

ALBERSON DA SILVA MIRANDA

**RECONCILIAÇÃO ÓTIMA PROBABILÍSTICA EM SÉRIES
TEMPORAIS HIERÁRQUICAS**

Projeto de dissertação apresentado ao Programa
de Pós-Graduação em Economia da Universidade
Federal do Espírito Santo.

Orientador: Prof. Dr. Guilherme A. A. Pereira

Vitória

2022

MIRANDA, Alberson da Silva. **Reconciliação Ótima Probabilística em Séries Temporais Hierárquicas**. 2022. 11 folhas. Projeto de dissertação (Programa de Pós-Graduação em Economia) – Universidade Federal do Espírito Santo, Vitória, 2022.

RESUMO

Minha pesquisa visa aplicar o método de estimação por regressão quantílica no problema da reconciliação ótima de previsões de séries temporais hierárquicas ou agrupadas. Além disso, pretendo evidenciar as vantagens e problemas do método o comparando com as abordagens por MQO, MQP e MinT, adotando tanto a mediana quanto quantis extremos. O experimento será realizado nas séries temporais de saldo de crédito do Banco do Estado do Espírito Santo SA, organizadas hierarquicamente por superintendências regionais e agências, e agrupadas por modalidades de crédito selecionadas.

Palavras-chave: Séries temporais hierárquicas. Previsão. Reconciliação. Regressão quantílica.

1 INTRODUÇÃO

Minha pesquisa visa abordar o problema da reconciliação ótima de previsões pontuais de séries temporais hierárquicas e agrupadas através de regressão quantílica. Após o desenvolvimento teórico, espero evidenciar *que vantagens e limitações a abordagem de reconciliação ótima por regressão quantílica proporciona*, aplicando o método no dados de crédito do Banco do Estado do Espírito Santo SA.

1.1 CONTEXTUALIZAÇÃO DA PESQUISA

Séries temporais hierárquicas são aquelas que podem ser agregadas ou desagregadas naturalmente em uma estrutura aninhada (HYNDMAN; ATHANASOPOULOS, 2021). Para ilustrar, tome a série do Pib brasileiro. Ela pode ser desagregada por estado que, por sua vez, pode ser desagregada por município.

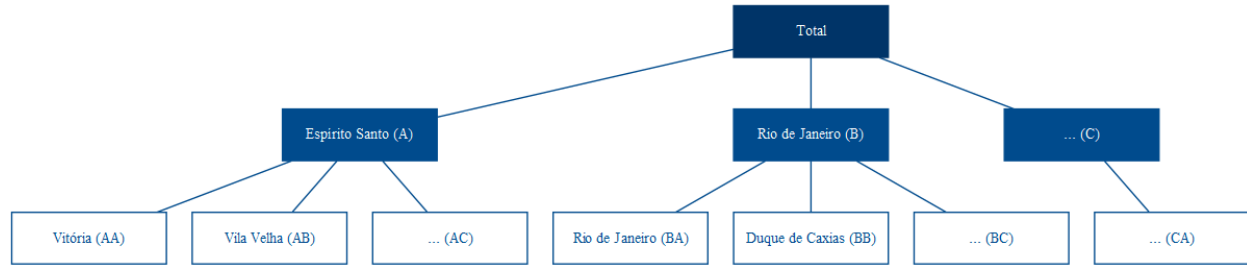


Figura 1 – Séries Hierárquicas

Essa estrutura pode ser representada por equações para qualquer nível de agregação. Assim, o agregado nacional pode ser representado apenas pelos agregados dos estados, através de (1), ou como o agregado dos municípios (2). Já o agregado para o estado do Espírito Santo é representado por (3).

$$y_t = y_{A,t} + y_{B,t} + y_{C,t} \quad (1)$$

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BC,t} + y_{CA,t} \quad (2)$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \quad (3)$$

Por outro lado, o Pib pode ser também desagregado de forma cruzada de acordo com a atividade econômica — lavoura, rebanho, indústria de transformação, extrativa, bens de capital, bens intermediários, comércio de vestuário, automotivos, serviços etc. Essa estrutura não pode ser desagregada naturalmente de uma única forma, como é a hierarquia de estados e municípios. Não pode ser aninhada por um atributo como a própria geografia. A esse tipo de estrutura dá-se o nome de séries agrupadas.

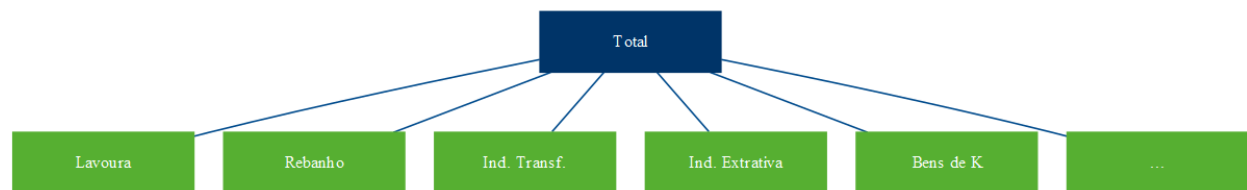


Figura 2 – Séries Agrupadas

Combinando as duas, temos a estrutura de séries hierárquicas agrupadas. Ao contrário da estrutura hierárquica, que só pode ser agregada de uma forma — como com os municípios abaixo dos estados —, a adição da estrutura agrupada pode ocorrer tanto acima quanto abaixo da hierárquica.

Na notação matricial, essas estruturas são representadas pelo vetor \mathbf{y}_t n -dimensional com todas as observações no tempo t para todos os níveis da hierarquia, pela matriz de soma \mathbf{S} de dimensão $n \times m$ que define as equações para todo nível de agregação e pela matriz \mathbf{b}_t composta pelas séries no nível mais desagregado.

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t \quad (4)$$

Talvez as formas mais intuitivas de se pensar em previsões para esses tipos de estrutura sejam as abordagens *top-down* e *bottom-up*. Na abordagem *top-down*, a previsão para os níveis mais desagregados da hierarquia são determinadas por uma proporção p_i do nível agregado. Para isso, temos de definir uma matriz com todos esses pesos, que, seguindo a formulação de [Hyndman e Athanasopoulos \(2021\)](#), chamo de \mathbf{G} . Já a abordagem *bottom-up* parte do raciocínio inverso e define as previsões de cada elemento da estrutura a partir das previsões dos elementos mais desagregados $\hat{\mathbf{y}}_t$. Para tanto, basta modificar a matriz \mathbf{G} . Então, \mathbf{G} define a abordagem — se *top-down* ou *bottom-up* —, e \mathbf{S} define a maneira da qual as previsões são somadas para formar as equações de previsão para cada elemento da estrutura.

Seja somando as previsões do nível mais desagregado para formar os níveis superiores da hierarquia (*bottom-up*) ou distribuindo proporcionalmente as previsões do nível mais agregado (*top-down*), o vetor $\tilde{\mathbf{y}}_t$ representa as previsões *coerentes*. Isso significa que as previsões “batem”, ou seja, são totalizadas corretamente — as previsões de cada elemento agregado corresponde ao somatório das previsões dos níveis inferiores da hierarquia. Isso é garantido pela multiplicação das matrizes $\mathbf{S}\mathbf{G}$.

Não fosse essa pré multiplicação, nada garantiria a coerência das previsões. Tomando a estrutura da figura 1 como exemplo, seria um acaso improvável que as previsões do agregado para o estado do

Espírito Santo sejam exatamente a soma das previsões individuais de seus municípios. Isso porque cada série pode seguir um processo diferente (e.g., arima) com erros e variâncias distintas.

Os métodos de gerar previsões coerentes a partir de previsões base são chamados de métodos de *reconciliação*. Os métodos de reconciliação tradicionais apresentados, *top-down* e *bottom-up*, utilizam informação limitada. No método *top-down*, utiliza-se apenas informações do nível mais agregado, enquanto na abordagem *bottom-up* utiliza-se apenas as informações dos níveis mais desagregados.

Alternativamente, podemos pensar numa matriz G qualquer que utilize toda a informação disponível e tenha algumas propriedades que garantam que as previsões coerentes tenham a menor diferença o possível em relação às previsões base. Esse é o problema de pesquisa trabalhado na *reconciliação ótima*.

1.2 JUSTIFICATIVA

O estado da arte em reconciliação ótima de previsões pontuais se encontra no método de MinT. Esse método, assim como toda estimação de média de parâmetros, está sujeita à interferência de *outliers* extremos que, no caso de previsão de séries hierárquicas, se apresentam na forma de elementos da hierarquia muito maiores que os demais. Isso pode significar em menor peso atribuído à elementos menores da hierarquia e, talvez, a estimação pela mediana possua alguma vantagem em determinadas situações. Além disso, pode ser do interesse do pesquisador a estimação em eventos extremos, o que exigiria usar quantis muito altos ou muito baixos. Nesse sentido, a estimação por regressão quantílica pode ser uma adição à literatura do tema.

2 OBJETIVOS

Meu objetivo geral para a dissertação é estudar o problema da reconciliação ótima de previsões pontuais a partir de regressões quantílicas.

Como objetivos específicos, tenho:

1. Estudar métodos para estimação da matriz G para qualquer quantil desejado, e não apenas a média;
2. Identificar possíveis vantagens e limitações da abordagem quantílica na reconciliação de previsões pontuais a partir de aplicação do método estudado na previsão de saldos de crédito do Banco do Estado do Espírito Santo SA.

Não está no escopo do estudo os métodos de se obter previsões probabilísticas, ou seja, a estimação da matriz \hat{y}_{T+h} para além de previsões pontuais.

3 REVISÃO DA LITERATURA

Previsões pontuais de séries temporais hierárquicas não é um assunto novo. Desde a década de 70, ao menos, pesquisas foram publicadas acerca de abordagens *bottom-up* e *top-down*, suas vantagens e desvantagens, e tentativas de se definir qual é o melhor método¹. Entretanto, é apenas em Hyndman, Ahmed et al. (2011) que é formalizada uma abordagem prática que utiliza toda a informação disponível, (i.e. as previsões de todos elementos de todos os níveis da hierarquia) a partir da estimação da matriz G via regressão linear por MQG.

Entretanto, para ser capaz de estimar o modelo por MQG, é necessária a matriz de variância-covariância dos erros. Hyndman, Ahmed et al. (2011) usam a matriz de erros de coerência, ou seja, a diferença entre as previsões reconciliadas e as previsões base, que tem posto incompleto e não identificada e, portanto, não pode ser estimada. Os autores contornam esse problema adotando no lugar da matriz de variância-covariância dos erros uma matriz diagonal constante, ou seja, assumem variância constante dos erros de reconciliação, e estimam a matriz G por MQO.

A estimação por esse método resulta numa reconciliação ótima que depende apenas da matriz S , ou seja, da estrutura hierárquica, e independe da variância e covariância das previsões base \hat{y}_{T+h} — o que não é uma conclusão satisfatória.

Hyndman, Lee e Wang (2016) tentam aperfeiçoar o método usando as variâncias das previsões base estimadas (dentro da amostra) como estimativa para a matriz de variância-covariância dos erros de

¹Uma revisão dessa literatura pode ser encontrada em Athanasopoulos, Ahmed e Hyndman (2009).

reconciliação, de forma a as utilizar como pesos e realizar a reconciliação ótima por MQP. Assim, previsões base mais acuradas têm peso maior do que as mais ruidosas. Entretanto, não fornecem justificativa teórica para usar a diagonal da matriz de variância-covariância de \hat{e}_t .

Wickramasuriya, Athanasopoulos e Hyndman (2019) argumentam que o que de fato interessa é que as previsões reconciliadas tenham o menor erro. Então, corrigem a abordagem de reconciliação ótima para o objetivo de minimização dos erros das previsões reconciliadas \tilde{y}_{t+h} , ao invés dos erros das previsões base \hat{y}_{t+h} . Dado que isso implica na minimização da variância de \tilde{e}_{t+h} , ou seja, na minimização do somatório da diagonal, o traço, da matriz de variância-covariância de \tilde{e}_{t+h} , eles chamaram esse método de MinT (*Minimum Trace*). Paralelamente, usam desigualdade triangular para demonstrar que as previsões reconciliadas obtidas por esse método são ao menos tão boas quanto as previsões base.

Panagiotelis et al. (2021) reinterpreta a literatura de coerência e reconciliação de previsões pontuais a partir de uma abordagem geométrica, trazendo provas alternativas para conclusões anteriores ao mesmo tempo em que fornece novos teoremas. Além disso, Panagiotelis et al. (2021) então estende essa interpretação geométrica para o contexto probabilístico, fornecendo métodos paramétricos e não paramétricos (via *bootstrapping*) para reconciliação de previsões probabilísticas, ou seja, para reconciliar previsões \hat{y}_t obtidas a partir de toda a distribuição, e não apenas a média.

4 METODOLOGIA

Para atingir os objetivos, a primeira etapa da pesquisa consistirá na pesquisa bibliográfica relacionada à reconciliação ótima de previsões de séries temporais hierárquicas e agrupadas e à regressão quantílica.

Utilizando a base de dados do Google Acadêmico, pesquisei pelas palavras-chave “*hierarchical forecast reconciliation*” para qualquer lugar no corpo do texto, encontrando 27.600 resultados. Utilizando um script para ordenar os resultados pelo número de citações², verifiquei que o trabalho mais citado é Hyndman e Athanasopoulos (2021).

²<https://github.com/WittmannF/sort-google-scholar>

Tabela 1 – Trabalhos mais citados com os termos “hierarchical forecast reconciliation”

Autor	Citações	Ano
Hyndman, G Athanasopoulos	5222	2018
Dellarocas, X Zhang	2082	2007
Hyndman, A Lee, E Wang, S Wickramasuriya	1023	2013
Badre, M D’esposito	974	2009
Hong, S Fan	912	2016
Tashman	798	2000

Utilizando essa obra como texto base, obtive os textos referenciados no capítulo 11 “*Forecasting Hierarchical and Grouped Time-Series*”, subcapítulo 3 “*Forecast Reconciliation*”, além de [Hyndman, Lee e Wang \(2016\)](#), onde o método por MPQ foi desenvolvido porém não está citado nas referências do capítulo.

Tabela 2 – Artigos de referência em Hyndman e Athanasopoulos (2021)

Autor	Ano
Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L.	2011
Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J.	2021
Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J.	2019
Rob J. Hyndman and Alan J. Lee and Earo Wang	2016

Adicionando o termo “*quantile*” e refinando a pesquisa para encontrar as palavras chave no título dos trabalhos, nenhum resultado foi encontrado. Modificando os termos para encontrar palavras semelhantes, a busca por “*allintitle: hierarchical OR hierarchically probabilistic OR quantile reconciliation OR reconciled*” retornou 5 títulos, dentre os quais dois são artigos de acesso livre e, portanto, foram adicionadas à pesquisa bibliográfica a ser realizada.

Tabela 3 – Trabalhados encontrados na busca estendida

Autor	Ano
Han, X., Dasgupta, S., & Ghosh, J.	2021
Roach, C.	2019

A segunda etapa do trabalho consistirá na aplicação e avaliação dos métodos estimados nas séries de saldo de empréstimos e financiamentos do Banco do Estado do Espírito Santo SA. A coleta dos dados para aplicação do método será realizada diretamente do sistema de planejamento orçamentário do Banestes. A importação, limpeza e tratamento dos dados será realizado utilizando o *software* R (R CORE TEAM, 2022) e o pacote {bees} (MIRANDA, 2022), que é um pacote privado que desenvolvi para esses fins. A disponibilização em repositório aberto, tanto da base de dados quanto do pacote para importação e tratamento deles, para fins de reprodutibilidade será proposta à instituição.

Para a comparação com os métodos MQO, MQP e MinT, os dados serão organizados na estrutura hierárquica agrupada utilizando o pacote {tsibble} (WANG et al., 2022), as previsões base realizadas com o pacote {fable} (O'HARA-WILD; HYNDMAN; WANG, 2022) e as reconciliações via MQO, MQP e MinT para comparação dos resultados serão realizadas com o pacote {fabletools} (O'HARA-WILD; HYNDMAN; WANG, 2021). Já a reconciliação via regressão quantílica será realizada com o pacote {quantreg} (KOENKER, 2022). Inicialmente, a métrica para avaliação será a raiz do erro quadrático médio, *RMSE* na sigla em inglês, para os erros de previsão das séries reconciliadas.

A estrutura hierárquica terá 3 níveis — total, superintendências regionais (norte, sul, centro e empresarial) e agências —, e a estrutura agrupada será formada por modalidades de crédito selecionadas, sendo algumas mais ruidosas que outras³ de forma a evidenciar vantagens e desvantagens em cada caso.

5 PLANO DE TRABALHO

Para a introdução, o planejado é detalhar os conceitos de séries temporais hierárquicas e agrupadas, coerência e reconciliação ótima na notação matricial. Além disso, devo introduzir o conceito envolvendo a matriz G e as abordagens de reconciliação clássicas (*bottom-up* e *top-down*), também na notação matricial. Paralelamente, já iniciarei a escolha das séries a serem trabalhadas e a formalização do uso e publicação da base de dados.

³Por exemplo, espera-se que a série de crédito imobiliário seja menos ruidosa que a série de crédito pessoal.

No primeiro capítulo, devo detalhar as abordagens de reconciliação ótima para previsões pontuais nos três métodos MQO, MQP e MinT, demonstrando a estimação, hipóteses, vantagens e limitações.

No segundo capítulo, vou trazer os conceitos da regressão quantílica e aplicá-los ao problema da reconciliação ótima, que é o objetivo principal do trabalho. Trabalharei tanto com a mediana quanto com quantis extremos.

No capítulo 3, aplicarei os métodos estudados aos dados de crédito do Banestes, comparando as performances entre eles enquanto analiso onde a reconciliação quantílica se mostra mais ou menos vantajosa.

Tabela 4 – Cronograma das atividades

Atividade	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1. Introdução	x											
Capítulo 1: Reconciliação OLS, WLS e MinT	x	x	x	x								
Capítulo 2: Regressão quantílica				x	x							
Capítulo 2: Reconciliação quantílica					x	x	x	x				
Capítulo 3: Breve exploratória dos dados								x				
Capítulo 3: Estimação e avaliação das abordagens								x				
Capítulo 3: Análise das vantagens e problemas do método								x	x			
Escolha das séries		x										
Obter autorizações do Banestes	x											
Qualificação						x						

REFERÊNCIAS

- ATHANASOPOULOS, George; AHMED, Roman A.; HYNDMAN, Rob J. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, v. 25, n. 1, p. 146–166, 2009. ISSN 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2008.07.004>.
- HAN, Xing; DASGUPTA, Sambarta; GHOSH, Joydeep. Simultaneously reconciled quantile forecasting of hierarchically related time series. In: PMLR. INTERNATIONAL Conference on Artificial Intelligence and Statistics. [S.l.: s.n.], 2021. P. 190–198.
- HYNDMAN, R.J.; AHMED, R.A. et al. Optimal Combination Forecasts for Hierarchical Time Series. *Computational Statistics and Data Analysis*, v. 55, p. 2579–2589, 2011.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. 3. ed. Melbourne, Australia, 2021. Acessado em 14/09/20. Disponível em: <<https://otexts.com/fpp3/>>.
- HYNDMAN, Rob J.; LEE, Alan J.; WANG, Earo. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, v. 97, p. 16–32, 2016. ISSN 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2015.11.007>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016794731500290X>>.
- KOENKER, Roger. *quantreg: Quantile Regression*. [S.l.], 2022. R package version 5.94. Disponível em: <<https://www.r-project.org>>.
- MIRANDA, Alberson. *bees: Helper R Functions to Banestes*. [S.l.], 2022. R package version 0.0.0.9005. Disponível em: <<https://github.com/albersonmiranda/bees>>.
- O’HARA-WILD, Mitchell; HYNDMAN, Rob; WANG, Earo. *fable: Forecasting Models for Tidy Time Series*. [S.l.], 2022. R package version 0.3.2. Disponível em: <<https://CRAN.R-project.org/package=fable>>.
- _____. *fabletools: Core Tools for Packages in the fable Framework*. [S.l.], 2021. R package version 0.3.2. Disponível em: <<https://CRAN.R-project.org/package=fabletools>>.

- PANAGIOTELIS, Anastasios et al. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, v. 37, n. 1, p. 343–359, 2021. ISSN 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2020.06.004>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207020300911>>.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>.
- ROACH, C. Reconciled boosted models for GEFCom2017 hierarchical probabilistic load forecasting. *International Journal of Forecasting*, v. 35, n. 4, p. 1439–1450, 2019.
- WANG, Earo et al. *tsibble: Tidy Temporal Data Frames and Tools*. [S.l.], 2022. R package version 1.1.3. Disponível em: <<https://tsibble.tidyverts.org>>.
- WICKRAMASURIYA, Shanika L.; ATHANASOPOULOS, George; HYNDMAN, Rob 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. DOI: [10.1080/01621459.2018.1448825](https://doi.org/10.1080/01621459.2018.1448825). eprint: <https://doi.org/10.1080/01621459.2018.1448825>. Disponível em: <<https://doi.org/10.1080/01621459.2018.1448825>>.