*
Robust						
HUBER (1)	1.030	1.035	1.032	1.020	1.020	1.020
HUBER (2)	1.052	1.060	1.057	1.038	1.052	1.048
Resistant						
LAD	**0.967**	**0.955**	**0.955**	*0.999*	*0.998*	*0.997*
Benchmarks						
BU	1.014	1.001	1.001	1.021	1.010	1.011
OLS	1.052	1.060	1.057	1.038	1.052	1.048
MinT-S	0.987	0.974	0.974	**0.986**	**0.984**	**0.984**
TDGSF	2.924	3.040	3.026	2.773	2.879	2.874
TDGSA	3.148	3.257	3.242	2.892	3.019	3.010

Table 4.3: *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* across all hierarchical levels for different forecast horizons. ARIMA as base forecasting method.

To further understand the gains originating from the use of LAD reconciliation, Table 4.4 illustrates the values of the forecast evaluation metrics obtained for each level of the hierarchy when the reconciliation strategies were applied to ETS base forecasts. As can be noted, the best improvements come from the bottom of the hierarchy, which contains the largest number of forecasts to be reconciled. The LAD estimator performed best according to every evaluation metric and for both forecast horizons at the bottom level, a noteworthy result. In the shorter-term, i.e., when reconciling forecasts up to three steps ahead ($h = 1 - 3$), the average improvement provided by the LAD estimator in terms of MAE and MAPE were 8%, and 6.8% in terms of RMSE. The LAD estimator also provided the second-best performance at the top (most aggregate) level of the hierarchy. In this case, the BU approach provided the most competitive results under ETS as base forecasting method. However, it is important to note that the number of forecasts benefited from LAD reconciliation were higher than any other proposed methodology.

The results presented in Table 4.5, in turn, show the improvements brought by reconciliation strategies at every level of the hierarchy when these strategies are applied to SARIMA base forecasts. In this case, the LAD estimator provided the most competitive results at the bottom and intermediate levels when reconciling SARIMA forecasts up to three months ahead, followed by the MinT-S reconciliation.

Reconciliation approach	Forecast horizon $h = 1 - 3$			Forecast horizon $h = 1 - 4$		
	AveRelMAE	*AveRelRMSE*	*AveRelMAPE*	*AveRelMAE*	*AveRelRMSE*	*AveRelMAPE*
Total						
Robust						
HUBER (1)	0.875	0.908	0.875	0.934	0.851	0.936
HUBER (2)	0.950	0.967	0.950	0.932	0.951	0.933
Resistant						
LAD	*0.859*	0.883	*0.859*	0.935	0.845	0.937
Benchmarks						
BU	**0.844**	**0.831**	**0.844**	0.958	0.830	0.961
OLS	0.950	0.967	0.950	0.932	0.951	0.933
MinT-S	0.860	*0.850*	0.860	0.972	0.820	0.976
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
Classes of Consumption						
Robust						
HUBER (1)	1.029	1.019	1.028	1.007	1.009	1.007
HUBER (2)	1.062	1.044	1.060	1.018	1.006	1.016
Resistant						
LAD	1.014	1.008	1.013	1.009	1.013	1.008
Benchmarks						
BU	**0.986**	**0.981**	**0.987**	1.025	1.055	1.025
OLS	1.062	1.044	1.060	1.018	1.006	1.016
MinT-S	*0.989*	*0.985*	*0.989*	1.009	1.028	1.009
TDGSF	3.411	3.140	3.392	3.612	3.274	3.624
TDGSA	3.972	3.635	3.950	4.060	3.604	4.063
Regions						
Robust						
HUBER (1)	1.037	1.029	1.037	0.992	0.999	0.990
HUBER (2)	1.075	1.062	1.074	1.060	1.053	1.058
Resistant						
LAD	**0.920**	**0.932**	**0.920**	**0.948**	**0.956**	**0.949**
Benchmarks						
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	1.075	1.062	1.074	1.060	1.053	1.058
MinT-S	1.000	*0.995*	1.000	*0.984*	*0.989*	*0.984*
TDGSF	3.439	3.201	3.433	3.971	3.711	3.965
TDGSA	3.497	3.253	3.488	4.097	3.875	4,087

Table 4.4: *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* for each hierarchical level and different forecast horizons. ETS as base forecasting method.

Reconciliation approach	Forecast horizon $h = 1 - 3$			Forecast horizon $h = 1 - 4$		
	AveRelMAE	*AveRelRMSE*	*AveRelMAPE*	*AveRelMAE*	*AveRelRMSE*	*AveRelMAPE*
Total						
Robust						
HUBER (1)	1.098	1.055	1.098	1.180	1.174	1.179
HUBER (2)	1.061	1.027	1.060	1.076	1.059	1.075
Resistant						
LAD	1.104	1.050	1.103	1.185	1.175	1.184
Benchmarks						
BU	1.259	1.191	1.258	1.321	1.325	1.321
OLS	1.061	1.027	1.060	1.076	1.059	1.075
MinT-S	1.258	1.180	1.256	1.302	1.301	1.301
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000
Classes of Consumption						
Robust						
HUBER (1)	0.918	0.920	0.917	*0.989*	**0.980**	*0.989*
HUBER (2)	0.977	0.948	0.975	1.001	0.985	0.999
Resistant						
LAD	**0.905**	**0.910**	*0.905*	0.991	*0.981*	0.991
Benchmarks						
BU	0.947	1.046	0.952	0.994	1.059	0.998
OLS	0.977	0.948	0.975	1.001	0.985	0.999
MinT-S	0.846	0.958	**0.850**	**0.935**	0.990	**0.938**
TDGSF	2.707	2.621	2.686	2.489	2.553	2.514
TDGSA	3.414	3.227	3.388	2.859	2.746	2.873
Regions						
Robust						
HUBER (1)	1.057	1.053	1.053	1.018	1.022	1.014
HUBER (2)	1.077	1.075	1.074	1.062	1.048	1.057
Resistant						
LAD	**0.958**	**0.975**	**0.958**	*0.990*	*0.995*	*0.990*
Benchmarks						
BU	1.000	1.000	1.000	1.000	1.000	1.000
OLS	1.077	1.075	1.074	1.062	1.048	1.057
MinT-S	*0.989*	*0.984*	*0.988*	**0.980**	**0.972**	**0.980**
TDGSF	3.289	3.154	3.276	3.125	2.967	3.112
TDGSA	3.423	3.317	3.408	3.225	3.081	3.210

Table 4.5: *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* for each hierarchical level and different forecast horizons. SARIMA as base forecasting method.

4.4.1.2
Error comparison analysis

The results depicted in the previous section concerned the values of the average relative metrics observed for each reconciliation method in the first set of experiments. As outlined, these metrics provide a straightforward illustration of the benefits (or hindrances) brought forth by each reconciliation method when comparing the accuracy of their reconciled forecasts with the accuracy of the base forecast method. Hence, the lower the values of the relative metrics, the more accurate the reconciliation strategy is suggested to be. Numbers higher than one indicate that the reconciliation approach worsens the accuracy of the base forecast after reconciliation.

Despite providing a quick and effective way of comparing the performance between the selected reconciliation approaches, average relative metrics do not provide an estimate of how much energy can 'saved' by opting for a more accurate forecast reconciliation approach in lieu of its competitors. To address this need, we computed the absolute values of the Root Mean Squared Errors (RMSEs) of each forecast reconciliation approach separately, for each series in the involved hierarchy. Then, we put forth a visual comparison of the RMSEs for each series at the bottom (most disaggregated) level of the hierarchy, thus providing a comparative overview on the accuracy of each method (benchmarks, state-of-the-art and proposed approach) independently.

The RMSEs for the first experiment conducted, i.e., reconciling ETS base forecasts generated for the period between September 2021 to November 2021, are depicted in Figure 4.11. For each time series at the bottom hierarchical level, a boxplot containing the RMSEs of the seven competing methods (all reconciliation methods considered except the LAD proposed approach) is presented, as well as the average RMSE across all these seven methods (blue dots). Then, the RMSE of the reconciled forecasts originated from the proposed LAD reconciliation approach is presented in red.

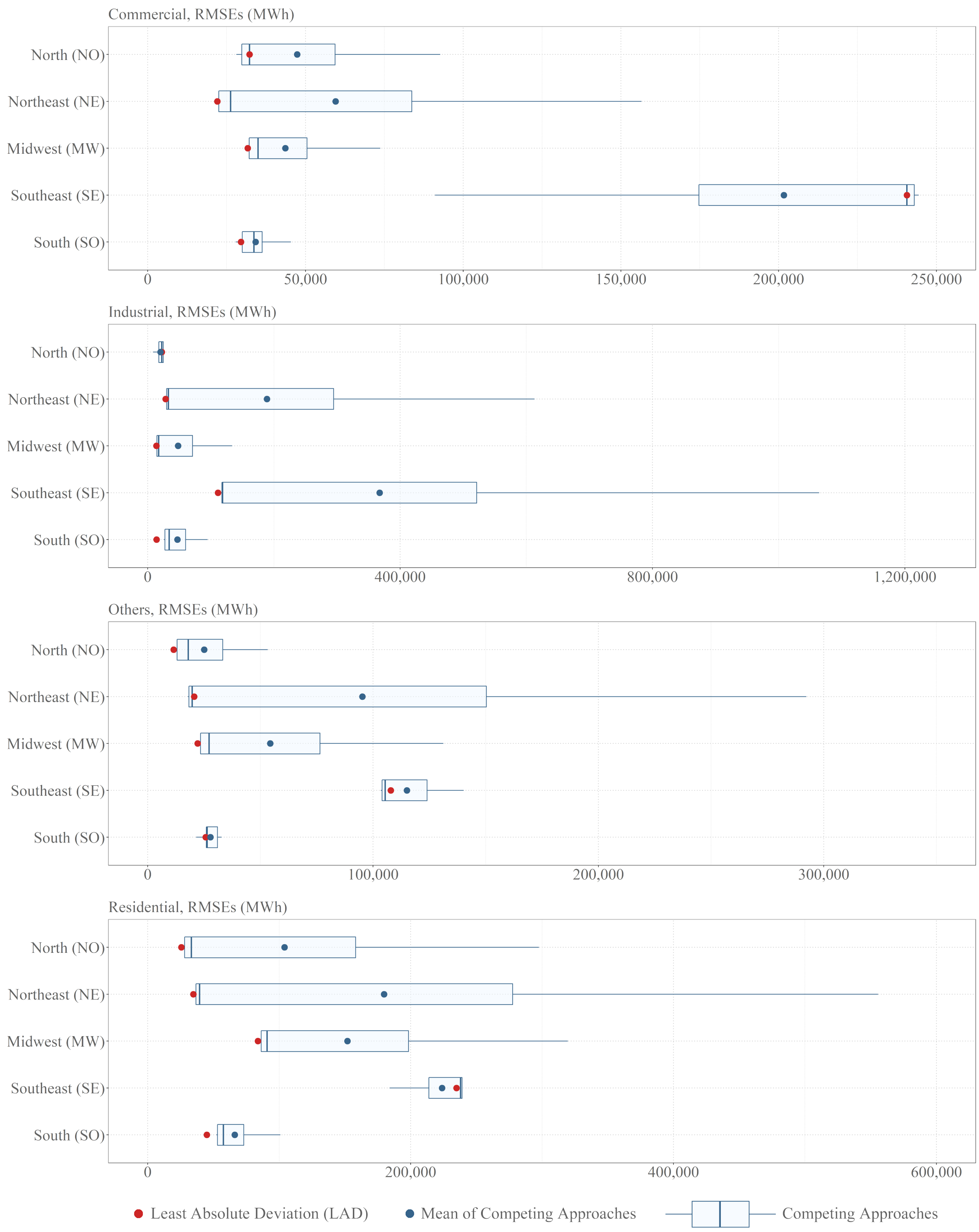

Figure 4.11: Root Mean Squared Errors (RMSEs) of the reconciliation strategies for the time series at the bottom level of the hierarchy (case 1, ETS as base forecasting method and forecasts up to three months ahead).

The results from Figure 4.11 demonstrate the benefits brought forth by the proposed approach in terms of providing more accurate forecasts than its competitors for most of the cases considered. Overall, the proposed LAD reconciliation approach provided lower RMSEs than the average RMSE of the competing methods in 17 of the 20 time series considered at the bottom level of the hierarchy. When the LAD was not the best method, it usually ranked among the best, outperforming the median of the competing approaches in 18 cases. In terms of absolute gains, the accumulated energy that could be 'saved' across all regions and classes of Brazilian National Interlinked System (SIN) by opting for the proposed LAD approach for forecast reconciliation revolves around 940,000 MWh. This amount is the sum of the differences between the average RMSEs of the competing reconciliation methods and the RMSEs from the proposed LAD approach across all the 20 time series at the bottom level of the hierarchy.

In terms of the suitability of the proposed (LAD) reconciliation approach to different demand profiles, we note that the method provided very competing (if not the best) results in most cases. The only two instances in which the proposed approach performed worse than the average of the competing methods were in the southeastern region for the commercial and residential classes. It should be noted, however, that in these two cases the LAD performed better than the median of the competing methods. This suggests that the proposed LAD approach was still competitive in these cases, but there was one specific reconciliation method that provided considerably lower RMSEs than all other methods. The best methods in these two cases were Top-Down approaches: Top-Down Gross-Sohl A for the commercial class in the southeast and Top-Down Gross-Sohl F for the residential class in the southeast. These methods have been proved in the forecasting literature to deliver biased forecasts across the hierarchy (HYNDMAN et al., 2011; WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2019). Therefore, Top-Down methods may sometimes provide accurate results for specific time series across the hierarchy, but this usually occurs by chance, and is not usually followed by other accurate forecast values in adjacent time series. Overall, our results indicate that the LAD reconciliation stands out as a balanced, accurate and robust choice for hierarchical forecasting reconciliation in every class and region of the Brazilian power system.

It is also interesting to note that the amount of energy that can be saved in the system increased with the number of necessary forecasting steps ahead. We conducted the same RMSE comparison analysis for the second case of the first set of experiments, i.e., reconciling ETS base forecasts generated for the horizon comprising the months from August 2021 to November 2021 (4 steps ahead). The overall gains, once again in terms of the differences between the RMSEs of the average of the competing reconciliation methods and the RMSEs from the proposed LAD approach across all 20 bottom level time series, were more than 976,000 MWh. In this case, the LAD reconciliation approach provided lower RMSEs than the average RMSEs of the competing methods in 18 of the 20 time series considered.

4.4.2 Results for the second set of experiments

To investigate the robustness of each reconciliation strategy when different forecast horizons are considered, we also conducted several cross-validation experiments using rolling forecast origins over the last six months of observations. Four different forecast horizons were considered: $h = 1, 2, 3, 4$. The results for the mean/average of the average relative metrics (*AveRelMAE*, *AveRelRMSE* and *AveRelMAPE*), computed across all hierarchical levels, when reconciling ETS base forecasts at different forecast origins, are presented in Table 4.6. They show consistent improvements on accuracy measures for LAD reconciliation for the forecast windows of lengths $h = 1$, $h = 3$ and $h = 4$, suggesting a consistent performance of resistant reconciliation over different forecast origins. The HUBER (1) robust reconciliation approach proposed in LILA et al. (2022) also presented competitive results, particularly in very short forecast horizons (one or two months ahead).

Table 4.7 depicts the results for the mean/average of the average relative metrics, computed across all hierarchical levels, when reconciling SARIMA base forecasts, at different forecast origins. In this case, the best results were observed for the HUBER (1), LAD and MinT-S reconciliation techniques, which ranked best or second-best in the majority of the cases considered.

Reconciliation	*AveRelMAE*				*AveRelRMSE*				*AveRelMAPE*			
Approach	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
Robust												
HUBER (1)	**0.976**	1.007	1.033	1.071	**0.987**	1.008	1.018	1.056	**0.976**	1.006	1.032	1.069
HUBER (2)	0.994	1.015	1.083	1.171	1.012	1.025	1.065	1.14	0.995	1.014	1.081	1.168
Resistant												
LAD	*0.993*	1.009	**0.969**	**0.950**	0.998	1.018	**0.979**	**0.949**	*0.992*	1.009	**0.969**	**0.950**
Benchmarks												
BU	0.999	**0.996**	0.995	0.98	0.995	*0.997*	0.992	0.974	0.999	*0.996*	0.995	0.981
OLS	0.994	1.015	1.083	1.171	1.012	1.025	1.065	1.14	0.995	1.014	1.081	1.168
MinT-S	0.994	*0.989*	*0.985*	*0.975*	*0.991*	**0.992**	*0.982*	*0.973*	0.994	**0.989**	*0.985*	*0.974*
TDGSF	4.334	3.361	3.442	3.291	3.898	3.246	3.293	2.998	4.323	3.349	3.407	3.262
TDGSA	4.155	3.226	3.292	3.214	3.741	3.108	3.128	2.902	4.149	3.219	3.262	3.189

Table 4.6: *AveRelMAE*, *AveRelRMSE* and *AveRelMAPE* metrics computed across all hierarchical levels when rolling forecast origins are considered via cross-validation. ETS as base forecasting method.

Reconciliation	*AveRelMAE*				*AveRelRMSE*				*AveRelMAPE*			
Approach	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
Robust												
HUBER (1)	1.004	1.021	**0.939**	**0.949**	1.005	1.009	**0.948**	**0.951**	1.002	1.019	**0.936**	**0.946**
HUBER (2)	1.057	1.05	*0.977*	0.993	1.051	1.036	*0.989*	*0.982*	1.056	1.049	*0.974*	*0.990*
Resistant												
LAD	1.013	**0.981**	0.998	*0.983*	1.017	**0.996**	1.018	1.003	1.011	**0.981**	0.998	0.983
Benchmarks												
BU	*0.990*	1.006	1.042	1.027	1.009	1.014	1.056	1.045	*0.991*	1.007	1.043	1.028
OLS	1.057	1.05	*0.977*	0.993	1.051	1.036	*0.989*	*0.982*	1.056	1.049	*0.974*	*0.990*
MinT-S	**0.982**	*0.991*	1.005	0.994	**0.992**	*0.997*	1.015	1.000	**0.982**	*0.992*	1.006	0.995
TDGSF	3.904	3.065	2.758	2.398	3.499	2.872	2.649	2.299	3.885	3.046	2.736	2.381
TDGSA	3.742	2.938	2.63	2.258	3.36	2.777	2.546	2.154	3.726	2.925	2.611	2.242

Table 4.7: Average of the *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* metrics computed across all hierarchical levels when rolling forecast origins are considered via cross-validation. SARIMA as base forecasting method.

As also conducted in the first set of experiments, we explored the contribution of each reconciliation technique at each level of the hierarchy in the rolling forecast origin experiments. The results for the case of ETS base forecasts are presented in Table 4.8. In this case, the LAD reconciliation provided improvements in terms of all assessment metrics and for all levels of the hierarchy. The most significant contribution occured at the bottom level. The method not only improved base forecasts, by overcoming the BU strategy, but also overcame the MinT-S according to the number of times the LAD strategy ranked best.

Finally, the results presented in Table 4.9 illustrates the average gains provided by reconciliation strategies applied to SARIMA base forecasts at different levels of the hierarchy and when considering the cross-validation of different forecast origins. The most relevant gains for the LAD estimator in this case occur at the intermediate level, in terms of RMSE. However, it should also be highlighted that the LAD strategy also ranked second-best in several occasions, suggesting a consistent performance in all forecast horizons considered when compared to other competing strategies (benchmarks and state-of-the-art reconciliation techniques).

Reconciliation Approach	*AveRelMAE* $h=1$	$h=2$	$h=3$	$h=4$	*AveRelRMSE* $h=1$	$h=2$	$h=3$	$h=4$	*AveRelMAPE* $h=1$	$h=2$	$h=3$	$h=4$
Total												
Robust												
HUBER (1)	0.947	0.911	**0.905**	0.805	0.939	0.899	0.84	0.736	0.944	0.911	**0.910**	0.805
HUBER (2)	0.996	0.944	0.928	0.88	0.984	0.961	0.942	0.911	0.995	0.943	*0.929*	0.879
Resistant												
LAD	*0.943*	*0.904*	*0.927*	0.788	*0.934*	0.882	0.841	0.714	*0.940*	**0.905**	0.933	0.787
Benchmarks												
BU	**0.933**	**0.903**	0.979	**0.718**	**0.928**	**0.849**	*0.825*	*0.624*	**0.930**	*0.906*	0.987	**0.719**
OLS	0.996	0.944	0.928	0.88	0.984	0.961	0.942	0.911	0.995	0.943	*0.929*	0.879
MinT-S	0.969	0.921	0.933	*0.724*	0.958	*0.866*	**0.786**	**0.623**	0.966	0.924	0.941	*0.725*
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Classes of Consumption												
Robust												
HUBER (1)	*0.972*	1.024	1.004	0.971	0.984	1.036	*0.984*	0.969	*0.972*	1.024	1.003	0.970
HUBER (2)	**0.970**	1.042	1.070	0.993	**0.979**	1.052	0.997	1.001	**0.971**	1.041	1.068	0.990
Resistant												
LAD	0.976	1.017	0.992	**0.954**	*0.983*	1.029	0.987	*0.960*	0.976	1.017	0.992	**0.954**
Benchmarks												
BU	1.011	*0.999*	*0.973*	*0.959*	0.990	1.021	0.997	**0.952**	1.012	1.000	*0.974*	*0.961*
OLS	**0.970**	1.042	1.070	0.993	**0.979**	1.052	0.997	1.001	**0.971**	1.041	1.068	0.990
MinT-S	0.997	**0.988**	**0.962**	0.982	0.985	1.011	**0.977**	0.963	0.997	**0.988**	**0.963**	0.983
TDGSF	4.923	3.447	3.593	3.705	3.862	3.251	3.159	3.154	4.960	3.457	3.566	3.672
TDGSA	4.505	3.119	3.192	3.157	3.647	3.038	2.865	2.703	4.556	3.139	3.177	3.127
Regions												
Robust												
HUBER (1)	**0.978**	1.008	1.045	1.108	**0.990**	1.008	1.035	1.094	**0.978**	1.008	1.044	1.105
HUBER (2)	0.998	1.014	1.094	1.227	1.020	1.023	1.086	1.183	1.000	1.013	1.092	1.225
Resistant												
LAD	0.999	1.012	**0.967**	**0.958**	1.005	1.023	**0.985**	**0.961**	0.998	1.013	**0.966**	**0.958**
Benchmarks												
BU	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
OLS	0.998	1.014	1.094	1.227	1.020	1.023	1.086	1.183	1.000	1.013	1.092	1.225
MinT-S	*0.995*	**0.993**	*0.992*	*0.988*	*0.994*	**0.995**	*0.994*	*0.998*	*0.995*	**0.992**	*0.992*	*0.987*
TDGSF	4.547	3.552	3.630	3.412	4.180	3.442	3.524	3.135	4.526	3.535	3.589	3.379
TDGSA	4.390	3.444	3.516	3.420	4.016	3.304	3.370	3.105	4.372	3.430	3.479	3.392

Table 4.8: *AveRelMAE*, *AveRelRMSE* and *AveRelMAPE* across all hierarchical levels for different forecast horizons via Cross-Validation in a rolling forecasting origin. ETS as base forecasting method.

Reconciliation	*AveRelMAE*				*AveRelRMSE*				*AveRelMAPE*			
Approach	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
Total												
Robust												
HUBER (1)	0.981	1.179	1.540	1.046	0.978	1.050	1.459	1.075	0.979	1.187	1.547	1.046
HUBER (2)	0.998	1.076	1.198	1.025	0.988	1.017	1.174	1.030	0.998	1.078	1.198	1.024
Resistant												
LAD	0.978	1.185	1.571	1.021	0.973	1.049	1.480	1.054	0.975	1.193	1.577	1.022
Benchmarks												
BU	**0.970**	1.267	1.860	1.123	**0.960**	1.110	1.865	1.282	**0.970**	1.275	1.868	1.121
OLS	0.998	1.076	1.198	1.025	0.988	1.017	1.174	1.030	0.998	1.078	1.198	1.024
MinT-S	*0.973*	1.274	1.757	1.112	*0.964*	1.114	1.805	1.225	*0.972*	1.282	1.765	1.111
TDGSF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TDGSA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Classes of Consumption												
Robust												
HUBER (1)	0.961	0.978	**0.984**	1.036	0.984	*0.988*	*0.992*	1.014	0.959	0.977	**0.982**	1.035
HUBER (2)	1.012	0.997	1.042	1.048	0.992	1.002	1.011	1.030	1.012	0.997	1.041	1.045
Resistant												
LAD	*0.955*	*0.966*	**0.984**	1.009	**0.978**	**0.977**	**0.984**	**0.987**	*0.953*	*0.965*	*0.983*	1.008
Benchmarks												
BU	**0.948**	0.979	1.109	1.148	1.070	1.062	1.207	1.237	**0.949**	0.982	1.112	1.152
OLS	1.012	0.997	1.042	1.048	0.992	1.002	1.011	1.030	1.012	0.997	1.041	1.045
MinT-S	0.958	**0.948**	1.019	1.055	1.028	1.016	1.113	1.133	0.958	**0.951**	1.022	1.059
TDGSF	3.944	2.873	2.685	2.144	3.566	2.722	2.555	2.068	3.969	2.889	2.678	2.128
TDGSA	3.694	2.587	2.351	1.838	3.377	2.576	2.391	1.716	3.723	2.612	2.359	1.822
Regions												
Robust												
HUBER (1)	1.014	1.023	**0.908**	**0.928**	1.011	1.011	**0.920**	**0.933**	1.012	1.020	**0.904**	**0.924**
HUBER (2)	1.069	1.059	0.954	0.981	1.066	1.043	0.977	0.970	1.068	1.058	*0.951*	0.977
Resistant												
LAD	1.027	**0.975**	0.978	*0.976*	1.027	*0.998*	1.005	1.004	1.025	**0.974**	0.978	*0.976*
Benchmarks												
BU	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
OLS	1.069	1.059	*0.954*	0.981	1.066	1.043	0.977	0.97	1.068	1.058	0.951	0.977
MinT-S	**0.988**	*0.988*	0.975	0.977	**0.986**	**0.988**	*0.969*	*0.966*	**0.988**	0.988	0.975	0.977
TDGSF	4.171	3.283	2.918	2.562	3.711	3.060	2.801	2.448	4.140	3.255	2.890	2.543
TDGSA	4.008	3.181	2.823	2.450	3.567	2.967	2.702	2.343	3.979	3.157	2.796	2.433

Table 4.9: *AveRelMAE*,*AveRelRMSE* and *AveRelMAPE* across all hierarchical levels for different forecast horizons via Cross-Validation in a rolling forecasting origin. SARIMA as base forecasting method.

4.4.3 Discussions

Forecasts of electricity consumption at distinct levels of a power system have become increasingly important to several stakeholders, ranging from power system operators, transmission companies in self-dispatching markets and investors operating in exchanges and over-the-counter markets using energy derivatives. This work proposed a novel reconciliation technique aimed at improving the accuracy of forecasts at all hierarchical levels within a power system. The methodology is designed to be resistant to outliers occurring in both extremes of the distribution from any time series involved in the hierarchy. The experiments conducted demonstrate the reliability and robustness of the proposed reconciliation approach, specifically when compared to benchmarks and state-of-the-art techniques. In short, for forecasts lead times of three or four (months) steps ahead, the proposed LAD reconciliation approach provided lower Root Mean Squared Errors (RMSEs) than the average RMSE of the competing methods in at least (for the three months ahead forecast experiments) 17 of the 20 time series considered at the bottom level of the hierarchy. Accurate forecasts at such level are paramount to achieving several Brazilian electricity sector goals, such as securing electricity supply, affordability of tariffs and universalization of access to public electricity services. When the LAD was not the best method, it usually ranked among the best, outperforming the median of the competing approaches in 18 cases, regardless of whether the forecast horizon comprised three or four months ahead. In terms of absolute gains, the accumulated energy that could be 'saved' across all regions and classes of Brazilian National Interlinked System (SIN) by opting for the proposed LAD approach for forecast reconciliation revolved around 940,000 MWh for three steps ahead forecasts and 976,000 MWh for four steps ahead. These amounts are the sum of the differences between the average RMSEs of the competing reconciliation methods and the RMSEs from the proposed LAD approach across all 20 time series at the bottom level of the hierarchy.

In terms of limitations, the only two cases in which the LAD approach performed worse than the average of the competing methods were in the southeastern region for the commercial and residential classes. It should be noted, however, that in these two cases the LAD performed better than the median of the competing methods. This suggests that the proposed LAD approach was still competitive in these cases, but there was one specific reconciliation method that provided considerably lower RMSEs than all other methods. The best methods in these two cases were Top-Down approaches. These methods have been proved in the forecasting literature to deliver biased forecasts across the hierarchy (HYNDMAN et al., 2011; WICKRAMASURIYA et al., 2019). Therefore, Top-Down methods may sometimes provide accurate results for specific time series across the hierarchy, but this usually occurs by chance, and is not usually followed by other accurate forecast values in adjacent time series. Overall, we can infer from the results that the LAD reconciliation stands out as a balanced, accurate and robust choice for hierarchical forecasting reconciliation in every class and region of the Brazilian power system. We also emphasize that it is possible to apply the proposed LAD approach to any set of time series organized in a hierarchical fashion.

The performance gains are remarkable as accurate electricity demand forecasts at distinct levels of a power system are decisive for assertive profit/cost management and investment decisions, as well as for the definition of sectoral policies in a local or national scale. The issue is even more crucial in the short/mid-run, where flexibilities such as storage facilities construction and diversification of energy sources are limited.

On a final note, we highlight that forecasting independently each time series may make more sense if the interest lies in a single or in a limited set of time series belonging to the same hierarchical level. However, when the interest lies in the comprehensive set of time series organized according to a defined hierarchical structure, as in the case of power systems, restricting the attention to a single forecasting approach for each time series and not taking into consideration the cross-sectional associations between these series will almost surely result in forecasts that do not add up across the hierarchy, losing interpretation of the overall results.

4.5 Conclusions, implications, and future works

Drawing from the fields of statistics, optimization and time series, this paper introduced a resistant-based reconciliation approach to improve the accuracy of the forecasts from multiple time series organized according to a defined hierarchical structure. The new approach was applied to the problem of forecasting electric energy demand at multiple levels of the Brazilian National Interlinked System (SIN), the largest and one of the most complex power systems in Latin America.

The results from multiple sets of experiments indicate that it is possible to improve the accuracy of electricity consumption forecasts by considering the cross-sectional associations between the time series comprised by the hierarchical structure of the power system. Furthermore, the proposed resistant reconciliation technique consistently provided more accurate forecasts than competing reconciliation methods that represent the state-of-the-art in terms of hierarchical forecasting reconciliation. The average gains brought forth by the proposed methodology, in terms of accumulated energy that can be ‘saved’ across all regions and classes of the Brazilian power system given more accurate forecasts, were greater than 900,000 MWh for forecast horizons of three and four months ahead. The results were particularly promising for forecasts of time series situated at the intermediate and bottom levels of the power system, i.e., the time series representing classes of consumption and geographic regions within the classes. Accurate forecasts at such levels are paramount to achieving several Brazilian electricity sector goals, such as securing electricity supply, affordability of tariffs and universalization of access to public electricity services.

To recapitulate our scientific hypothesis, presented in the deductive reasoning section, the results from the multiple experiments conducted provided consistent numerical evidence that the proposed reconciliation approach was able to circumvent any source of data contamination in the independent (base) forecasts. The approach was able to proper reconcile the base forecasts and generate coherent and accurate final forecasts for all the 25 time series that represent the electric energy consumption across all divisions of the Brazilian power system.

As policy implications, we argue that the proposed methodology can be considered in future updates of the forecasting tools used or recommended by official institutions, such as the Brazilian Electricity Regulatory Agency (ANEEL), the Brazilian Energy Research Company (EPE) and the Electric System National Operator (ONS). The proposed approach could also be particularly useful in the supply side of electricity in Brazil. For instance, resistant reconciliation techniques could be applied to increase the accuracy of stream flow forecasts, given that several rivers that are suitable for hydropower generation in Brazil form hierarchical structures, which are usually addressed as energy equivalent reservoirs.

Future research agenda includes considering an extension to non-negative reconciled forecasts. When performing reconciliation, we address coherency, i.e, ensuring the additive properties of the time series across all hierarchical levels, as the central issue. However, when an optimal solution is found, there is no way to ensure that the reconciled forecasts are always non-negative, an important, and often overlooked issue when forecasting electricity demand. The use of alternative forecasting techniques to generate the base forecasts is also a natural extension of this work. To allow straightforward comparison with recent papers in the field of hierarchical forecasting, we restricted our attention to two widely used family of forecasting models: exponential smoothing and SARIMA formulations. Future studies might benefit, for instance, from more sophisticated approaches, such as Artificial Neural Networks (ANNs), Support Vector Regressions (SVRs), Recurrent and Vector Singular Spectrum Analysis (RSSA/VSSA), among others, to generate the base forecasts. Finally, one may also consider the use of alternative hierarchies to represent the Brazilian power system, such as the one which uses the Subsystems from the National

Interlinked System (SIN) instead of geographic regions.

5
Third contribution: An extension to Minimal Trace Reconciliation using robust covariance estimators

5.1
Introduction

Hierarchical time series forecasting is the process of generating coherent forecasts, by reconciling incoherent ones, forecasted individually, but incorporating the relationships within the hierarchy of the original data (HYNDMAN et al., 2011). The goal of forecast reconciliation is to produce a more accurate and coherent aggregation by leveraging the information from the different sources and levels of aggregation in the hierarchy. In essence, it involves adjusting the forecasts to ensure that they add up to the forecast at the highest level of aggregation, while also minimizing any inconsistencies or errors between the forecasts at the lower levels. Reconciliation methods have been shown to improve forecast accuracy (HYNDMAN et al., 2011; HYNDMAN et al., 2016; WICKRAMASURIYA et al., 2019), specially the regression based approaches, which apply linear transformations to forecasts for all levels of the hierarchy, generating coherent ones at the lower level.

From a theoretical point of view, the recent reconciliation techniques were proposed to guarantee that reconciled forecasts are at least as good the base forecasts. An enhanced version can be constructed when a covariance structure is well estimated and can be incorporated into the process. In this case, the reconciled forecasts will also have minimum variance amongst all possible combinations of forecasts, as in the case of the Minimum Trace (MinT) reconciliation (WICKRAMASURIYA et al., 2019). In MinT reconciliation, the individual forecasts are adjusted using a linear combination of their errors, which are weighted according to their covariance. The weights are chosen to minimize the trace of the covariance matrix of the forecast errors, subject to constraints that ensure that the reconciled forecasts are consistent with the forecasts at the upper aggregation levels.

In practice, covariance structures are estimated from the *in-sample* one step ahead residuals. These covariance estimators are very sensitive to outlying residuals, which can produce undesirable effects on the weights of the linear transformations.

The idea of estimating robust covariance structures is the main focus of this Chapter. Instead of using the original covariance estimator, we propose two robust covariance estimators. In this case, we use the Minimum Volume Ellipsoid (MVE) and the Minimum Covariance Determinant (MCD) methods (ROUSSEEUW, 1985), which are a highly robust estimators of multivariate location and scatter. While the MVE seeks to find the g-subset of the data points that produces an ellipsoid of minimum volume, the MCD searches for the sample covariance matrix which has the lowest possible determinant. These methods can provide an original contribution to MinT reconciliation. The advantage of robust covariance estimation methods is that they can improve the reliability and accuracy of statistical models by reducing the influence of outliers and/or other departures from some expected elliptical distribution of the data. These methods can also be computationally intensive and may require more data than traditional methods to obtain reliable estimates.

5.1.1 How reconciliation is conducted

The idea of reconciliation consists of adjusting the forecasts so that, theoretically, better estimates of a given process can be found than using only independent forecasting methods. The process of base forecasts generation and forecast reconciliation in HTS methods can be represented in the flowchart described in Figure 5.1. It is important to emphasize that given the number of nodes and hierarchical levels, the amount of time series to forecast can increase exponentially. In this case, it is reasonable to adopt automated routines for model selection.

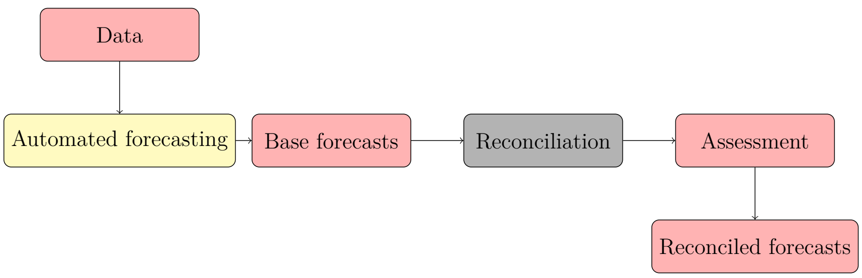

Figure 5.1: Forecast reconciliation process in HTS settings.

In several applications, time series data may contain outliers or other anomalies that can distort the underlying patterns and relationships in the data. Outliers can be caused by a variety of factors, such as measurement errors or unexpected events. These observations can significantly affect the accuracy and reliability of time series models leading to poor biased forecasts. In these cases, automated robust techniques play an important role when producing base forecasts which will be further reconciled. This would lead to the flowchart illustrated in Figure 5.2. In this case, the primary concern is to address measurement issues occurring in the original time series. Several articles address the topic of attenuating the effects of outliers and its detection in time series during the modeling stage (BARROW et al., 2020; ROUSSEEUW et al., 2019; CROUX et al., 2010).

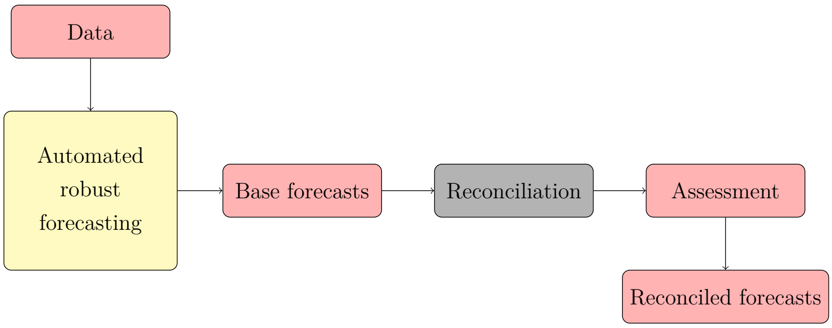

Figure 5.2: Robust base forecasts generation and forecast reconciliation.

The present Chapter focuses on reconciliation techniques that rely on the underlying covariance structure of the hierarchical time series to optimally combine base forecasts into reconciled ones. Overall, covariance-based reconciliation techniques offer a powerful tool for improving the accuracy and reliability of forecasts by taking into account the complex dependencies between different base forecasts. When modeling these relationships through their covariance structure, these techniques can produce reconciled forecasts that are both more accurate and more robust than any of the individual base forecasts produce independently. We propose a robust covariance estimation method to be implemented prior to reconciliation. In this case, the reconciliation processes can be described in the flowchart presented in Figure 5.3.

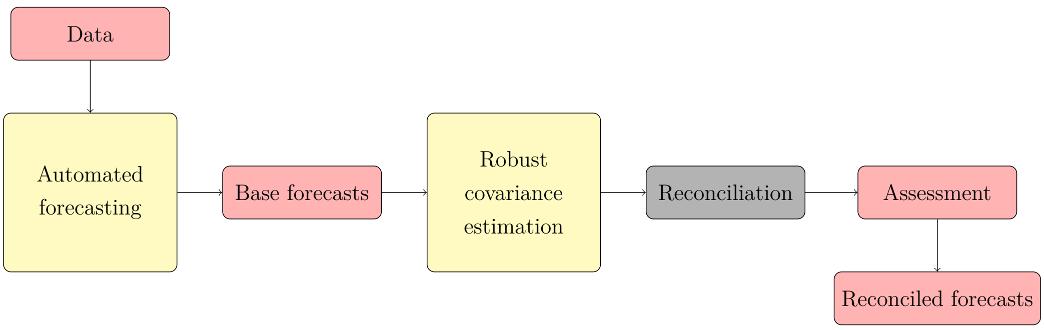

Figure 5.3: Base forecasts generation and robust covariance estimation prior to forecast reconciliation.

In addition, this study can also be viewed as an extension to the work of LILA et al. (2022) and MEIRA et al. (2023), providing doubly robust settings to the reconciliation process. In the context of reconciliation, doubly robust strategies can be particularly useful because they allow for more accurate and reliable reconciled forecasts. In this case, we considered the use of automated robust techniques to produce base forecasts and robust estimators for hierarchical reconciliation, which represents another contribution to the state-of-the-art techniques. Our idea of doubly robust reconciliation is depicted in the flowchart of Figure 5.4.

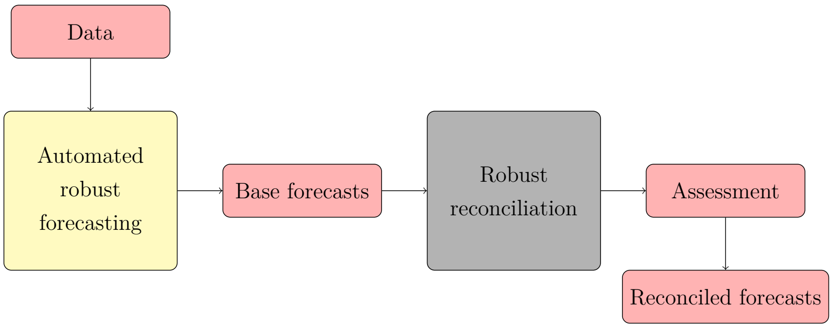

Figure 5.4: Robust base forecasts generation and robust forecast reconciliation.

To demonstrate the potential and validity of the proposed strategies, we set forth an application using hierarchical data on monthly electric energy consumption in Brazil, using a similar hierarchy as presented in MEIRA et al. (2023), exchanging levels. In addition, we also consider thew data on tourism from the original contribution of WICKRAMASURIYA et al. (2019) aggregated in a different fashion as presented in HYNDMAN & ATHANASOPOULOS (2018) in order to validate our approach. The results from this contribution will be tested against some of state-of-the-art methods which can be seen from the perspective of a regression model in order to validate the empirical findings.

5.2 Hierarchical Time Series and forecast reconciliation

Hierarchical Time Series (HTS) stand for a set of time series that can be aggregated at different levels, according to a well-defined hierarchical structure. It is possible to classify hierarchies into two basic structures: balanced and unbalanced. Balanced hierarchies have a regular structure and the same number of items or subgroups at each level. Unbalanced hierarchies, on the other hand, lack a regular structure, and there may be different levels of variability or uncertainty within the hierarchy. Consider a hierarchical time series with the following unbalanced structure at a given instant t.

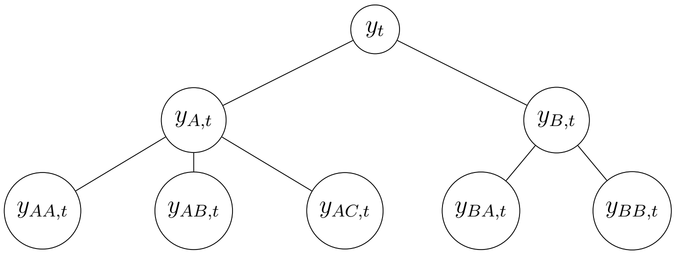

Figure 5.5: Unbalanced hierarchy at time t

The structure presented in Figure 5.5 can be represented in matrix notation. Let $\mathbf{y_t}$ be a vector of size m, comprising observations from all hierarchical levels at time t. It is possible to define an appropriate matrix $\mathbf{S}$ of dimension $m \times n$ such that,

$$\mathbf{y_t} = \mathbf{S}\mathbf{y_t^b} \tag{5-1}$$

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

$$\underbrace{\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}}_{\mathbf{y_t}} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}}_{\mathbf{y_t^b}} \tag{5-2}$$

In mathematical terms, the process of forecast reconciliation can be represented as follows. First, consider $\hat{\mathbf{y}}_{t+h|t}$ a vector of h steps ahead base forecasts, generated using independent methods, with the same arrangement as $\mathbf{y_t}$. Thus, for a given matrix $\mathbf{P}$ of dimension $n \times m$, we have the following equation

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{\mathbf{t+h|t}} \tag{5-3}$$

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

5.2.1 The optimal combination and its variants

In the work of HYNDMAN et al. (2011) introduced the idea of optimally combining forecasts from a regression-based perspective. The optimal reconciliation approach can be expressed according to the following regression model:

$$\hat{\mathbf{y}}_{t+h|t} = S\beta_{\mathbf{t+h|t}} + \epsilon_{\mathbf{t+h|t}} \tag{5-4}$$

where $\beta_{\mathbf{t+h|t}} = \mathbf{E}\left[\mathbf{y^b_{t+h}}|\mathcal{I}_\mathbf{t}\right]$, $\mathcal{I}_t = y_1, y_2, \ldots, y_t$ and $V\left(\epsilon_{t+h|t}|\mathcal{I}_t\right) = \Sigma_h$.

In HYNDMAN et al. (2016), the authors proposed the use of the Weighted Least Squares (WLS) estimator ignoring the elements outside the diagonal. Recent studies proposed alternative methods to obtain this matrix. In the work of WICKRAMASURIYA et al. (2019), it was introduced the Minimum Trace (MinT) reconciliation approach, which aims to find a matrix $\mathbf{P}$ that minimizes $tr(\mathbf{SPW_hP'S'})$ subject to $\mathbf{SPS} = \mathbf{S}$, the unbiasedness condition. Later, WICKRAMASURIYA et al. (2020), in turn, reconsidered the least squares minimization problem with non-negativity constraints to ensure that the coherent forecasts are strictly non-negative.

In order to use MinT reconciliation, it is necessary to estimate $\mathbf{W_h}$, the variance-covariance matrix of the h-step-ahead base forecast errors. Several approaches were proposed, giving birth to variants of the optimal combination approach depicted in eq. (5-4). Regardless of the approach selected for estimating $\mathbf{W_h}$, the optimal reconciled forecasts are given by

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S'W_h^{-1}S})^{-1}\mathbf{S'W_h^{-1}}\hat{\mathbf{y}}_\mathbf{h}. \tag{5-5}$$

In this case we have $\mathbf{P} = (\mathbf{S'W_h^{-1}S})^{-1}\mathbf{S'W_h^{-1}}$, to produce the reconciled forecasts. In this work we considered the following approaches for benchmarks to be compared with the new proposal as listed in LILA et al. (2022).

- The OLS approach takes $\mathbf{W_h} = \mathbf{k_h I}$, for a given constant $k_h > 0$. This is the simplest hypothesis since that the matrix $\mathbf{P} = \mathbf{S}(\mathbf{S'S})^{-1}\mathbf{S'}$ is independent of the data.
- The WLS(v) which stands for Weighted Least Squares estimator with variance scaling : $\mathbf{W_h} = \mathbf{k_h}\text{diag}(\hat{\mathbf{W}}_\mathbf{1})$ where $k_h > 0$ and $\hat{\mathbf{W}}_1$ is the unbiased sample covariance estimator of the in-sample one-step-ahead base forecast errors, represented as follows:

$$\hat{\mathbf{W}}_1 = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e_t(1)e_t(1)}' \tag{5-6}$$

$$\mathbf{W} = \begin{bmatrix} \hat{\sigma}^2_{tot} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\sigma}^2_{A} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{\sigma}^2_{B} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{\sigma}^2_{AA} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{\sigma}^2_{AB} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2_{AC} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2_{BA} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2_{BB} \end{bmatrix} \tag{5-7}$$

- The WLS (s) Weighted Least Squares estimator with structural scaling: $\mathbf{W_h} = \mathbf{k_h}\mathbf{\Lambda}$ where $k_h > 0$, $\mathbf{\Lambda} = \text{diag}(\mathbf{S1})$ and $\mathbf{1}$ is a $n \times 1$ vector of ones. This strategy uses the number of time series that organize the hierarchy in order to provide a hypothetical covariance structure illustrated in the following matrix.

$$\mathbf{W} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5-8}$$

- The MinT-Sample takes $\mathbf{W_h} = \mathbf{k_h}\hat{\mathbf{W}}_\mathbf{1}$ which considers the entire unbiased sample covariance estimator of the in-sample one-step-ahead base forecast errors. In this case, the matrix incorporates information of the whole hierarchy in the the reconciliation process.

$$\mathbf{W} = \begin{bmatrix} \hat{\sigma}^2_{tot} & \hat{\sigma}_{tot,A} & \hat{\sigma}_{tot,B} & \hat{\sigma}_{tot,AA} & \hat{\sigma}_{tot,AB} & \hat{\sigma}_{tot,AC} & \hat{\sigma}_{tot,BA} & \hat{\sigma}_{tot,BB} \\ \hat{\sigma}_{tot,A} & \hat{\sigma}^2_{A} & \hat{\sigma}_{A,B} & \hat{\sigma}_{A,AA} & \hat{\sigma}_{A,AB} & \hat{\sigma}_{A,AC} & \hat{\sigma}_{A,BA} & \hat{\sigma}_{A,BB} \\ \hat{\sigma}_{tot,B} & \hat{\sigma}_{A,B} & \hat{\sigma}^2_{B} & \hat{\sigma}_{B,AA} & \hat{\sigma}_{B,AB} & \hat{\sigma}_{B,AC} & \hat{\sigma}_{B,BA} & \hat{\sigma}_{B,BB} \\ \hat{\sigma}_{tot,AA} & \hat{\sigma}_{A,AA} & \hat{\sigma}_{B,AA} & \hat{\sigma}^2_{AA} & \hat{\sigma}_{AA,AB} & \hat{\sigma}_{AA,AC} & \hat{\sigma}_{AA,BA} & \hat{\sigma}_{AA,BB} \\ \hat{\sigma}_{tot,AB} & \hat{\sigma}_{A,AB} & \hat{\sigma}_{B,AB} & \hat{\sigma}_{AA,AB} & \hat{\sigma}^2_{AB} & \hat{\sigma}_{AB,AC} & \hat{\sigma}_{AB,BA} & \hat{\sigma}_{AB,BB} \\ \hat{\sigma}_{tot,AC} & \hat{\sigma}_{A,AC} & \hat{\sigma}_{B,AC} & \hat{\sigma}_{AA,AC} & \hat{\sigma}_{AB,AC} & \hat{\sigma}^2_{AC} & \hat{\sigma}_{AC,BA} & \hat{\sigma}_{AC,BB} \\ \hat{\sigma}_{tot,BA} & \hat{\sigma}_{A,BA} & \hat{\sigma}_{B,BA} & \hat{\sigma}_{AA,BA} & \hat{\sigma}_{AA,BA} & \hat{\sigma}_{AC,BA} & \hat{\sigma}^2_{BA} & \hat{\sigma}_{BA,BB} \\ \hat{\sigma}_{tot,AA} & \hat{\sigma}_{A,BB} & \hat{\sigma}_{B,BB} & \hat{\sigma}_{AA,BB} & \hat{\sigma}_{AA,BB} & \hat{\sigma}_{AC,BB} & \hat{\sigma}_{BA,BB} & \hat{\sigma}^2_{BB} \end{bmatrix} \tag{5-9}$$

- The shrinkage estimator, so called MinT-Shrink (WICKRAMASURIYA et al, 2019), takes $\mathbf{W_h} = \mathbf{k_h}\lambda\hat{\mathbf{W}}_\mathbf{1,D} + (\mathbf{1} - \lambda)\hat{\mathbf{W}}_\mathbf{1}$ where $k_h > 0$, $\hat{\mathbf{W}}_1$ is the unbiased sample covariance estimator of the in-sample one-step-ahead base forecast errors and $\hat{\mathbf{W}}_{1,D} = diag(\hat{\mathbf{W}}_1)$. In this case, the estimator of the covariance matrix aims to reduce the importance of elements outside the main diagonal of $\hat{\mathbf{W}}_1$. The shrinkage parameter λ is a function of the in-sample correlations and is estimated as follows.

$$\hat{\lambda} = \frac{\sum\limits_{i \neq j} \hat{V}(\hat{r}_{ij})}{\sum\limits_{i \neq j} \hat{r}_{ij}^2} \qquad (5\text{-}10)$$

where $\hat{r}_{ij}$ corresponds to ij-element of $\hat{\mathbf{R}}_1$, the one-step-ahead in-sample correlation matrix.

The Minimum Volume Ellipsoid (MVE) estimator is based on the smallest volume ellipsoid that covers g of the T observations (in-sample one-step-ahead base forecast errors).This is one of the oldest robust covariance estimators that is affine equivariant and has a positive breakdown value (AELST; ROUSSEEUW, 2009). From its definition, for a fixed g with $[T-m-1]/2 \leq g \leq T$, the MVE estimator of location $\hat{\mu}$ and scatter $W_h = \hat{\Sigma}$ is the solution of

$$(\hat{\mu}, \hat{\Sigma}) = \arg\min_{\mu, \Sigma} |\hat{\Sigma}| \qquad (5\text{-}11)$$

over all real μ and symmetric positive definite Σ that satisfy

$$\#\{i; d_i = \sqrt{(\mathbf{e_i(1)} - \hat{\mu})' \hat{\mathbf{\Sigma}}^{-\mathbf{1}} (\mathbf{e_i(1)} - \hat{\mu})} \leq c^2\} \geq g, \quad i = 1, \ldots, T \qquad (5\text{-}12)$$

where c is a constant, usually set as $c = \sqrt{\chi^2_{m,\alpha}}$, for $\alpha = g/T$. In this work , we made the assumption that there exists a small number of outliers by setting $\alpha = 0.95$.

The Minimum Covariance Determinant (MCD) estimator is one of the first affine equivariant and highly robust estimators of multivariate location and scatter (ROUSSEEUW, 1985; HUBERT et al., 2018). In other words, MCD is a method for estimating the mean and the covariance matrix whilst trying to minimize the influence of outliers. The idea is to estimate these parameters from a subset of the data that has been chosen to not contain the data imperfections. The basic idea, is to create all possible subsets of the data, of a specified size g and Estimate the mean and covariance matrix for each subset. Then, keep the estimates for the subset whose covariance matrix has the smallest determinant. Since its inception, many algorithms have provided efficient solutions in order to efficiently estimate the covariance structure without and exhaustive enumeration of all possible determinants values, which can be prohibitive depending on the data (ROUSSEEUW; DRIESSEN, 1999).

5.2.2 Robust and resistant reconciliation

From the perspective of a regression model, the reconciliation process can be written in another fashion as follows.

$$\tilde{\mathbf{y}}_h = \mathbf{S}\tilde{\beta}_{\mathbf{h}}. \tag{5-13}$$

Let the residuals from the reconciliation process be defined as:

$$\mathbf{y_h} - \tilde{\mathbf{y}}_{\mathbf{h}} = \epsilon(\tilde{\beta}_{\mathbf{h}}) \tag{5-14}$$

where $\mathbf{y_h}$ are the actual (true) values of time series and $\tilde{\mathbf{y}}_h$ are the reconciled forecasts. One of the problems that we find when reconciling forecasts is that we do not observe the reconciled residuals at the estimation stage. In this case, the distance between the reconciled forecasts and the independent ones can be used:

$$\hat{\mathbf{y}}_h - \tilde{\mathbf{y}}_h = \eta(\tilde{\beta}_h) \tag{5-15}$$

The work of LILA et al. (2022) considered the use of M-Estimators for $\tilde{\beta}_h$ in the context of hierarchical forecast reconciliation and proposed a robust-based approach applied to unemployment data from multiple labor force surveys. The approach can be summarized as follows: let ρ be a function having the following properties: nonnegative, i.e., $\rho(z) \geq 0$; $\rho(0) = 0$; symmetric, $\rho(z) = \rho(-z)$ and monotone in $|Z_i|$, $\rho(z_i) \geq \rho(z_{i'})$ for $|z_i| > |z_{i'}|$. Then, the robust M-estimator based on Equation (5-13) is given as:

$$\tilde{\beta}_{M,h} = \arg\min_{\tilde{\beta}_h} \sum_{i=1}^{n} \rho(\eta_i(\tilde{\beta}_h)). \tag{5-16}$$

In light of its desirable properties for computational convergence, LILA et al., (2022) considered the (HUBER, 1964) function, as follows:

$$\rho(z) = \begin{cases} z^2, & \text{if } |z| < c; \\ |2z|c - c^2, & \text{if } |z| \geq c \end{cases} \tag{5-17}$$

for a given constant c. Since some robust estimators are influenced by the scale of the residuals, a scale-invariant version of the M-estimator was used:

$$\tilde{\beta}_{\mathrm{M},h} = \arg\min_{\tilde{\beta}_h} \sum_{i=1}^{m} \rho\left(\frac{\eta_i(\tilde{\beta}_h)}{\sigma}\right), \tag{5-18}$$

In this study, we considered a version based on the Mean Absolute Deviation (MAD), represented below:

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{median\{|\eta_i(\tilde{\beta}_h)|\}}{0.6745} \tag{5-19}$$

Whilst robust reconciliation may provide accurate reconciled forecasts on several occasions, robust regression methods are only capable of dampening the influence of outliers. In other words, these approaches do not drop outliers, but instead reduce their effects during regression. In the context of hierarchical forecast reconciliation, some resistant techniques like the Least Absolute Deviations (LAD) can be used and provide reliable outcomes. In the work of MEIRA et al. (2023) the LAD-estimator based on the quantities from the equation (5-15) is given as

$$\tilde{\beta}_{LAD,h} = \arg\min_{\beta_h} \sum_{i=1}^{m} |\eta_i(\tilde{\beta}_h)|. \quad (5\text{-}20)$$

The solution of this minimization problem invokes a variant of the BARRODALE & ROBERTS (1974) simplex algorithm described in KOENKER & D'OREY (1987).

5.3 Base forecasting methods

The reconciliation strategy consists of reconciling, i.e., combining, a set of base forecasts into coherent ones by finding a solution which minimises the aggregate reconciliation error. This involves finding an orthogonal or oblique projection of the (incoherent) base forecasts onto a coherent subspace, then aggregating the information according a hierarchical structure, such that the upper level forecasts are obtained in an additive fashion from the most granular ones.

The choice of a base forecasting strategy depends on how data behave over time, or whether explanatory variables exist to improve the predictive power of a particular model. To allow straightforward comparisons with previously published papers in hierarchical forecasting reconciliation, in this work, we considered three forecasting approaches to generate the base forecasts: exponential smoothing models, robust exponential smoothing models and Seasonal Autoregressive Integrated Moving Average (SARIMA) formulations. These strategies are detailed in the next three sections.

5.3.1 Exponential Smoothing via ETS formulations

Exponential smoothing models belong to a class of forecasting methods whose forecasts correspond to weighted averages of past observations (OLIVEIRA, 2020). The weights decay exponentially, given the time difference between observations. Exponential smoothing formulations were first presented in seminal works (HOLT, 1957; BROWN, 1959; WINTERS, 1960). The selection of the best-fit exponential smoothing for a given time series is usually conducted by the identification of its error, trend and seasonal patterns (PEGELS, 1969; TAYLOR, 2003). In this work, we consider an automatic model selection routine commonly known as ETS – an acronym for Error, Trend and Seasonality, three components that are allowed to vary across exponential smoothing formulations. The ETS approach was addressed and placed in the form of state space by HYNDMAN et al. (2002). Each model consists of an observation equation, which describes the data, and one or more state equations, which describe the components of level ℓ_t, trend b_t and seasonality s_t, where t is equivalent to time instants. It provides a flexible framework for model selection selection. According to the taxonomy proposed by PEGELS (1969) and extended by GARDNER & MCKENZIE (1985), the possibilities for the trend and seasonal components in the ETS framework are depicted in Table 5.1. In addition, the error term can also vary between additive or multiplicative. That way, a total of 30 different formulations can be achieved as presented in (HYNDMAN; ATHANASOPOULOS, 2021).

		Seasonal Component		
Trend Component		N (None)	A (Additive)	M (Multiplicative)
N	(None)	(N,N)	(N,A)	(N,M)
A	(Additive)	(A,N)	(A,A)	(A,M)
A_d	(Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)
M	(Multiplicative)	(M,N)	(M,A)	(M,M)
M_d	(Multiplicative damped)	(M_d,N)	(M_d,A)	(M_d,M)

Table 5.1: Exponential smoothing methods

In practice, an automated algorithm, implemented in the `ets()` function of the `forecast` (HYNDMAN et al., 2020) package in the software `R` is used for the identification of the best-fit ETS formulation for each time series in the hierarchy.

5.3.2 Robust ETS

The Robust ETS models use a combination of robust estimation techniques and model structures to improve the accuracy and robustness of the forecasts. The main idea is to use a robustified likelihood estimator, which reduces the influence of extreme values in the estimation process. The robust ETS provides an alternative set of tools for estimating the ETS formulations as presented in 5.1. The strategy adopted in this work is based on CREVITS & CROUX (2017), which provides an outlier robust estimation procedure for estimating the vector of parameters θ of smoothing parameters. First, the procedure replaces the outlying observations in the following fashion in the case of additive error models.

$$\hat{y}_t^* = \psi\left[\frac{y_t - \hat{y}^*_{t|t-1}}{\hat{\sigma}_t}\right]\hat{\sigma}_t + \hat{y}^*_{t|t-1} \tag{5-21}$$

where ψ is the Huber influence function (HUBER, 1964), given as

$$\psi(z) = \begin{cases} z, & \text{if } |z| < c; \\ c[\text{sgn}(z)], & \text{if } |z| \geq c \end{cases} \tag{5-22}$$

The parameter σ_t^2 can be estimated recursively as presented in GELPER et al. (2010), for a determined constant λ_σ. As stated previously, assuming an additive error model the estimate of σ_t^2 is obtained as follows.

$$\hat{\sigma}_t^2 = \lambda_\sigma \rho\left[\frac{y_t - \hat{y}^*_{t|t-1}}{\hat{\sigma}_{t-1}}\right]\hat{\sigma}_{t-1} + (1-\lambda_\sigma)\hat{\sigma}^2_{t|t-1} \tag{5-23}$$

where

$$\rho(z) = \begin{cases} c_k\left\{1 - (1 - (\frac{z}{k})^2)^3\right\}, & \text{if } |z| < k; \\ c_k, & \text{if } |z| \geq k \end{cases}$$

is the Tukey's Biweight function, adjusted for c_k and k as presented in CREVITS & CROUX (2017). In the case of a multiplicative error model, outliers are replaced as follows.

$$\hat{y}_t^* = \left(1 + \psi\left[\frac{y_t - \hat{y}^*_{t|t-1}}{\hat{y}^*_{t|t-1}\hat{\sigma}_t}\right]\hat{\sigma}_t\right)\hat{y}^*_{t|t-1} \tag{5-24}$$

In this case, the estimation of σ_t^2 is given recursively by

$$\hat{\sigma}_t^2 = \lambda_\sigma \rho\left[\frac{y_t - \hat{y}^*_{t|t-1}}{\hat{y}^*_{t|t-1}\hat{\sigma}_{t-1}}\right]\hat{\sigma}_{t-1} + (1-\lambda_\sigma)\hat{\sigma}^2_{t|t-1} \tag{5-25}$$

Assuming that errors are normally distributed, it is possible to estimate the vector of parameters $\boldsymbol{\theta}$ by maximizing a robust version of the likelihood function. In this case the estimator $\hat{\boldsymbol{\theta}}$ is obtained as follows.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} -\frac{T}{2} log\left[\frac{s_T^2(\boldsymbol{\theta})}{T}\sum_{i=1}^{T}\rho\left[\frac{y_t - \hat{y}^*_{t|t-1}(\boldsymbol{\theta})}{s_T^2(\boldsymbol{\theta})}\right]\right] \tag{5-26}$$

where $s_T^2(\boldsymbol{\theta}) = 1.4826 \underset{t}{median} |y_t - \hat{y}^*_{t|t-1}(\boldsymbol{\theta})|$.

In the case of a multiplicative error model, the estimator $\hat{\boldsymbol{\theta}}$ is obtained in the following fashion.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} -\frac{T}{2} log \left[\frac{s_T^2(\boldsymbol{\theta})}{T} \sum_{i=1}^{T} \rho \left[\frac{y_t - \hat{y}^*_{t|t-1}(\boldsymbol{\theta})}{\hat{y}^*_{t|t-1}(\boldsymbol{\theta}) s_T^2(\boldsymbol{\theta})} \right] \right] - \sum_{i=1}^{T} log |\hat{y}^*_{t|t-1}(\boldsymbol{\theta})| \quad (5\text{-}27)$$

where $s_T^2(\boldsymbol{\theta}) = 1.4826 \underset{t}{median} \left| \frac{y_t - \hat{y}^*_{t|t-1}(\boldsymbol{\theta})}{\hat{y}^*_{t|t-1}(\boldsymbol{\theta}) s_T^2(\boldsymbol{\theta})} \right|$.

In practice, CREVITS & CROUX (2016) converted the function `ets` in the `forecast` package of HYNDMAN et al. (2008) to a robust version, an automated algorithm, implemented in the `robets()` function of the `robets` package in the software `R`.

5.3.3 SARIMA formulations

The Seasonal Autoregressive Integrated Moving Average (SARIMA) formulations are an integral part of the so-called BOX & JENKINS (BOX; JENKINS, 1970b) family of models for estimating and forecasting univariate time series data. Devised as alternative approaches to traditional exponential smoothing methods, SARIMA models are similar to the latter in the sense that they are adaptive, can model trends and seasonal patterns, and can be automated. Conversely, they are based on autocorrelations (patterns in time) rather than a structural view of level, trend and seasonality (as in ETS formulations).

The Autoregressive Integrated Moving Average (ARIMA) models explain a univariate time series as a combination of autoregressive and moving average components that explore the existing autocorrelation within the time series. In addition, the integration order depends on the number of consecutive times that the series has been differenced so as to obtain a stationary process. The class of seasonal ARIMA models, so called SARIMA of order $(p, d, q) \times (P, D, Q)_s$, is composed by a non-seasonal part (p, d, q) and a seasonal $(P, D, Q)_s$ one. These models can be written in a compact fashion, as follows:

$$\phi(B)\Phi(B)\nabla^d\nabla_s^D y_t = \theta(B)\Theta(B)a_t \quad (5\text{-}28)$$

where:

y_t is the variable of interest over time;

a_t is the error term;

$\phi(B)$ is the autoregressive operator of order p ;

$\Phi(B)$ is the autoregressive seasonal operator of order$P$;

∇^d is the non-seasonal difference operator;
∇_s^D is the seasonal difference operator;
$\theta(B)$ is the moving average operator of order q; and
$\Theta(B)$ is the seasonal moving average operator of order Q.

SARIMA models can be implemented in R by means of the `arima()` function from `forecast` in `R`. In our case, we opt to select the best SARIMA model for each time series in the hierarchy via the `auto.arima()` function from the same package, which implements a variation of the Hyndman-Khandakar algorithm (HYNDMAN; KHANDAKAR, 2008). The algorithm combines unit root tests, minimization of the lowest Akaike Information Criteria with corrections (AICc) (SUGIURA, 1978) and Maximum Likelihood Estimation (MLE) to obtain best-fit SARIMA model.

5.3.4 Assessment metrics

To obtain the overall accuracy of the reconciled forecasts, we investigated the improvements on three different assessment metric as follows.

– The Root Mean Squared Error (RMSE).

$$\sqrt{\frac{\sum_{t=1}^{h}(y_t - \hat{y}_t)^2}{h}} \tag{5-29}$$

– The Mean Absolute Error (MAE).

$$\frac{1}{h}\sum_{t=1}^{h}|y_t - \hat{y}_t| \tag{5-30}$$

– The Mean Absolute Percentage Error (MAPE).

$$\frac{100}{h}\sum_{t=1}^{h}\frac{|y_t - \hat{y}_t|}{|y_t|} \tag{5-31}$$

where $\hat{y}_t$ are the real (actual) and forecasted values of the underlying series, respectively; h is the forecasting horizon (number of forecasting steps ahead).

Given that multiple reconciliation strategies are considered to generate forecasts for the whole hierarchy, relative measures from reconciled forecasts are obtained with respect to independent ones. The geometric mean of these relative measures within each level of the hierarchy (or for the entire hierarchy) provides a summarized measure of improvement in terms of MAE, RMSE and MAPE. These quantities can be obtained through the following equations:

$$AveRelMAE = \sqrt[\#L]{\prod_{i \in L} \frac{MAE_{i,h}^{rec}}{MAE_{i,h}^{base}}} \tag{5-32}$$

$$AveRelRMSE = \sqrt[\#L]{\prod_{i \in L} \frac{RMSE_{i,h}^{rec}}{RMSE_{i,h}^{base}}} \tag{5-33}$$

$$AveRelMAPE = \sqrt[\#L]{\prod_{i \in L} \frac{MAPE_{i,h}^{rec}}{MAPE_{i,h}^{base}}} \tag{5-34}$$

where the denominator corresponds to the metric associated to the independent forecasts, also called base forecasts, for the series i of the hierarchy at the forecast horizon h, and the numerator is the metric obtained for the same series and time horizon after the reconciliation process. For equations (5-34)-(5-34), L is the corresponding level of the hierarchy.

In order to gauge the percentage of improvement on RMSE, we use $(1 - AveRelRMSE) \times 100$ as shown in WICKRAMASURIYA et al. (2019). The same applies to MAE and MAPE.

5.4 The Experimental Setup

5.4.1 First set of experiments

The first set of experiments considers hierarchical monthly data of electric energy consumption across the Brazilian National Interlinked System (SIN). The data comprise a three-level balanced hierarchy. The data is organized by five geographic regions that provide administrative divisions of the Brazilian territory and four classes of consumption. These data are officially com piled in megawatt-hours (MWh) by the Brazilian Energy Research Company (EPE, 2022a) and spans from January 2004 to November 2021, the last official date available at the time of collection. Geographic regions are divided into: North (NO), Northeast (NE), Midwest (MW), Southeast (SE) and South (SO). Classes of consumption within regions in the Brazilian power system are:Commercial, Industrial, Residential and Others (mainly rural, public service and public lighting). Figure 5.6 illustrates the hierarchical structure considered to represent the Brazilian power system in this study.

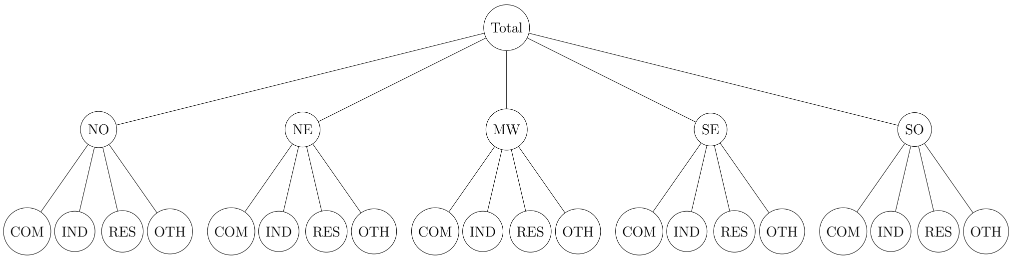

Figure 5.6: Hierarchical structure for a three-level hierarchy of regions and classes of consumption within regions.

The charts from Figure 5.8 present similar trends to the chart from Figure 5.7, suggesting that regional data present similar behaviors to national electricity consumption. In addition, it is possible to identify some indicative of a multiplicative seasonal component as the volatility of the last months increases compared to the initial points of the series.

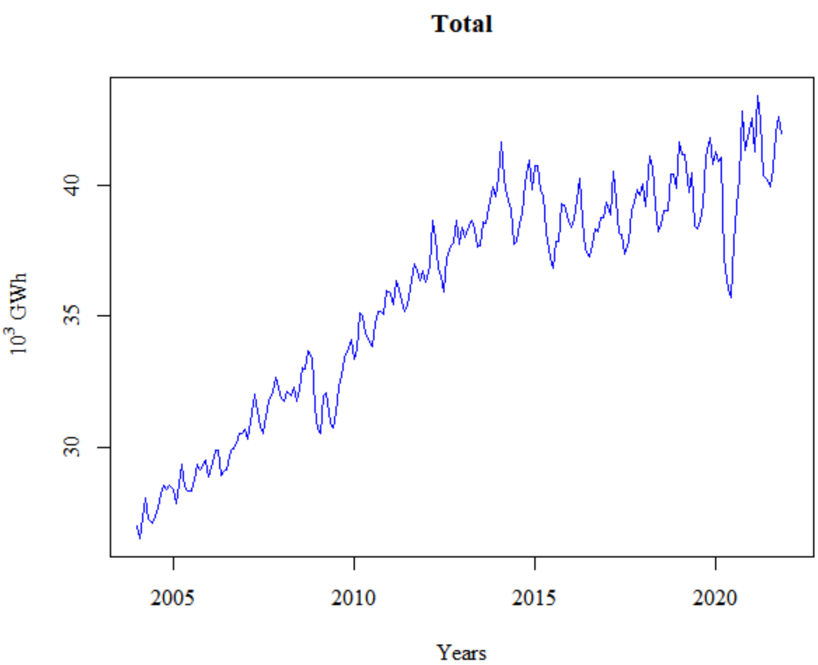

Figure 5.7: Total electric energy consumption in the SIN.

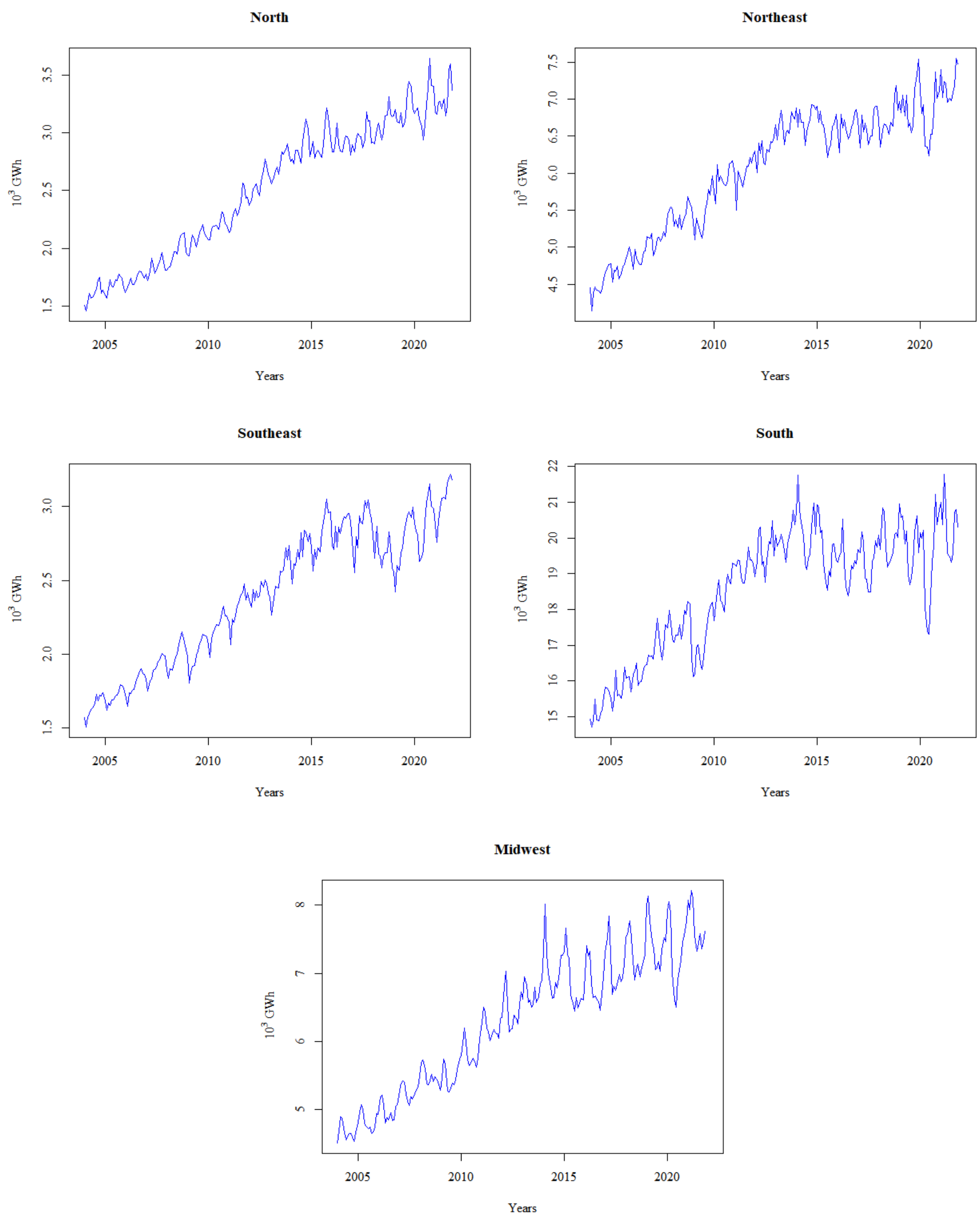

Figure 5.8: Total electric energy consumption in the SIN by geographical regions.

Figure 5.9 illustrates the behavior of the time series at the most granular level of the hierarchy, i.e., classes of consumption within geographic regions . In most cases, the trend and seasonality components are also observed, sharing similar behaviors with the more aggregated levels of the hierarchy. One can also observe a substantial difference in the behavior of the time series referring to the Industrial sector, whose historical records are considerably different than those observed in other geographic regions.

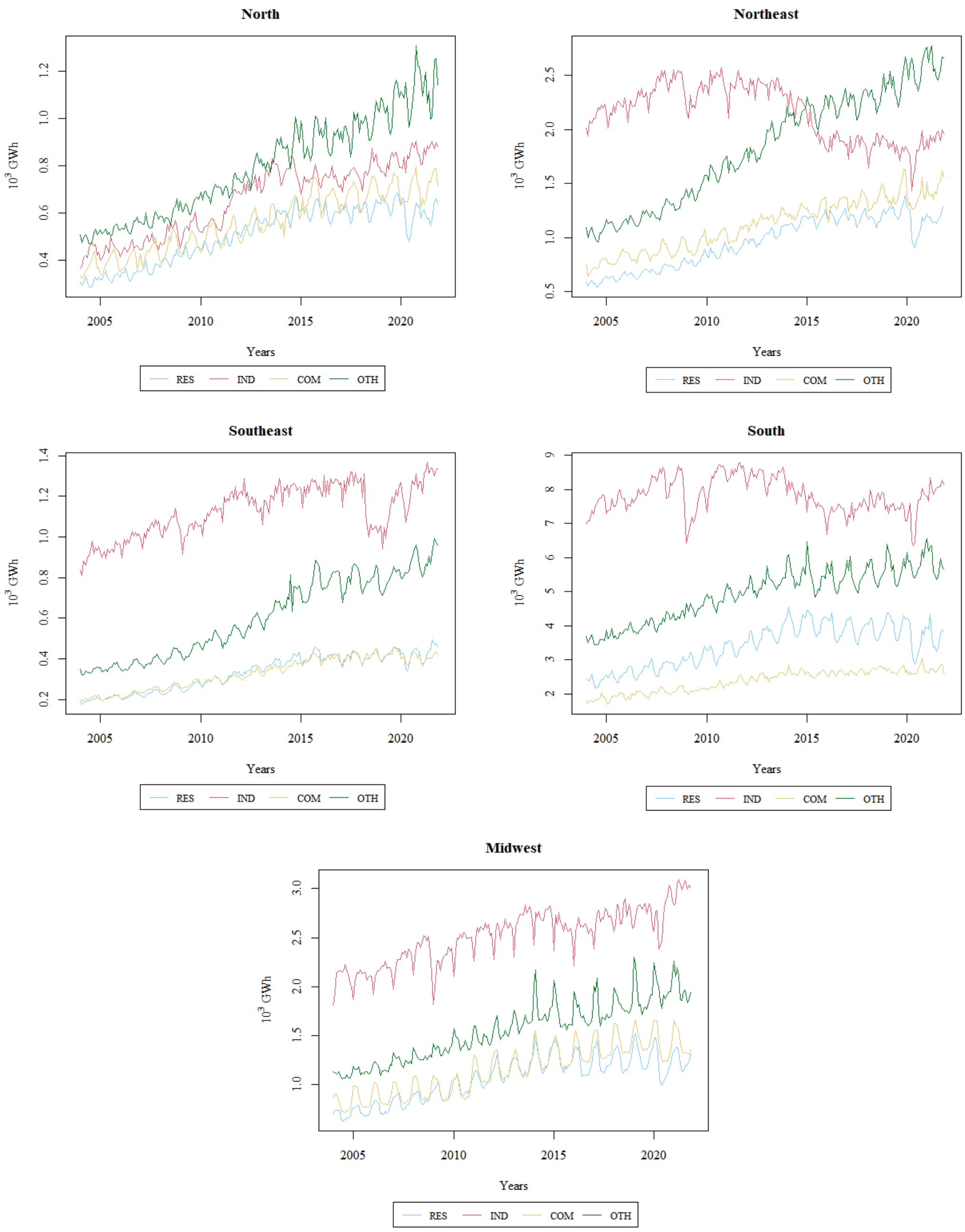

Figure 5.9: Total electric energy consumption in the SIN by classes of consumption within geographical regions.

These experiments explore the forecasting accuracy of the reconciliation approaches in two fixed medium-term forecasting horizons. These correspond to the last six months (experiment 1.a) and the last twelve months (experiment 1.b) of official data available in the test set, i.e., from June 2021 to November 2021 ($h = 1 - 6$) and from December 2020 to November 2021 ($h = 1 - 12$) as shown in Figure 5.10.

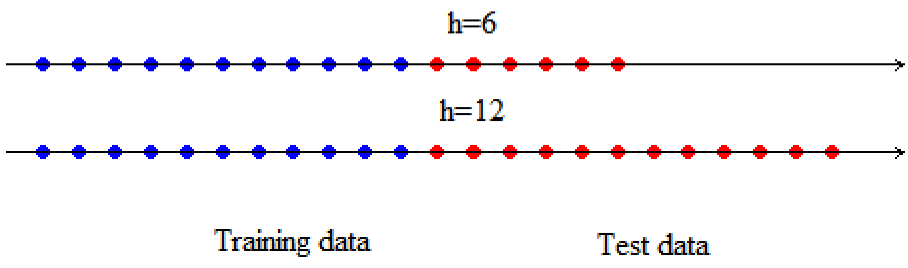

Figure 5.10: Defining the train and test sets in the first set of forecasting experiments.

5.4.1.1 Results from the first set of experiments

In order to provide a better understanding and guidance to the results, numbers highlighted in **bold** indicate the methods that provide the best forecasting performance in each forecasting horizon, while numbers in *italics* represent the methods that rank as second best reconciliation strategies.

When analysing the results presented in Table 5.2, considering ETS formulation to produce base forecasts, it is possible to identify the contribution from the MCD covariance estimator in the forecast horizon $h = 1 - 6$. In this horizon, the MCD estimator offered an improvement on assessment metrics varying from 6.67% to 7.85%. Although it ranked as the second best strategy for reconciliation, it outperformed both MinT-Shrink and Mint-Sample, whose reconciliation process relies on a well estimated covariance structure. The results for the longer forecasting horizon show that reconciling using a conditional median, given the hierarchical structure, provides improvements on the assessment metrics varying from 3.17% to 3.59%, suggesting that robust settings can produce substantial gains on accuracy.

Reconciliation approach	Forecast horizon $h = 1 - 6$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	0.65	0.43	0.27	-0.66	-0.81	-0.7
LAD	-0.48	-0.49	-0.39	**3.17**	**3.32**	**3.59**
OLS	**11.67**	**12.6**	**11.56**	-10.98	-11.77	-11.42
WLS(v)	1.1	1.31	1.32	1.42	2.05	2.11
MinT-Shrink	1.95	2.04	1.97	*2.21*	*2.53*	*2.49*
MinT-Sample	5.18	6.24	6.2	-1.15	-1.45	-1.54
WLS(s)	4.23	4.24	3.89	-2.07	-2.46	-2.29
MCD	*6.67*	*7.74*	*7.85*	-0.11	-0.24	-0.38
MVE	-77.46	-88.55	-101.11	-112.73	-131.27	-129.53

Table 5.2: Improvement, in percentage points (%), on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. ETS as base forecasting method.

The results presented in Table 5.3 show the benefits of considering robust covariance estimates in the process of reconciling forecasts and SARIMA formulations. Although for the shorter forecasting horizon the results were not promising, when forecasting longer periods, the improvement on the accuracy measures were twice as good when compared with other reconciliation strategies. The observed improvement on MAPE reached 8.26% while the improvements on RMSE and MAE were 7.77% and 8.11%, respectively.

Reconciliation approach	Forecast horizon $h = 1 - 6$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	*0.77*	-2.97	-2.68	-3.3	-2.87	-2.45
LAD	0.01	-0.89	-0.7	-0.32	0.11	0.43
OLS	-1.42	-4.11	-3.72	-15.8	-14.31	-13.73
WLS(v)	**2.45**	**1.62**	**1.63**	1.37	0.92	1.01
MinT-Shrink	-0.14	-0.35	-0.34	*3.84*	3.74	3.72
MinT-Sample	-7.53	-6.48	-6.47	3.56	*3.95*	*4.05*
WLS(s)	0.51	-2.4	-2.18	-3.9	-3.65	-3.27
MCD	-9.21	-7.94	-7.97	**7.77**	**8.11**	**8.26**
MVE	-89.72	-95.47	-94.63	-78.98	-92.47	-93.62

Table 5.3: Improvement, in percentage points (%), on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. SARIMA as base forecasting method.

Finally, the results presented in Table 5.4 show the important contribution from robust settings in both endings. When forecasting using the Robust ETS to produce base forecasts and the MCD for covariance estimation, the results can be considered the best for all scenarios. The MCD was dominant for two assessment metrics in both forecasting horizons.

Reconciliation approach	Forecast horizon $h = 1 - 6$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	-9.58	-9.11	-8.85	-6.66	-5.71	-5.43
LAD	-5.47	-5.37	-5.32	-6.85	-6.43	-6.53
OLS	-10.67	-12.68	-12.24	-8.47	-8.21	-8.05
WLS(v)	-1.41	-2.58	-2.55	1.74	2.36	2.46
MinT-Shrink	1.08	-0.36	-0.27	*2.06*	*2.77*	*2.82*
MinT-Sample	**6.92**	*5.39*	*5.6*	-1.38	-0.06	-0.04
WLS(s)	-7.23	-7.52	-7.22	-1.07	0.13	0.41
MCD	*5.3*	**5.7**	**5.8**	**2.49**	**3.73**	**3.84**
MVE	-59.85	-67.87	-67.59	-59.76	-69.86	-69.28

Table 5.4: Improvement, in percentage points (%), on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. Robust ETS as base forecasting method.

5.4.2 Second set of experiments

To understand the effects of the proposed methodological approach to other hierarchical time series, we consider the compiled aggregated data set from HYNDMAN & ATHANASOPOULOS (2018) that was also used to validate the MinT approach with a more extended hierarchy. The data set contains information on total quarterly visitor nights (in millions) that Australians spend away from home, from 1998-2016 for twenty regions of Australia within six states. The states are: New South Wales (NSW), Queensland (QLD), South Australia (SAU), Victoria (VIC), Western Australia (WAU), and Other (OTH). The states are aggregation of zones that are organized in an unbalanced fashion as show in Figure 5.11. The zones within states are: Metro (ME), North Coast (NC), South Coast (SC), South Inner (SI), North Inner (NI), Central (CE), Coastal (CO), Inner (IN), West Coast (WC), East Coast (EC) and Non-Metro (NM).

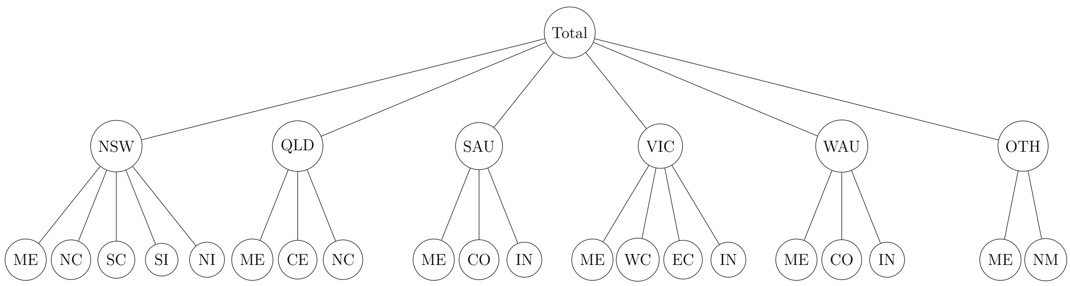

Figure 5.11: Hierarchical structure for a three-level unbalanced hierarchy of states and zones within states.

Figure 5.12 illustrates the behavior of the Australian tourism data for the whole data set. I is possible to identify a strong seasonal component over the entire period and a slightly increasing trend starting from 2008.

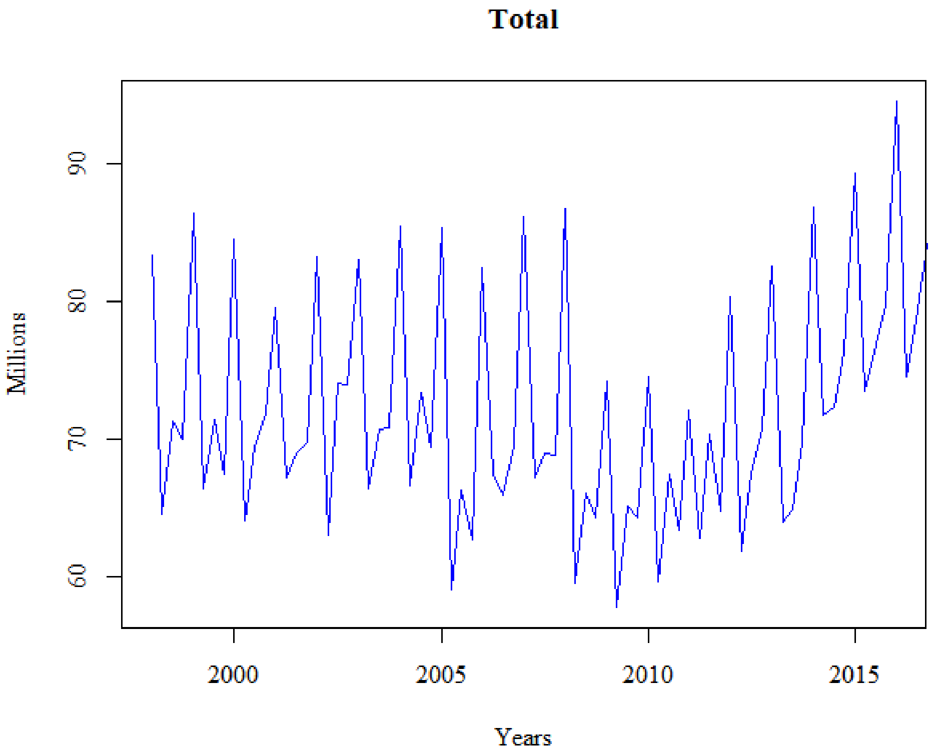

Figure 5.12: Australian domestic visitor nights.

Figure 5.13 shows how different the time series are across states. Although there is a clear seasonal component as shown in the most aggregated hierarchical level, there is a shift in the level and some states show a more upward trend.

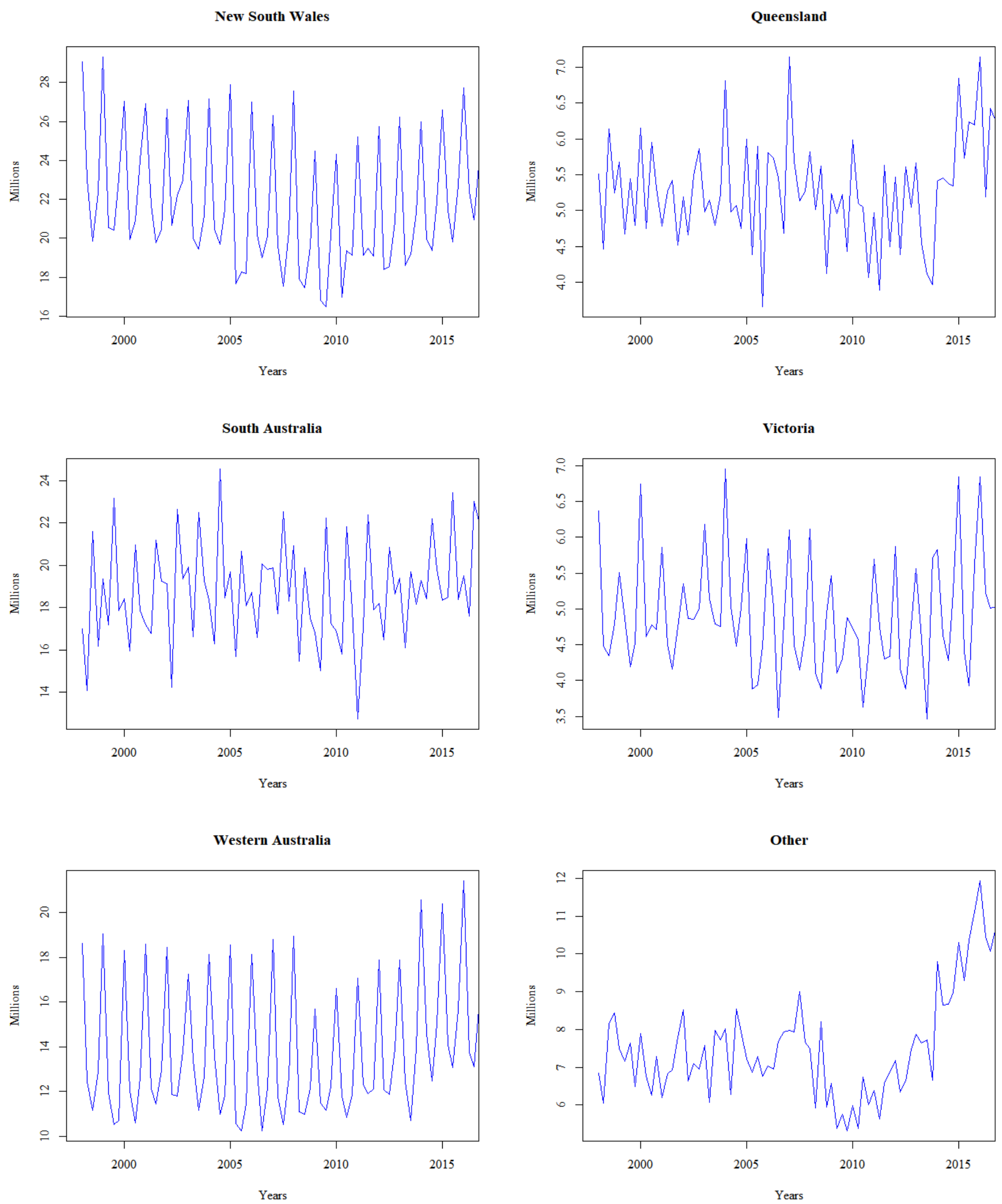

Figure 5.13: Australian domestic visitor nights by States.

The series illustrated in Figure 5.14 show within states differences due to zones. The seasonal components and the level changes are the most significant changes depicted in this group of time series.

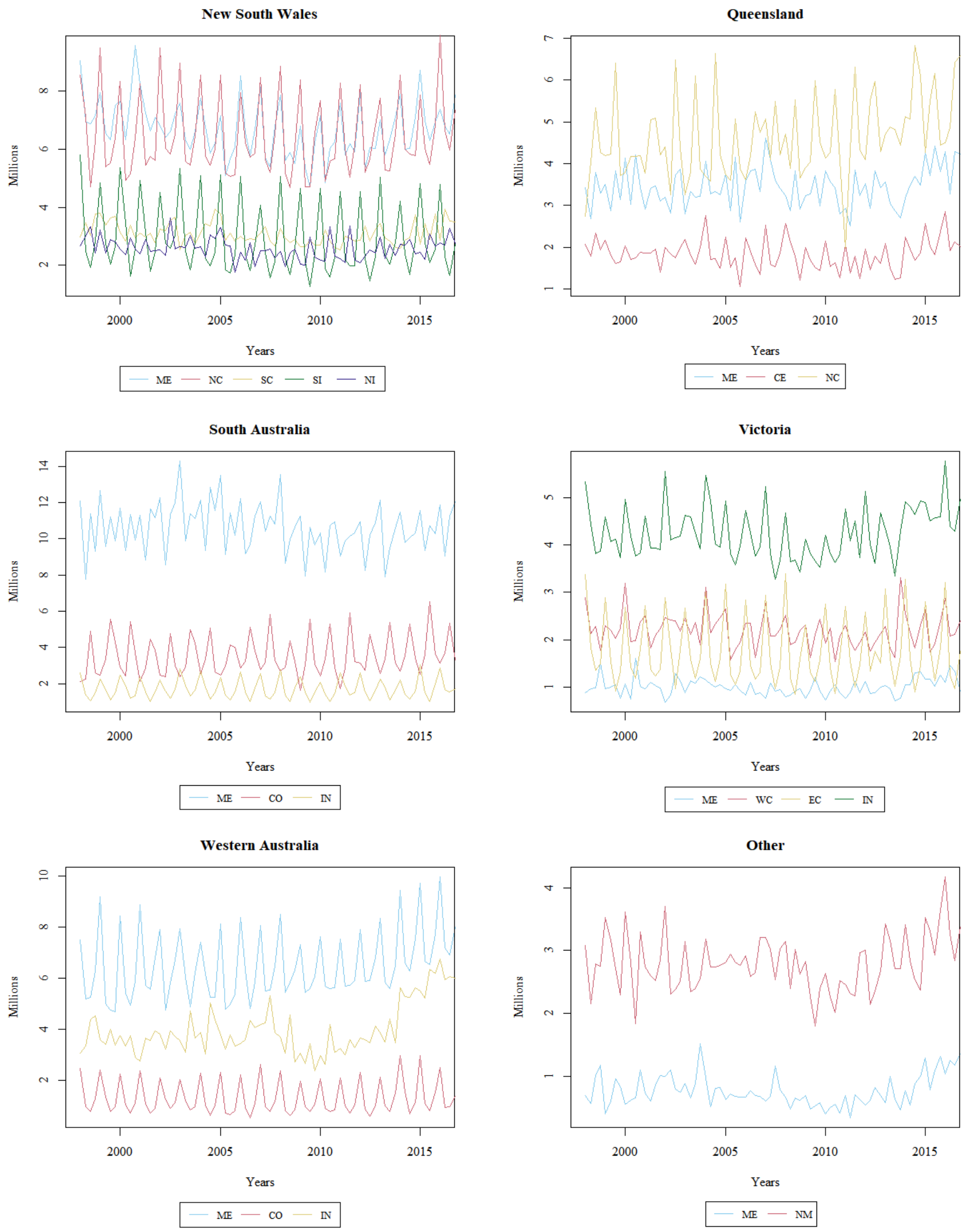

Figure 5.14: Australian domestic visitor nights by Zones within States.

This second set of experiments explores the forecasting accuracy of the reconciliation approaches in two fixed forecasting horizons. These correspond to the last eight quarters (experiment 2.a) and the last twelve quarters (experiment 2.b) available in the test set, i.e., from the first quarter of 2015 to the last quarter of 2016 ($h = 1 - 8$) and from the first quarter of 2014 to the last quarter of 2016 ($h = 1 - 12$), as shown in Figure 5.15.

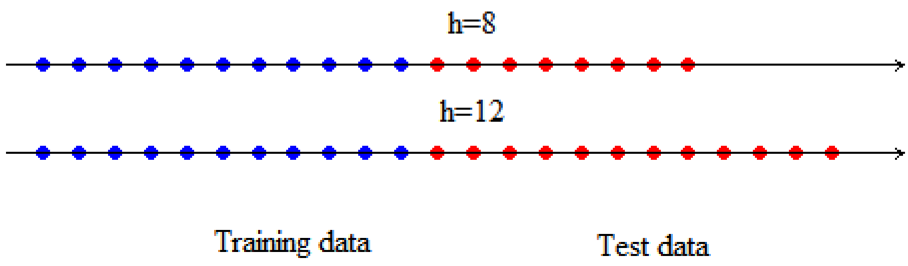

Figure 5.15: Defining the train and test sets in the second set of forecasting experiments.

5.4.2.1
Results from the second set of experiments

The results presented in Table 5.5 show that the applicability of the proposed approaches can be extended to other hierarchical data sets. Considering ETS as the base forecasting method, the MCD estimator ranked as the second best strategy for the shorter horizon, offering improvements on assessment metrics varying from 6.67% to 7.85%. When considering the forecasting horizon $h = 1 - 12$, the MCD estimator indicates the best overall performance. In this scenario the improvement on the assessment metrics have a similar magnitude, varying from 4.14% in terms of RMSE and 4.65% for MAPE.

Reconciliation approach	Forecast horizon $h = 1 - 8$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	0.65	0.43	0.27	2.84	*3.31*	3.21
LAD	-0.48	-0.49	-0.39	2.17	2.61	2.58
OLS	**11.67**	**12.6**	**11.56**	*2.91*	3.3	*3.22*
WLS(v)	1.1	1.31	1.32	0.63	0.78	0.76
MinT-Shrink	1.95	2.04	1.97	0.91	1.12	1.09
MinT-Sample	5.18	6.24	6.2	1.95	1.87	1.82
WLS(s)	4.23	4.24	3.89	1.59	1.72	1.64
MCD	*6.67*	*7.74*	*7.85*	**4.14**	**4.65**	**4.39**
MVE	-77.46	-88.55	-101.11	-155.28	-173.62	-181.83

Table 5.5: Improvement, in percentage points (%), on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. ETS as base forecasting method.

When considering SARIMA formulations, the MCD estimator also performed better that the regression-based approaches that depend on covariance structures, outperforming both Mint-Shrink and Mint-Sample. The MCD provided significant gains in all assessment metrics for both forecasting horizons.

Reconciliation approach	Forecast horizon $h = 1 - 8$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	*9.56*	9.53	*9.08*	6.68	7.45	*7.41*
LAD	2.68	2.08	1.41	5.24	6.53	6.8
OLS	**14.58**	**14.26**	**13.03**	**9.24**	**9.43**	**8.65**
WLS(v)	0.9	1.44	1.67	1.46	1.7	1.75
MinT-Shrink	1.16	1.67	1.78	3.88	4.19	3.82
MinT-Sample	6.14	4.67	2.72	4.63	4.55	2.97
WLS(s)	5.48	6.73	6.8	3.4	3.54	3.32
MCD	8.58	*11.11*	7.36	*8.05*	*8.2*	6.88
MVE	-14.2	-16.79	-23.49	-24.81	-27.81	-33.51

Table 5.6: Improvement, in percentage points (%), on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. SARIMA as base forecasting method.

Finally, Table 5.7 presents the results from Robust ETS formulations as base forecasts for the reconciliation process. In this case it is possible to infer the benefits of doubly robust settings. The robust reconciliation approach was capable of improving the final forecasts showing improvements of 6.27% in terms of MAE, 5.46% and 5.9% for RMSE and MAPE respectively. For the second forecasting horizon (experiment 2.b) horizon $h = 1 - 12$ the doubly robust settings offered the second best performance according to the assessment metrics.

Reconciliation approach	Forecast horizon $h = 1 - 8$			Forecast horizon $h = 1 - 12$		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HUBER	**5.46**	**6.27**	**5.9**	*1.76*	*2.38*	*2.08*
LAD	-1.99	-1.63	-2.16	-2.68	-3.3	-4.37
OLS	*4.91*	*5.68*	*5.06*	**2.4**	**2.64**	**2.51**
WLS(v)	-0.06	-0.29	-0.3	0.28	0.6	0.49
MinT-Shrink	-2.43	-2.37	-2.34	-0.6	-0.32	-0.51
MinT-Sample	-26.31	-28.81	-29.31	-12.07	-12.69	-14.11
WLS(s)	2.14	1.57	1.3	1.54	1.95	1.9
MCD	-22.68	-24.85	-25.29	-10.86	-11.1	-12.78
MVE	-48.95	-53.45	-59.86	-28.43	-32.08	-36.19

Table 5.7: Improvement(%) on RMSE, MAE and MAPE across all hierarchical levels for different forecast horizons. Robust ETS as base forecasting method.

5.5 Conclusions

The main contribution of this work addresses to how reconciliation processes based on regression models can benefit from robust covariance structures. The MCD covariance estimator provided more stable results for all hierarchical levels.

The choice between MCD and Minimum Volume Ellipsoid MVE methods for robust covariance estimation depends on specific characteristics of the data. The MCD approach for covariance estimator is generally considered to be more suitable than MVE in situations when the underlying data distribution is multivariate normal.

There are different approaches to estimate robust covariances in the most up-to-date literature, which extends the application of these methods in the context of forecast reconciliation for HTS such as M-estimators, the Stahel-Donoho estimator, S-estimators and MM-estimators, subject to the unbiaseness condition.

The presented results indicated that the MCD approach performed well in both balanced and unbalanced hierarchies, providing reliable estimates for both settings.

Although, the robust methods have shown considerable gains on accuracy measures, it is important to investigate in which occasions these methods performed better, from hierarchical levels to forecasting horizons.

The framework provided in WICKRAMASURIYA et al. (2019) can still be considered as the cornerstone for different strategies of reconciling forecasts. The combination of these methods with more sophisticated forecasting approaches may produce even better gains on accuracy measures .

A natural extension to this work addresses the production of prediction intervals with robust settings, using both parametric and non-parametric approaches.

6
Achievements, conclusions and next steps

This thesis is driven by the increased interest in producing dependable forecasts for hierarchical data in recent years, primarily due to their growing use in policy-making, government funding allocation, and regional planning. As a result, HTS has emerged as a promising research area, with a focus on developing new studies to meet these demands.

Despite the considerable progress made in hierarchical time series forecasting, previous research had not explored the use of robust estimators to generate reconciled forecasts from a regression-based perspective. This is a significant oversight, given the potential for outliers to occur in multiple time series within a hierarchy due to a variety of reasons, including inaccurate or misleading information, measurement errors, and incorrect data processing.

This thesis addresses this gap in Chapter 3, by proposing two robust methods for hierarchical time series and applying them to numerous unemployment time series hierarchies in Brazil, at both monthly and quarterly frequencies. Our approach is motivated by the fact that robust regression is less sensitive to outliers and influential point forecasts. By using an M-estimator based regression, we can mitigate the undesirable effects of high leverage data points. Our empirical evaluation demonstrates promising results in support of the proposed robust reconciliation methods. While these estimators may not be uniformly dominant across all scenarios, they rank among the top-performing approaches in almost every case, competing with state-of-the-art techniques in hierarchical forecasting reconciliation. We observe significant improvements in accuracy, as measured by assessment metrics for various hierarchical levels.

In cases where there are no outliers, reconciled forecasts that take into account error structure dependencies will have the smallest variance compared to other possible forecast combinations. However, this strategy may not be effective when there are outlying observations in the reconciliation process. To address this issue, we introduce the concept of hierarchical forecast reconciliation based on resistant regression in Chapter 4 and develop a modified resistant-based strategy for electricity consumption time series in Brazil. By incorporating resistant statistics into the forecast reconciliation process, we make another valuable contribution to the field of HTS forecasting. Specifically, our method substantially reduces the impact of outliers on the reconciled forecasts. We conducted several experiments comparing our approach to traditional and innovative benchmarks and found that it provided superior forecasting accuracy in various scenarios. Consequently, our approach is well-suited to support decision-making in the energy and related sectors.

Reconciliation strategies that combine base forecasts using regression models may compromise the weighting system if the covariance structures are not well estimated, leading to distortions in the reconciliation process. To address this issue, we proposed in Chapter 5 the use of robust estimation of the covariance structure for hierarchical forecast reconciliation. The proposed approach has been demonstrated to be suitable for supporting decision-making in both public and private sectors. Additionally, the developed methodology is flexible and can be applied to other sets of hierarchical time series without restrictions.

As observed throughout the Chapters, the findings of this research have made original contributions to the area of HTS as the publication of two original researchs articles, the first in the Socio-Economic Planning Sciences journal (ISSN: 0038-0121) – LILA, M. F.; MEIRA, E.; OLIVEIRA, F. L. C. F (2022), transcribed in full in Chapter 3 – and a second article published at Energy (ISSN 0360-5442) - MEIRA, E. LILA,; M. F.; OLIVEIRA, F. L. C. F (2023), transcribed in full in Chapter 4.

Besides, the original articles, in order to be in compass with the academic community in the field HTS forecasting, the methodological core of this thesis was submitted for presentation in one of the most prestigious conferences on forecasting, the International Symposium on Forecasting (ISF) on three occasions. All contributions were accepted for oral presentation receiving feedbacks from valuable researchers in this specific area of HTS.

The first presentation was in the year of 2020 at the ISF 2020: 40th International Symposium on Forecasting, when the idea of robust reconciliation was first presented, as a application to unemployment time series. The second presentation occurred at ISF 2021 : 41st International Symposium on Forecasting, when a follow-up on robust reconciliation was presented. The audience provided excellent insights that culminated with the publication of the article at Socio-Economic Planning Sciences. At the ISF 2022: 42nd International Symposium on Forecasting, the idea of resistant reconciliation was first introduced and discussed. During the section on hierarchical forecasting, a fruitful discussion with participants was promoted. The insights derived from the discussions provided directions to a third contribution, which applies a modification in the covariance structure imposed by the MinT reconciliation.

The research was also presented for the Brazilian audience at the 53th Brazilian Symposium on Operations Research (SBPO 2021), whose work was entitled "Avaliação de técnicas de reconciliação em séries hierárquicas de consumo de energia". This work combined hierarchical time series and Bagging (a semi-supervised machine learning algorithm in the context of energy demand in Brazil. In addition, another paper was presented at SBPO 2022, entitled "Previsão de consumo hierárquico brasileiro de energia elétrica por métodos de reconciliação robusta", when the potential of robust forecast reconciliation in the context of energy demand in Brazil was explored.

We concentrated our efforts on forecast reconciliation for point forecasts, which may prove useful on several occasions. However, the evaluation of these point forecasts was made using specific assessment metrics to understand the predictive power of our proposals. These numbers do not give us any information about the uncertainty of the forecast. In this sense, prediction intervals, provides a information of how confident we are in our forecasts and can be useful when we need to make decisions based on the degree of uncertainty involved. A natural extension to the main core of this thesis is to offer prediction intervals for robust-based reconciled forecasts.

Another point worth noting is that the essays presented in this research, focused on cross-sectional hierarchies. As described in the Introduction, forecast reconciliation in a cross-sectional fashion yields coherent forecasts for a given hierarchical structure. The combination of robust reconciliation settings with temporal and cross-temporal aggregations may produce another set of original scientific contributions, derived from this work.

As mentioned in Chapter 4, we focussed our attention on two widely used family of forecasting models to generate the base forecasts, which are further reconciled using specific methods: exponential smoothing and SARIMA formulations. In Chapter 5 we extended the set of methods to generate base forecasts to robust exponential smoothing, providing doubly robust settings. Future studies might benefit, for instance, from more sophisticated approaches, such as Artificial Neural Networks, Support Vector Regressions, Recurrent and Vector Singular Spectrum Analysis, among others, to generate the base forecasts.

We also intend to extend the original research of WICKRAMASURYIA et al. 2019 to other robust covariance estimators such as M-estimators, the Stahel-Donoho estimator, S-estimators and MM-estimators, subject to the unbiaseness condition.

7
References

AELST, S. V.; ROUSSEEUW, P. Minimum volume ellipsoid. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley Online Library, v. 1, n. 1, p. 71–82, 2009.

ATHANASOPOULOS, G.; AHMED, R. A.; HYNDMAN, R. J. Hierarchical forecasts for australian domestic tourism. **International Journal of Forecasting**, Elsevier, v. 25, n. 1, p. 146–166, 2009.

ATHANASOPOULOS, G. et al. Hierarchical forecasting. In: ______. **Macroeconomic Forecasting in the Era of Big Data: Theory and Practice**. Cham: Springer International Publishing, 2020. p. 689–719. ISBN 978-3-030-31150-6. Disponível em: <https://doi.org/10.1007/978-3-030-31150-6_21>.

ATHANASOPOULOS, G. et al. Forecasting with temporal hierarchies. **European Journal of Operational Research**, v. 262, n. 1, p. 60 – 74, 2017. ISSN 0377-2217.

ATHANASOPOULOS, G. et al. The tourism forecasting competition. **International Journal of Forecasting**, v. 27, n. 3, p. 822–844, 2011. ISSN 0169-2070. Special Section 1: Forecasting with Artificial Neural Networks and Computational Intelligence Special Section 2: Tourism Forecasting.

BAGCHI, A.; PAUL, J. A. Youth unemployment and terrorism in the MENAP (middle east, north africa, afghanistan, and pakistan) region. **Socio-Economic Planning Sciences**, Elsevier, v. 64, p. 9–20, 2018.

BARNEA, A.; LAKONISHOK, J. An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. **Decision Sciences**, Wiley Online Library, v. 11, n. 1, p. 17–26, 1980.

BARNICHON, R.; GARDA, P. Forecasting unemployment across countries: The ins and outs. **European Economic Review**, Elsevier, v. 84, p. 165–183, 2016.

BARRODALE, I.; ROBERTS, F. Solution of an overdetermined system of equations in the l 1 norm [f4]. **Communications of the ACM**, ACM New York, NY, USA, v. 17, n. 6, p. 319–320, 1974.

BARROW, D. et al. Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning. **Expert Systems with Applications**, v. 160, p. 113637, 2020. ISSN 0957-4174.

BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. **Computational Statistics & Data Analysis**, Elsevier, v. 120, p. 70–83, 2018.

BOX, G. E.; JENKINS, G. M. **Time series analysis: Forecasting and control**. [S.l.]: Holden-Day San Francisco, 1970. 498 p.

BOX, G. E. P.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. San Francisco, CA, USA: Holden-Day, Inc., 1970. ISBN 0816211043.

BROWN, R. G. **Statistical forecasting for inventory control**. [S.l.]: McGraw/Hill, 1959.

CASTELLI, M.; VANNESCHI, L.; De Felice, M. Forecasting short-term electricity consumption using a semantics-based genetic programming framework: The south italy case. **Energy Economics**, v. 47, p. 37–41, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0140988314002539>.

COX, D. R. Prediction by exponentially weighted moving averages and related methods. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 23, n. 2, p. 414–422, 1961.

CREVITS, R.; CROUX, C. Forecasting with robust exponential smoothing with damped trend and seasonal components. **SSRN: KBI_1741 https://dx. doi. org/10.2139/ssrn**, v. 3068634, 2016.

CREVITS, R.; CROUX, C. Forecasting using robust exponential smoothing with damped trend and seasonal components. **KBI_1741**, 2017.

CROUX, C.; GELPER, S.; MAHIEU, K. Robust exponential smoothing of multivariate time series. **Computational statistics & data analysis**, Elsevier, v. 54, n. 12, p. 2999–3006, 2010.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine learning research**, v. 7, n. Jan, p. 1–30, 2006.

ENERDATA. **World Energy & Climate Statistics - Total Energy Consumption in 2021**. 2022. <https://yearbook.enerdata.net/total-energy/world-consumption-statistics.html>. Accessed on 21 Aug 2022.

EPE. **Brazilian Energy Expansion Plan 2031**. 2022. <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/plano-decenal-de-expansao-de-energia-pde>. Accessed on 21 Jun 2022.

EPE. **Monthly electric energy consumption by classes (regions and subsystems)**. 2022. <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica>. Accessed on 26 May 2022.

ERVEN, T. v.; CUGLIARI, J. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In: **Modeling and stochastic learning for forecasting in high dimensions**. [S.l.]: Springer, 2015. p. 297–317.

FLIEDNER, E. B.; MABERT, V. A. Constrained forecasting: some implementation guidelines. **Decision Sciences**, Wiley Online Library, v. 23, n. 5, p. 1143–1161, 1992.

FLIEDNER, G. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. **Computers & Operations Research**, Elsevier, v. 26, n. 10-11, p. 1133–1149, 1999.

FOGARTY, D.; BLACKSTONE, J.; HOFFMANN, T. **Production and Inventory Management**. [S.l.]: Cengage Learning, 1991.

GARDNER, E. S.; MCKENZIE, E. Forecasting trends in time series. **Management Science**, INFORMS, v. 31, n. 10, p. 1237–1246, 1985. ISSN 00251909, 15265501. Disponível em: <http://www.jstor.org/stable/2631713>.

GELPER, S.; FRIED, R.; CROUX, C. Robust forecasting with exponential and holt–winters smoothing. **Journal of forecasting**, Wiley Online Library, v. 29, n. 3, p. 285–300, 2010.

GROSS, C. W.; SOHL, J. E. Disaggregation methods to expedite product line forecasting. **Journal of Forecasting**, Wiley Online Library, v. 9, n. 3, p. 233–254, 1990.

GRUNFELD, Y.; GRILICHES, Z. Is aggregation necessarily bad? **The Review of Economics and Statistics**, JSTOR, p. 1–13, 1960.

HOLLYMAN, R.; PETROPOULOS, F.; TIPPING, M. E. Understanding forecast reconciliation. **European Journal of Operational Research**, Elsevier BV, v. 294, n. 1, p. 149–160, oct 2021.

HOLT, C. **Forecasting trends and seasonals by exponentially weighted averages. Carnegie Institute of Technology**. [S.l.], 1957.

HUBER, P. J. Robust estimation of a location parameter. **Ann. Math. Statist.**, The Institute of Mathematical Statistics, v. 35, n. 1, p. 73–101, 03 1964.

HUBERT, M.; DEBRUYNE, M.; ROUSSEEUW, P. J. Minimum covariance determinant and extensions. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley Online Library, v. 10, n. 3, p. e1421, 2018.

HYNDMAN, R. et al. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, p. 439–454, 02 2002.

HYNDMAN, R. et al. **Forecasting with exponential smoothing: the state space approach**. [S.l.]: Springer Science & Business Media, 2008.

HYNDMAN, R. J. et al. Optimal combination forecasts for hierarchical time series. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 9, p. 2579–2589, 2011.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [S.l.]: OTexts, 2018.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3. ed. OTexts: Melbourne, Australia, 2021. Disponível em: <OTexts.com/fpp3>.

HYNDMAN, R. J. et al. forecast: Forecasting functions for time series and linear models. 2020. Disponível em: <https://cran.r-project.org/web/packages/forecast/forecast.pdf>.

HYNDMAN, R. J.; KHANDAKAR, Y. **Automatic Time Series Forecasting: TheforecastPackage forR**. [S.l.]: Foundation for Open Access Statistic, 2008. v. 27.

HYNDMAN, R. J.; LEE, A. J.; WANG, E. Fast computation of reconciled forecasts for hierarchical and grouped time series. **Computational statistics & data analysis**, Elsevier, v. 97, p. 16–32, 2016.

JEON, J.; PANAGIOTELIS, A.; PETROPOULOS, F. Probabilistic forecast reconciliation with applications to wind power and electric load. **European Journal of Operational Research**, Elsevier BV, v. 279, n. 2, p. 364–379, dec 2019. Disponível em: <https://doi.org/10.1016/j.ejor.2019.05.020>.

KABOLI, S. H. A. et al. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. **Energy**, Elsevier BV, v. 126, p. 144–164, may 2017. Disponível em: <https://doi.org/10.1016/j.energy.2017.03.009>.

KARMY, J. P.; MALDONADO, S. Hierarchical time series forecasting via support vector regression in the european travel retail industry. **Expert Systems with Applications**, v. 137, p. 59 – 73, 2019. ISSN 0957-4174.

KINNEY, W. R. Predicting earnings: entity versus subentity data. **Journal of Accounting Research**, JSTOR, p. 127–136, 1971.

KOENKER, R. W.; D'OREY, V. Algorithm as 229: Computing regression quantiles. **Applied statistics**, JSTOR, p. 383–393, 1987.

KOURENTZES, N. tsutils: Time series exploration, modelling and forecasting. **R Package Version 0.9. 0**, 2019.

KOURENTZES, N.; ATHANASOPOULOS, G. Cross-temporal coherent forecasts for australian tourism. **Annals of Tourism Research**, Elsevier, v. 75, p. 393–409, 2019.

LILA, M. F.; MEIRA, E.; OLIVEIRA, F. L. C. Forecasting unemployment in brazil: A robust reconciliation approach using hierarchical data. **Socio-Economic Planning Sciences**, Elsevier, p. 101298, 2022.

LITTERMAN, R. B. et al. The forecasting accuracy of major time series methods. **Journal of the American Statistical Association**, JSTOR, v. 81, n. 393, p. 262, mar 1986.

MANCUSO, P.; PICCIALLI, V.; SUDOSO, A. M. A machine learning approach for forecasting hierarchical time series. 2020. Disponível em: <https://arxiv.org/abs/2006.00630>.

MANCUSO, P.; PICCIALLI, V.; SUDOSO, A. M. A machine learning approach for forecasting hierarchical time series. **Expert Systems with Applications**, Elsevier, v. 182, p. 115102, 2021.

MEIRA, E.; LILA, M. F.; OLIVEIRA, F. L. C. A novel reconciliation approach for hierarchical electricity consumption forecasting based on resistant regression. **Energy**, Elsevier, p. 126794, 2023.

MEIRA, E.; OLIVEIRA, F. L. C.; MENEZES, L. M. D. Point and interval forecasting of electricity supply via pruned ensembles. **Energy**, Elsevier BV, v. 232, p. 121009, 2021.

MEIRA, E.; OLIVEIRA, F. L. C.; MENEZES, L. M. de. Forecasting natural gas consumption using bagging and modified regularization techniques. **Energy Economics**, v. 106, p. 105760, 2022.

NARASIMHAN, S. L.; MCLEAVEY, D. W.; BILLINGTON, P. **Production planning and inventory control**. [S.l.]: Prentice Hall Englewood Cliffs, 1995.

NENOVA, Z. D.; MAY, J. H. Determining an optimal hierarchical forecasting model based on the characteristics of the data set. **Journal of Operations Management**, Elsevier, v. 44, p. 62–68, 2016.

NYSTRUP, P. et al. Dimensionality reduction in forecasting with temporal hierarchies. **International Journal of Forecasting**, Elsevier, v. 37, n. 3, p. 1127–1146, 2021.

OLIVEIRA, E. **Getting the most out of the wisdom of the crowds: Improving fore-casting performance through ensemble methods and variable selection techniques. Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil**. [S.l.]: PUC-Rio, 2020.

OLIVEIRA, E. M. D. et al. Dynamic relationships between crude oil prices and socially responsible investing in brazil: evidence for linear and non-linear causality. **Applied Economics**, Informa UK Limited, v. 49, n. 22, p. 2125–2140, 2017.

OLIVEIRA, E. M. de; OLIVEIRA, F. L. C. Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. **Energy**, Elsevier BV, v. 144, p. 776–788, feb 2018.

OLIVEIRA, F. L. C.; SOUZA, R. C.; MARCATO, A. L. M. A time series model for building scenarios trees applied to stochastic optimisation. **International Journal of Electrical Power &amp Energy Systems**, Elsevier BV, v. 67, p. 315–323, may 2015.

ORCUTT, G. H.; WATTS, H. W.; EDWARDS, J. B. Data aggregation and information loss. **The American Economic Review**, JSTOR, p. 773–787, 1968. Disponível em: <https://www.jstor.org/stable/1815532>.

PANAGIOTELIS, A. et al. Forecast reconciliation: A geometric view with new insights on bias correction. **International Journal of Forecasting**, Elsevier, v. 37, n. 1, p. 343–359, 2021.

PEGELS, C. C. Exponential forecasting: some new variations. **Management Science**, JSTOR, p. 311–315, 1969. Disponível em: <https://www.jstor.org/stable/2628137>.

PETROPOULOS, F. et al. Forecasting: theory and practice. **International Journal of Forecasting**, 2022. ISSN 0169-2070.

PRITULARGA, K. F.; SVETUNKOV, I.; KOURENTZES, N. Stochastic coherency in forecast reconciliation. **International Journal of Production Economics**, Elsevier, v. 240, p. 108221, 2021.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <https://www.R-project.org/>.

ROUSSEEUW, P. et al. Robust monitoring of time series with application to fraud detection. **Econometrics and statistics**, Elsevier, v. 9, p. 108–121, 2019.

ROUSSEEUW, P.; YOHAI, V. Robust regression by means of s-estimators. **Springer Lecture Notes in Statistics**, v. 26, p. 256–272, 06 1984.

ROUSSEEUW, P. J. Multivariate estimation with high breakdown point. **Mathematical statistics and applications**, v. 8, n. 283-297, p. 37, 1985.

ROUSSEEUW, P. J.; DRIESSEN, K. V. A fast algorithm for the minimum covariance determinant estimator. **Technometrics**, Taylor & Francis, v. 41, n. 3, p. 212–223, 1999.

SILVA, F. L. da; OLIVEIRA, F. L. C.; SOUZA, R. C. A bottom-up bayesian extension for long term electricity consumption forecasting. **Energy**, v. 167, p. 198–210, 2019. ISSN 0360-5442.

SMUCLER, E.; YOHAI, V. J. Robust and sparse estimators for linear regression models. **Computational Statistics & Data Analysis**, v. 111, n. C, p. 116–130, 2017.

SPILIOTIS, E. et al. Hierarchical forecast reconciliation with machine learning. 2020. Disponível em: <https://arxiv.org/abs/2006.02043>.

STRATIGAKOS, A. et al. End-to-end learning for hierarchical forecasting of renewable energy production with missing values. 2022.

SUGIURA, N. Further analysts of the data by akaike' s information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Taylor Francis, v. 7, n. 1, p. 13–26, 1978. Disponível em: <https://doi.org/10.1080/03610927808827599>.

SYNTETOS, A. A. et al. Supply chain forecasting: Theory, practice, their gap and the future. **European Journal of Operational Research**, Elsevier, v. 252, n. 1, p. 1–26, 2016.

TAIEB, S. B.; TAYLOR, J. W.; HYNDMAN, R. J. Coherent probabilistic forecasts for hierarchical time series. In: JMLR. ORG. **Proceedings of the 34th International Conference on Machine Learning-Volume 70**. [S.l.], 2017. p. 3348–3357.

TANG, T. et al. GM(1,1) based improved seasonal index model for monthly electricity consumption forecasting. **Energy**, v. 252, p. 124041, 2022. ISSN 0360-5442.

TAYLOR, J. M.; MCSHARRY, P. E. Short-term load forecasting methods: An evaluation based on european data. **IEEE Transactions on Power Systems**, v. 22, n. 4, p. 2213–2219, 2007. Disponível em: <https://doi.org/10.1109/tpwrs.2007.907583>.

TAYLOR, J. W. Exponential smoothing with a damped multiplicative trend. **International Journal of Forecasting**, v. 19, n. 4, p. 715 – 725, 2003. ISSN 0169-2070.

TOBIAS, J.; ZELLNER, A. A note on aggregation, disaggregation and forecasting performance. **Journal of Forecasting**, v. 19, 09 2000.

TORRINI, F. C. et al. Long term electricity consumption forecast in brazil: a fuzzy logic approach. **Socio-Economic Planning Sciences**, Elsevier, v. 54, p. 18–27, 2016.

VEEN, A. van der; EVERS, G. A simultaneous model for regional labor supply, incorporating labor force participation, commuting and migration. **Socio-Economic Planning Sciences**, Elsevier, v. 17, n. 5-6, p. 239–250, 1983.

VELASQUEZ, C. E. et al. Analysis of time series models for brazilian electricity demand forecasting. **Energy**, v. 247, p. 123483, 2022. ISSN 0360-5442.

WANG, L. The l1 penalized lad estimator for high dimensional linear regression. **Journal of Multivariate Analysis**, Elsevier, v. 120, p. 135–151, 2013.

WICKRAMASURIYA, S. L.; ATHANASOPOULOS, G.; HYNDMAN, R. J. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. **Journal of the American Statistical Association**, Taylor & Francis, v. 114, n. 526, p. 804–819, 2019.

WICKRAMASURIYA, S. L.; TURLACH, B. A.; HYNDMAN, R. J. Optimal non-negative forecast reconciliation. **Statistics and Computing**, Springer, v. 30, n. 5, p. 1167–1182, 2020.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management science**, INFORMS, v. 6, n. 3, p. 324–342, 1960.