

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES
TEMPORAIS HIERÁRQUICAS E AGRUPADAS

VITÓRIA

2023

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS
HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia.

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

VITÓRIA

2023

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA RECONCILIAÇÃO ÓTIMA DE
SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS/ ALBERSON DA SILVA MI-
RANDA. – VITÓRIA, 2023-

60p. : il. (algumas color.) ; 30 cm.

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

Dissertação (Mestrado) – UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA, 2023.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Universidade
xxx. III. Faculdade de xxx. IV. Título

ALBERSON DA SILVA MIRANDA

MÉTODOS DE MACHINE LEARNING PARA
RECONCILIAÇÃO ÓTIMA DE SÉRIES TEMPORAIS
HIERÁRQUICAS E AGRUPADAS

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Economia.

Aprovada em xx de xx de 20xx.

COMISSÃO EXAMINADORA

Prof. Dr. Guilherme A. A. Pereira
Universidade Federal do Espírito Santo
Orientador

Professor
Instituição

Professor
Instituição

RESUMO

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Palavras-chave: palavra-chave1. palavra-chave2. palavra-chave3.

ABSTRACT

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Keywords: keyword1. keyword2. keyword3.

LISTA DE FIGURAS

Figura 1 – Séries Hierárquicas	29
Figura 2 – Séries Agrupadas	30
Figura 3 – Séries Hierárquicas Agrupadas (a)	31
Figura 4 – Séries Hierárquicas Agrupadas (b)	31
Figura 5 – Validação k -fold aleatória	40
Figura 6 – Validação k -fold com origem móvel	40
Figura 7 – Validação k -fold não-dependente	41
Figura 8 – Série temporal do agregado de crédito do Banestes no ES	44
Figura 9 – Série temporal do agregado de crédito do Banestes por mesorregião do ES	44
Figura 10 – Série temporal do agregado de crédito do Banestes por microrregião do ES	45
Figura 11 – Verbetes no agregado do ES	45
Figura 12 – Verbetes por mesorregião do ES	46
Figura 13 – Modelo de dados	60

LISTA DE QUADROS

Quadro 1 – Variáveis do dataset	42
---	----

LISTA DE TABELAS

Tabela 1 – Estrutura do dataset	42
Tabela 2 – Microrregiões por mesorregião	42
Tabela 3 – Municípios por microrregião	43
Tabela 4 – Contagem de únicos no dataset ESTBAN	43

LISTA DE ABREVIATURAS E SIGLAS

MinT	<i>Minimum Trace</i>
MCRL	Modelo Clássico de Regressão Linear
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados
MQGF	Mínimos Quadrados Generalizados Factíveis
ANN	<i>Artificial Neural Network</i>
SVR	<i>Support Vector Regression</i>
SFN	Sistema Financeiro Nacional
Favar	<i>Factor Augmented Vector Autoregression</i>
Lasso	<i>Least Absolute Shrinkage and Selection Operator</i>

LISTA DE SÍMBOLOS

t	Tempo dentro da amostra
T	Último tempo dentro da amostra, quantidade de observações numa série
h	Horizonte de previsão, tempo fora da amostra
Ω	Conjunto de dados dentro da amostra
y	Série temporal dentro da amostra
\hat{y}	Série temporal estimada
\tilde{y}	Série temporal reconciliada
n	Número de séries na hierarquia
m	Número de séries no menor nível da hierarquia
k	Número de níveis na hierarquia
\mathbf{S}	Matriz de soma
\mathbf{G}	Matriz de reconciliação
$\{\dots\}$	Conjunto
$ \{\dots\} $	Cardinalidade de um conjunto

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Previsão de saldos de crédito de instituições financeiras	21
1.2	Previsão de séries temporais hierárquicas	22
1.3	Objetivos	24
1.4	Outline da dissertação	25
2	REVISÃO DE LITERATURA	25
2.1	Previsão de saldos de crédito de instituições financeiras	25
2.2	Previsão de séries temporais hierárquicas e agrupadas	26
2.2.1	Abordagens de nível único	26
2.2.2	Reconciliação ótima de séries temporais hierárquicas	27
3	MÉTODOS PARA RECONCILIAÇÃO DE SÉRIES TEMPORAIS HI- ERÁRQUICAS E AGRUPADAS	29
3.1	Notação algébrica	29
3.2	Abordagens top-down, bottom-up e middle-out	31
3.3	Coerência e reconciliação	35
3.4	Métodos analíticos de reconciliação ótima	36
3.5	Métodos de reconciliação ótima baseados em aprendizado de máquina	37
3.5.1	O método elastic net	37
3.5.2	O processo de ajuste e sobreajuste	37
3.5.3	Reamostragem	38
3.5.4	Validação cruzada k -fold em séries temporais	39
4	METODOLOGIA	41
4.1	Dados e variáveis	41
4.2	Análise exploratória dos dados	42
4.3	Previsões base	45
4.4	Design do experimento	46
4.4.1	Métodos analíticos de reconciliação	46
4.4.2	Métodos de reconciliação ótima baseados em aprendizado de máquina	46
4.4.3	Espaço de busca dos hiperparâmetros	47
5	RESULTADOS	47
	REFERÊNCIAS	49

APÊNDICES	53
APÊNDICE A – DEMONSTRAÇÕES	55
ANEXOS	57
ANEXO A – CÓDIGO PARA CONSTRUÇÃO DA BASE DE DADOS	59

1 INTRODUÇÃO

1.1 Previsão de saldos de crédito de instituições financeiras

Embora no séc. XX ainda houvesse espaço para uma gestão guiada apenas por instinto (WALLANDER, 1999), atualmente é impensável um banco não realizar previsões de seus resultados e comunicar suas expectativas ao mercado. Nesse documento, ou *guidance*, a projeção da carteira de crédito — o total de empréstimos e financiamentos, dentre outros itens — é frequentemente a primeira informação fornecida, uma vez que é um dos principais elementos para o planejamento dos bancos comerciais. Juntamente com as projeções de depósitos, provisões para créditos de liquidação duvidosa, eficiência operacional, entre outros indicadores-chave, essas projeções determinam a temperatura das expectativas da instituição em relação a elementos cruciais como rentabilidade, dividendos e posição no mercado (*market-share*), e isso é essencial para os acionistas e investidores. Essas projeções precisam ser tão precisas quanto possível para que se possa calcular o risco de transacionar com a instituição financeira, seja como investidor ou cliente.

Ainda que não existam penalidades específicas para instituições financeiras que erram (por uma boa margem) em suas projeções, elas podem sofrer consequências negativas em outros aspectos, como na avaliação de seus desempenhos por parte dos investidores e clientes. Os investidores e clientes podem considerar as projeções equivocadas como um sinal de falta de competência ou confiança na instituição financeira, o que pode afetar negativamente a reputação e a imagem da instituição. Isso pode levar a uma redução no número de investimentos e depósitos, o que irá afetar diretamente sua saúde financeira.

Além disso, nos casos em que algum grupo se sentir lesado, as instituições financeiras podem enfrentar ações judiciais se suas projeções forem consideradas enganosas ou fraudulentas. Por exemplo, se uma instituição financeira fizer projeções excessivamente otimistas para incentivar os investidores a comprar seus títulos e, posteriormente, as projeções se mostrarem incorretas, ela pode ser acusada de fraude¹ ou, ao menos, gestão temerária² — ambos caracterizados como crime contra o Sistema Financeiro Nacional.

Não são raros os casos em que bancos manipulam seus demonstrativos para se apresentarem mais saudáveis do que realmente o são para atrair clientes e investidores. Em 2001, o Banco Santos era classificado como o oitavo maior banco brasileiro. Em 2004 se encontrava sob intervenção do Bacen: ao mesmo tempo em que continuava a expandir seus negócios, o banco escondia um rombo de cerca de R\$ 2,2 bi em créditos que já haviam sido liquidado mas eram mantidos no balanço (MOURA, 2007). De forma semelhante, em 2009, o Banco PanAmericano

¹ Art. 3º: Divulgar informação falsa ou prejudicialmente incompleta sobre instituição financeira. Pena: Reclusão, de 2 (dois) a 6 (seis) anos, e multa. Art. 4º: Gerir fraudulentamente instituição financeira. Pena: Reclusão, de 3 (três) a 12 (doze) anos, e multa (BRASIL, 1986).

² Art. 4º, parágrafo único: Se a gestão é temerária: Pena: Reclusão, de 2 (dois) a 8 (oito) anos, e multa (BRASIL, 1986).

publicou em suas demonstrações contábeis carteira de crédito de R\$ 9,9 bi. Entretanto, cerca de 25% já havia sido liquidado e era mantido no balanço artificialmente como objetivo de evitar a redução da atratividade do banco no mercado (COELHO et al., 2015).

Por isso, é importante que as instituições financeiras sejam transparentes e precisas em suas projeções, fornecendo informações confiáveis e atualizadas para seus clientes e investidores. No entanto, há também motivações estratégicas para essa atividade. Beccalli et al. (2015) mostraram que, em uma amostra de 55 bancos europeus, a utilização de *guidance* está associada a um aumento de 15% na probabilidade do banco atingir ou superar as expectativas de mercado. Isso, por sua vez, está associado a um incremento de até 5% no retorno por ação em relação aos bancos que não alcançaram ou superaram as expectativas.

1.2 Previsão de séries temporais hierárquicas

No que concerne a elaboração dessas previsões, os bancos, assim como em diversas outras indústrias, se enquadram em uma categoria de negócio que requerem previsões de múltiplas séries temporais correlacionadas que são resultados de agregação. Por exemplo, o total de empréstimos de uma instituição financeira corresponde ao agregado dos empréstimos de cada uma de suas agências; o total de vendas de uma rede nacional de farmácias corresponde ao agregado de vendas de suas unidades em cada estado; o total da produção de uma petrolífrica multinacional corresponde ao total produzido em cada país por cada uma de suas plataformas. A essas estruturas naturais de agregação dá-se o nome de *séries temporais hierárquicas*.

Pode-se realizar previsões individualmente para todos os níveis da estrutura. No caso de uma instituição financeira, isso significa realizar previsões, por exemplo, para cada agência, para o agregado de cada região (ou superintendência regional) e para o agregado da instituição. Infelizmente, não há qualquer razão, exceto para métodos de previsão muito simples, para que essas previsões sejam *coerentes* (i.e. que a soma das previsões individuais seja igual à previsão do agregado). Além disso, realizar as previsões individualmente ignoraria os relacionamentos existentes entre as séries temporais na estrutura. Para fazer com que essas previsões se tornem coerentes entre si é que foram desenvolvidos os chamados métodos de *reconciliação*, sendo os mais simples o *top-down*, *bottom-up* e uma combinação das duas, a *middle-out*.

A prática usual em *budgeting*³, principalmente para empresas com muitas filiais, é a *top-down*, ou seja, realizar previsões para o total agregado e então distribuí-las para cada unidade seguindo alguma lógica proporcional. No caso dos bancos de varejo, com muitas agências espalhadas pelo território, especialmente em um país grande como o Brasil, esse método pode ser muito prático.

Esse é o caso do Banestes. Com 96 agências distribuídas pelos 78 municípios capixabas,

³ O orçamento é um documento no qual é definido o planejamento financeiro de um empresa, geralmente para o ano seguinte, estabelecendo metas e objetivos. Nele são projetadas as expectativas da empresa e é base de comparação para saber como os resultados estão se desviando da performance esperada.

realizar o *budgeting* para R\$ 5,5 bi de faturamento⁴ não é uma tarefa trivial. Além de uma estrutura hierárquica de alta dimensionalidade por conta da quantidade de agências, se tratando de um banco múltiplo⁵ que opera com diversas carteiras, as n modalidades de crédito⁶ expandem a estrutura para um total de $n \times 96$ séries temporais a serem estimadas.

Dada tal complexidade, a abordagem *top-down* se coloca como uma opção viável em termos de tempo de processamento e análise. No entanto, conforme descemos na hierarquia, menos precisa ela se torna e, além disso, as características individuais das séries temporais do menor nível hierárquicos são ignoradas. Isso significa que, se no agregado a previsão para uma carteira de crédito for de crescimento de 10%, todas as 96 agências devem seguir a mesma estimativa, divergindo apenas na proporção de participação de cada uma no total.

Tomando o caminho inverso, a abordagem *bottom-up* consiste em realizar previsões para cada série temporal individualmente e, então, agregá-las para obter a previsão para o total. Essa abordagem pode ser mais precisa, pois leva em consideração as características individuais de cada série temporal do nível mais desagregado. No entanto, ela é mais custosa em termos de tempo de processamento e análise. Nesse sentido, cabe ao analista avaliar o *trade-off* entre os ganhos de precisão percebidos com a geração de previsões individuais e a economia de tempo e processamento em realizar o contrário (GROSS; SOHL, 1990).

Além disso, ambas são abordagens de nível único, isto é, são realizadas as previsões para um único nível e então os demais níveis são obtidos agregando ou desagregando. O problema com esses tipos de abordagem é que elas utilizam informação incompleta (HYNDMAN; ATHANASOPOULOS, 2021). Por exemplo, suponha-se que se escolha estimar modelos para cada uma das 96 agências e agregá-las (*bottom-up*). Nesse caso, ignora-se a influência que os níveis mais agregados — aqui a carteira de crédito da região ou de todo o estado — pode ter na estimação do saldo de crédito de cada agência. Por outro lado, se escolher estimar modelos para os níveis mais agregados (*top-down*), ignora-se a informação individual de cada agência.

Uma terceira possibilidade é a *reconciliação ótima*. Ela é uma abordagem que busca resolver esse problema e consiste em realizar previsões para todos os níveis hierárquicos e, então, estimar um modelo para reescrever as previsões do nível mais desagregado como uma combinação linear de todos os elementos da hierarquia. Obtidas as novas previsões no menor nível, ela são então agregadas, gerando previsões coerentes nos níveis superiores. Dessa forma, a informação de todos os níveis é utilizada na estimação dos modelos e na geração das previsões, ao mesmo tempo em que a variância do erro de previsão é minimizado (HYNDMAN; AHMED et al., 2011).

⁴ Conforme demonstrativos publicados referentes ao exercício de 2022 (BANCO DO ESTADO DO ESPÍRITO SANTO, 2022).

⁵ Para ser classificado como banco múltiplo, a instituição financeira deve operar com, no mínimo, duas carteiras dentre: comercial; investimento ou desenvolvimento; crédito imobiliário; de crédito, financiamento e investimento, e; arrendamento mercantil (CONSELHO MONETÁRIO NACIONAL, 1994).

⁶ Crédito consignado, rural, imobiliário, pessoal, capital de giro, desconto de títulos etc.

Atualmente, os métodos analíticos baseados na minimização do traço da matriz da variância-covariância dos erros, desenvolvidos em [Wickramasuriya, Athanasopoulos e Hyndman \(2019\)](#), são os mais populares na literatura da reconciliação ótima. Esses métodos divergem apenas na forma da qual se dá o relacionamento entre os diferentes elementos da hierarquia: se os erros de previsão são descorrelacionados e equivariantes ao longo de toda estrutura (MQO), o que é impossível em séries temporais hierárquicas; se os erros são descorrelacionados e equivariantes apenas dentro do mesmo nível hierárquico (MQP estrutural); se os erros são descorrelacionados mas ponderados pela variância da série (MQP); ou se são correlacionados e variantes ao longo de toda a estrutura (estimadores *MinT Sample* e *MinT Shrink*).

Entretanto, tais métodos são sujeitos a uma série de restrições, como as do MCLR, e têm sua capacidade preditiva reduzida quando suas hipóteses são violadas. Em previsões de séries temporais, o objetivo na maioria dos casos é prever valores futuros com a maior acurácia possível. Em vista disso, métodos de *machine learning* são mais gerais, no sentido de permitir parâmetros não lineares e poderem aproximar virtualmente qualquer função. Além disso, são focados na capacidade preditiva, muitas vezes em detrimento da explicativa. Espera-se, portanto, que esses métodos alcancem melhor performance no problema da reconciliação ótima, justificando a pesquisa e atenção ao tema. Nesse sentido, trabalhos como [Spiliotis et al. \(2021\)](#) desenvolvem métodos de reconciliação ótima de séries temporais utilizando algoritmos de *machine learning*, especificamente o *XGBoost* e *Random Forest*, obtendo resultados superiores aos métodos analíticos tradicionais, especialmente quando as séries não possuem características semelhantes.

Tomando como ponto de partida as conclusões de [Spiliotis et al. \(2021\)](#), este trabalho busca continuar e estender a exploração de métodos de *machine learning* para a tarefa de reconciliação ótima, especificamente o *elastic net*, verificando se sua performance se mantém superior à dos métodos analíticos em um contexto de séries temporais financeiras de alta dimensionalidade.

1.3 Objetivos

O objetivo geral da dissertação é estudar o problema da reconciliação ótima de previsões pontuais a partir de métodos de *machine learning*.

Os objetivos específicos são:

1. Aplicar o método *elastic net* para reconciliação ótima de previsões pontuais de séries temporais hierárquicas e agrupadas e atestar sua validade em dados de alta dimensionalidade;
2. Comparar a performance, em termos de funções de custo e tempo de processamento, do método *elastic net* nas configurações $\alpha = 0$ (*lasso*), $\alpha = 1$ (*ridge*) e uma combinação de ambos obtida via reamostragem;

3. Comparar a performance, em termos de funções de custo e tempo de processamento, do método *elastic net* com os métodos analíticos *bottom-up*, *top-down* e *MinT Shrink*;
4. Comparar a performance, em termos de funções de custo e tempo de processamento, das estratégias de reamostragem validação cruzada *k-fold* e *holdout* para o método *elastic net*;
5. Verificar se a aplicação de regressão regularizada resulta em algum padrão reconhecível na estrutura (e.g., se a regularização tende a favorecer os coeficientes de séries temporais que compartilham o mesmo nó pai ou, em outras palavras, se tende a zerar os coeficientes de séries temporais que não compartilham o mesmo nó pai na hierarquia);

1.4 Outline da dissertação

[Escrever *outline* da dissertação]

2 REVISÃO DE LITERATURA

2.1 Previsão de saldos de crédito de instituições financeiras

A nível macroeconômico, a previsão do agregado de crédito das instituições financeiras é uma preocupação de bancos centrais ao redor do mundo. No Brasil, [Bader, Koyama, Sérgio Mikio e Tsuchida, Marcos Hiroyuki \(2014\)](#) aprimoram o método FAVAR (*Factor Augmented Vector Autoregression*) com uma etapa de análise de correlação canônica para identificar as melhores, em termos de correlação com as variáveis de crédito do SFN, combinações lineares de componentes principais. Esse método, que chamaram de FAVAR canônico, alcançou resultado superior aos FAVAR em 1 e 2 estágios na previsão das variáveis de crédito utilizadas, que foram: a concessão de crédito total com recursos livres, o saldo da carteira de crédito total com recursos livres, o saldo da carteira de crédito total com recursos direcionados, a taxa de inadimplência da carteira de crédito total com recursos livres e a taxa média de juros das operações de crédito total com recursos Livre. O trabalho abordou apenas o nível mais agregado, no total do SFN.

[Çolak et al. \(2019\)](#) produzem uma série de indicadores para monitoramento dos períodos de expansão e desaceleração moderada ou excessiva de crédito no setor bancário turco. Os autores utilizam séries filtradas do agregados de crédito, crédito comercial, crédito direto ao consumidor, crédito imobiliário e financiamento de veículos, além de diversos setores da economia, como agricultura, manufatura, construção, comércio, dentre outros, para prever os ciclos de crédito no sistema bancário turco. Os autores concluem que os indicadores com maior poder de explicação para as variáveis macroeconômicas são a taxa de crescimento real do crédito e a taxa de resposta ao impulso do crédito.

Já para níveis abaixo do agregado de crédito (i.e. o total de crédito da economia), poucos trabalhos foram encontrados. Tangenciando o tema da previsão de saldos de crédito, outros tópicos da economia bancária foram objeto de estudo para previsão de séries temporais. [Sezer,](#)

Gudelek e Ozbayoglu (2019) produziram revisão de literatura de trabalhos publicados entre 2005 e 2019 que realizaram previsão de séries temporais financeiras utilizando *deep learning* e os agruparam em preços de ações individuais, índices (e.g., IBovespa, Dow Jones), preços de commodities, tendência e volatilidade de ativos, preços de títulos, câmbio e preços de criptomoedas. Apesar da extensa revisão de literatura realizada pelos autores, nenhum dos trabalhos citados combinaram estruturas hierárquicas com *machine learning*.

Gorodetskaya, Gobareva e Koroteev (2021) propõem o que chamaram de “uma metodologia universal” para aplicação automática de *machine learning* na previsão de séries temporais do setor bancário, que poderia ser aplicada em qualquer tipo de problema. A metodologia consiste em obter preditores a partir da própria variável defasada, das estatísticas básicas da variável (máximo, mínimo, média, variância etc.), e de anomalias periódicas detectadas, e selecioná-las pela medida de importância. Realizaram uma revisão de literatura recente sobre o assunto e apresentaram sua abordagem para o problema da previsão da demanda por moeda em caixas eletrônicos.

No que diz respeito à previsão de séries temporais em largas hierarquias, Prayoga, Suhartono e Rahayu (2017) trabalharam na previsão do fluxo de caixa do Banco da Indonésia, utilizando uma hierarquia de 3 níveis — 40 agências no nível mais desagregado, as 6 grandes ilhas do país como nível intermediário e o total no nível mais agregado. Os autores realizaram um *benchmark* de 5 modelos para previsão da série no nível mais agregado e utilizando o método *top-down* para obter as previsões no nível mais desagregado, concluindo pela efetividade do método *top-down* por proporções históricas. Entretanto, os autores não incluíram reconciliação ótima, a estimativa *bottom-up* ou mesmo outros métodos *top-down* para efeito de comparação, o que limita o alcance do trabalho.

2.2 Previsão de séries temporais hierárquicas e agrupadas

2.2.1 Abordagens de nível único

Uma abordagem de nível único é uma abordagem em que as previsões são realizadas para um único nível da hierarquia. A partir dessas previsões, os demais níveis são obtidos, ou desagregando (no caso dos níveis inferiores), ou agregando (no caso dos níveis superiores) essas informações (HYNDMAN; ATHANASOPOULOS, 2021). Os métodos *top-down*, *bottom-up* e *middle-out* são abordagens de nível único.

Enquanto há apenas uma única forma de se agregar níveis na hierarquia (*bottom-up*), a desagregação (*top-down*) pode ser realizada de, ao menos, duas dezenas de maneiras (GROSS; SOHL, 1990). Dois dos métodos mais intuitivos são a média das proporções históricas e a proporção das médias históricas.

Na média das proporções históricas, cada proporção p_j , com $j = 1, \dots, m$, consiste em tomar a média das proporções da série desagregada $y_{j,t}$ em relação ao agregado $y_{1,t}$:

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_{1,t}} \quad (1)$$

Já a proporção das médias históricas consiste em tomar a proporção das médias das séries desagregadas em relação à média do agregado⁷.

$$p_j = \frac{\sum_{t=1}^T \frac{y_{j,t}}{T}}{\sum_{t=1}^T \frac{y_{1,t}}{T}} \quad (2)$$

Athanasopoulos, Ahmed e Hyndman (2009) desenvolvem o método proporções de previsão, que consiste em um método *top-down* em que os pesos são calculados a partir das proporções das previsões fora da amostra ao invés do passado. A vantagem do método é que os pesos estarão os mais próximos das características mais recentes da série, ao invés de serem baseados em dados históricos. A desvantagem é que se deve realizar previsões para toda a hierarquia, perdendo o ganho de agilidade dos demais métodos *top-down*.

$$p_j = \prod_{i=0}^{K-1} \frac{\hat{Y}_{j,t+h}^i}{\sum \hat{Y}_{j,t+h}^{(i+1)}}$$

Li et al. (2016) compararam dois algoritmos de *machine learning* para previsão da produção de energia solar no estado da Flórida/EUA: ANN (*Artificial Neural Network*) e SVR (*Support Vector Regression*). Argumentando que tradicionalmente as previsões nesse problema são realizadas com os dados de produção total da planta, eles propõem uma abordagem hierárquica *bottom-up*, com previsões a partir de cada inversor solar. Os autores concluem que a abordagem hierárquica *bottom-up* é mais precisa do que a previsão do agregado, ao menos na previsão um passo a frente. Embora os autores utilizem algoritmos de *machine learning* para as previsões, eles não utilizam esses algoritmos para a reconciliação ótima, caracterizando a abordagem do trabalho ainda como nível único.

2.2.2 Reconciliação ótima de séries temporais hierárquicas

Previsões pontuais de séries temporais hierárquicas não é um assunto novo. Ao menos desde a década de 70, pesquisas foram publicadas acerca de abordagens *bottom-up* e *top-down*, suas vantagens e desvantagens, e tentativas de se definir qual é o melhor método⁸. 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

⁷ Isso é equivalente a tomar a proporção direta entre os somatórios das séries. Note que, pelas propriedades do operador de somatório, $\sum_{t=1}^T \frac{y_t}{T} = \frac{y_1}{T} + \dots + \frac{y_T}{T} = \frac{y_1 + \dots + y_T}{T} = \frac{\sum_{t=1}^T y_t}{T}$. Então, a equação Equação 2 pode ser simplificada para $p_j = \frac{\sum_{t=1}^T y_{j,t}}{\sum_{t=1}^T y_{1,t}}$.

⁸ Uma revisão dessa literatura pode ser encontrada em Athanasopoulos, Ahmed e Hyndman (2009).

da hierarquia) a partir da estimação de uma matriz de pesos via regressão linear por mínimos quadrados generalizados (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 individuais, 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 de pesos por mínimos quadrados ordinários (MQO).

A estimação por esse método resulta numa reconciliação ótima que depende apenas da estrutura hierárquica e independe da variância e covariância das previsões individuais — o que não é uma conclusão satisfatória, uma vez que é impossível que uma série hierárquica seja descorrelatada, pois um nível superior é, necessariamente, o agregado do inferior.

Hyndman, Lee e Wang (2016) tentam aperfeiçoar o método usando as variâncias das previsões individuais estimadas (dentro da amostra) como estimativa para a matriz de variância-covariância dos erros de reconciliação, de forma a as utilizar como pesos e realizar a reconciliação ótima por mínimos quadrados ponderados (MQP). Assim, previsões individuais 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 individuais \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 Traço Mínimo (*MinT*, na sigla em inglês). Paralelamente, usam desigualdade triangular para demonstrar que as previsões reconciliadas obtidas por esse método são ao menos tão boas quanto as previsões individuais.

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, os autores estendem 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.

Spiliotis et al. (2021) propõem a utilização de *machine learning* para a reconciliação ótima de séries temporais, especificamente os algoritmos de árvore de decisão *Random Forest* e *XGBoost*. Os autores descrevem como vantagens desse método em relação aos anteriores a de-

scrição de relacionamentos não lineares, performance preditiva e a desnecessidade da utilização de todos os elementos da hierarquia na combinação ótima. Para o conjunto de dados utilizados, os autores afirmam que os métodos de *machine learning*, especialmente o XGBoost, alcançaram, em média, melhor performance que as abordagens de nível único e o *MinT*. Além disso, concluíram que quanto maior é a diferença entre as séries, em todos os níveis hierárquicos, maior são os benefícios da abordagem por *machine learning*.

3 MÉTODOS PARA RECONCILIAÇÃO DE SÉRIES TEMPORAIS HIERÁRQUICAS E AGRUPADAS

3.1 Notação algébrica

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 de um país fictício com três estados, cada um com dois municípios. Essa série pode ser desagregada por estado que, por sua vez, pode ser desagregada por município (Figura 1).

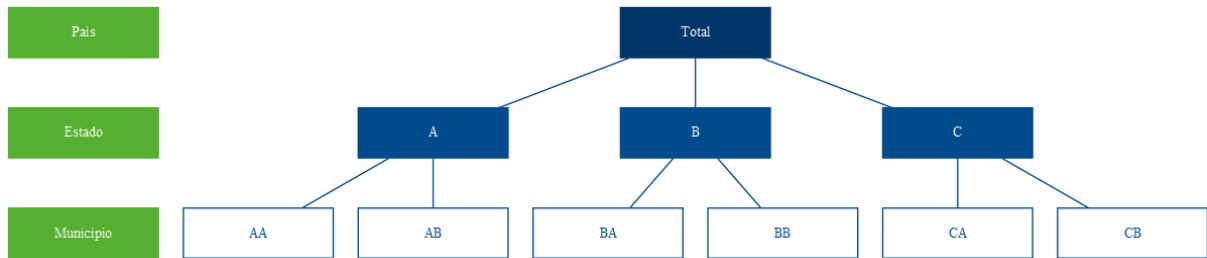


Figura 1 – Séries Hierárquicas

Essa estrutura pode ser representada através de equações para qualquer nível de agregação. Dessa forma, o agregado nacional pode ser descrito pelos agregados dos estados, Equação (3), ou como o agregado dos municípios, Equação (4). Já o agregado para o estado A é representado pela Equação (5).

$$y_t = y_{A,t} + y_{B,t} + y_{C,t} \quad (3)$$

$$y_t = y_{AA,t} + y_{AB,t} + y_{BA,t} + y_{BB,t} + y_{CA,t} + y_{CB,t} \quad (4)$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} \quad (5)$$

Alternativamente, podemos descrever a estrutura completa de forma matricial:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{BA,t} \\ y_{BB,t} \\ y_{CA,t} \\ y_{CB,t} \end{bmatrix}_{m \times 1} \quad (6)$$

Uma outra forma de desagregarmos o PIB é por atividade econômica — agricultura, indústrias extrativas, indústria de transformação, eletricidade e gás, construção etc. Essa estrutura não pode ser desagregada naturalmente de uma única maneira, 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.

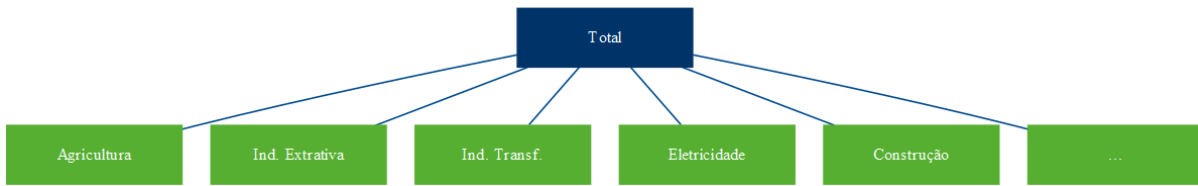


Figura 2 – Séries Agrupadas

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 estados⁹, a adição da estrutura agrupada pode ocorrer tanto acima (Figura 3) quanto abaixo (Figura 4) da hierárquica.

Na notação matricial, a estrutura da Figura 4 é representada como abaixo. Formalmente, o primeiro membro da igualdade é composto pelo vetor y_t n -dimensional com todas as observações no tempo t para todos os níveis da hierarquia. O segundo membro é composto pela matriz de soma S de dimensão $n \times m$ que define as equações para todo nível de agregação, e pelo vetor b_t composta pelas séries no nível mais desagregado.

$$y_t = S b_t \quad (7)$$

⁹ Essa estrutura é única no sentido que o somatório dos municípios totaliza o estado, mas não se pode somar estados para totalizar um município. Outro exemplo de estrutura hierárquica é a série de vendas de uma empresa: pode-se agregar as vendas de cada filial para obter o total, mas não o contrário.

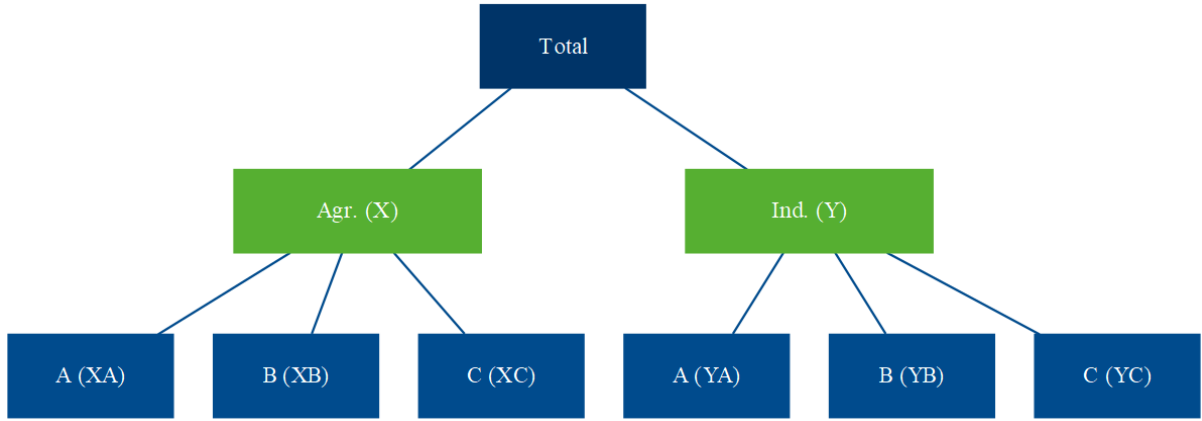


Figura 3 – Séries Hierárquicas Agrupadas (a)

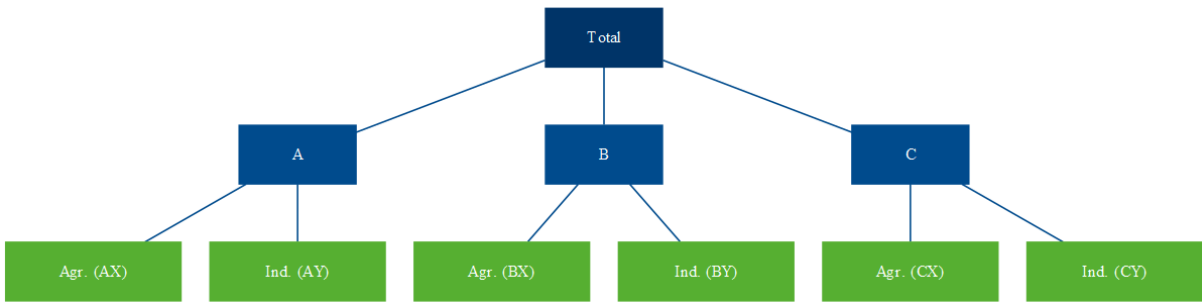


Figura 4 – Séries Hierárquicas Agrupadas (b)

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{CX,t} \\ y_{CY,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \\ y_{CX,t} \\ y_{CY,t} \end{bmatrix}_{m \times 1} \quad (8)$$

3.2 Abordagens top-down, bottom-up e middle-out

Talvez as formas mais intuitivas de se pensar em previsões para esses tipos de estrutura sejam as abordagens top-down e bottom-up. Tome a estrutura descrita na Figura 1, por exemplo. Podemos realizar a previsão para o horizonte de tempo h do agregado do PIB, representado no topo da hierarquia por *Total* (Equação 9), e então distribuir os valores previstos proporcionalmente entre os estados e municípios.

$$\hat{\mathbf{y}}_{T+h|T} = E[\mathbf{y}_{T+h}|\Omega_T] \quad (9)$$

Essa é a abordagem top-down. Nela, a previsão para os níveis mais desagregados da hierarquia são determinadas por uma proporção p_i do nível agregado. Por exemplo, as previsões para o município AA são dadas pela Equação 10.

$$\tilde{\mathbf{y}}_{AA,T+h|T} = p_1 \hat{\mathbf{y}}_{T+h|T} \quad (10)$$

Para isso, temos de definir uma matriz com todos esses pesos, que, seguindo a formulação de Hyndman e Athanasopoulos (2021), chamamos de \mathbf{G} :

$$\mathbf{G} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \quad (11)$$

\mathbf{G} é uma matriz $m \times n$ que multiplica o vetor $\hat{\mathbf{y}}_{T+h|T}$ que chamamos de *previsões base*, isto é, as previsões individuais para todos os níveis de agregação. A equação para a abordagem top-down será, então:

$$\tilde{\mathbf{y}}_{T+h|T} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{T+h|T} \quad (12)$$

Na notação matricial para a estrutura da Figura 1, temos:

$$\begin{bmatrix} \tilde{\mathbf{y}}_t \\ \tilde{\mathbf{y}}_{A,t} \\ \tilde{\mathbf{y}}_{B,t} \\ \tilde{\mathbf{y}}_{C,t} \\ \tilde{\mathbf{y}}_{AA,t} \\ \tilde{\mathbf{y}}_{AB,t} \\ \tilde{\mathbf{y}}_{BA,t} \\ \tilde{\mathbf{y}}_{BB,t} \\ \tilde{\mathbf{y}}_{CA,t} \\ \tilde{\mathbf{y}}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \begin{bmatrix} \hat{\mathbf{y}}_{T+h|T} \\ \hat{\mathbf{y}}_{A,T+h|T} \\ \hat{\mathbf{y}}_{B,T+h|T} \\ \hat{\mathbf{y}}_{C,T+h|T} \\ \hat{\mathbf{y}}_{AA,T+h|T} \\ \hat{\mathbf{y}}_{AB,T+h|T} \\ \hat{\mathbf{y}}_{BA,T+h|T} \\ \hat{\mathbf{y}}_{BB,T+h|T} \\ \hat{\mathbf{y}}_{CA,T+h|T} \\ \hat{\mathbf{y}}_{CB,T+h|T} \end{bmatrix}_{n \times 1} \quad (13)$$

Multiplicando as matrizes \mathbf{G} e $\hat{\mathbf{y}}$ obtemos o vetor \mathbf{b}_t ¹⁰, que contém as previsões para os níveis mais desagregados, agora definidas como proporção do total.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \end{bmatrix}_{m \times 1} \quad (14)$$

Note que, por se tratar de um método *top-down*, é necessário apenas o primeiro elemento do vetor de previsões base, ou seja, a previsão do nível mais agregado (Equação 14). Sendo essa exatamente uma das vantagens do método *top-down*, na prática, podemos anular os demais elementos de $\hat{\mathbf{y}}$.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \mathbf{S}_{n \times m} \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \begin{bmatrix} \hat{y}_{T+h|T} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{n \times 1} \quad (15)$$

Substituindo a matriz \mathbf{S} , temos as equações que definem as previsões reconciliadas.

¹⁰ Ver Equação 7.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} p_1 \hat{y}_{T+h|T} \\ p_2 \hat{y}_{T+h|T} \\ p_3 \hat{y}_{T+h|T} \\ p_4 \hat{y}_{T+h|T} \\ p_5 \hat{y}_{T+h|T} \\ p_6 \hat{y}_{T+h|T} \end{bmatrix}_{m \times 1} \quad (16)$$

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. Para tanto, basta modificar a matriz \mathbf{G} .

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{m \times n} \quad (17)$$

Portanto, \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. Portanto, chamamos \mathbf{G} de matriz de reconciliação.

$$\begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{A,t} \\ \tilde{y}_{B,t} \\ \tilde{y}_{C,t} \\ \tilde{y}_{AA,t} \\ \tilde{y}_{AB,t} \\ \tilde{y}_{BA,t} \\ \tilde{y}_{BB,t} \\ \tilde{y}_{CA,t} \\ \tilde{y}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} \hat{y}_{AA,T+h|T} \\ \hat{y}_{AB,T+h|T} \\ \hat{y}_{BA,T+h|T} \\ \hat{y}_{BB,T+h|T} \\ \hat{y}_{CA,T+h|T} \\ \hat{y}_{CB,T+h|T} \end{bmatrix}_{m \times 1} \quad (18)$$

Quando m — a quantidade de elementos do nível mais desagregado — é muito grande, tornando muito custoso obter $\hat{\mathbf{y}}_t$, e não se deseja uma abordagem estritamente *top-down*, pode-se combinar as duas formas:

$$\mathbf{G} = \begin{bmatrix} 0 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_6 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{m \times n} \quad (19)$$

Esse método é chamado de *middle-out*. Nele, o vetor \mathbf{b}_t é reescrito como proporções de algum nível intermediário arbitrariamente escolhido, ao invés de proporções do total. Isso permite uma abordagem mais econômica, em termos de custo computacional e de tempo, ao mesmo tempo em que mantém em algum grau as características individuais das hierarquias.

$$\begin{bmatrix} \tilde{\mathbf{y}}_t \\ \tilde{\mathbf{y}}_{A,t} \\ \tilde{\mathbf{y}}_{B,t} \\ \tilde{\mathbf{y}}_{C,t} \\ \tilde{\mathbf{y}}_{AA,t} \\ \tilde{\mathbf{y}}_{AB,t} \\ \tilde{\mathbf{y}}_{BA,t} \\ \tilde{\mathbf{y}}_{BB,t} \\ \tilde{\mathbf{y}}_{CA,t} \\ \tilde{\mathbf{y}}_{CB,t} \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{n \times m} \begin{bmatrix} p_1 \hat{\mathbf{y}}_{A,T+h|T} \\ p_2 \hat{\mathbf{y}}_{A,T+h|T} \\ p_3 \hat{\mathbf{y}}_{A,T+h|T} \\ p_4 \hat{\mathbf{y}}_{B,T+h|T} \\ p_5 \hat{\mathbf{y}}_{B,T+h|T} \\ p_6 \hat{\mathbf{y}}_{B,T+h|T} \end{bmatrix}_{m \times 1} \quad (20)$$

3.3 Coerência e reconciliação

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 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{SG} .

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 não há qualquer razão para que cada série siga o mesmo processo estocástico.

Os métodos de gerar previsões coerentes ($\tilde{\mathbf{y}}_t$) a partir de previsões base ($\hat{\mathbf{y}}_t$) são chamados de métodos de *reconciliação*. Os métodos de reconciliação tradicionais apresentados, *top-down*, *bottom-up* e *middle-out*, utilizam informação limitada. No método *top-down*,

utiliza-se apenas informações do nível mais agregado — por isso, apenas a primeira coluna em (Equação 11) é diferente de zero. Já na abordagem *bottom-up*, utiliza-se apenas as informações dos níveis mais desagregados, o que resulta na submatriz identidade $m \times m$ na Equação 17, enquanto as colunas que representam os níveis mais agregados são nulas. Por fim, a abordagem *middle-out* não utiliza a mesma, mas utiliza a mesma quantidade de informação que a *top-down* (Equação 19).

Alternativamente, podemos pensar numa matriz \mathbf{G} qualquer que utilize toda a informação disponível e tenha algumas propriedades que garantam que as previsões coerentes tenham o menor erro o possível. Esse é o problema de pesquisa trabalhado na *reconciliação ótima*.

3.4 Métodos analíticos de reconciliação ótima

Os métodos analíticos de reconciliação ótima são aqueles que estimam a matriz de reconciliação, \mathbf{G} , através de regressão linear. Isso resulta na redefinição das previsões do nível mais desagregado como uma combinação linear¹¹ das previsões de todos os elementos de todos os níveis da hierarquia, utilizando, assim, toda a informação disponível.

O estado-da-arte para esse tipo de método é o *MinT* (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2019). Nele, o objetivo é minimizar o erro das previsões reconciliadas:

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t} \quad (21)$$

Essa equação pode ser reescrita como $\tilde{\mathbf{e}}_t = \mathbf{S}\mathbf{G}\hat{\mathbf{e}}_t$, que tem variância dada por¹²

$$\text{Var}[\tilde{\mathbf{e}}] = \mathbf{S}\mathbf{G}\hat{\mathbf{W}}_{t+h|t}\mathbf{G}'\mathbf{S}' \quad (22)$$

em que $\hat{\mathbf{W}}_{t+h|t}$ é a matriz de variância-covariância dos erros de previsão base.

A abordagem consiste então em se obter um valor de \mathbf{G} que minimize o traço de $\text{Var}[\mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t}]$. Isso resultaria no melhor (variância mínima) estimador linear não viesado¹³ (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2019).

$$\mathbf{G} = (\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger\mathbf{S})^{-1}\mathbf{S}'\hat{\mathbf{W}}_{t+h|t}^\dagger \quad (23)$$

Essa formulação corresponde a um problema de regressão por mínimos quadrados generalizados, considerando \mathbf{S} como a matriz de preditores e \mathbf{G} os coeficientes a serem estimados. Consequentemente, as previsões ótimas reconciliadas são dadas por

¹¹ Por essa razão, esses métodos são também chamados de métodos de combinação.

¹² Ver demonstração A.3.

¹³ A ausência de viés é garantida pela *condição de ausência de viés em $\tilde{\mathbf{y}}$* .

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_{t+h|t} \quad (24)$$

3.5 Métodos de reconciliação ótima baseados em aprendizado de máquina

3.5.1 O método elastic net

O *elastic net* (ZOU; HASTIE, 2005) é um método de regressão regularizada que combina as normas L_1 e L_2 , as penalidades do *lasso* e do *ridge*, respectivamente. A função objetivo a ser minimizada é dada por

$$L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1 \quad (25)$$

em que λ_1 e λ_2 são os parâmetros de regularização e β é o vetor de coeficientes a serem estimados. A solução para essa função objetivo é dada por¹⁴

$$\hat{\beta} = \arg \min_{\beta} |\mathbf{y} - \mathbf{X}\beta|^2, \text{ sujeito a } (1 - \alpha)|\beta|_1 + \alpha|\beta|^2 \leq t \quad (26)$$

com $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ e $t \in \mathbb{R}^+$.

A função $(1 - \alpha)|\beta|_1 + \alpha|\beta|^2$ é a penalidade *elastic net*, uma combinação das penalidades *lasso* e *ridge*. O parâmetro α controla a mistura das duas penalidades, incluindo os casos extremos. Note que $\alpha = 0 \implies \lambda_2 = 0$, resultando em uma penalidade exclusivamente *lasso*, enquanto $\alpha = 1 \implies \lambda_1 = 0$, e a penalidade é apenas do tipo *ridge*.

Portanto o *elastic net* é um método de *shrinkage*, uma vez que a penalidade *ridge* reduz o tamanho dos coeficientes, e de *seleção de variáveis*, uma vez que a penalidade *lasso* tende a anular os coeficientes de variáveis irrelevantes. Essas propriedades são desejáveis para a reconciliação de séries temporais, uma vez que a estrutura hierárquica pode conter séries insignificantes para a previsão de outras séries.

Diferentemente dos métodos analíticos estudados, o *elastic net* não possui uma solução fechada. Portanto, é necessário utilizar métodos iterativos para encontrar o valor ótimo de $\hat{\beta}$ e Zou e Hastie (2005) utilizam validação cruzada *k*-fold para encontrar quais os valores de λ_1 e λ_2 que minimizam o resíduo. Nesse sentido, dado a metodologia de processo iterativo envolvendo calibragem de hiperparâmetros e reamostragem, podemos classificar o *elastic net* como um método de *machine learning*.

3.5.2 O processo de ajuste e sobreajuste

Considere uma função de ajuste f , um conjunto de pontos $D = d_1, \dots, d_n$ com $d_i = (\mathbf{x}_i \ \mathbf{y}_i)'$, variáveis de decisão ou parâmetros $\mathbf{x}_i \in \mathbb{R}^m$ e imagem $\mathbf{y}_i = f(\mathbf{x}_i) \in \mathbb{R}$. Diferente-

¹⁴ Sob o valor otimizado ainda é aplicada correção de escala na forma $(1 + \lambda_2)\hat{\beta}$. Ver Zou e Hastie (2005).

mente da abordagem do modelo clássico de regressão linear em que, cumpridas certas hipóteses, há um modelo teórico de coeficientes estimados por mínimos quadrados ordinários (MQO) que é garantido pelo teorema de Gauss-Markov ser o melhor estimador linear não viesado (BLUE), em *machine learning* o objetivo é encontrar, de forma iterativa, um modelo que melhor aproxima a função f usando a informação contida em D , ou seja, queremos ajustar uma função de regressão \hat{f}_D aos nossos dados D de forma que $\hat{\mathbf{y}} = \hat{f}_D(\mathbf{x}, \varepsilon)$ tenha o menor erro de aproximação ε (BISCHL et al., 2012).

Para verificar o quão bem o modelo \hat{f}_D se aproxima da função real f , é necessário uma função de perda $L(\mathbf{y}, \hat{f}(\mathbf{x}))$ que, no caso de regressão, será a perda quadrática $(\mathbf{y} - \hat{f}(\mathbf{x}))^2$ ou a perda absoluta $|\mathbf{x} - \hat{f}(\mathbf{x})|$. Esses valores são agregados pela média para formar as funções de custo erro médio quadrático (MSE) e erro médio absoluto (MAE).

Calculando o custo sobre a amostra D usada para ajustar o modelo, teremos o chamado erro de resubstituição (Equação 27). Nesse caso, estaríamos usando o mesmo conjunto de dados tanto para treinar o preditor quanto para estimar o erro, o que nos levaria a uma estimativa enviesada do erro de generalização. Caso usássemos essa estimativa para seleção de modelos, esse viés favoreceria modelos mais adaptados à amostra.

$$\widehat{GE}(\hat{f}_D, D) = \sum_{(x, y)' \in D} \frac{L(y, \hat{f}_D(x))}{|D|} \quad (27)$$

Dadas suficientes iterações, o erro de resubstituição tende a zero. Isso acontece porque conforme o preditor se adapta cada vez mais aos dados de treinamento ele irá memorizar a relação entre o conjunto de pontos D e a imagem $f(\mathbf{x}_i)$, ou seja, irá se ajustar perfeitamente ao formato da função a ser modelada. E não necessariamente um modelo perfeitamente ajustado se traduz na capacidade de predição de dados futuros (fora da amostra). De forma geral, espera-se que o preditor reduza seu viés durante o treino apenas o suficiente para que seja capaz de generalizar sua predição para fora da amostra em um nível ótimo de acurácia. A partir desse ponto, a redução no viés é penalizada com o aumento da variância, ou seja, com a redução de sua capacidade de prever dados futuros. A esse processo se dá o nome de *overfitting* ou sobreajuste. Isso quer dizer que não podemos considerar a performance do preditor em D se desejamos estimar honestamente a performance real do modelo (BISCHL et al., 2012).

3.5.3 Reamostragem

Uma forma de se corrigir esse problema é dividindo a amostra em um conjunto para treino D^ϕ e outro conjunto para teste D^θ de forma que $D^\phi \cup D^\theta = D$ e $D^\phi \cap D^\theta = \emptyset$. Assim, pode-se treinar um meta-modelo em D^ϕ para se obter \hat{f}_{D^ϕ} e calcular seu erro de generalização usando os dados de D^θ . Essa abordagem é chamada de *hold-out* e ela é de simples implementação e utilização, uma vez que as observações do conjunto teste são completamente independentes das observações com as quais o modelo foi treinado. Então, podemos estimar o

erro de generalização do modelo, que consiste no cômputo do custo de \hat{f}_D^ϕ aplicada à amostra de teste D^θ .

$$\widehat{GE}(\hat{f}_{D^\phi}, D^\theta) = \sum_{(x, y)' \in D^\theta} \frac{L(y, \hat{f}_{D^\phi}(x))}{|D^\theta|} \quad (28)$$

Como esse método mais simples pode não ser suficiente para detectar a variância e instabilidade de modelos mais complexos, foram desenvolvidas diferentes técnicas de reamostragem ao longo do tempo. Uma das mais populares é a validação cruzada (STONE, 1974), que consiste em gerar repetidamente i subconjuntos de treino D_i^ϕ e teste D_i^θ com o conjunto de dados disponível, ajustar um meta-modelo com cada conjunto de treino e atestar sua qualidade no conjunto de teste correspondente. A estimativa do erro de generalização então se torna:

$$\widehat{GE} = \frac{1}{k} \sum_{i=1}^k \widehat{GE}(\hat{f}_{D_i^\phi}, D_i^\theta) \quad (29)$$

Dividindo a amostra em k subconjuntos, utilizando $k - 1$ para ajustar um meta-modelo e validando a performance no subconjunto restante — e repetindo esse procedimento para todas as possibilidades de subconjuntos —, temos a validação cruzada k -fold. Esta é uma técnica de reamostragem que permite estimar o erro de generalização de um modelo de forma mais robusta e confiável que o *hold-out* simples. Isso porque, ao contrário do *hold-out*, a validação cruzada permite que todos os dados sejam usados tanto para treino quanto para teste, o que reduz a variância da estimativa do erro de generalização.

3.5.4 Validação cruzada k -fold em séries temporais

Para dados *cross-section*, essa definição de validação cruzada k -fold é suficiente para qualquer caso. Entretanto, no caso de séries temporais, o analista deve tomar alguns cuidados na escolha da abordagem de validação cruzada. Isso porque, ao contrário de dados *cross-section*, os dados de séries temporais são dependentes no tempo. Isso traz dois aparentes problemas: primeiramente, ao dividir a amostra em k subconjuntos aleatórios, o meta-modelo será treinado em períodos descontínuos e com dados futuros aos dados de teste em, ao menos, $k - 1$ subconjuntos. Em segundo lugar, realizando o *split* treino-teste em y_t , com y_{t-1} na amostra treino, significa que os subconjuntos de treino e teste são dependentes (Figura 5).

Para contornar esse problema, outros métodos de validação cruzada foram desenvolvidos pensando em séries temporais. Tomando o conjunto de validação com dados exclusivamente posteriores aos dados de treino, temos o método conhecido como validação cruzada com origem móvel (Figura 6).

Já quando excluimos as defasagens dependentes do subconjunto de treino (em relação ao conjunto de validação), temos a validação cruzada não dependente (Figura 7). Para ilustrar esse

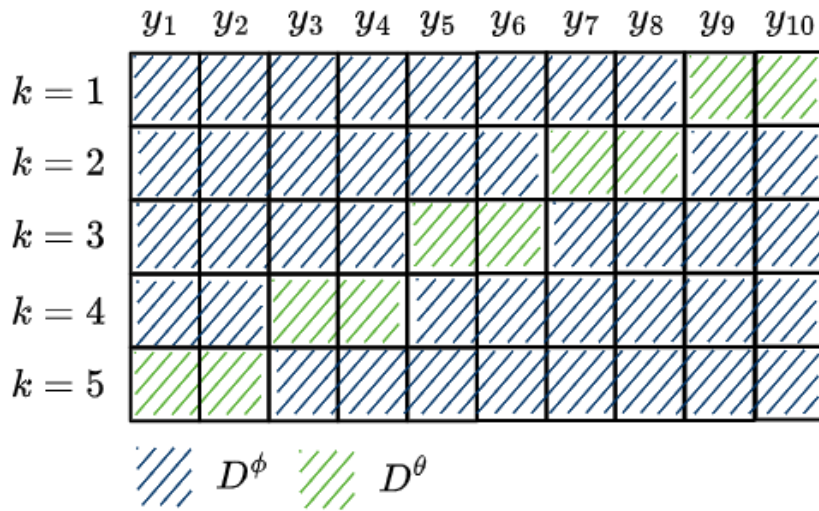


Figura 5 – Validação k -fold aleatória

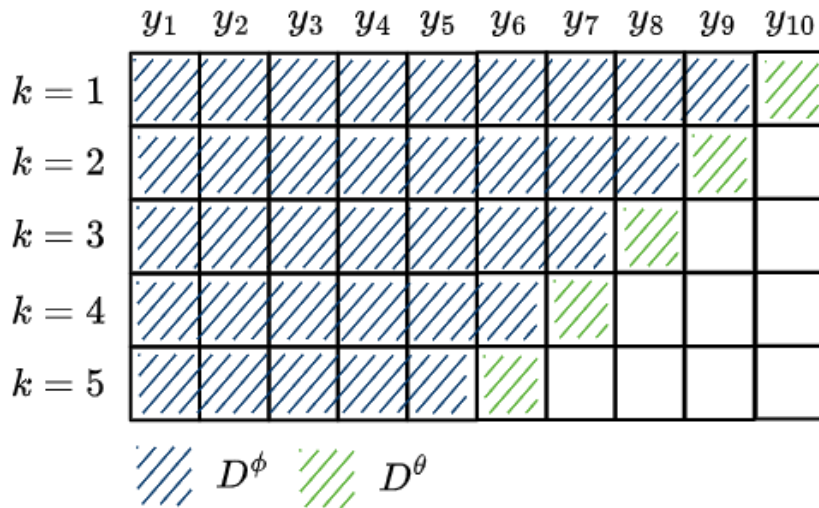


Figura 6 – Validação k -fold com origem móvel

caso, tome um processo AR(3). Excluir as defasagens dependentes significa que, se o conjunto de validação começa em y_t , então o conjunto de treino pode apenas ir até y_{t-4} . O problema evidente dessa abordagem é que, dependendo do tamanho da estrutura de autocorrelação da série, muitas defasagens são excluídas, podendo inviabilizar o processo de validação.

Entretanto, esses problemas são apenas aparentes. [Bergmeir e Benítez \(2012\)](#) comparam uma série de métodos de validação em séries temporais e concluem que, considerando séries estacionárias, os problemas teóricos relacionados à dependência não produzem impactos empíricos. [Bergmeir, Hyndman e Koo \(2018\)](#) vão além e abordam esse problema teórico e empiricamente, concluindo que não apenas é possível o uso de validação cruzada k -fold quando a série não apresenta autocorrelação serial nos resíduos, mas também é uma melhor escolha do que a validação fora da amostra¹⁵.

¹⁵ *Out-of-sample evaluation* é o método padrão na literatura de séries temporais e consiste na separação da porção final da série — geralmente entre 20% e 30% — para validação.

cionados a crédito e mantidas apenas as agências atualmente em atividade. Quanto ao período, há dados disponíveis desde 1988. Entretanto, utilizaremos os dados a partir de 2010 pois, se tratando de uma hierarquia larga, o custo computacional deve ser levado em conta.

Por fim, as variáveis mantidas no *dataset* foram as descritas no Quadro 1. Dessa forma, temos séries mensais de saldos de créditos de 96 agências bancárias, distribuídas por 70 municípios, a partir de janeiro de 2010, totalizando 43.923 observações.

Quadro 1 – Variáveis do dataset

Variável	Descrição
ref	Data de referência do relatório ESTBAN
nome_mesorregiao	Nome da mesorregião do ES
nome_microrregiao	Nome da microrregião do ES
verbete	Descrição da rubrica do balancete
nome	Nome do município
cnpj_agencia	CNPJ da agência bancária
valor	Saldo do verbete

Os dados foram organizados de forma hierárquica por estado, mesorregião, microrregião, município e agência bancária; e de forma agrupada, por verbete.

O detalhamento da construção do *dataset* se encontra no Anexo A.

4.2 Análise exploratória dos dados

Após limpeza, o *dataset* adquiriu a estrutura apresentada na Tabela 1. As microrregiões que compõem cada mesorregião são apresentadas na Tabela 2 e os municípios que compõem cada microrregião são apresentados na Tabela 3.

Tabela 1 – Estrutura do dataset

Data de referência	Mesorregião	Microrregião	Município	Agência	Verbetes	Saldo
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603014802	empréstimos e títulos descontados	367.233.178
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603014802	financiamentos	1.223.990
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603014802	financiamentos rurais	655.171
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603016767	empréstimos e títulos descontados	29.293.860
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603016767	financiamentos	651.290
2023-02-01	Central Espírito-santense	Vitória	Vitória	28127603016767	financiamentos rurais	1.165.244

Tabela 2 – Microrregiões por mesorregião

Mesorregião	Microrregiões
Central Espírito-santense	Afonso Cláudio, Guarapari, Santa Teresa, Vitória
Litoral Norte Espírito-santense	Linhares, Montanha, São Mateus
Noroeste Espírito-santense	Barra de São Francisco, Colatina, Nova Venécia
Sul Espírito-santense	Alegre, Cachoeiro de Itapemirim, Itapemirim

O tamanho de uma estrutura hierárquica, em termos de observações, é determinada por seu nível mais desagregado. Assim, sendo 156 meses e 96 agências, a estrutura hierárquica

Tabela 3 – Municípios por microrregião

Microrregiões	Municípios
Afonso Cláudio	Afonso Cláudio , Brejetuba , Conceição do Castelo , Domingos Martins , Laranja da Terra , Marechal Floriano , Venda Nova do Imigrante
Alegre	Alegre , Dorcas do Rio Preto , Guaçuí , Ibatiba , Ibitirama , Irupí , Iúna , Muniz Freire
Barra de São Francisco	Água Doce do Norte , Barra de São Francisco , Ecoporanga , Mantenópolis
Cachoeiro de Itapemirim	Apiacá , Atilio Vivacqua , Bom Jesus do Norte , Cachoeiro de Itapemirim , Castelo , Jerônimo Monteiro , Mimoso do Sul , Muqui , São José do Calçado , Vargem Alta
Colatina	Alto Rio Novo , Baixo Guandu , Colatina , Pancas
Guarapari	Alfredo Chaves , Anchieta , Guarapari , Iconha , Piúma , Rio Novo do Sul
Itapemirim	Itapemirim , Marataízes , Presidente Kennedy
Linhares	Aracruz , Fundão , Ibiracú , João Neiva , Linhares , Rio Bananal , Sooretama
Montanha	Mucurici , Pinheiros , Ponto Belo
Nova Venécia	Águia Branca , Boa Esperança , Nova Venécia , Vila Valério
Santa Teresa	Itaguaçu , Itarana , Santa Leopoldina , Santa Maria de Jetibá , Santa Teresa , São Roque do Canaã
São Mateus	Jaguaré , Pedro Canário , São Mateus
Vitória	Cariacica , Serra , Viana , Vila Velha , Vitória

deve contar com $156 \times 96 = 14.976$ observações. Sendo também uma estrutura agrupada por 3 verbetes, a quantidade de observações é multiplicada pela quantidade de níveis transversais, totalizando 44.928 observações. Como há apenas, 43.923 observações, podemos concluir que há dados faltantes, originados de agências que foram criadas no período e, portanto, não possuem a série completa desde 2010. Para permitir a estimação, esses dados serão imputados com zero.

Tabela 4 – Contagem de únicos no dataset ESTBAN

	Únicos
ref	156
nome_mesorregiao	4
nome_microrregiao	13
nome	70
cnpj_agencia	96
verbeta	3

A série temporal do agregado de crédito no Banestes no Espírito Santo é apresentada na Figura 8. Podemos observar tendência de crescimento a partir de 2020, indicando que a série seja não estacionária e que talvez seja interessante adicionar um regressor externo para o período da pandemia, quando foram implementados programas de crédito emergenciais no país.

A concentração na mesorregião Central Espírito-santense (Figura 9) pode ser um indicativo de que os elementos das demais regiões tenham menor peso na reconciliação.

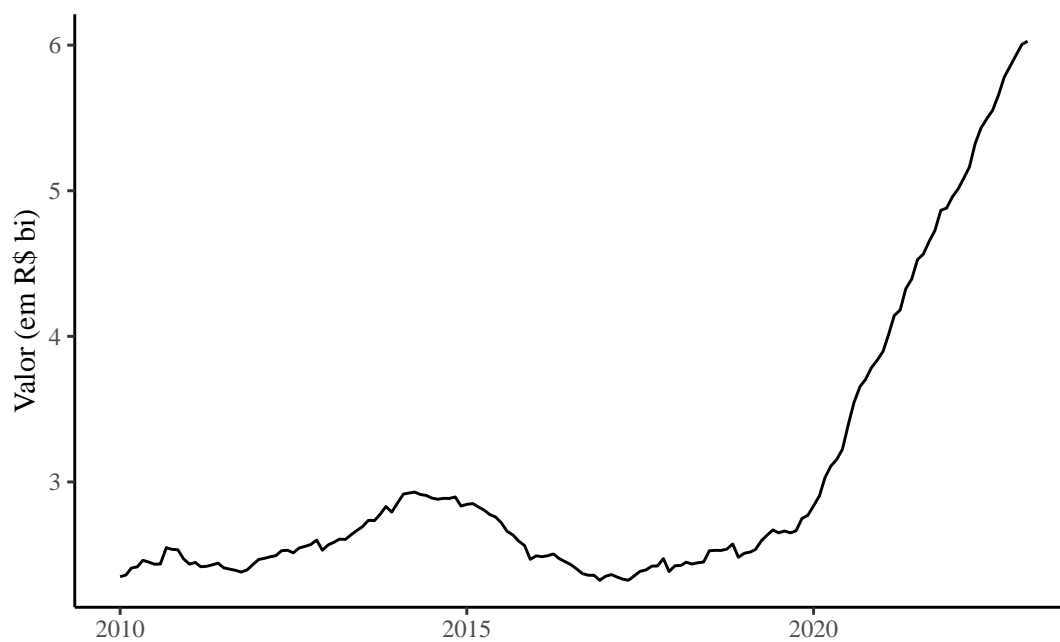


Figura 8 – Série temporal do agregado de crédito do Banestes no ES

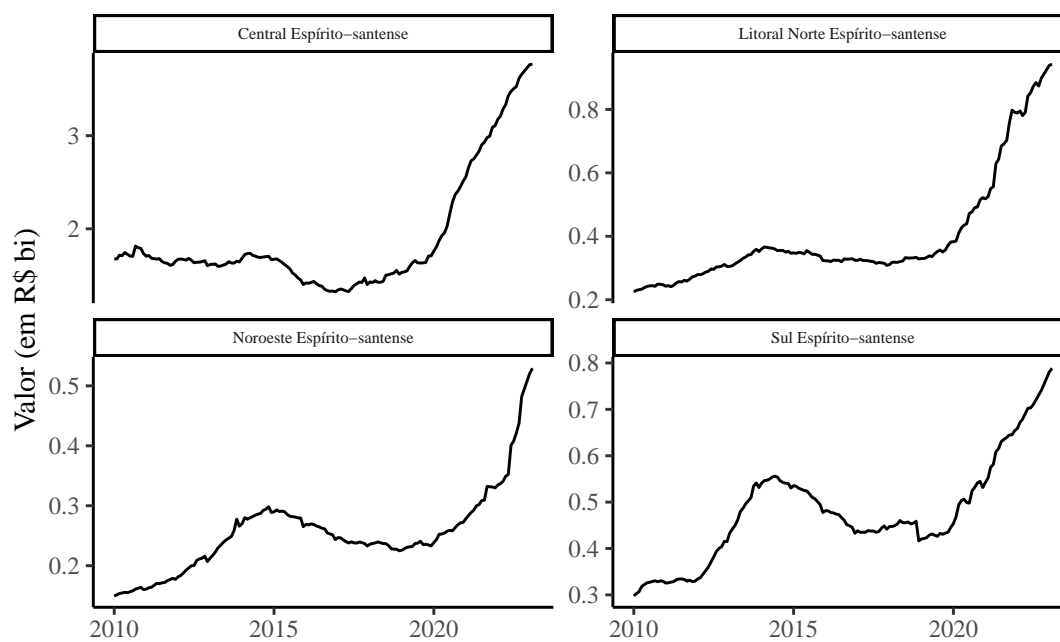


Figura 9 – Série temporal do agregado de crédito do Banestes por mesorregião do ES



Figura 10 – Série temporal do agregado de crédito do Banestes por microrregião do ES

Analisando o agregado por verbetes, a Figura 11 indica que o crescimento é liderado pelas alíneas empréstimos e títulos descontados.

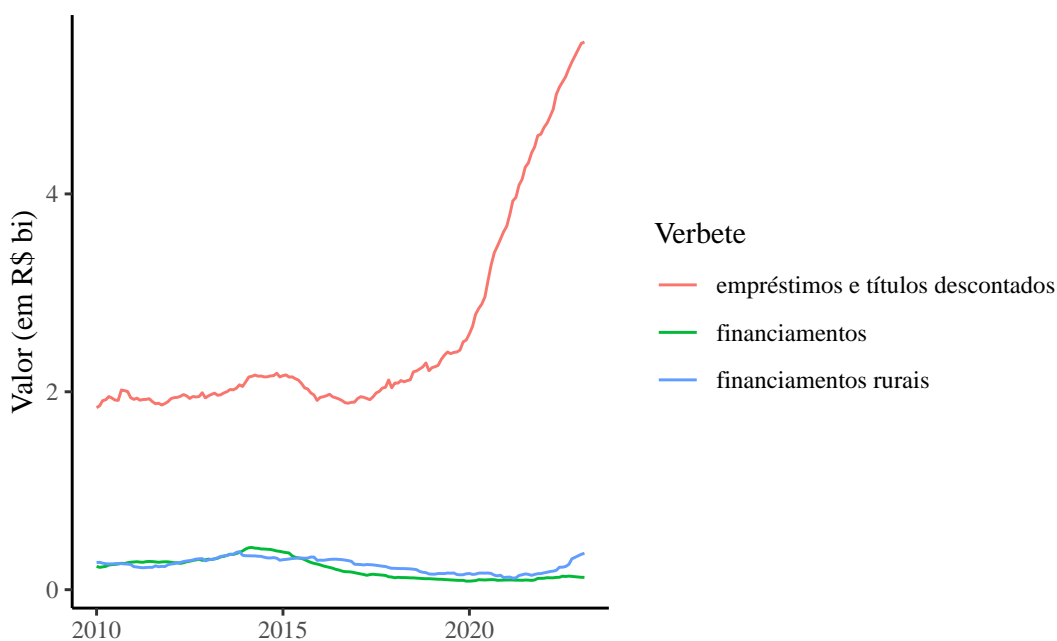


Figura 11 – Verbetes no agregado do ES

4.3 Previsões base

Uma vez que o foco deste trabalho é no ganho proporcionado pela reconciliação ótima, as previsões base serão obtidas por meio dos métodos básicos para previsão de séries tempo-

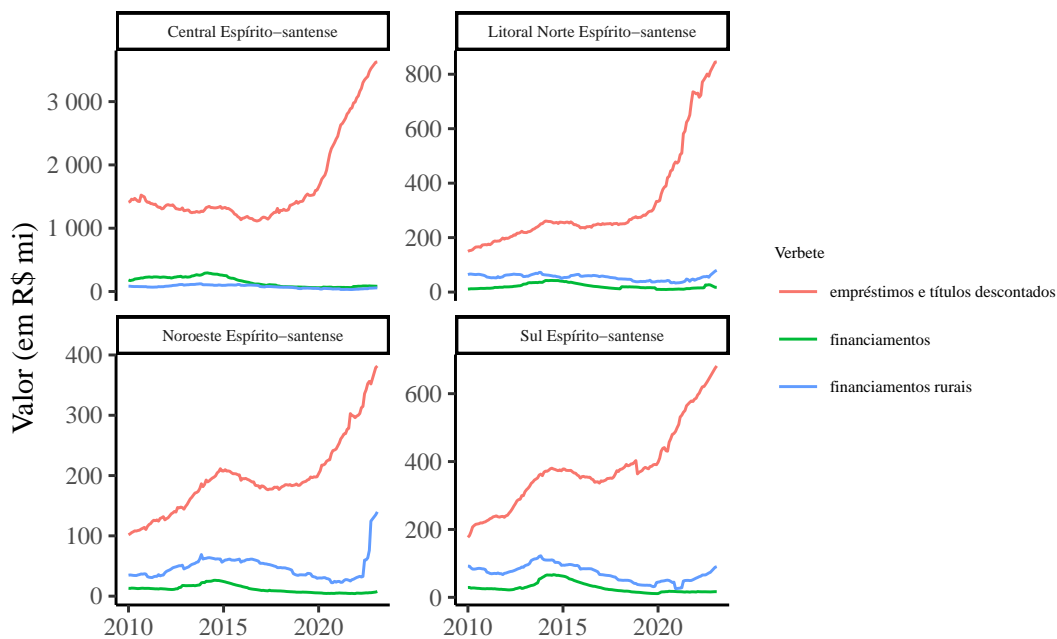


Figura 12 – Verbete por mesorregião do ES

rais, como algum tipo de suavimento exponencial ou pelo algoritmo de Hyndman-Khandakar (HYNDMAN; ATHANASOPOULOS, 2021). Esses métodos são amplamente utilizados na literatura de séries temporais e, portanto, servem como *benchmark* para a avaliação dos métodos de reconciliação ótima.

4.4 Design do experimento

4.4.1 Métodos analíticos de reconciliação

Os métodos analíticos de reconciliação utilizados no *benchmark* serão os de nível único *bottom-up* e o *top-down*; os de reconciliação ótima *MinT-OLS*, *MinT-WLS* e *MinT-shrink*. Essa escolha se justifica para que seja possível comparar os métodos de reconciliação ótima com os métodos de nível único, principalmente em relação ao tempo de processamento, que pode revelar um *trade-off* a ser analisado pelo pesquisador.

4.4.2 Métodos de reconciliação ótima baseados em aprendizado de máquina

A metodologia para obtenção das previsões reconciliadas por métodos baseados em *machine-learning* será semelhante ao de Spiliotis et al. (2021):

1. dividir a amostra em treino e teste;
2. treinar um modelo de previsão na amostra treino e obter previsões um passo a frente para a amostra teste;
3. treinar um modelo de *machine learning* para cada série do menor nível da hierarquia, em que os parâmetros são as previsões obtidas no passo 2 e a variável explicada são os

valores observados. Isso resulta em um modelo de reconciliação ótima para cada elemento do menor nível da hierarquia, combinando informações disponíveis de todos os níveis hierárquicos;

4. obter as previsões base \hat{y}_t ;
5. passar as previsões base ao modelo treinado no passo 3 para se obter as previsões reconciliadas para o menor nível da hierarquia;
6. agregar as previsões reconciliadas para se obter as previsões nos demais níveis hierárquicos.

Esse procedimento será realizado três vezes, uma para cada configuração do método *elastic net*, sendo $\alpha = 0$ (*lasso*), $\alpha = 1$ (*ridge*) e uma combinação de ambos obtida via reamostragem. Espera-se que todos os três modelos propostos nesse trabalho sejam capazes de superar os métodos analíticos de reconciliação.

4.4.3 Espaço de busca dos hiperparâmetros

[Descrever o espaço de hyperparâmetros]

5 RESULTADOS

[Escrever o *benchmarking* da performance dos métodos de reconciliação]

REFERÊNCIAS

- ATHANASOPOULOS, G.; AHMED, R. A.; HYNDMAN, R. J. Hierarchical forecasts for Australian domestic tourism. en. **International Journal of Forecasting**, v. 25, n. 1, p. 146–166, jan. 2009. ISSN 0169-2070. DOI: [10.1016/j.ijforecast.2008.07.004](https://doi.org/10.1016/j.ijforecast.2008.07.004). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207008000691>>. Acesso em: 11 jan. 2023. Citado na p. 27.
- BADER, F. L. C.; KOYAMA, SÉRGIO MIKIO; TSUCHIDA, MARCOS HIROYUKI. Modelo favar canônico para previsão do mercado de crédito. pt. **Banco Central do Brasil**, v. 369, p. 38, nov. 2014. ISSN 1519-1028. Citado na p. 25.
- BANCO DO ESTADO DO ESPÍRITO SANTO. **Demonstrações Financeiras**. Vitória, 2022. Disponível em: https://www.banestes.com.br/ri/arquivos/informacoes/demoFinanceiras/2022_2S_demonFinanceiras.pdf>. Acesso em: 4 jun. 2023. Citado na p. 23.
- BECCALLI, E. et al. Earnings management, forecast guidance and the banking crisis. **The European Journal of Finance**, v. 21, n. 3, p. 242–268, fev. 2015. ISSN 1351-847X. DOI: [10.1080/1351847X.2013.809548](https://doi.org/10.1080/1351847X.2013.809548). Disponível em: <https://doi.org/10.1080/1351847X.2013.809548>>. Acesso em: 7 mai. 2023. Citado na p. 22.
- BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. en. **Information Sciences**, v. 191, p. 192–213, mai. 2012. ISSN 0020-0255. DOI: [10.1016/j.ins.2011.12.028](https://doi.org/10.1016/j.ins.2011.12.028). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025511006773>>. Acesso em: 30 mai. 2023. Citado na p. 40.
- BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. en. **Computational Statistics & Data Analysis**, v. 120, p. 70–83, abr. 2018. ISSN 01679473. DOI: [10.1016/j.csda.2017.11.003](https://doi.org/10.1016/j.csda.2017.11.003). Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167947317302384>>. Acesso em: 28 mai. 2023. Citado na p. 40.
- BISCHL, B. et al. Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. **Evolutionary Computation**, v. 20, n. 2, p. 249–275, jun. 2012. Conference Name: Evolutionary Computation. ISSN 1063-6560. DOI: [10.1162/EVCO_a_00069](https://doi.org/10.1162/EVCO_a_00069). Citado na p. 38.
- BRASIL. **Lei nº 7.492, de 16 de junho de 1986**. Brasília, DF: Presidência da República, jun. 1986. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/17492.htm>. Citado na p. 21.
- COELHO, A. N. B. et al. A responsabilidade da auditoria externa na fraude contábil do banco panamericano. pt. **RAGC**, v. 3, n. 7, set. 2015. Number: 7. ISSN 2317-0484. Disponível em:

<<https://revistas.fucamp.edu.br/index.php/rage/article/view/604>>. Acesso em: 31 mai. 2023. Citado na p. 22.

ÇOLAK, M. S. et al. **TCMB - Monitoring and Forecasting Cyclical Dynamics in Bank Credits: Evidence from Turkish Banking Sector**. en. Ankara, Turquia, 2019. Disponível em: <<https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Papers/2019/19-29>>. Acesso em: 6 mar. 2023. Citado na p. 25.

CONSELHO MONETÁRIO NACIONAL. **Resolução nº 2.099, de 17 de agosto de 1994**. Brasília, DF: Banco Central do Brasil, ago. 1994. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/1994/pdf/res_2099_v1_O.pdf>. Citado na p. 23.

_____. **Resolução nº 2.682, de 21 de dezembro de 1999**. Brasília, DF: Banco Central do Brasil, dez. 1999. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf>. Citado na p. 41.

DAHIS, R. et al. Data Basis (Base Dos Dados): Universalizing Access to High-Quality Data. en. **SSRN Electronic Journal**, 2022. ISSN 1556-5068. DOI: [10.2139/ssrn.4157813](https://doi.org/10.2139/ssrn.4157813). Disponível em: <<https://www.ssrn.com/abstract=4157813>>. Acesso em: 18 jun. 2023. Citado na p. 41.

GORODETSKAYA, O.; GOBAREVA, Y.; KOROTEEV, M. A Machine Learning Pipeline for Forecasting Time Series in the Banking Sector. en. **Economies**, v. 9, n. 4, p. 205, dez. 2021. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. ISSN 2227-7099. DOI: [10.3390/economies9040205](https://doi.org/10.3390/economies9040205). Disponível em: <<https://www.mdpi.com/2227-7099/9/4/205>>. Acesso em: 27 fev. 2023. Citado na p. 26.

GROSS, C. W.; SOHL, J. E. Disaggregation methods to expedite product line forecasting. en. **Journal of Forecasting**, v. 9, n. 3, p. 233–254, 1990. ISSN 1099-131X. DOI: [10.1002/for.3980090304](https://doi.org/10.1002/for.3980090304). Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980090304>>. Acesso em: 18 jun. 2023. Citado nas pp. 23, 26.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3. ed. Melbourne, Austrália: OTexts, 2021. Disponível em: <<https://otexts.com/fpp3/>>. Citado nas pp. 23, 26, 29, 32, 46.

HYNDMAN, R. J.; AHMED, R. A. et al. Optimal combination forecasts for hierarchical time series. en. **Computational Statistics & Data Analysis**, v. 55, n. 9, p. 2579–2589, set. 2011. ISSN 0167-9473. DOI: [10.1016/j.csda.2011.03.006](https://doi.org/10.1016/j.csda.2011.03.006). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167947311000971>>. Acesso em: 11 jan. 2023. Citado nas pp. 23, 27, 28.

HYNDMAN, R. J.; LEE, A. J.; WANG, E. Fast computation of reconciled forecasts for hierarchical and grouped time series. en. **Computational Statistics & Data Analysis**, v. 97, p. 16–32, mai. 2016. ISSN 0167-9473. DOI: [10.1016/j.csda.2015.11.007](https://doi.org/10.1016/j.csda.2015.11.007). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016794731500290X>>. Acesso em: 11 jan. 2023. Citado na p. 28.

LI, Z. et al. A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting. en. **Energies**, v. 9, n. 1, p. 55, jan. 2016. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. ISSN 1996-1073. DOI: [10.3390/en9010055](https://doi.org/10.3390/en9010055). Disponível em: <https://www.mdpi.com/1996-1073/9/1/55>. Acesso em: 8 abr. 2023. Citado na p. 27.

MOURA, D. d. **Análise dos fatores de convencimento do juízo brasileiro quanto à ocorrência de fraude contábil: um estudo de caso Múltiplo da Gallus, da Encol e do Banco Santos**. 2007. Dissertação de mestrado – Fundação Getúlio Vargas, Rio de Janeiro. Accepted: 2009-11-18T19:01:34Z. Disponível em: <http://bibliotecadigital.fgv.br:80/dspace/handle/10438/4038>. Acesso em: 31 mai. 2023. Citado na p. 21.

PANAGIOTELIS, A. et al. Forecast reconciliation: A geometric view with new insights on bias correction. en. **International Journal of Forecasting**, v. 37, n. 1, p. 343–359, jan. 2021. ISSN 0169-2070. DOI: [10.1016/j.ijforecast.2020.06.004](https://doi.org/10.1016/j.ijforecast.2020.06.004). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207020300911>. Acesso em: 15 jan. 2023. Citado na p. 28.

PRAYOGA, I.; SUHARTONO, S.; RAHAYU, S. Top-down forecasting for high dimensional currency circulation data of Bank Indonesia. **International Journal of Advances in Soft Computing and its Applications**, v. 9, p. 62–74, jan. 2017. Citado na p. 26.

SEZER, O. B.; GUDELEK, M. U.; OZBAYOGLU, A. M. **Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019**. Turquia: arXiv, nov. 2019. arXiv:1911.13288 [cs, q-fin, stat]. Disponível em: <http://arxiv.org/abs/1911.13288>. Acesso em: 7 mar. 2023. Citado na p. 25.

SPILOTIS, E. et al. Hierarchical forecast reconciliation with machine learning. en. **Applied Soft Computing**, v. 112, p. 107756, nov. 2021. ISSN 1568-4946. DOI: [10.1016/j.asoc.2021.107756](https://doi.org/10.1016/j.asoc.2021.107756). Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494621006773>. Acesso em: 11 jan. 2023. Citado nas pp. 24, 28, 46.

STONE, M. Cross-validation and multinomial prediction. **Biometrika**, v. 61, n. 3, p. 509–515, dez. 1974. ISSN 0006-3444. DOI: [10.1093/biomet/61.3.509](https://doi.org/10.1093/biomet/61.3.509). Disponível em: <https://doi.org/10.1093/biomet/61.3.509>. Acesso em: 28 mai. 2023. Citado na p. 39.

WALLANDER, J. Budgeting — an unnecessary evil. en. **Scandinavian Journal of Management**, v. 15, n. 4, p. 405–421, dez. 1999. ISSN 0956-5221. DOI: [10.1016/S0956-5221\(98\)00032-3](https://doi.org/10.1016/S0956-5221(98)00032-3). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0956522198000323>. Acesso em: 8 mai. 2023. Citado na p. 21.

WICKRAMASURIYA, S. L.; ATHANASOPOULOS, G.; HYNDMAN, R. J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. en. **Journal of the American Statistical Association**, v. 114, n. 526, p. 804–819, abr. 2019. ISSN 0162-1459, 1537-274X. DOI: [10.1080/01621459.2018.1448825](https://doi.org/10.1080/01621459.2018.1448825). Disponível em: <https://doi.org/10.1080/01621459.2018.1448825>.

[//www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825](https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1448825)>. Acesso em: 18 jun. 2023. Citado nas pp. 24, 28, 36.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. en. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301–320, 2005. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2005.00503.x>. ISSN 1467-9868. DOI: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x). Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x>>. Acesso em: 5 jun. 2023. Citado na p. 37.

Apêndices

APÊNDICE A – DEMONSTRAÇÕES

Proposição A.1 (condição de ausência de viés em $\tilde{\mathbf{y}}$). *Se as previsões reconciliadas são não viesadas, então $\mathbf{SGS} = \mathbf{S}$, ou seja, \mathbf{G} é inversa generalizada de \mathbf{S} .*

Demonstração.

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.1})$$

Se $\hat{\mathbf{y}}_{t+h|t}$ é não viesado, então

$$\mathbb{E}[\hat{\mathbf{y}}_{t+h|t}] = \mathbb{E}[\mathbf{y}_{t+h|t}] = \mathbf{S}\mathbf{b}_t \quad (\text{A.2})$$

Da mesma forma, se espera-se que as previsões reconciliadas não sejam viesadas,

$$\mathbb{E}[\tilde{\mathbf{y}}_{t+h|t}] = \mathbb{E}[\mathbf{y}_{t+h|t}] = \mathbf{S}\mathbf{b}_t \quad (\text{A.3})$$

Substituindo (A.2) em (A.1), temos

$$\tilde{\mathbf{y}}_{t+h|t} = \mathbf{SGS}\mathbf{b}_t \quad (\text{A.4})$$

Logo, para manter a igualdade entre (A.1) e (A.4), $\mathbf{SGS} = \mathbf{S}$

□

Proposição A.2. $\tilde{\mathbf{e}}_t = \mathbf{SG}\hat{\mathbf{e}}_t$.

Demonstração.

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \tilde{\mathbf{y}}_{t+h|t} \quad (\text{A.5})$$

Substituindo (A.1) em (A.5),

$$\tilde{\mathbf{e}}_{t+h|t} = \mathbf{y}_{t+h} - \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.6})$$

Lembrando que, por definição, $\mathbf{y}_{t+h} = \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t}$, então

$$\tilde{\mathbf{e}}_{t+h|t} = \hat{\mathbf{y}}_{t+h|t} + \hat{\mathbf{e}}_{t+h|t} - \mathbf{SG}\hat{\mathbf{y}}_{t+h|t} \quad (\text{A.7})$$

$$= \hat{\mathbf{e}}_{t+h|t} + \hat{\mathbf{y}}_{t+h|t}(\mathbf{I} - \mathbf{SG}) \quad (\text{A.8})$$

Usando a definição novamente, temos que

$$\tilde{e}_{t+h|t} = \hat{e}_{t+h|t} + (y_{t+h} - \hat{e}_{t+h|t})(I - SG) \quad (\text{A.9})$$

$$= y_{t+h} - SGy_{t+h|t} + SG\hat{e}_{t+h|t} \quad (\text{A.10})$$

$$= y_{t+h}(I - SG) + SG\hat{e}_{t+h|t} \quad (\text{A.11})$$

Substituindo (7) em (A.11), temos

$$\tilde{e}_{t+h|t} = Sb_{t+h}(I - SG) + SG\hat{e}_{t+h|t} \quad (\text{A.12})$$

$$= Sb_{t+h} - Sb_{t+h}SG + SG\hat{e}_{t+h|t} \quad (\text{A.13})$$

$$= Sb_{t+h} - (G'S')(b'_{t+h}S') + SG\hat{e}_{t+h|t} \quad (\text{A.14})$$

$$= Sb_{t+h} - SG Sb_{t+h} + SG\hat{e}_{t+h|t} \quad (\text{A.15})$$

Finalmente, pela [condição de ausência de viés em \$\tilde{y}\$](#) , temos que

$$\tilde{e}_{t+h|t} = Sb_{t+h} - Sb_{t+h} + SG\hat{e}_{t+h|t} \quad (\text{A.16})$$

$$= SG\hat{e}_{t+h|t} \quad (\text{A.17})$$

□

Proposição A.3. $Var[\tilde{e}_t] = SG\hat{W}G'S'$.

Demonstração. Por [A.2](#), temos que

$$Var[\tilde{e}] = \mathbb{E}[(SG\hat{e})(SG\hat{e})'] \quad (\text{A.18})$$

$$= \mathbb{E}[SG\hat{e}\hat{e}'G'S'] \quad (\text{A.19})$$

$$= SG\hat{W}G'S' \quad (\text{A.20})$$

Em que \hat{W} é a matriz de variância-covariância dos erros de previsão base. □

Anexos

ANEXO A – CÓDIGO PARA CONSTRUÇÃO DA BASE DE DADOS

```
# pacotes
library(magrittr, include.only = "%>%")

# municípios x regiões imediatas
municipios = basedosdados::read_sql("
SELECT
  id_municipio
  , nome
  , nome_mesorregiao
  , nome_microrregiao
  , nome_regiao
  , nome_regiao_imediata
  , nome_regiao_intermediaria
FROM `basedosdados.br_bd_diretorios_brasil.municipio`
WHERE sigla_uf = 'ES'
")

# estban
estban = basedosdados::read_sql("
SELECT
  CAST(ano AS STRING) AS ano
  , CAST(mes AS STRING) AS mes
  , id_municipio
  , cnpj_agencia
  , CASE
    WHEN id_verbete = '160' THEN 'operações de crédito'
    WHEN id_verbete = '161' THEN 'empréstimos e títulos descontados'
    WHEN id_verbete = '162' THEN 'financiamentos'
    WHEN id_verbete = '163' THEN 'financiamentos rurais'
    WHEN id_verbete = '169' THEN 'financiamentos imobiliários'
    WHEN id_verbete = '172' THEN 'outros créditos'
    WHEN id_verbete = '174' THEN 'provisão para operações de crédito'
    ELSE 'outros'
  END AS verbete
  , valor
```

```

FROM `basedosdados.br_bcb_estban.agencia`

WHERE
  -- CNPJ do Banestes
  cnpj_basico = '28127603'
  -- filtrando verbetes de interesse
  AND id_verbete IN ('161', '162', '163', '169')
")

# formatando datas
estban = within(estban, {
  mes = formatC(as.numeric(mes), format = "d", width = 2, flag = "0")
  ref = as.Date(paste(ano, mes, "01", sep = "-"))
})

# identificando agências em atividade
agencias_fim = subset(estban, ref == max(ref), select = cnpj_agencia) |>
  ((x) unique(x$cnpj_agencia))()

# filtrando apenas agências em atividade
estban = subset(
  estban,
  cnpj_agencia %in% agencias_fim
)

# mesclando com tabela municípios
estban = merge(estban, municipios, by = "id_municipio")

```

Figura 13 – Modelo de dados

