



**Eduardo Takamine Correia**

**SELEÇÃO AUTOMÁTICA DE MODELOS PARA DADOS EM  
PAINEL: UMA APLICAÇÃO PARA DEFINIÇÃO DAS METAS  
DE PERDAS NÃO TÉCNICAS DAS DISTRIBUIDORAS  
BRASILEIRAS**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para  
obtenção do grau de Mestre pelo Programa de Pós-  
Graduação em Metrologia da PUC-Rio.

Orientador: Prof. Rodrigo Flora Calili  
Co-orientador: Prof. José Francisco Moreira Pessanha

Rio de Janeiro  
abril de 2021



**Eduardo Takamine Correia**

**SELEÇÃO AUTOMÁTICA DE MODELOS PARA DADOS EM  
PAINEL: UMA APLICAÇÃO PARA DEFINIÇÃO DAS METAS  
DE PERDAS NÃO TÉCNICAS DAS DISTRIBUIDORAS  
BRASILEIRAS**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Metrologia da PUC-Rio. Aprovada pela Comissão examinadora abaixo.

**Prof. Rodrigo Flora Calili**

Orientador

Pontifícia Universidade Católica (PUC-Rio)

**Prof. José Francisco Moreira Pessanha**

Co-orientador

Centro de Pesquisas de Energia Elétrica (CEPEL)

**Prof. Amaro Olimpio Pereira Junior**

Universidade Federal do Rio de Janeiro (UFRJ)

**Prof. Reinaldo Castro Souza**

Pontifícia Universidade Católica (PUC-Rio)

**Paulo Fernando Mahaz Simões**

Instituto Brasileiro de Geografia e Estatística (IBGE)

Rio de Janeiro, 30 abril de 2021

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Eduardo Takamine Correia**

Graduou-se em Estatística pela Universidade Federal Fluminense em 2017.

#### Ficha Catalográfica

Correia, Eduardo Takamine

Seleção automática de modelos para dados em painel: uma aplicação para definição das metas de perdas não técnicas das distribuidoras brasileiras / Eduardo Takamine Correia; orientador: Rodrigo Flora Calili ; co-orientador: José Francisco Moreira Pessanha. – Rio de Janeiro PUC, Programa de Pós-graduação em Metrologia, 2021.

86 f.; il. ; 30 cm

1. Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Centro Técnico Científico, Programa de Pós-Graduação em Metrologia.

Inclui referências bibliográficas.

1. Metrologia – Teses. 2. Metrologia. 3. Perdas não técnicas. 4. Distribuidoras de energia elétrica. 5. Regulação econômica. 6. Dados em painel. I. Calili, Rodrigo Flora. II. Pessanha, José Francisco Moreira. III. Pontifícia Universidade Católica do Rio de Janeiro. Centro Técnico Científico. Programa de Pós-Graduação em Metrologia. IV. Título.

CDD: 389.1

A minha família,  
Pelo apoio e confiança.

## Agradecimentos

Primeiramente gostaria de agradecer a Deus.

Aos meus orientadores Prof. Rodrigo Flora Calili e o Prof. José Francisco Moreira Pessanha por aceitar conduzir o meu trabalho de pesquisa.

A todos os meus professores do curso de Pós-Graduação em Metrologia pela PUC Rio pela excelência da qualidade técnica de cada um.

Aos meus pais Maria Nancy Takamine Correia e Antero Joaquim de Freitas Correia que sempre estiveram ao meu lado me apoiando ao longo de toda a minha trajetória.

À minha esposa Lídia Pinto Takamine Correia pela compreensão e paciência demonstrada durante o período do projeto.

Aos meus colegas da PUC-Rio.

Aos professores que participaram da Comissão examinadora.

A todos funcionários do Departamento pelos ensinamentos e pela ajuda.

O presente trabalho foi desenvolvido dentro do Projeto de P&D PUC-Rio e Enel Brasil que busca estabelecer modelos alternativos para definição de metas de inadimplência e perdas não técnicas regulatórias. Assim agradeço a equipe de P&D da Enel e a todos os pesquisadores envolvidos no projeto pelo apoio.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Resumo

Correia, Eduardo Takamine; Calili, Rodrigo Flora; Pessanha, José Francisco Moreira. **Seleção automática de modelos para dados em painel: uma aplicação para definição das metas de perdas não técnicas das distribuidoras brasileiras.** Rio de Janeiro, 2021. 86p. Dissertação de Mestrado - Programa de Pós-graduação em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro.

A perdas não técnicas de energia elétrica (PNT) constituem, atualmente, um dos principais problemas enfrentados pelas concessionárias de distribuição de eletricidade no Brasil. Entende-se que parte desses fatores não são gerenciáveis pelas distribuidoras, e assim, são repassadas aos consumidores através de tarifas de eletricidade, seguindo os limites definidos pela Agência Nacional de Energia Elétrica (ANEEL). Apesar do progresso realizado pela ANEEL, a metodologia atualmente empregada pela Agência utiliza-se de dados em painel para formulações desses limites, no entanto, define-se algumas medidas *ad-hoc* que acabam resultando em metas de perdas pouco condizentes com a realidade das distribuidoras brasileiras. Essa dissertação tem por objetivo selecionar modelos de forma automática para dados em painel, avaliando sua aplicabilidade na definição das metas de perdas não técnicas. Os modelos propostos foram empregados a um painel de dados contendo observações anuais de 62 concessionárias de distribuição de energia elétrica brasileira no período de 2007 a 2017, gerando, assim, 1.097.789 modelos com respeitando os tipos de regressões em dados em painel, resultando em 3 modelos potenciais com metas de PNT mais aderentes às PNT reais das distribuidoras brasileiras.

## Palavras-chaves

Metrologia, Perdas não técnicas; Distribuidoras de energia elétrica; Regulação econômica; Dados em Painel

## Abstract

Correia, Eduardo Takamine; Calili, Rodrigo Flora (Advisor); Pessanha, José Francisco Moreira (Co-Advisor). **Automatic selection of models for panel data: an application for setting non-technical losses' targets of Brazilian electricity distribution utilities**. Rio de Janeiro, 2021. 86p. Dissertação de Mestrado - Programa de Pós-graduação em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro.

Non-technical electricity losses (NTL) are currently one of the main problems faced by electricity distribution utilities in Brazil. It is understood that part of these factors is not manageable by the utilities, and thus, they are passed on to consumers through electricity tariffs, following the limits defined by the Brazilian Electric Energy Agency (ANEEL). Despite the progress made by ANEEL, the methodology currently used by the Agency uses panel data to formulate these limits, however, some ad hoc measures are defined which end up resulting in loss targets that are not consistent with the reality of Brazilian distribution utilities. This dissertation aims to select models automatically for panel data, evaluating their applicability in setting non-technical loss targets. The proposed models were used in a panel data containing annual observations from 62 Brazilian electricity distribution utilities in the period from 2007 to 2017, thus generating 1,097,789 models with respect to the types of regressions in panel data, resulting in 3 potential models with NTL targets more adherent to the real possibilities of the Brazilian distribution utilities.

## KEYWORDS

Metrology, Non-technical losses; Electricity utilities; Economic regulation; Panel data

## Sumário

1	Introdução	13
1.1	Definição do problema de pesquisa	18
1.2	Objetivos: geral e específicos	18
1.3	Metodologia	19
1.3.1	Fase exploratória e descritiva	20
1.3.2	Fase de pesquisa aplicada	23
1.3.3	Fase conclusiva	23
1.4	Estrutura da dissertação	23
2	Perdas não técnicas: uma análise crítica do modelo brasileiro de energia elétrica	25
2.1	A regulação e a teoria da competição por padrões ( <i>Yardstick Competition</i> )	25
2.2	Perdas Não Técnicas no Brasil: O problema estrutural	29
2.3	A regulação das Perdas Não Técnicas no Brasil	32
2.3.1	Histórico da metodologia empregada pela ANEEL	34
2.4	Metodologia atualmente adotada pela ANEEL	35
2.4.1	Variáveis utilizadas na composição do Índice de Complexidade	35
2.4.2	Críticas as variáveis utilizadas	37
2.4.3	Modelos utilizados atualmente pela ANEEL	41
2.4.4	Críticas aos Modelos utilizados ANEEL	42
2.4.5	Matriz de Probabilidade para fins de cálculo das metas de PNT	46
2.4.6	Crítica à Matriz de Probabilidade	49
2.4.7	Consulta Pública nº 029 de 2020	49
2.5	Considerações sobre o capítulo	54
3	Proposta de seleção automática de modelos para dados em painel	55
3.1	Seleção automática de modelos	55
3.2	Dados em Painel	57
3.3	Considerações sobre o capítulo	61
4	Metodologia para seleção automática de modelos para dados em painel	63
4.1	Visão geral da metodologia proposta	63
4.2	Fase 1: Geração de possíveis modelos.	65
4.3	Fase 2: Classificação dos modelos em dados em painel	65



4.4 Fase 3: Requisitos de escolha dos modelos	65
4.5 Considerações sobre o capítulo	68
5 Demonstração da aplicabilidade da metodologia proposta para o problema da perda não técnica no Brasil	70
5.1 Resultado da metodologia seleção automática	70
5.2 Utilização dos pesos de Akaike	72
5.3 Predição dos modelos	73
5.4 Discussão dos resultados	76
6 Conclusões	78
7 Referências bibliográficas	80

## Lista de Figuras

Figura 1.1– Representatividade dos Custos das Perdas Não Técnicas Regulatórias sobre a Receita Requerida	14
Figura 1.2 - Desenho da pesquisa, seus componentes e métodos	20
Figura 1.3 - Mapa conceitual da pesquisa	22
Figura 2.1 - Perdas Não Técnicas no Brasil (% da Energia Injetada Total)	29
Figura 2.2 - Participação das perdas não técnicas regulatória sobre a Parcela B (2018)	30
Figura 2.3 - Perdas Não Técnicas no Brasil (% da Energia Injetada Total)	31
Figura 2.4 – Procedimentos para a escolha dos modelos de dados em painel	43
Figura 2.5 – Processo de decisão dos modelos de dados em painel utilizados em PNT	44
Figura 2.6 – Processo de inferência dos modelos de PNT	45
Figura 3.1 – Pesquisa exaustiva: modelos alternativos a serem avaliados em números diferentes de covariáveis inicialmente admissíveis	56
Figura 3.2 - Processo de decisão da escolha adequada do modelo de dados em painel	60
Figura 4.1 - Processo de decisão da escolha adequada do modelo de dados em painel	64
Figura 5.1 – Fluxograma da seleção dos modelos	70
Figura 5.2 – Importância das variáveis explicativas	72

## Lista de Quadros

Quadro 2.1 – Variáveis utilizadas nos modelos ANEEL.....	36
Quadro 2.2 – Variáveis utilizadas para a construção do índice de complexidade (IC) .....	37
Quadro 2.3 – Resumo das variáveis disponibilizadas na CP 29/2020 .....	50
Quadro 2.4 - Variáveis Socioeconômicas testadas pela ANEEL na CP 029/2020 .....	51
Quadro 2.5 – Critério para a definição do Ponto de Partida das PNT regulatórias sobre o mercado de baixa tensão (BT) proposto na CP 029/2020 .....	53
Quadro 2.6 - Critério para a definição dos limites anuais de redução das PNT regulatórias sobre o mercado de baixa tensão (BT) proposto na CP 029/2020 .....	53
Quadro 4.1 – Correlação com o sinal esperado inconsistente .....	67
Quadro 5.1 – Sínteses das distribuidoras que tiveram os melhores e piores resultados dos modelos ao comparar suas perdas 2017 .....	75

## Lista de Tabelas

Tabela 2.1 – Base de dados das variáveis da ENEL Rio	37
Tabela 2.2 – Percentual do Saldo de operações de créditos para PF e PJ, assim como suas respectivas taxas de inadimplência, tanto para o RJ, quanto para a média nacional	39
Tabela 2.3 – Índice de violência na área de concessão da ENEL Rio	40
Tabela 2.4 – Critério de seleção de modelos adotado pela ANEEL na CP 029/2020	51
Tabela 4.1 – Resumo das variáveis e os sinais esperados dos coeficientes	66
Tabela 5.1 – Resultado do conjunto de variáveis significativas	71
Tabela 5.2 – Resultado dos modelos com efeitos aleatórios	72
Tabela 5.3 – Resultado dos modelos com efeitos fixos	72
Tabela 5.4 – Resultado preditivo das distribuidoras de energia elétrica	73

# 1 Introdução

Os sistemas elétricos de potência são tradicionalmente organizados em três segmentos: geração, transmissão, comercialização e distribuição. As perdas referem-se à energia elétrica gerada que passa pelas linhas de transmissão e redes da distribuição, mas que não chegam a ser comercializadas, seja por motivos técnicos ou comerciais.

As perdas na rede de distribuição são aquelas que ocorrem dentro do próprio sistema de distribuição e podem ser divididas em duas categorias, conforme sua causa: as perdas técnicas (PT), que são as perdas inerentes ao transporte da energia elétrica na rede, sendo relacionadas à transformação de energia elétrica em energia térmica nos condutores (efeito Joule), perdas nos núcleos dos transformadores, perdas dielétricas etc., podendo ser entendidas como o consumo dos equipamentos responsáveis pela distribuição de energia; e as perdas não técnicas (PNT), que correspondem à diferença entre as perdas totais e as perdas técnicas, considerando, portanto, todas as demais perdas associadas à distribuição de energia elétrica, tais como furtos de energia, erros de medição, erros no processo de faturamento, unidades consumidoras sem equipamento de medição etc. Esse tipo de perda está diretamente associado à gestão comercial da distribuidora.

Atualmente, a redução das perdas não técnicas de energia constitui um dos maiores desafios enfrentado por muitas concessionárias de distribuição de energia elétrica em países em desenvolvimento, em particular em alguns países da África, América Latina e sul da Ásia (Penin, 2008; Smith, 2004; Amin *et al.*, 2015; WORLD BANK, 2009; Millard, Emmerton e Emmerton, 2009; Fourie, 2005; Arango *et al.*, 2017), onde os níveis atuais dessas perdas representam um risco não apenas para a solvência das empresas de distribuição de eletricidade, mas também para a própria segurança do fornecimento de energia (Depuru *et al.*, 2011; Smith, 2004).

As distribuidoras de energia elétrica de todo o mundo estão perdendo cerca de US\$ 96 bilhões a cada ano por furto de energia elétrica, fraude e erros de

faturamento. Os efeitos das chamadas perdas não técnicas podem ser "irreversíveis" e levar ao aumento dos preços da eletricidade para clientes, gerando, assim, instabilidade financeira/operacional para as distribuidoras. (Northeast Group, 2017)

No Brasil, em 2018, o custo das perdas técnicas, valoradas pelo preço médio da energia nos processos tarifários, sem considerar tributos, é da ordem de R\$ 7,1 bilhões. Essas perdas, inevitáveis em qualquer sistema de distribuição, são repassadas aos consumidores, já se considerando a operação eficiente das redes e, portanto, não são passíveis de maiores reduções (ANEEL, 2019).

As perdas não técnicas reais no país representaram um custo de aproximadamente R\$ 6,6 bilhões. No entanto, as perdas não técnicas regulatórias, que são calculadas conforme a metodologia da Agência Nacional de Energia Elétrica (ANEEL), alcançou um custo da ordem de R\$ 5 bilhões ao ano, o que representa aos consumidores cerca de 3% do valor da tarifa de energia elétrica, variando por distribuidora, conforme apresentado na Figura 1.1 (ANEEL, 2019).

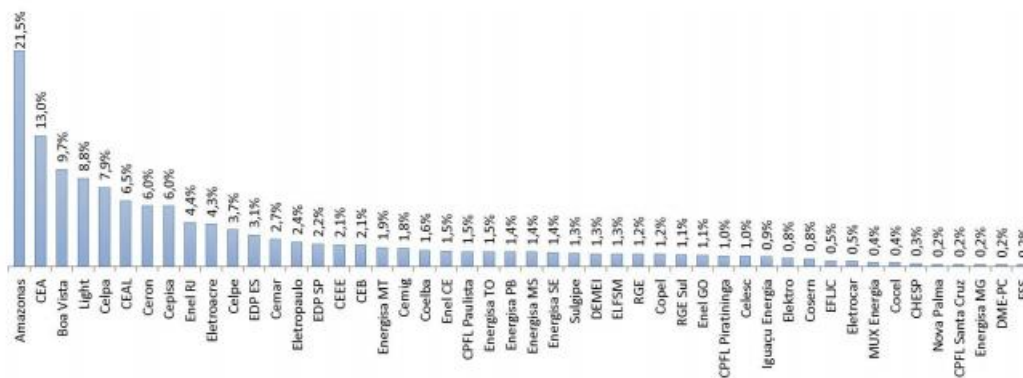


Figura 1.1– Representatividade dos Custos das Perdas Não Técnicas Regulatórias sobre a Receita Requerida

Fonte: ANEEL, 2019

A redução das PNT é um grande desafio enfrentado pelas concessionárias de distribuição de energia elétrica, sobretudo nos países em desenvolvimento. Uma das soluções para o problema abrange a implementação de inovações tecnológicas como as Redes Elétricas Inteligentes (REI) e a *Advanced Metering Infrastructure* (AMI) (Gopi e Lalu, 2016) (Mohassel *et al.*, 2014). Além disso, envolve a formulação de novos mecanismos regulatórios (Simões *et al.*, 2020), baseados em

análise de dados. Por exemplo, a Nota Técnica (NT) nº 106/2015 SGT/SRM/ANEEL, na qual a Agência Nacional de Energia Elétrica (ANEEL), responsável pela regulação do setor elétrico brasileiro, descreve a metodologia e critérios adotados no cálculo regulatório das metas de PNT para as distribuidoras de energia elétrica do Brasil.

Embora a metodologia usada pela ANEEL represente um avanço notável e introduza inovações na gestão das concessionárias de distribuição, inclusive pelo uso de modelos econométricos para dados em painel, a metodologia usada pela agência reguladora ainda possui elementos *ad-hoc*, nos quais observa-se a atuação do poder discricionário do regulador (Calili *et al.*, 2017).

Na metodologia de PNT, utilizam-se variáveis socioeconômicas e de mercado para o período de 2003 a 2013, formando assim, uma base de dados em painel para todas as concessionárias visando a obtenção de três modelos econométricos (C, G e K) para determinação do Índice de Complexidade (IC) que reflita a dificuldade do combate às PNT.

A ANEEL, no submódulo 2.6 do PRORET (procedimento de regulação tarifária), que trata das regras das perdas não técnicas, coloca que há espaços metodológicos que tentaram ser preenchidos, inicialmente, com a Consulta Pública (CP) nº 11/2013 e nas 1ª e 2ª etapas da Audiência Pública (AP) nº 023/2014, uma vez que no primeiro ciclo de revisão tarifária foi utilizado um procedimento “univariado”, ou seja, a meta de cada distribuidora foi estabelecida a partir da média aritmética simples da série temporal anual das PNT dos últimos anos. Tal metodologia reconheceu níveis mínimos aceitáveis de PNT ao estabelecer as metas a patamares observados historicamente (Calili *et al.*, 2017).

A ANEEL (2010, 2011, 2013a, 2014a, 2014b, 2020a) tem tratado o tema das perdas não técnicas seguindo os princípios preconizados na regulação por comparação (*Yardstick Competition*) de forma sistemática ao longo da última década. Para tentar distinguir as ineficiências das distribuidoras nas ações de combate a perdas não técnicas das limitações impostas pelas condições socioeconômicas de cada área de concessão, a agência desenvolveu um conjunto de modelos estatísticos baseado em dados longitudinais que permite, entre outras

coisas, comparar as perdas não técnicas das diversas distribuidoras do país e estabelecer qual volume deve ser reconhecido para cada distribuidora e repassado aos consumidores nas tarifas de energia elétrica. No entanto, de acordo como Calili *et al.* (2017), a metodologia empregada pela ANEEL apresenta diversas limitações que acabam por resultar em distorções relevantes no processo de reconhecimento tarifário das perdas não técnicas. A principal crítica apontada pelos autores do referido estudo é que à metodologia atualmente empregada pela ANEEL é excessivamente complexa, frágil do ponto de vista do metodológico e, em certa medida incompleta, o que impõem ao órgão regulador a adoção de uma série de medidas *ad hoc* que acaba resultando em metas de perdas pouco condizentes com a realidade das distribuidoras.

Em abril de 2020 a ANEEL abriu uma consulta pública (ANEEL, 2020a) com o intuito de reformular e atualizar os modelos atualmente empregados na definição das metas regulatórias de PNT. Embora reconheça as limitações do modelo atual, a proposta da ANEEL altera apenas de forma marginal a metodologia, mantendo-se validas as críticas apontadas por Calili *et al.* (2017). Já Antmann (2009) reforça esse problema, no Brasil, há 63 empresas de distribuição de eletricidade fiscalizadas pelo mesmo regulador nacional e que enfrentam enormes diferenças no tamanho de mercado, cobertura geográfica e outros parâmetros-chaves. A menor concessionária, por exemplo, atende cerca de 10.000 clientes em pequenas cidades ou áreas rurais, enquanto as duas maiores atendem mais de 7 milhões de clientes. (Antmann, 2009).

Uma das críticas sobre os métodos adotados pelo regulador brasileiro para calcular perdas não técnicas é o mesmo tratamento que se dá a todas as empresas de distribuição no processo de revisão de tarifário, embora a obtenção de perdas menores possa não ser viável para algumas empresas, considerando as áreas específicas de atuação e características do sistema (Mimmi *et al.*, 2015)

Recentemente, (Simões *et al.*, 2020) analisaram a evolução das PNT experimentadas pelas distribuidoras de energia brasileiras ao longo do período de 2004 a 2012. A análise foi conduzida por meio de modelos de regressão para dados longitudinais com efeitos mistos e considerou um amplo conjunto de potenciais variáveis explicativas e diferentes alternativas de modelagem da estrutura de autocorrelação dos erros, por exemplo, erros condicionalmente independentes e



com parametrizações ARMA [p,q]. Em outro trabalho, Leite *et al.*, (2020a), apresentam um modelo de fronteira estocástico (SFA, que advém do acrônimo em língua inglesa *Stochastic Frontier Analysis*) para dados em painel cuja equação é especificada para fornecer os limites toleráveis para as PNT, as metas regulatórias.

Assim, com o intuito de reformular e atualizar os modelos regulatórios atualmente, a ANEEL abriu a Consulta Pública nº 29/2020, disponibilizando uma base de dados com 22 variáveis no período de 2004 a 2017. Neste sentido, a busca exhaustiva de combinações de variáveis explicativas que possam explicar o problema das perdas não técnicas, bem como a definição das metas de PNT requer avaliação de um número enorme de modelos por meio de testes de hipóteses e métricas de qualidade de ajuste disponíveis na literatura. Desta forma, pautada pelas variáveis disponibilizadas pela ANEEL, é possível fazer uma busca exaustivas das variáveis e avaliar o modelo ótimo que se relaciona a problemática de PNT.

Dado um critério de seleção, procedimentos simples de seleção de variável, como seleção direta, eliminação reversa e procedimento gradual (Draper e Smith, 1998) estão disponíveis para análise de regressão linear em muitos pacotes de *software* estatístico padrão. Miller (1990) forneceu detalhes desses procedimentos. Ao examinar apenas uma pequena parte de todos os possíveis modelos, esses procedimentos são fáceis e rápidos. Mas o modelo selecionado muitas vezes está longe do um ótimo. Por outro lado, a busca exhaustiva de subconjuntos sempre pode encontrar o modelo ideal sob um determinado critério, mas a complexidade envolvida rapidamente se torna intransponível à medida que o número de variáveis explicativas aumenta (Cai *et al.* 2009).

Ripley (2003) já relatava a dificuldade de lidar com muitas variáveis no modelo de regressão. Com o aumento exponencial nas capacidades de memória e processamento, a pesquisa da busca exhaustiva passou a ser mais utilizada. Alguns autores, Igarashi *et al.* (2018) e Wang *et al.* (2019), criaram pacotes computacionais que utilizam a busca exhaustiva para encontrar modelos econométricos ótimos. A metodologia proposta foi inspirada no pacote *glmulti* (Calcagno e Mazancourt, 2010) – *Software R* – tem como objetivo a seleção automática de modelos de regressão. Ressalta-se que uma abordagem similar utilizada para este fim foi

proposta por (Silva e Pessanha, 2020) na busca por indicadores contábeis e financeiros relevantes para a previsão de insolvência das distribuidoras.

Diante do amplo número de potenciais variáveis explicativas capazes de descrever o fenômeno de perdas não técnicas, a presente pesquisa tem por objetivo cobrir essa lacuna, propondo uma metodologia para seleção automática de modelos para dados em painel podendo ser aplicável para a definição das metas de perdas não técnicas.

Esta dissertação insere-se na linha de pesquisa “Redes Inteligentes (*Smart Grid*)” do Programa de Pós-graduação em Metrologia (PósMQI) da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

## **1.1 Definição do problema de pesquisa**

Considerando a dimensão problemática crescente de perdas nas redes de distribuição de energia elétrica em todo o mundo, o desafio de contribuir para avaliação mais robusta das variáveis utilizadas e a seleção de um ou mais modelos para regulação de perdas não técnicas, definiu-se a seguinte questão principal a ser respondida ao longo da pesquisa:

“Como selecionar as variáveis mais relevantes para o fenômeno de PNT e definir modelos que possam auxiliar no estabelecimento das metas regulatórias de redução de perdas não técnicas para as distribuidoras de energia elétrica brasileiras, equilibrando o princípio da regulação por incentivo e a complexidade da área de atuação de cada distribuidora?”

## **1.2 Objetivos: geral e específicos**

Buscando dar uma contribuição para analisar de forma mais assertiva o cálculo das metas de perdas não técnicas de energia elétrica, o objetivo geral da dissertação é selecionar as variáveis mais relevantes correlacionadas com o fenômeno de perdas não técnicas, por meio da busca exaustiva de modelos de dados em painel.

Em termos específicos, a dissertação busca:

- Identificar as principais limitações que podem ser identificadas nos modelos atuais para definição das metas regulatórias de perdas não técnicas de energia elétrica no Brasil;
- Criar uma rotina para busca exaustiva de modelos de dados em painel, visando selecionar as mais relevantes variáveis que causam maior impacto no fenômeno de perdas não técnicas;
- Desenvolver modelo(s) auxiliares(s) ao utilizado pela ANEEL para definição das metas regulatórias de perdas não técnicas para as distribuidoras de energia elétrica no Brasil.
- Demonstrar aplicabilidade rotina para busca exaustiva de modelos de dados em painel no contexto do setor de distribuição elétrica no Brasil, explicitando seus diferenciais em relação ao modelo atualmente adotado pela ANEEL e propondo recomendações para a agência.

### **1.3 Metodologia**

Segundo a taxonomia proposta por Vergara (2002), a pesquisa pode ser considerada aplicada, descritiva e metodológica (quanto aos fins). Quanto aos meios de investigação, a metodologia compreende:

- Pesquisa bibliográfica, documental, experimental e estudo de caso.
- Automação de variáveis para modelos
- A partir de uma seleção exaustiva de combinações possíveis das variáveis, desenvolver modelo(s) auxiliares(s) em dados em painel, para definição das metas de PNT regulatória das distribuidoras de energia elétrica do Brasil.

Estudo empírico visando demonstrar a aplicabilidade do modelo no contexto de um projeto de P&D Enel-PUC-Rio.

A Figura 1.2 apresenta o desenho da pesquisa, destacando-se seus componentes e métodos, de acordo com três fases principais: (i) exploratória e descritiva; (ii) pesquisa aplicada; e (iii) conclusiva.

Detalham-se, a seguir, o desenvolvimento de cada fase e os resultados esperados em cada bloco da Figura 1.2. Destaca-se que o desenho da pesquisa foi desenvolvido em total alinhamento com os objetivos enunciados na seção 1.2 deste capítulo.



Figura 1.2 - Desenho da pesquisa, seus componentes e métodos

Fonte: Elaboração própria.

### 1.3.1 Fase exploratória e descritiva

A fase exploratória e descritiva foi iniciada com pesquisa bibliográfica e documental, com o objetivo de levantar trabalhos conceituais e documentos de referência para delimitação do tema central da pesquisa. Em seguida, aprofundou-se a revisão bibliográfica, buscando identificar estudos sobre a busca exaustiva das combinações de variáveis focalizando-se nos modelos em dados em painel que estavam sendo adotados com esse propósito e os objetos de análise. Além disso, avaliou-se as utilizações dos testes em modelos em dados de painel e estatísticas de

aderência para estar em parcimônia com os resultados. Por fim, a concepção preliminar de um modelo para auxiliar a definição das metas regulatórias de perdas não técnicas das distribuidoras de energia elétrica no Brasil.

Apresenta-se na Figura 1.3 uma visão geral e esquemática dos resultados desta primeira fase.

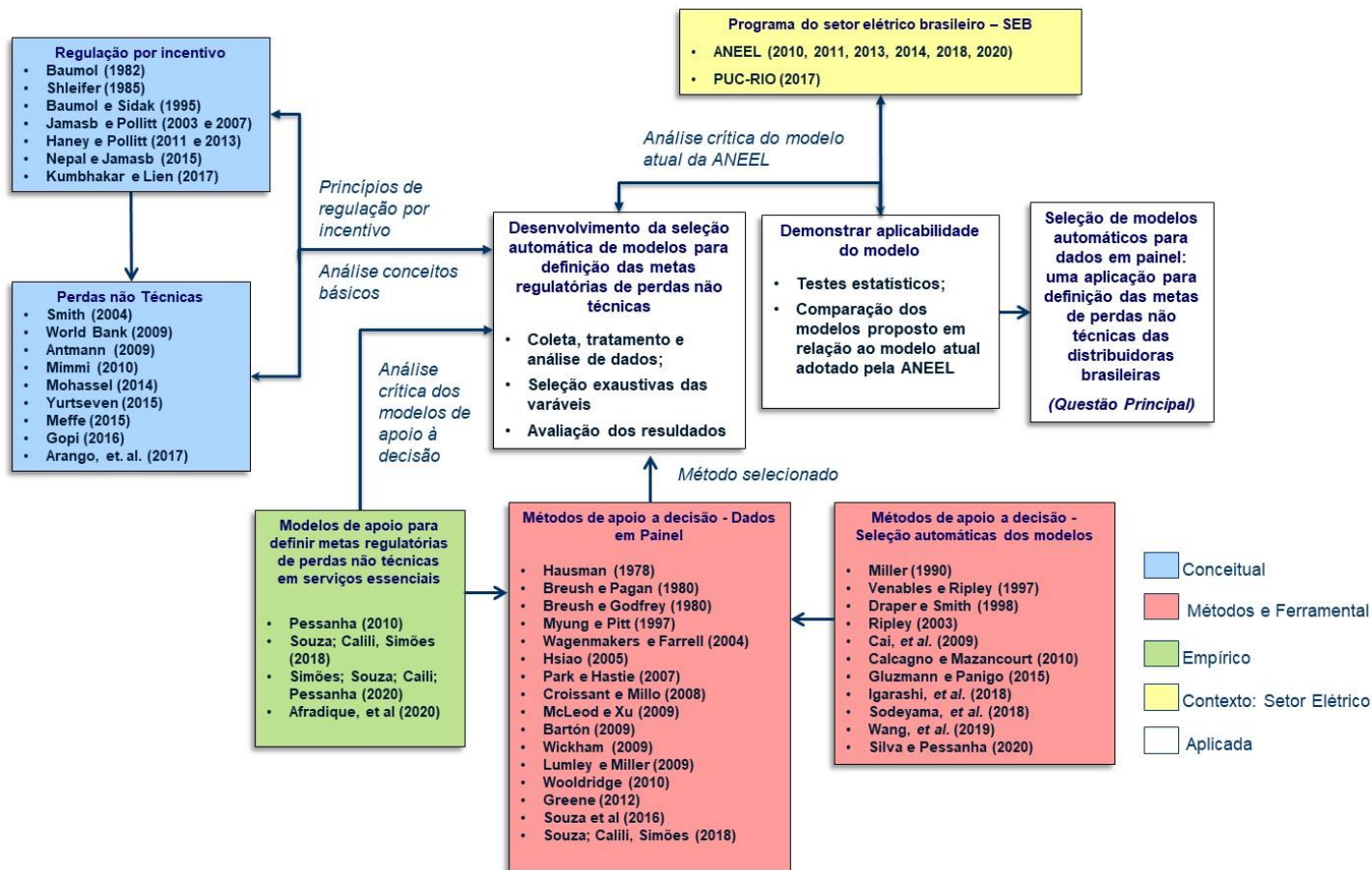


Figura 1.3 - Mapa conceitual da pesquisa  
 Fonte: Elaboração própria.

O referencial teórico constituiu uma orientação conceitual para a pesquisa, ao compor o vocabulário especializado e organizar o conhecimento sobre avaliação e seleção de tecnologias para geração de energia elétrica a partir de fontes renováveis.

### **1.3.2 Fase de pesquisa aplicada**

Desenvolveu-se a fase da pesquisa aplicada propriamente dita, a partir dos resultados da revisão bibliográfica e documental, que nortearam a definição dos objetivos da pesquisa e o desenvolvimento do modelo de perdas não técnicas. Esta fase foi realizada para o caso brasileiro, considerado os dados definidos na consulta pública nº 029/2020. A partir da coleta, tratamento e análise das variáveis disponibilizada pela ANEEL, avalia-se combinações para modelo dados em painel sendo empilhado, aleatório ou fixo, a depender dos testes de hipóteses, e verificando-se o resultado por meio das estatísticas Akaike e  $R^2$  para selecionar os modelos com maior relevância. Modelos mais relevantes. Por fim, estima-se os resultados dos modelos selecionados e compara com os valores reais.

### **1.3.3 Fase conclusiva**

Na terceira fase, elaboraram-se a conclusão geral e as específicas em relação a cada um dos objetivos enunciados na seção 1.2, formularam-se um conjunto de recomendações aos diversos atores interessados na aplicação do modelo de perdas não técnicas.

## **1.4 Estrutura da dissertação**

Além desta introdução, capítulo 1, a dissertação encontra-se estruturada em seis capítulos.

O capítulo 2 apresenta uma breve revisão do conceito que norteia a regulação por incentivo, assim como os principais conceitos e desafios em relação ao tema de perdas não técnicas, em particular no Brasil. É apresentado também o modelo utilizado atualmente pela Agência Nacional de Energia Elétrica (ANEEL) para a definição das metas regulatórias de perdas não técnicas, assim como suas principais limitações.

No capítulo 3, expõem-se a metodologia proposta na busca exaustiva das variáveis e a verificação dos modelos em dados em painel. Além disso, apresenta-se o referencial teórico que fundamentou a proposição de um modelo alternativo para auxiliar a definição das metas regulatórias de perdas não técnicas para as distribuidoras de energia elétrica.

O capítulo 4 apresenta a metodologia proposta da busca exaustiva para os modelos em dados em painel e todas suas etapas até as propostas para auxílio da definição das metas regulatórias de perdas não técnicas para as distribuidoras de energia elétrica, construído a partir do referencial teórico apresentado no capítulo 3. No capítulo 5, apresenta-se a aplicabilidade apresentada no capítulo 4 e os principais resultados dos modelos propostos ao caso brasileiro, destacando-se a aderência desses resultados à realidade local.

No capítulo 6, apresentam-se as conclusões do estudo, demonstrando a aplicabilidade do modelo proposto ao Brasil, assim como suas vantagens em relação ao atual cenário real. Além disso, nesse capítulo, são apresentadas recomendações para estudos futuros.



## 2

### **Perdas não técnicas: uma análise crítica do modelo brasileiro de energia elétrica**

Nesta seção, apresentam-se os principais pontos sobre a regulação por incentivo e os desafios do agente regulador. Na segunda parte é feita uma contextualização do cenário de perdas não técnicas (PNT) no Brasil, sua evolução na metodologia de regulação e os principais desafios para o seu enfrentamento. Por fim, é apresentado o modelo com o qual o ANEEL vem enfrentando o tema de PNT no Brasil e suas principais limitações.

#### **2.1**

##### **A regulação e a teoria da competição por padrões (*Yardstick Competition*)**

Em qualquer atividade econômica, parte das perdas no processo produtivo, é repassada ao preço final dos produtos/serviço. Em um ambiente competitivo, a capacidade de um determinado agente de repassar os seus custos individuais com as perdas é limitada pelo nível médio de perdas praticado pelos demais agentes. Essa limitação decorre do simples fato de que em um ambiente competitivo, os consumidores poderiam, a princípio, optar entre diversos fornecedores (Baumol, 1982; Baumol e Sidak, 1995; ANEEL, 2010).

A “ameaça” imposta pela concorrência (*perfect contestable markets*) cria um incentivo permanente para que os agentes busquem um aumento da eficiência produtiva e, conseqüentemente, a redução do nível de perdas, assegurando a maximização da eficiência econômica (eficiência produtiva e alocativa) e conseqüentemente do bem-estar social (Baumol, 1982; Baumol e Sidak, 1995).

A Teoria dos Mercados Contestáveis desenvolvida por Baumol (1982), no entanto, pressupõe uma série de condições para que de fato esse mecanismo de

mercado funcione, entre elas a inexistência de custos irrecuperáveis (*sunk cost*) (Baumol, 1982; Baumol e Sidak, 1995).

Em uma indústria com características de “rede”, como é o caso do setor elétrico, a introdução da competição para toda a extensão da cadeia produtiva não é viável (Baumol, 1982). A presença de elevadas economias de escala (custos fixos altos e custos marginais baixos) e de custos irrecuperáveis (*sunk cost*), principalmente nas etapas de transmissão e distribuição de energia elétrica, impedem a entrada de novos concorrentes, inviabilizando o funcionamento adequado do mecanismo de “auto regulação” do mercado exercido pela concorrência (Baumol, 1982).

Em cenários como esse, segundo Baumol (1982), é necessário a criação de mecanismos adicionais que estimulem as empresas a se tornarem mais eficientes, de forma a se obter um resultado próximo ao que seria obtido em um mercado competitivo (Haney e Pollitt, 2009, 2011 e 2013). Assim, a regulação do setor, através da intervenção do Estado, seria então fundamental (Baumol *et al.*, 1982; Shleifer, 1985; Baumol e Sidak, 1995, J; Haney e Pollitt, 2009, 2011 e 2013).

A regulação por incentivo busca introduzir o componente “produtividade” associado à redução nos custos como um dos principais objetivos das entidades reguladoras. Há muitos mecanismos de regulação que buscam atender a esse objetivo. Alguns mecanismos usualmente utilizados estão mais direcionados para a eficiência produtiva, como o *Price Cap* (em que são estabelecidos preços máximos a serem cobrados dos consumidores finais), e outros mais para a eficiência alocativa, como *Revenue Cap* (em que é definida uma receita máxima a ser auferida pelo agente detentor da concessão).

Na maior parte dos países, inclusive o Brasil, no entanto, o que tem ocorrido é adoção de um modelo híbrido, por parte do regulador, conjugando características tanto do modelo de *Price Cap* e *Revenue Cap*. No entanto, independente do mecanismo, cabe ao regulador definir as tarifas máximas que serão cobradas dos consumidores finais, definindo em última instancia o nível máximo de custos que pode ser repassado aos consumidores.

O desafio nesse processo é como definir esses preços de forma a manter os incentivos econômicos que sejam capazes de estimular os agentes regulados a reduzirem custos, inovar e melhorar a qualidade dos serviços (Haney e Pollitt, 2009, 2011 e 2013).

O reconhecimento tarifário integral dos custos incorridos pelas empresas reguladas nas tarifas a serem cobradas dos consumidores finais reduz os estímulos das empresas reguladas na busca da maior eficiência (Shleifer, 1985; Kumbhakar e Lien, 2017). Por outro lado, o não reconhecimento dos custos incorridos pelas empresas reguladas, particularmente aqueles não gerenciáveis por essas empresas, pode reduzir a atratividade do setor, conduzindo a um cenário de redução de investimento e conseqüentemente perda da qualidade do serviço prestado pelas concessionárias (Shleifer, 1985, Kumbhakar e Lien, 2017). Neste regime, a empresa regulada pode recuperar os custos incorridos enquanto os clientes estão protegidos de pagar a mais para os serviços de monopólio. Como qualquer regime de regulamentação, entretanto, a regulamentação da taxa de retorno tem suas deficiências (Kuosmanen e Nguyen, 2019). O efeito Averch-Johnson é resultado da determinação de uma taxa de retorno aos investimentos superior ao custo de oportunidade do mercado. Nesta lógica, o custo de oportunidade do capital torna-se menor do que o incorrido nos demais investimentos, dada a vantagem da remuneração garantida (Kuosmanen e Nguyen, 2019).

A Teoria de competição por padrões (*Yardstick Competition*) desenvolvida por Shleifer (1985) sugere que a melhor forma de definir de forma eficiente o repasse de custos das empresas para os consumidores seria através da comparação, seja entre os diversos agentes regulados do setor, seja na comparação com uma empresa “fictícia”.

O modelo de comparação emularia a dinâmica de mercado em regime de concorrências, evitando o problema clássico de assimetria de informações entre reguladores e agentes regulados, assegurando a maximização da eficiência econômica (Baumol, 1982; Shleifer, 1985; Baumol e Sidak, 1995; Haney e Pollitt, 2009, 2011 e 2013).

Ao definir os custos a serem repassados ao consumidor com base nos custos médios obtidos através de um modelo de comparação (*Benchmarking*), o regulador

criaria um mecanismo de incentivo claro a eficiência, punindo as empresas menos eficientes (com a redução dos seus ganhos) e beneficiando as empresas mais eficientes (com ampliação dos seus ganhos). Shleifer (1985) demonstra que mesmo onde haja uma relativa heterogeneidade entre as empresas reguladas, o modelo de competição por comparação ainda pode ser adotado desde que essas heterogeneidades sejam tratadas pelo regulador dentro do processo (Shleifer, 1985)

Os métodos de regulação por comparação (*Benchmarking*) vêm sendo bastante empregados por reguladores de todo mundo durante os processos de definição ou revisão de tarifas dos mais diversos setores (Haney e Pollitt, 2009, 2011 e 2013).

A forma, no entanto, como os resultados desses processos vem sendo aplicados varia muito de regulador para regulador (Haney e Pollitt, 2011, Jamasb e Pollitt, 2002, 2007; Haney e Pollitt, 2009). Há também uma grande variedade de técnicas de comparação empregadas nesse processo, bem como de parâmetros de comparação adotados.

Segundo Haney e Pollitt (2009,2011) os métodos que vem sendo mais empregados pelos reguladores no mundo são: o uso do *Data Envelopment Analysis* (DEA), Mínimos Quadrados Corrigidos (COLS), Análise de Fronteira Estocástica (SFA), Análise de Dados em Painel e o método de benchmarking por processo/atividade (ex. modelo de Empresa de Referência que foi utilizado no Brasil até 2010).

A escolha do método é, no entanto, bastante complexa e depende em geral do contexto do setor e do país onde ele está sendo aplicado (Agrell e Bogetoft, 2016; Haney e Pollitt, 2011). Vale ressaltar que o emprego de diferentes métodos de comparação pode gerar diferentes resultados.

Uma discussão sobre a metodologia empregada pela Agência Nacional de Energia Elétrica (ANEEL), que utiliza a regulação por incentivo, bem como suas implicações para o contexto das perdas não técnicas de energia no Brasil será apresentada nas próximas seções desse capítulo.

## 2.2

### Perdas Não Técnicas no Brasil: O problema estrutural

No Brasil, as perdas não técnicas de energia elétrica (PNT) tem se mantido relativamente estáveis ao longo da última década apesar dos fortes investimentos realizados pelas distribuidoras do país para combatê-las. Em 2018 as PNT no país totalizaram 6,5% da energia total injetada no sistema de distribuição, valor bastem similar aos 6,4% observados em 2008, uma década antes (Figura 2.1).

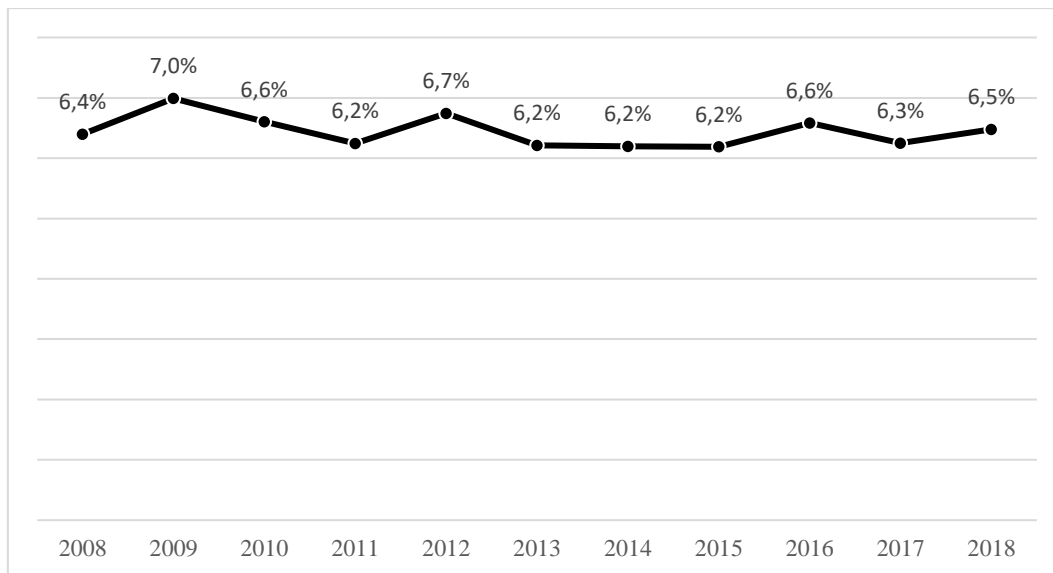


Figura 2.1 - Perdas Não Técnicas no Brasil (% da Energia Injetada Total)

Fonte: ANEEL (2019a)

Embora a situação do Brasil seja bem menos dramática que a observada em outros países em desenvolvimento e, em termos globais, o país se encontram em níveis de PNT na média, próximo aos 7% observados nos países da OECD (Yurtseven, 2015), as PNT têm um impacto financeiro bastante relevante tanto para as distribuidoras quanto para os consumidores brasileiros. Na Figura 2.2, apresenta a participação das perdas não técnicas regulatória sobre a Parcela B no ano de 2018. Percebam, que a parcela B representa os custos diretamente gerenciáveis pela distribuidora. São custos próprios da atividade de distribuição que estão sujeitos ao controle ou influência das práticas gerenciais adotadas pela empresa

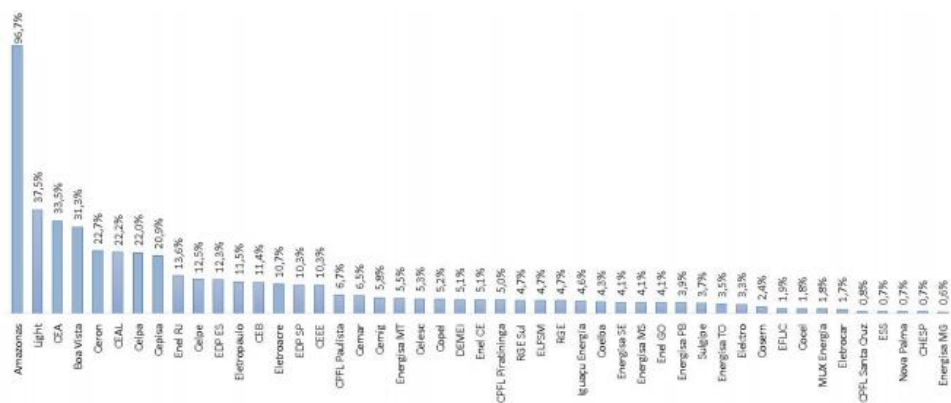


Figura 2.2 - Participação das perdas não técnicas regulatória sobre a Parcela B (2018)  
 Fonte: ANEEL (2019a)

Isso porque além dos custos da própria energia gerada e não paga, os consumidores e/ou as distribuidoras de energia tem que arcar com todo o custo da infraestrutura de transmissão e distribuição associado a essas perdas<sup>1</sup>.

Um ponto que torna o tratamento regulatório do tema bastante complexo no Brasil, é a relativa heterogeneidade das PNT observadas entre as distribuidoras de energia elétrica do país. Em 2018, por exemplo, das 54 principais distribuidoras de energia elétrica, 25 registraram PNT inferiores a 3% da energia injetada total, valores bastante próximos aos valores de PNT observado em países como EUA e Canadá (Yurtseven, 2015). Por outro lado, 11 dessas distribuidoras apresentaram PNT superiores a 10% da energia injetada total (ANEEL, 2019a), chegando em alguns casos a atingir mais de 30%, valores comparáveis aos observado em países como Índia e Bangladesh (Yurtseven, 2015), países que registram os piores resultados de PNT no mundo (Figura 2.2).

<sup>1</sup> Tomando-se a estrutura tarifária atual média do Brasil (Figura 2), observa-se que o custo de compra com a energia elétrica representa em média 57% da tarifa. Tomando-se o volume de PNT de energia elétrica registrados em 2018 no Brasil de 33,2 TWh e o preço médio de compra de energia das distribuidoras de R\$ 191,47 por MWh, tem-se que o custo médio das PNT somente com a compra de energia foi de R\$ 6,4 bilhões e os custos indiretos de R\$ 4,8 bilhões, perfazendo um total de R\$ 11,2 bilhões (ANEEL, 2019).

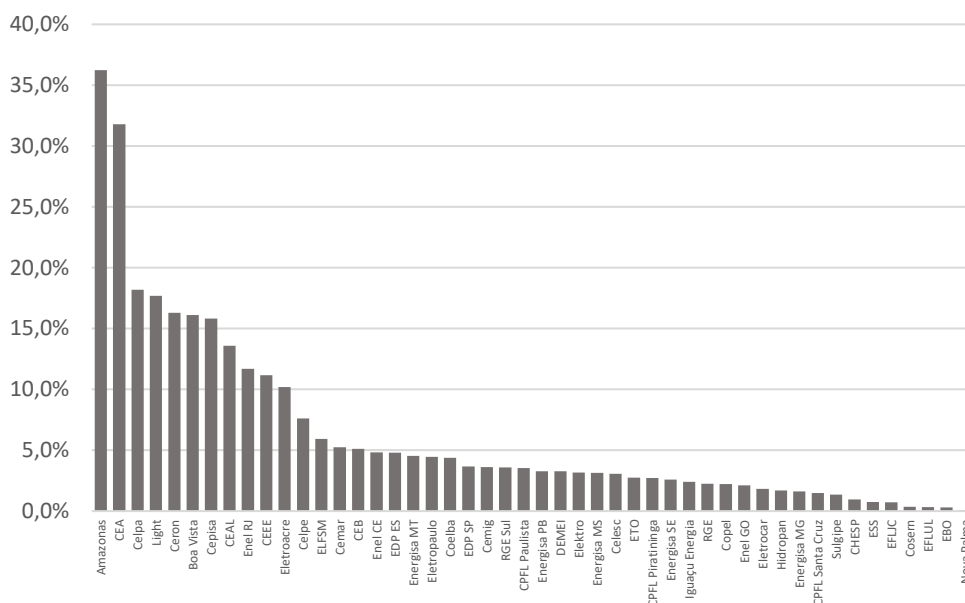


Figura 2.3 - Perdas Não Técnicas no Brasil (% da Energia Injetada Total)

Fonte: ANEEL (2019a)

Um ponto que se destaca nessa heterogeneidade do nível de PNT observado entre as distribuidoras brasileira, parece ser a baixa correlação entre as PNT observada e a renda per capita de cada região, o que sugere que o problema de PNT no Brasil é mais complexo que um problema meramente de nível de renda das unidades consumidoras de uma concessionária de distribuição.

Na Figura 2.2 é possível observar, por exemplo, que entre as dez distribuidoras de energia elétrica do país com maior nível de PNT, encontram-se as duas principais distribuidoras de energia elétrica do estado do Rio de Janeiro, o estado com o segundo maior PIB per capita do País. Por outro lado, distribuidoras de estados entre os menores níveis de PIB per capita do país, como por exemplo as distribuidoras da Paraíba e do Rio Grande do Norte (respectivamente Energisa Paraíba e Cosern), apresentam níveis de PNT bastante baixos (inferiores a 4% da energia injetada total).

De fato, essa baixa correlação entre o nível de PNT e renda per capita do estado ou macrorregião, não é uma particularidade apenas do Brasil (Smith, 2004; Millard e Emmerton e Emmerton, 2009). Países como a Indonésia, Peru e a

Tailândia, e, em certa medida a China, apesar de apresentarem PIB per capita relativamente baixos frente a média mundial, apresentam nível de PNT bastante próximos aos observados em países desenvolvidos como EUA e Canadá (Millard e Emmerton, 2009).

Esse fato tem levado alguns autores a proporem um espectro mais amplo de variáveis para explicar o fenômeno das PNT. Variáveis como a força das instituições públicas locais, o nível de corrupção, a aceitação social do furto de energia, o nível de escolaridades, o nível de desemprego, o nível de violência, entre outras variáveis vem sendo propostas para o estudo do fenômeno de perda não técnica (Smith, 2004; Winther, 2012; Yurtseven, 2015; Yakubu *et al.*, 2018).

Apesar desse relativo consenso entorno do tema, a forma, de medir esses fenômenos e estimar o seu real impacto sobre as PNT parece continuar sendo um tema em aberto na literatura especializada.

No Brasil, como será visto nas próximas seções deste capítulo, a ANEEL tem tentado tratar esse tema com a adoção de um conjunto mais amplos de variáveis que buscam capturar o mais amplo conjunto possível de fatores que possam impactar a capacidade das distribuidoras de combater as PNT.

## 2.3

### **A regulação das Perdas Não Técnicas no Brasil**

O modelo regulatório brasileiro divide as tarifas de energia elétrica a serem cobradas pelas distribuidoras em duas partes: parcela A e parcela B.

A parcela A corresponde aos custos não gerenciáveis pela distribuidora, onde, em geral, a distribuidora exerce um papel de mero “arrecadador”. Os custos da parcela A englobam os custos de compra de energia elétrica dos geradores, os custos com o uso do sistema de transmissão, os encargos setoriais, as perdas de energia elétrica e, em alguns casos, a inadimplência.

A parcela B corresponde, por outro lado, aos custos gerenciáveis pelas distribuidoras, tipicamente os custos com a operação e manutenção do sistema de distribuição de energia e os investimentos na rede.



Por se tratar de custos não gerenciáveis pelas distribuidoras, os custos relativos aos itens da parcela A são em geral repassados integralmente para os consumidores, com duas importantes exceções: as PNT e a Inadimplência (ou Receita Irrecuperáveis na denominação adotada pela ANEEL).

As PNT nunca serão totalmente erradicadas (Smith, 2004). Há um limite mínimo tolerável para essas perdas, a partir do qual os custos para reduzi-las superam os benefícios de sua redução. É até certo ponto consensual que as características da área na qual a distribuidora de energia elétrica opera podem influenciar dramaticamente a sua capacidade de combater este problema. Características socioeconômicas da região, como a presença de criminalidade, acentuada desigualdade, precariedade de infraestrutura, ausência de presença do Estado e suas instituições, dentre outras (ANEEL, 2010; Leite *et al.*, 2020; Simões *et al.*, 2020; Smith, 2004; Winther, 2012; Yurtseven, 2015; Yakubu *et al.* 2018), tornam o combate PNT em algumas áreas bem mais complexo que em outras.

Isso significa que esse limite mínimo tolerável para as PNT tende a variar de distribuidora para distribuidora em função da complexidade da sua área de atuação. Por outro lado, o esforço empreendido pela concessionária no combate às PNT, afeta o seu nível. Algumas concessionárias têm alcançado substanciais reduções de PNT mesmo em condições mais adversas por meio de inovação tecnológica, melhorias de gestão e investimentos na rede elétrica (ANEEL, 2013a).

Se fosse possível, através da observação, decompor os níveis de PNT de cada área concessão em duas parcelas, uma decorrente das ações realizadas ou não pela distribuidora (parcela gerenciável) e uma segunda relacionada ao ambiente no qual a distribuidora está inserida (parcela não gerenciável), o problema do regulador se resumiria a determinar o nível ótimo de PNT da parcela gerenciável assegurando o repasse integral da parcela não gerenciável. Entretanto, não é possível efetuar essa decomposição de forma direta e precisa, portanto, o regulador precisa definir um método que permita-lhe estimar essas parcelas de forma separada (Leite *et al.* 2020b).

Na seção seguinte será apresentado um breve histórico da evolução do tratamento dado as PNT pela ANEEL. Será apresentado também um detalhamento

da metodologia atual empregada pela ANEEL, culminando nas discussões recentes para sua atualização propostas na Consulta Pública 029/2020 promovida pela ANEEL.

### 2.3.1

#### **Histórico da metodologia empregada pela ANEEL**

Até 2003 as perdas elétricas consideradas no processo tarifário das distribuidoras de energia elétrica brasileiras correspondiam às perdas reais praticadas pelas distribuidoras. Apesar da parcela de perdas nem sempre ser explicitada no processo tarifário, tais perdas estavam embutidas nos custos de compra de energia reconhecidos nas tarifas (ANEEL, 2010).

A partir do ano 2003, com o início do Primeiro Ciclo de Revisões Tarifárias Periódicas – 1º CRTP, o reconhecimento tarifário das perdas de energia passou a ser limitado pelos patamares de perdas observados historicamente pela própria distribuidoras. Distribuidoras que vinham registrando elevações dos seus níveis de perdas passaram a enfrentar restrições no repasse desses custos à tarifa (ANEEL, 2010).

A partir de 2007, com o início do 2º CRTP, a definição das perdas regulatórias passou a ser segmentada, com o desenvolvimento de uma metodologia específica para a definição de perdas técnicas (PT) e outra para as perdas não técnicas (PNT). Para as PNT foi empregado um modelo de *benchmarking* para definição do referencial regulatório. A meta de cada concessionária passou a ser definida a partir da observação dos níveis praticados por outras distribuidoras mais eficientes e diretamente comparáveis. A comparação foi viabilizada pela criação de um “índice de complexidade”, para cada distribuidora baseado em um conjunto de variáveis socioeconômicas relacionadas com o fenômeno PNT (ANEEL, 2013a).

No 3º CRTP, iniciado em 2011, foram incorporadas algumas alterações na forma de construção do “índice de complexidade”. A principal mudança foi em relação à avaliação da velocidade da trajetória de redução das PNT resultante da identificação da empresa benchmark. O dimensionamento dos recursos necessários para cumprimento das metas regulatórias de PNT utilizado no 2º CRTP foi

abandonado, e os limites de redução passaram a ser definidos também com base na análise de *benchmarking* (ANEEL, 2013a e 2014a).

Em 2014 e 2015, essa metodologia foi atualizada, sofrendo pequenas alterações, sendo adotada até o presente momento.

Em abril de 2020 a ANEEL (2020a) abriu uma consulta pública propondo algumas alterações pontuais na metodologia atualmente empregada, principalmente associada a inclusão/substituição de algumas variáveis explicativas e nos ajustes dos limites máximos de redução anual. Uma síntese análise das mudanças propostas será apresentada na seção 2.4.7.

Na seção a seguir será apresentado um breve histórico sobre a metodologia atualmente empregada pela ANEEL para definição das metas regulatórias de perdas não técnicas.

## **2.4 Metodologia atualmente adotada pela ANEEL**

Nesta seção são descritos os procedimentos adotados pela ANEEL no âmbito das Perdas Não Técnicas e suas críticas. Resumidamente, para as PNT, são coletadas variáveis que explicam as perdas, assim estimando regressões de dados em painel, obtendo coeficientes destas variáveis em relação às PNT. Destes coeficientes, cria-se um Índice de Complexidade, compondo assim uma matriz de complexidade, que por sua vez fará a comparação entre as empresas, para, por fim, definir as metas regulatórias de PNT de distribuidoras.

### **2.4.1 Variáveis utilizadas na composição do Índice de Complexidade**

A principal vantagem da aplicação dos modelos de dados em painel na metodologia de perdas é a possibilidade de acompanhar, para um mesmo conjunto de distribuidoras, as perdas não técnicas associadas a cada área de concessão ao longo do tempo. Na relação que existe entre perdas e características socioeconômicas, a estimação por painel considera tanto os fatores que relacionam os patamares de perdas com as condições socioeconômicas como fatores associados a mudanças estruturais no tempo.

A Nota Técnica nº 494/2013, de 12 de novembro de 2013, teve por objetivo dar publicidade e reprodutibilidade às bases de dados de PNT, apresentando sua característica, fonte e metodologia de construção. Oriunda desta publicação, o Quadro 2.1 apresenta as variáveis utilizadas nos modelos da ANEEL, assim como sua fonte em que estas variáveis são disponibilizadas.

Quadro 2.1 – Variáveis utilizadas nos modelos ANEEL

<b>Código</b>	<b>Variáveis socioeconômicas</b>	<b>Fonte</b>
vio	Violência - Óbitos por Agressão a cada 100 mil habitantes	DATASUS
pob2	% de pessoas com renda per capita inferior a ½ salário mínimo	IBGE/CENSO
gini	Gini	IBGE/CENSO
sub2	Precariedade - % de pessoas em domicílios subnormais	IBGE/CENSO
lixo.u	% Domicílios com Coleta de lixo – urbano	IBGE/CENSO
inad	Inadimplência do setor de crédito – BACEN	BACEN
Mbr.Mb1Mbr	Mercado baixa renda/(Mercado B1 total)	SAMP
Mbr.Mbt	Mercado baixa renda/(Mercado BT total)	SAMP

Fonte: Elaboração própria a partir da Nota Técnica nº 106/2015-SGT/SRM/ANEEL.

A título de exemplo, na

Tabela 2.1, é feita uma sumarização da base de dados das variáveis utilizadas pela ANEEL para a ENEL Rio no período de 2003 a 2013. Além disto, também é apresentado o nível de PNT medida sobre o mercado de baixa tensão. Percebe-se que na primeira coluna, representada pela cor verde, tem-se o ano. Em azul, apresentam-se as variáveis independentes, utilizadas para o modelo econométrico. E por fim, em laranja, a PNT verificada na empresa no período, i.e., a variável dependente de interesse.

Tabela 2.1 – Base de dados das variáveis da ENEL Rio

ano	vio	lixo.u	pob2	gini	sub2	inad	Mbr.Mb1 Mbr	Mbr.Mbt	PNT
2003	48,99	0,83	0,25	0,56	0,07	NA	0,12	0,08	0,39
2004	44,18	0,83	0,24	0,55	0,07	0,06	0,23	0,16	0,39
2005	50,91	0,83	0,28	0,56	0,07	0,05	0,26	0,18	0,37
2006	47,84	0,83	0,29	0,56	0,07	0,04	0,26	0,18	0,35
2007	45,85	0,83	0,26	0,55	0,07	0,02	0,23	0,16	0,32
2008	38,94	0,83	0,25	0,54	0,08	0,02	0,19	0,13	0,30
2009	37,87	0,83	0,25	0,55	0,08	0,03	0,19	0,13	0,33
2010	37,27	0,91	0,27	0,54	0,08	0,02	0,18	0,12	0,33
2011	33,12	0,90	0,18	0,53	0,08	0,02	0,17	0,12	0,28
2012	33,31	0,90	0,20	0,53	0,08	0,03	0,10	0,07	0,28
2013	29,80	0,90	0,19	0,53	0,08	0,02	0,10	0,07	0,26

Fonte: Elaboração Própria

## 2.4.2 Críticas as variáveis utilizadas

O Quadro 2.2 explicita a dimensão, código utilizado pela ANEEL, descrição das variáveis, suas respectivas fontes de coleta, além das fontes de atualização. Desta forma, a primeira crítica sobre os procedimentos de PNT são centradas na natureza das variáveis que sustentam os modelos de dados em painel, utilizados atualmente pela ANEEL.

Quadro 2.2 – Variáveis utilizadas para a construção do índice de complexidade (IC)

Dimensão	Código	Descrição	Fonte (Abrangência)	Atualização (Abrangência)
Violência	vio	Violência - Óbitos por Agressão por 100 mil habitantes.	DATASUS (Municipais)	DATASUS (Municipais)
Desigualdade	pob2	% de pessoas com renda <i>per capita</i> inferior a ½ salário mínimo	IBGE/IPEA (Municipais)	PNAD (Estaduais)
	gini	Índice de Gini.	IBGE/IPEA (Estaduais)	IBGE (Estaduais)
Precariedade	sub2	% de pessoas em domicílios subnormais (Censo 2010).	IBGE/IPEA (Municipais)	DATASUS (Pop. Municipal)

<b>Comprometimento da Renda</b>	<b>inad</b>	Inadimplência do setor de crédito.	BACEN (Estaduais)	BACEN (Estaduais)
<b>Infraestrutura</b>	<b>Lixo.u</b>	% de domicílios urbano com coleta de lixo.	IBGE (Municipais)	PNAD (Estaduais)
<b>Consumo, Fatura e Baixa Renda</b>	<b>Mbr.Mb1Mbr</b>	Mercado Baixa Renda / Mercado B1 total.	SAMP (Distribuidora)	SAMP (Distribuidora)
	<b>Mbr.Mbt</b>	Mercado Baixa Renda / Mercado BT total.	SAMP (Distribuidora)	SAMP (Distribuidora)

Fonte: Elaboração própria a partir da Nota Técnica nº 106/2015-SGT/SRM/ANEEL.

Cabe ressaltar que nem sempre a fonte de coleta das informações é a de atualização. As variáveis construídas com base das informações do Instituto Brasileiro de Geografia e Estatística (IBGE) têm como seu principal alicerce os Censos Demográficos, que por sua vez são disponibilizados a cada dez anos.

A grande vantagem das informações dos Censos é que estas são desagregadas por município, auxiliando em uma inferência mais fidedigna. Porém, com a periodicidade de 10 anos, as variáveis disponíveis no Censo carecem de extrapolações lineares (entre 2000 e 2010) e atualizações (após 2010). De acordo com o Quadro 2.1, as variáveis coletadas do IBGE, são atualizadas com base na Pesquisa Nacional de Amostra de Domicílios PNAD, que por sua vez possuem abrangências estaduais, com a exceção da variável “sub2” que é atualizada apenas pela população municipal.

A principal crítica para estes conjuntos de variáveis é que, apesar das semelhanças das definições entre as informações da PNAD e do Censo, as extrapolações das variáveis do Censo, pelo crescimento das variáveis da PNAD, tende a distorcer a informação com o distanciamento anual do Censo. Desta forma, as variáveis utilizadas, quando atualizadas em um período temporal distante de 2010 (último Censo), podem não traduzir às realidades dos municípios.

Existe também a crítica sobre o fato de que algumas variáveis só possuem abrangência estadual, notadamente, inadimplência e índice de Gini. Desta forma, estas variáveis podem não refletir de maneira correta as situações de complexidade de concessões com menores abrangências, como distribuidoras que atuam em poucos municípios do Estado, caso que pode alterar a complexidade de algumas distribuidoras.

Além das análises supracitadas, ressalta-se que duas variáveis podem ter vieses em suas definições para explicar os fenômenos de PNT, sendo elas, inadimplência e violência. No que tange a inadimplência, frisa-se que esta é conceitualmente definida como: “Taxa segregada em pessoas físicas e jurídicas, medida pela razão entre o saldo dos contratos em que há pelo menos uma prestação, integral ou parcial, com atraso superior a noventa dias, e o saldo total das operações (exceto saldo de operações de cartão de crédito – compras à vista)” (ANEEL, 2013b). Porém, na ótica das Perdas Não-Técnicas, evidencia-se que a maior parte dessas perdas são oriundas do mercado de baixa tensão, que, por sua vez, é composto, majoritariamente, por pessoas físicas. Assim, a agregação da inadimplência de pessoas físicas (PF) com a inadimplência de pessoas jurídicas (PJ), tende a não refletir a verdadeira pressão do endividamento da população sobre as perdas não técnicas.

Por exemplo, na área de concessão da ENEL Rio, no Estado do Rio de Janeiro, o peso gerado por pessoas jurídicas pode comprometer a avaliação da inadimplência no Estado. De acordo com a Tabela 2.2, aproximadamente 68% do montante total de crédito disponibilizado no Estado foi para pessoas jurídicas, ante uma média nacional de 46% (Tabela 2.2). Ou seja, para o Rio de Janeiro, existe um maior peso da inadimplência das pessoas jurídicas, ante as pessoas físicas. Somase a este fato que a inadimplência de pessoas jurídicas no Estado é bastante incipiente, enquanto a taxa de pessoas físicas é alta. Assim, a inadimplência utilizada pela ANEEL gera um peso maior à menor inadimplência do estado do Rio de Janeiro, subestimando a inadimplência com maior correlação com as PNT, a inadimplência de pessoas físicas.

Tabela 2.2 – Percentual do Saldo de operações de créditos para PF e PJ, assim como suas respectivas taxas de inadimplência, tanto para o RJ, quanto para a média nacional

<b>Descrição</b>	<b>TOTAL</b>	<b>Pessoa Física</b>	<b>Pessoa Jurídica</b>
<b>% Saldo Operações Crédito (RJ)</b>	100%	32,16%	67,84%
<b>% Saldo Operações Crédito (Média Nacional)</b>	100%	53,82%	46,18%
<b>Taxa Inadimplência (RJ)</b>	2,37%	5,35%	0,96%
<b>Taxa Inadimplência (Média Nacional)</b>	3,53%	4,59%	2,38%

Fonte: Elaboração própria a partir das informações disponíveis no BACEN.

Outra variável que pode ser sub-representada nos modelos econométricos é a variável que representa a violência. Esta variável é definida como: “óbitos por agressão por 100 mil habitantes”. Apesar da descrição da variável ser fidedigna com o conceito de violência, percebe-se Tabela 2.3 que os índices de violência na área de concessão da ENEL Rio, por exemplo, vem diminuindo substancialmente, ou seja, de acordo com a variável utilizada pela ANEEL, a violência nos municípios do Rio de Janeiro estão com uma tendência de melhora, o que não condiz com a realidade observada.

A explicação para a diminuição do índice de violência no Rio de Janeiro pode estar atrelada à expansão do controle armado. Nos locais controlados por poderes paralelos os níveis de violência podem ser elevados, porém, as estatísticas oficiais tendem a não captar este tipo de violência, dado que em territórios dominados, talvez haja a omissão dessas mortes. Assim, cria-se um *under-reporting* da violência devido a pressão para a subnotificação dos crimes (Szwarcwald *et al.*, 1999). Desta forma, os índices de violência tendem a ser inversamente proporcional ao aumento de áreas controladas pelo tráfico ou milícias, que por sua vez, são áreas onde inexistente o controle sobre as PNT.

De acordo com a Tabela 2.3, a área de concessão da ENEL Rio foi apenas a 11ª mais violenta do Brasil (ante 33 áreas de concessão), em 2016. Este fato não condiz com a realidade, pois em 2018 o governo federal viu a necessidade de intervenção na segurança do Estado do Rio de Janeiro, devido ao panorama de violência generalizada.

Tabela 2.3 – Índice de violência na área de concessão da ENEL Rio

Ano	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Vio	44,2	50,9	47,8	45,9	38,9	37,9	37,3	33,1	33,3	35,7	38,5	32,4	39,3

Fonte: ANEEL, 2020a.

Sumarizando, as variáveis possuem limitações no que tange à atualização e abrangência. Além disto, as variáveis inadimplência e violência podem não representar a correta interação com as PNT, devido detalhes supracitados.



Na seção a seguir, serão descritos os modelos para definição de metas de PNT atualmente adotados pela ANEEL, considerando às variáveis descritas nesta seção.

### 2.4.3 Modelos utilizados atualmente pela ANEEL

A ANEEL, por meio da Nota Técnica nº 106/2015, de 22 de abril de 2015, homologou a metodologia de tratamento regulatório para PNT de energia. De forma resumida, pode-se dizer que o Regulador, utilizando-se de variáveis socioeconômicas e de mercado Tabela 2.1 para o período de 2003 a 2013 (dados em painel, para todas concessionárias do Brasil), obteve três modelos econométricos<sup>2</sup> (C, G e K) para determinação do Índice de Complexidade (IC) que refletisse a dificuldade do combate às perdas comerciais de acordo com as características de cada área de concessão.

A Agência considerou que os efeitos específicos presentes na regressão são do tipo efeito aleatórios<sup>3</sup>, controlando as heterogeneidades não inseridas na equação, invariantes no tempo, de cada empresa. As três especificações dos modelos utilizados (C, G e K) para a estimação dos coeficientes das variáveis que explicam as perdas não técnicas, são apresentadas nas Equações 2.1 (modelo C), 2.2 (modelo G) e 2.3 (modelo K).

$$y_{2,i} = 0,144 + \mu_i + 1,98sub2_{it} - 0,17lixo.u_{it} + 0,24pob2_{it} - 0,10Mbr.Mb1Mbr_{it} + 1,39inad_{it} + \varepsilon_{it} \quad (2.1)$$

$$y_{2,i} = 0,160 + \mu_i + 1,96sub2_{it} - 0,26lixo.u_{it} + 0,18gini_{it} + 1,47inad_{it} + \varepsilon_{it} \quad (2.2)$$

$$y_{2,i} = 0,175 + \mu_i + 2,21sub2_{it} - 0,18lixo.u_{it} + 0,20pob2_{it} - 0,19Mbr.Mbt_{it} + 0,001vio_{it} + \varepsilon_{it} \quad (2.3)$$

Em que:

<sup>2</sup> Na nota técnica Nº 188/2014-SER/ANEEL, foram apresentados os modelos denominados entre as letras de A a L (12 modelos), desses modelos, a Agência escolheu os modelos C, G e K como os melhores resultados.

<sup>3</sup> A Agência não explicita a utilização dos testes de *Breusch-Pagan* e *Hausman* para pautar a escolha dos modelos. Esses testes se fazem necessário para a robustez da metodologia selecionada.

$y_{2,i}$  é o índice de perdas não técnicas;

$sub2_{it}$  é o percentual de pessoas que vivem em domicílios subnormais, na área de concessão da empresa  $i$ , no ano  $t$ ;

$lixo.u_{it}$  é o percentual dos domicílios atendidos pela coleta de lixo, na área de concessão da empresa  $i$ , no ano  $t$ ;

$pob2_{it}$  é o percentual da população de que vive com menos de meio salário mínimo, residentes na área de concessão da empresa  $i$ , no ano  $t$ ;

$Mbr.Mb1Mbr_{it}$  é o mercado de baixa renda em relação ao mercado B1 (residencial) total, na área de concessão da empresa  $i$ , no ano  $t$ ;

$inad_{it}$  é a inadimplência do setor de crédito, dos residentes na área de concessão da empresa  $i$ , no ano  $t$ ;

$gini_{it}$  é o índice de Gini, na área de concessão da empresa  $i$ , no ano  $t$ ;

$Mbr.Mbt_{it}$  é o mercado de baixa renda em relação ao mercado BT (Baixa Tensão) total, na área de concessão da empresa  $i$ , no ano  $t$ ;

$vio_{it}$  é o índice de mortes por agressão a cada cem mil habitantes, na área de concessão da empresa  $i$ , no ano  $t$ ;

$\alpha$  é o intercepto da regressão;

$\varepsilon_{it}$  são os erros idiossincráticos da regressão; e

$\mu_i$  são as heterogeneidades de cada região de concessão  $i$ , invariante no tempo.

Definidos os modelos econométricos, os Índices de Complexidade (IC) são calculados como a soma das variáveis em 2013, ponderadas pelos respectivos coeficientes estimados. Cabe ressaltar que a heterogeneidade  $\mu_i$  estimada para cada área de concessão não é acrescentada ao IC, pressupondo que tal heterogeneidade é tratada como uma métrica de ineficiência.

#### **2.4.4 Críticas aos Modelos utilizados ANEEL**

O objeto central na metodologia de Perdas Não Técnicas é a estimação econométrica dos modelos de dados em painel. As boas práticas econométricas são constituídas de uma série de testes específicos, a fim de auxiliar na escolha dos

vários modelos disponíveis, além de indicar possíveis problemas oriundos dos conjuntos de dados. Apesar da formulação de modelos econométricos contar com a ajuda de testes para alcançar a especificação correta, as escolhas dos modelos para explicar as PNT, apresentados na NT nº 106/2015, não fazem menção a quaisquer testes. No caso dos dados em painel, sugere-se dois tipos de procedimentos, conforme a Figura 2.4.

Procedimento	Descrição	Testes	Impacto do IC
1) Processo de decisão apropriada da escolha do modelo	Escolha entre dados empilhados, efeitos fixos e efeitos aleatórios	Teste de Hausman e BP	Altera valores dos coeficientes
2) Processo de inferência das hipóteses	Teste de homocedasticidade e autocorrelações contemporâneas e seriais	Teste BPt, BPLM, BGW	Altera significância dos coeficientes

Figura 2.4 – Procedimentos para a escolha dos modelos de dados em painel

Fonte: Elaboração Própria.

O primeiro procedimento diz respeito à a decisão de qual a estimação econométrica é a mais adequada para o conjunto de dados disponíveis. No segundo procedimento, as escolhas dos modelos de dados em painel se baseiam nas inferências sobre a heterogeneidades, assim, existindo dois testes que direcionam a escolha parcimoniosa das estimações, apresentados na Figura 2.5 (Greene, 2012).

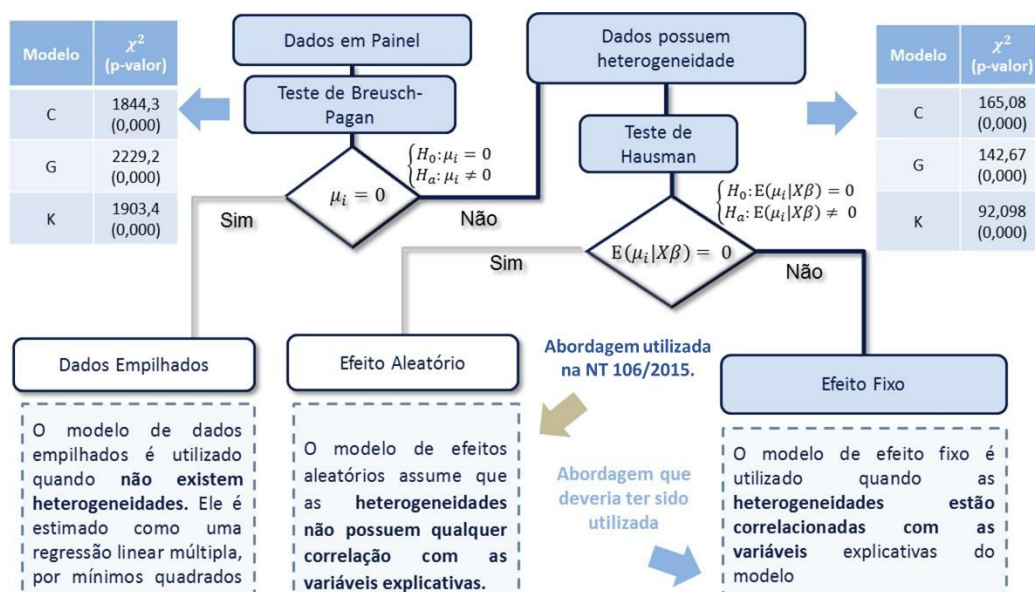


Figura 2.5 – Processo de decisão dos modelos de dados em painel utilizados em PNT

Fonte: Elaboração própria.

O primeiro teste, denominado teste de Breush Pagan, é utilizado para avaliar existência de heterogeneidades, assim pautando a escolha entre efeitos fixos ou aleatórios e o modelo *Pooled* (dados empilhados). Conforme apresentado na Figura 2.6 os modelos C, G e K da ANEEL possuem heterogeneidades, sendo excluída a utilização do modelo *Pooled*. Dado os resultados do teste de Breush Pagan, indica-se o teste de Hausman, a fim de subsidiar as escolhas entre efeitos fixos e efeitos aleatórios. De acordo com a Figura 2.6 e a Figura 2.5, o teste de Hausman indica a utilização de modelos de efeitos fixos.

Apesar do teste não indicar os modelos de efeitos aleatórios, a ANEEL utiliza para estimar os modelos C, G e K, sob pena de vieses econométricos. Conforme já abordado, a utilização dos modelos de efeitos aleatórios nos três modelos da ANEEL, quando o teste de Hausman indica o modelo de efeitos fixos, segundo Greene (2012), tende a gerar parâmetros inconsistentes e ineficientes. Assim, os *benchmarks* utilizados para o cálculo dos índices de complexidades possuem problemas de consistência, prejudicando a construção do índice e, futuramente, da matriz de probabilidades e metas resultantes. Desta forma, os modelos C, G e K possuem graves problemas de inferência.

Ademais, após a escolha do correto modelo econométrico de dados em painel, a literatura econométrica indica um segundo processo, no qual pauta-se a verificação das hipóteses iniciais do modelo, notadamente, homocedasticidade e ausência de autocorrelações seriais e contemporâneas. Cabe ressaltar que estes testes visam problemas econométricos sobre a inferência das variáveis (critérios de rejeição da hipótese nula), assim, não possuem a capacidade de indicar alterações nos valores dos coeficientes, diferentemente dos testes de processo de decisão. Os resultados dos testes BPt, BGW e BPLM são sumarizados na Figura 2.6.

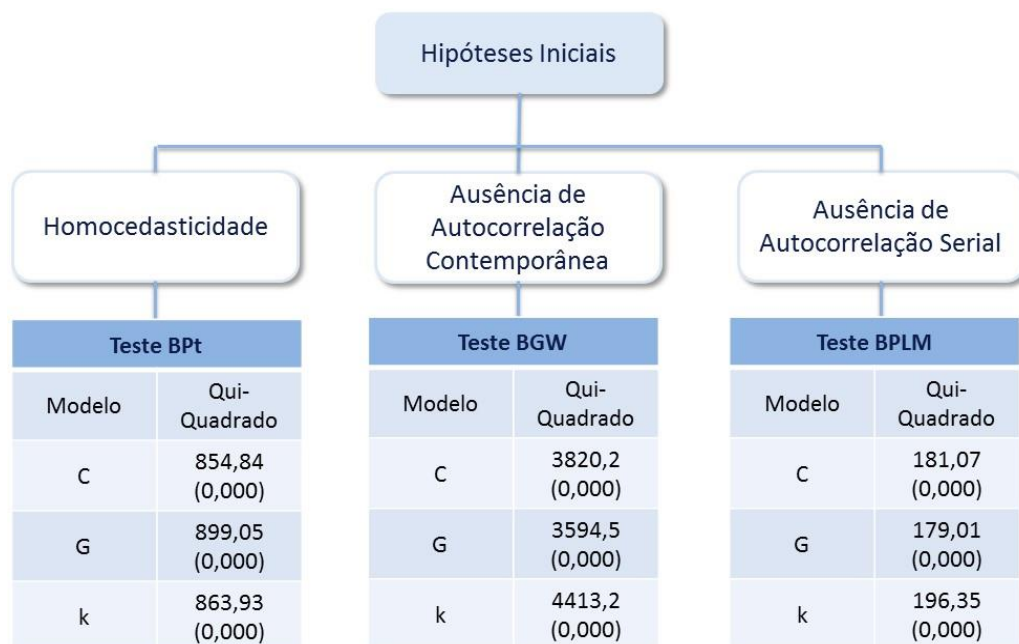


Figura 2.6 – Processo de inferência dos modelos de PNT

Fonte: Elaboração própria.

Os modelos C, G e K da ANEEL, de acordo com os testes de BPT, BGW e BPLM, possuem os problemas de heterocedasticidade, correlação serial e contemporânea. Estes problemas econométricos, apesar de não alterarem o valor dos coeficientes, alteram o intervalo de confiança, que por sua vez, é utilizado para a construção da matriz de complexidade. Desta forma, estes problemas econométricos não devem ser ignorados, sob pena de gerar uma matriz de complexidade que induza a *benchmarkings* errôneos. Para contornar os problemas de inferência citados, indica-se a reestimação da matriz de variância-covariância, como o método de correção de White de acordo com MacKinnon e White (1985).

Sumarizando, existem evidências de que os modelos C, G e K da ANEEL deveriam seguir a estimação de efeitos fixos, corrigindo os problemas de heterocedasticidade e autocorrelações serial e contemporânea. Desta forma, os coeficientes dos modelos atuais são inconsistentes e ineficientes, assim, prejudicando toda a metodologia de PNT.

Além dos problemas econométricos de estimação, a ANEEL constrói o IC de maneira a excluir as heterogeneidades dos cálculos. Esta exclusão é pautada no argumento de que a heterogeneidade é causada, em grande parte, pela ineficiência

das concessionárias em diminuir as perdas não técnicas. Porém, conforme já explicitado, as heterogeneidades são todos os efeitos não modelados, que são intrínsecos às distribuidoras e invariantes no tempo.

Desta forma, inferir que toda a heterogeneidade é oriunda de ineficiência é um pressuposto forte, face que nem todos os determinantes das PNT estão inseridos nos modelos. Além disto, não é razoável supor que as ineficiências das empresas possam ser imutáveis temporalmente, assim, sendo improvável que os efeitos específicos sejam somente dados pela ineficiência no combate às PNT.

#### 2.4.5

#### **Matriz de Probabilidade para fins de cálculo das metas de PNT**

Construído o IC, torna-se possível a construção de uma matriz de complexidade. A Agência Reguladora parte de uma premissa inicial de que quanto mais complexa é a área de concessão, maior é a perda aceitável. Deste modo, se a empresa que atua em uma área menos complexa do que outra e ainda assim tem uma PNT maior, a ANEEL entende que as perdas poderiam ser reduzidas, de modo que seja inferior ou igual à empresa que atua em uma área mais complexa. Assim, se uma empresa atua em uma área mais complexa e tem perdas baixas, presume-se que esta empresa é eficiente no combate às perdas.

Para medir a complexidade de uma área, presume-se que existem variáveis observáveis comuns à todas as empresas que auxiliam na construção de uma medida comum e comparável. No caso, conforme já colocado, estima-se uma regressão de dados em painéis com efeitos aleatórios, assim encontrando a magnitude da influência de cada variável escolhida na representatividade da complexidade (coeficientes da regressão ou  $\beta$ 's). Desta forma, estima-se, estatisticamente, valores que representam quão complexa é cada área de concessão, criando o IC.

Pode-se argumentar que comparar as empresas através de um ordenamento do *ranking* do índice de complexidade pode gerar erros de inferência, pois, como o índice é criado através de um modelo estatístico, existe a possibilidade de que as empresas estejam retratadas em um ordenamento errôneo. Assim, existe a possibilidade de que, por exemplo, apesar do ordenamento gerado apresentar a

empresa “A” mais complexa do que a empresa “B” (e assim “A” ser a meta regulatória de “B”), a empresa “B” pode ser mais complexa que “A”, assim, gerando metas intangíveis para “B”.

Visando diminuir a chance de erro no ordenamento de complexidade, a ANEEL sugere a criação de uma matriz de probabilidade de comparação. Quando se estima uma regressão, como se trata de resultados probabilísticos, existem valores que podem se desviar do valor estimado, calculado pelo desvio padrão da estimativa. Dado o desvio padrão, cria-se um “intervalo de confiança” para cada índice de complexidade calculado, mostrando que existe uma “banda” em que o valor da complexidade se situa.

Do intervalo de confiança é que se deriva a matriz de probabilidade. Tal matriz compara os intervalos de confiança de cada empresa e, via um teste t-Student (calcula-se um tipo de p-valor<sup>4</sup>), calcula a probabilidade do Índice de Complexidade de uma empresa ser maior do que a outra. Cabe ressaltar que as concessionárias são divididas em dois grupos<sup>5</sup>, grandes e pequenas, assim, criando distintos *rankings* de complexidade, além de distintas matrizes de complexidade (ANEEL, 2016)

Assim, na matriz de probabilidade, para cada empresa, calcula-se a probabilidade da outra empresa ser mais complexa. Desta matriz, a ANEEL define a empresa *benchmark* com o maior potencial de redução de perdas não técnicas. O cálculo é feito a partir da média ponderada da perda praticada pelo *benchmark* e da perda praticada pela concessionária em processo de revisão, elucidado pela Equação 2.4.

---

<sup>4</sup> No caso, infere-se que existe um teste de hipótese do tipo: 
$$\begin{cases} H_0: \hat{Y}_{Bench} \geq \hat{Y}_{Empr} \\ H_a: \hat{Y}_{Bench} < \hat{Y}_{Empr} \end{cases}$$

Nos testes de hipóteses são gerados p-valores, O p-valor (também chamado de nível descritivo ou probabilidade de significância), é a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, sob a hipótese nula. Desta forma, tal p-valor, ou o valor da probabilidade do Benchmarking estar em uma área realmente mais complexa é dada pelo seguinte teste t – Student:  $t_{cal} = (\hat{Y}_{Bench} - \hat{Y}_{emp}) / \left( \sqrt{V(\hat{Y}_{Bench}) + V(\hat{Y}_{emp})} \right)$ . Desta forma, calcula-se, via tabela de distribuição de t de student a  $\Pr(t_{cri} > t_{cal})$ .

<sup>5</sup> As empresas serão comparadas conforme o seu tamanho, sendo consideradas grandes (Grupo 1) aquelas que possuam mercado de baixa tensão maior que 1.000 GWh/ano e atendam mais que 500 mil unidades consumidoras ou que possuam mais do que 15.000 km de rede elétrica. As demais empresas são consideradas do Grupo 2.

$$E(\text{MetaEmpresa}_i)_j = \text{Prob}(\cdot)_j \times \text{Perdas}(\text{Benchmark}) + [1 - \text{Prob}(\cdot)_j] \times \text{Perdas}(\text{Empresa}_i) \quad (2.4)$$

Em que:

$E(\text{MetaEmpresa}_i)_j$  = potencial de redução da empresa  $i$ , de acordo com o modelo  $j$ ;

$\text{Prob}(\cdot)_j$  = probabilidade do *benchmark* estar em área de concessão mais complexa de acordo com o modelo  $j$ ;

$\text{Perdas}(\text{Benchmark})$  = perda não técnicas sobre mercado de baixa tensão da empresa *benchmark*;

$\text{Perdas}(\text{Empresa}_i)$  = perdas não técnicas sobre mercado de baixa tensão da empresa analisada.

Ressalta-se que o *benchmarking* regulatório final de cada empresa é a distribuidora que minimiza a meta calculada na Equação 2.4. Para cada modelo C, G e K foram considerados potenciais de redução de perdas diferenciados e, portanto, a meta de referência para cada empresa  $i$  é igual a média aritmética das metas oriundas de cada modelo, conforme Equação 2.5.

$$\text{Meta}_i = \frac{\sum_{j=1}^n \text{Meta}_{(i,j)}}{n} \quad (2.5)$$

Em que:

$\text{Meta}_i$  = potencial de redução da empresa  $i$  [%];

$\text{Meta}_{(i,j)}$  = meta de perdas da empresa  $i$  [%] conforme modelo  $j$ ;

$n$  = número de *rankings* selecionados.

Há algumas críticas sobre a matriz de probabilidade que foram abordadas, mas que são aprofundadas na próxima seção.



## 2.4.6 Crítica à Matriz de Probabilidade

Ressalta-se que todos os problemas econométricos analisados na subseção anterior são refletidos na matriz de probabilidade. Desta forma, o fato da ANEEL ter utilizado modelos viesados, já invalida a utilização da matriz de complexidade vigente.

Além disto, como a probabilidade da empresa *benchmarking* ser mais complexa que outra empresa é gerada pelo desvio padrão do índice de complexidade, que por sua vez é gerada pela regressão de dados em painéis de 2003-2013, infere-se que a matriz de metas é intrinsecamente relacionada à equação de 2003-2013. Desta forma, caso ocorra algum evento entre 2014-2019 que aumenta a complexidade da empresa, a probabilidade do *benchmarking* ser mais complexo também deveria se alterar. Um exemplo dessa situação é que com o aumento de milícias nas áreas de concessão da ENEL Rio (Junior, 2016), a complexidade da empresa aumentou no período de 2014 a 2019, porém a matriz ainda se mantém estática, não recalculando a probabilidade de seu *Benchmarking* ser mais complexo, que por sua vez, ficou menor. Desta forma, a matriz não considera a trajetória de complexidade das áreas de concessão, podendo designar *benchmarks* com realidades distantes, assim, podendo gerar metas intangíveis.

Com vistas a sanar alguns dos problemas aqui abordados e que haviam sido apontados por Calili *et al.* (2017), a ANEEL lançou em 2020 a Consulta Pública nº029/2020 que levou para as discussões as possíveis mudanças da metodologia para definição das metas regulatórias de PNT.

## 2.4.7 Consulta Pública nº 029 de 2020

Como colocado, em 10 de abril de 2020, a ANEEL pôs em consulta pública (CP) uma proposta de alteração da metodologia atual utilizada pela agência para a definição das metas regulatórias de PNT.

Assim, com o intuito de reformular e atualizar os modelos atualmente empregados no cálculo do Índice de Complexidade Socioeconômico (IC), a ANEEL disponibilizou uma base de dados utilizadas com 22 variáveis no período

de 2004 a 2017. No entanto, na base de dados disponibilizada pela ANEEL verifica-se a ausência de informações para a variável *lixo.u* nos anos de 2016 e de 2017 e a variável violência no ano de 2017. Portanto, os modelos que utilizam a variável *coleta de lixo urbano* estão limitadas até período de 2015 e, para variável de Homicídio (óbitos por agressão), limitado até 2016. No Quadro 2.3, apresentam-se as variáveis disponibilizadas pela ANEEL na CP 29/2020.

Quadro 2.3 – Resumo das variáveis disponibilizadas na CP 29/2020

Variável	Código
Densidade de moradores por dormitório	Admd
Taxa de evasão escolar do ensino médio	Eem
Percentual de pessoas abaixo da linha da pobreza	Pob3
Percentual de domicílios com rede geral de água encanada	Rga
Percentual de domicílios urbanos com coleta de lixo	Lixo.u
Índice de desigualdade social	Gini2
Percentual de pessoas que vivem em domicílios subnormais	Sub2
Inadimplência geral no setor de crédito do SFN	Inad
Inadimplência de pessoas físicas no setor de crédito do SFN	Inad.pf
Homicídio doloso	Homi_dolo
Intervenções policiais e operações de guerra	Int_pol
Furto de veículos	Furto_v
Roubo de veículos	Foubo_v
Latrocínio	Latro
Agressão por arma de fogo	Agraf
Homicídio (óbitos por agressão)	Vio
Homicídios (mortes violentas)	H.mvci
Unidade consumidoras de baixa renda/unidades consumidoras de BT	Ucbr.Mb1UCbr
Mercado de baixa renda/mercado B1 Total	Mbr.Mb1Mbr
Mercado de baixa renda/mercado BT Total	Mbr.Mbt
PIB per capita da área de concessão	PIB.PC
Tempo em anos	Time

Fonte: ANEEL (2020a)

Ao comparar com as variáveis utilizadas pela ANEEL nos modelos C, G e K; verifica-se uma presença maior das variáveis de infraestrutura, renda e violência que demonstram um maior impacto no fenômeno de PNT. Além disso, a ANEEL estimou 127 modelos de dados em painel com efeitos aleatórios correlacionado o

percentual de PNT sobre o mercado de baixa tensão medido das distribuidoras e um conjunto de 7 variáveis socioeconômicas das áreas de concessão. As variáveis socioeconômicas testadas foram as descritas no Quadro 2.4.

Quadro 2.4 - Variáveis Socioeconômicas testadas pela ANEEL na CP 029/2020

<b>Código</b>	<b>Variáveis socioeconômicas</b>	<b>Fonte</b>
Vio	Violência - Óbitos por Agressão a cada 100 mil habitantes	DATASUS
Pob3	% de domicílios com renda <i>per capita</i> inferior a ½ salário mínimo	IBGE/IPEA
Inad	Inadimplência do setor de crédito – BACEN	BACEN
sub2	Precariedade - % de pessoas em domicílios subnormais (Censo 2010)	IBGE
Admd	Nº de moradores por dormitório por domicílio	IBGE
PIB.PC	PIB <i>per capita</i>	IBGE
EficOper	% de eficiência operacional das Distribuidoras	ANEEL

Fonte: ANEEL (2020a)

Os 127 modelos foram avaliados posteriormente pela ANEEL de forma sequencial por 4 critérios previamente estabelecidos e listados na Tabela 2.4, de forma que ao final a ANEEL obtivesse 3 modelos (X, Y e Z).

Tabela 2.4 – Critério de seleção de modelos adotado pela ANEEL na CP 029/2020

<b>Filtro</b>	<b>Critérios Utilizados</b>	<b>Nº de Modelos Resultantes</b>
1º	Poder Explicativo $R^2 > 0,10$	90
2º	Significância Estatística p-valor < 0,10	31
3º	Sinal do Coeficiente coerente	14
4º	Ranking ( $R^2$ )	3

Fonte: ANEEL (2020a)

As especificações dos 3 modelos propostos (X, Y e Z) para a estimação dos coeficientes das variáveis que explicam as PNT, são apresentadas nas Equações 2.6 (modelo X), 2.7 (modelo Y) e 2.8 (modelo Z).

$$y_{2,i} = 0,3756 + 2,926 \text{ sub}2_{it} - 1,441 \text{ Admd}_{it} - 0,00000303 \text{ PIB}_{it} - 0,449 \text{ EficOper}_{it} + \varepsilon_{it} + \mu_i \quad (2.6)$$

$$y_{2,i} = 0,3214 + 2,735 \text{ sub}2_{it} - 1,502 \text{ Admd}_{it} + 0,001 \text{ vio}_{it} - 0,454 \text{ EficOper}_{it} + \varepsilon_{it} + \mu_i \quad (2.7)$$

$$y_{2,i} = 0,3482 + 2,928 \text{ sub}2_{it} - 1,449 \text{ Admd}_{it} - 0,4629 \text{ EficOper}_{it} + \varepsilon_{it} + \mu_i \quad (2.8)$$

Onde:

$y_{2,i}$  é o índice de perdas não técnicas estimados pelos modelos X, Y ou Z;

$\text{sub}2_{it}$  é o percentual de pessoas que vivem em domicílios subnormais, na área de concessão da empresa  $i$ , no ano  $t$ ;

$\text{Admd}_{it}$  é o N° de moradores por dormitório por domicílio, na área de concessão da empresa  $i$ , no ano  $t$ ;

$\text{EficOper}_{it}$  é o percentual % de eficiência operacional da empresa  $i$ , no ano  $t$ ;

$\text{PIB.pc}$  é o PIB *per capita* na área de concessão da empresa  $i$ , no ano  $t$ ;

$\text{vio}_{it}$  é o índice de mortes por agressão a cada cem mil habitantes, na área de concessão da empresa  $i$ , no ano  $t$ ;

$\alpha$  é o intercepto da regressão;

$\varepsilon_{it}$  são os erros idiossincráticos da regressão; e

$\mu_i$  são as heterogeneidades de cada área de concessão  $i$ , invariante no tempo.

Assim, pela proposta da ANEEL, esses modelos substituiriam os modelos atuais C, G e K (descritos nas equações 2.1, 2.2 e 2.3) no cálculo do IC, mantendo-se, no entanto, as demais etapas da metodologia atual.

Nos novos modelos, três novas variáveis socioeconômicas (ADMD, PIB.pc e EficOper) passam a compor o IC substituindo quatro variáveis atualmente empregadas (lixo.u, pob2, Mbr.Mb1Mbr e o Mbr.Mbt).

Além da alteração dos parâmetros utilizados no cálculo do IC, a ANEEL também propôs uma alteração no critério de separação das empresas, passando a considerar no grupo 1 (grande porte) as distribuidoras com mercado de baixa tensão superior à 700 GWh/ano (ANEEL, 2020).

A regra para definição do ponto de partida (meta do primeiro ano) também seria alterada passando a ser definida a partir de um único critério independente do porte das distribuidoras. Esse novo critério proposto é apresentado no Quadro 2.5.

Quadro 2.5 – Critério para a definição do Ponto de Partida das PNT regulatórias sobre o mercado de baixa tensão (BT) proposto na CP 029/2020

<b>Grupo</b>	<b>Condição Analisada</b>	<b>Ponto de partida</b>
Todas	Meta do novo Ciclo > Meta do Ciclo anterior	Meta do novo Ciclo
	Média 3 anos ( $PNT_A$ ) > 7,5%	Mínimo (Meta Ciclo Anterior, média dos últimos 3 anos)

Fonte: ANEEL (2020a)

Por fim, a proposta da ANEEL também altera os limites de redução anual das metas de PNT regulatórias o que é apresentado o Quadro 2.6.

Quadro 2.6 - Critério para a definição dos limites anuais de redução das PNT regulatórias sobre o mercado de baixa tensão (BT) proposto na CP 029/2020

<b>Grupo</b>	<b>Condição Analisada</b>	<b>Limite de redução</b>
1	Meta Final $PNT_A > 5,0\%$	$\% 7 * Meta Final PNT_A / 50 - 7 / 10$
	Meta Final $PNT_A < 5,0\%$	Sem Trajetória
2	Meta Final $PNT_A > 8,6\%$	$\% 7 * P Meta Final PNT_A / 50 - 7 / 10$
	$2,5\% < Meta Final PNT_A < 8,6\%$	0,50% ao ano
	Meta Final $PNT_A < 2,50\%$	Sem Trajetória

Fonte: ANEEL (2020a)

A nova proposta ainda inclui a possibilidade de se incorporar um tratamento ad hoc para as áreas de risco das distribuidoras de grande porte que apresentem o maior nível de perdas não técnicas (terceiro quartil considerando as perdas não técnicas sobre o mercado de BT medido).

Embora a proposta da ANEEL altera alguns parâmetros da metodologia atual empregada na definição das metas de PNT regulatórias, a Agência não altera a metodologia que vem sendo aplicada desde 2011. Assim, as mesmas críticas em relação a metodologia atual permanecem validas para o modelo proposto. Na seção 2.4 são apresentadas as principais críticas ao tratamento regulatório das PNT no Brasil.

Apesar da metodologia proposta pela ANEEL ainda estar em consulta pública e, portanto, a sua aplicação ainda seja incerta, neste trabalho optou-se por testar e incorporar algumas das melhorias propostas pela consulta pública. Assim no capítulo 5 apresenta-se uma metodologia para selecionar as variáveis mais relevantes na explicação do fenômeno das PNT, um aspecto fundamental para a definição de modelos que possam auxiliar no estabelecimento das metas regulatórias de redução de perdas não técnicas para as distribuidoras de energia elétrica brasileiras.

## **2.5**

### **Considerações sobre o capítulo**

Como demonstrado nesse capítulo, a excessiva complexidade da metodologia atualmente empregada pela ANEEL na definição das metas de PNT regulatória, assim como a adoção de um conjunto grande critérios ad hoc, põem em risco um dos principais parâmetros que norteiam a regulação por incentivo: o estabelecimento de incentivos econômicos corretos para busca contínua pela melhoria da eficiência.

Os problemas apontados, em particular os relativos à especificação do modelo, apresentam indícios claros de que há falhas no processo de comparação das PNT entre as distribuidoras no país que podem estar conduzindo tanto a um não reconhecimento tarifário das PNT de algumas distribuidoras quanto o sobre reconhecimento da PNT em outras.

Uma alternativa capaz de resolver parte dos problemas aqui apresentados no modelo da ANEEL corresponde à busca exaustiva de combinações das variáveis disponibilizadas, seguida pela avaliação da significância estatística dos modelos identificados.

### 3

## Proposta de seleção automática de modelos para dados em painel

Nesse capítulo, apresenta-se um breve resumo dos métodos e modelos utilizados nesta dissertação. Na primeira parte, é feita uma explanação do arcabouço teórico sobre procedimentos automatizados de seleção de modelos. Por fim, é apresentado todo arcabouço metodológico sobre a regressão para dados em painel e os testes geralmente empregados na literatura para obtenção de modelos mais robustos.

### 2.6

#### Seleção automática de modelos

É cada vez mais comum lidar com muitas variáveis explicativas na formulação de modelos de regressão (Ripley, 2003, Cai *et al.* 2009). Isso geralmente acontece em análises exploratórias ou em estudos experimentais que abordam sistemas complexos. Em tais situações, o número de modelos candidatos pode ser bastante grande, tornando a seleção de variáveis explicativas um processo iterativo de complexificação/simplificação tedioso e ineficiente (Ripley 2003). Isso exige procedimentos automatizados de seleção de modelos (Venables e Ripley, 1997).

O aumento exponencial da capacidade computacional do *hardware*, acompanhada pelo desenvolvimento de *software*, abriu a oportunidade para implementar a busca exaustiva na seleção da melhor especificação de um modelo de regressão linear, conforme apresentado na Figura 3.1.

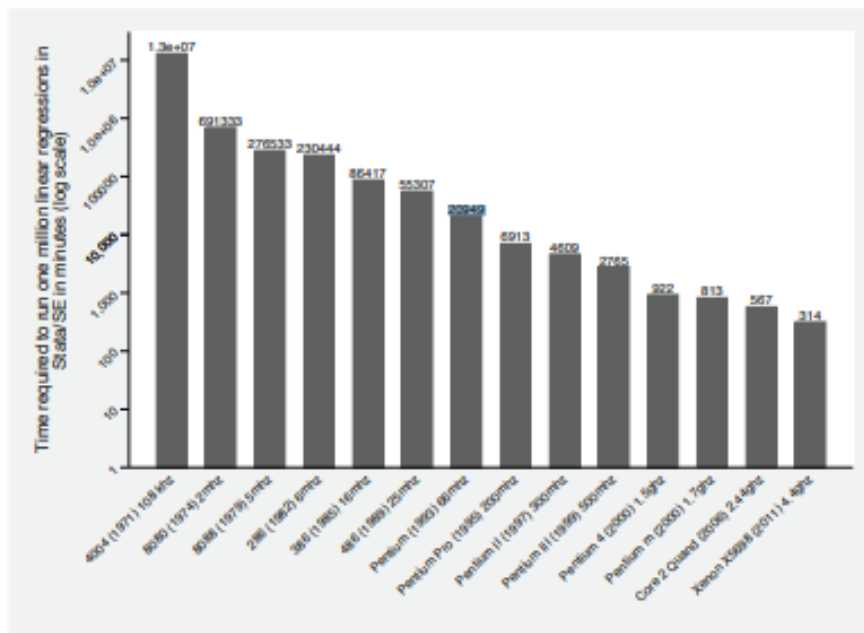


Figura 3.1 – Pesquisa exaustiva: modelos alternativos a serem avaliados em números diferentes de covariáveis inicialmente admissíveis

Fonte: Gluzmann e Panigo, 2015.

A seleção de variáveis explicativas em um modelo de regressão linear é essencialmente um problema de otimização combinatória que busca a solução ótima ao longo de  $2^m$  possíveis modelos de regressão linear, em que  $m$  é o número total de preditores candidatos (Cai *et al.* 2009).

Neste sentido, Calcagno e Mazancourt (2010) desenvolveram o pacote *glmulti* para seleção automática de modelos de regressão, a partir da busca exaustiva em uma lista com todas as combinações possíveis de variáveis explicativas e, opcionalmente, suas interações.

Igarashi *et al.* (2018) propôs um método de busca exaustiva *K-sparse* (ES-K) e um método de busca exaustiva aproximada *K-sparse* (AES-K) para selecionar variáveis explicativas em modelos de regressão linear. Com esses métodos, combinações de variáveis explicativas são testadas exaustivamente.

Sodeyama *et al.* (2018) examinou a precisão e a eficiência de três métodos de ajuste de modelos de regressão linear múltipla: mínimos quadrados ordinários, com operador de seleção (LASSO) e busca exaustiva. A busca exaustiva forneceu as melhores estimativas entre os procedimentos testados.



Para modelos lineares generalizados (MLG), Wang *et al.* (2019), apresentaram o *APproximated Exhaustive Search* (APES), que permite a busca rápida e exaustiva de variáveis em Modelos Lineares Generalizados. Enquanto a busca exaustiva permanece como o padrão ouro em muitos contextos de seleção de modelos, a busca exaustiva sofre de problemas de viabilidade computacional. Mais precisamente, muitas vezes há um alto custo associado ao cálculo das estimativas de máxima verossimilhança para todos os subconjuntos de MLG. Para evitar o elevado custo computacional, o método APES aprende com pesos observacionais em um modelo de regressão linear generalizada ótimo e reformula o problema, MLG, como um problema de regressão linear. Desta forma, o APES pode aproximar-se de uma busca verdadeiramente exaustiva no espaço original do MLG.

A partir das considerações acima, a metodologia proposta para seleção automática de modelos para dados em painel baseia-se no pacote *glmulti* (Calcagno e Mazancourt, 2010), apesar do pacote não considerar tal estrutura de dados.

## 2.7 Dados em Painel

No âmbito da econometria, existem muitas maneiras de estimar a resposta de uma variável dependente em relação às variáveis explicativas, em que cada maneira vai depender da estrutura de dados. Quando as informações estão disponíveis por unidade *cross section* (indivíduo, empresa, unidade geográfica, famílias etc.) e por unidade de tempo (ano, mês, dia), diz-se que estas informações estão organizadas na forma de dados em painel, justamente o mesmo tipo de dados disponibilizados pela ANEEL. A vantagem de se utilizar dados em painel consiste no maior tamanho da amostra e na capacidade de alcançar uma melhor estimativa das relações entre as variáveis, mesmo diante da heterogeneidade entre as unidades, i.e., as concessionárias de distribuição<sup>6</sup>.

Dados em painel permitem modelar aspectos comportamentais mais complexos (Hsiao, 2005). Pode-se, por exemplo, investigar os efeitos de políticas

---

<sup>6</sup> Os efeitos individuais capturam a heterogeneidade entre as distribuidoras deve-se à um conjunto de características únicas da área de concessão, por exemplo, o relevo, a vegetação, fatores climáticos, cultura da população etc., considerados invariantes com o tempo.

públicas quanto ao antes e depois, além de controlar o efeito das variáveis omitidas e desvelar relações dinâmicas entre as variáveis. Na Equação 3.1, apresenta-se o equacionamento genérico de dados em painel.

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \varepsilon_{it} \quad (3.1)$$

em que,

$y_{it}$  = Variável dependente na unidade  $i$  no ano  $t$ ;

$\alpha$  = Intercepto;

$\beta$  = vetor com coeficientes de regressão;

$x_{it}$  = Variáveis explicativas da unidade  $i$  no ano  $t$ ;

$\mu_i$  = efeito individual da unidade  $i$ , invariante no tempo;

$\varepsilon_{it}$  = Desvios idiossincráticos de um modelo de regressão.

Com a utilização de variáveis em formato de dados em painel, três principais tipos de modelos podem ser especificados: o modelo empilhado, o modelo de efeitos fixos e o modelo de efeitos aleatórios (escolhido pela ANEEL). As escolhas entre estes modelos são pautadas de acordo com o comportamento dos efeitos  $\mu_i$ .

O modelo para dados empilhados – ou *pooled* – pressupõe que a inexistência dos efeitos individuais  $\mu_i$ . Já os modelos de efeitos fixos e aleatórios assumem a existência dos efeitos individuais  $\mu_i$ . Porém, a diferença entre os dois modelos é que, enquanto o modelo de efeitos aleatórios assume que não existe correlação da heterogeneidade com nenhuma variável explicativa ( $E(\mu_i x b) = 0$ ), o modelo de efeitos fixos assume tal correlação exista ( $E(\mu_i x b) \neq 0$ ) (Greene, 2012).

A escolha do modelo inadequado compromete a inferência do modelo. Os principais problemas apresentados pela literatura são:

1. A utilização do modelo para dados empilhados, quando existem efeitos individuais (heterogeneidades), leva à “má especificação” do modelo. Na Equação 3.2, apresenta-se o modelo resultante.

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it} \quad (3.2)$$

2. A utilização do modelo de efeitos fixos pretende controlar os efeitos das variáveis omitidas que variam entre indivíduos e permanecem constantes ao longo do tempo. Dessa forma, supõe a presença de um efeito individual  $\mu_i$  que varia de um indivíduo para o outro, mas é constante ao longo do tempo. Na Equação 3.3, apresenta-se o modelo de efeito fixo, em que os efeitos individuais são os coeficientes de variáveis *dummies*.

$$y_{it} = \alpha + \mu_i + \beta x_{it} + \varepsilon_{it} \quad (3.3)$$

3. Para os efeitos aleatórios são utilizadas as mesmas suposições do modelo de efeito fixo, contudo o modelo considera que os indivíduos sobre os quais dispõe-se de dados são amostras aleatórias de uma população maior de indivíduos. Neste caso, os efeitos individuais são modelados por meio de uma variável aleatória normalmente distribuída. Na Equação 3.4, apresenta-se o modelo de efeito aleatório.

$$y_{it} = \bar{\alpha} + \mu_i + \beta x_{it} + \varepsilon_{it} \quad (3.4)$$

A escolha da especificação mais adequada do modelo é realizada por meio testes de hipóteses de Breusch-Pagan (BP) e Hausman, conforme indicado nas equações 3.5 e 3.6. Inicialmente deve-se aplicar o teste de Breusch-Pagan (BP), cujas hipóteses são apresentadas a seguir:

$$\begin{cases} H_0: \mu_i = 0; \text{utilizar o modelo empilhado} \\ H_a: \mu_i \neq 0; \text{utilizar efeitos fixos ou aleatórios} \end{cases} \quad (3.5)$$

A hipótese nula em 3.5 corresponde ao modelo para dados empilhados. Pautado em um teste de multiplicador de Lagrange, a estatística teste tem distribuição Qui-Quadrado, podendo ser acessada em Breush e Pagan (1980).

Verificada a presença da heterogeneidade (rejeição da hipótese nula do teste de Breush-Pagan), aplica-se o teste de Hausman (Hausman, 1978), cujas hipóteses são descritas a seguir:

$$\begin{cases} H_0: E(\mu_i|x) = 0; \text{utilizar modelos de efeitos aleatórios} \\ H_a: E(\mu_i|x) \neq 0; \text{utilizar modelos de efeitos fixos} \end{cases} \quad (3.6)$$

Na Figura 3.2, resume o processo de decisão da escolha adequada dos modelos de dados em painel.

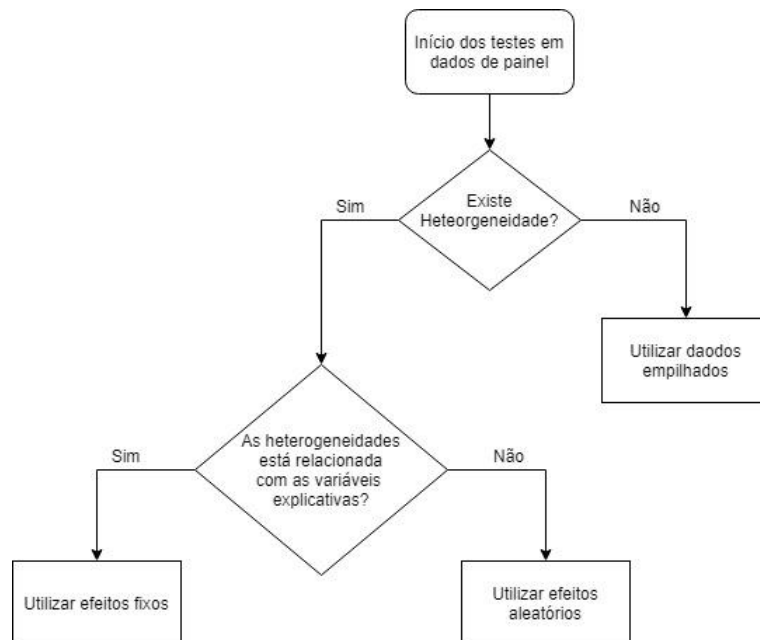


Figura 3.2 - Processo de decisão da escolha adequada do modelo de dados em painel

Adicionalmente, deve-se verificar possíveis violações das premissas básicas assumidas no ajuste do modelo, mais especificamente deve-se avaliar a presença de heterocedasticidade, correlação contemporânea e correlação serial dos erros. Embora tais violações não resultem em viés na estimação dos coeficientes de regressão, a presença delas implica em estimativas viesadas dos erros padrão e assim comprometer a inferência estatística dos coeficientes de regressão<sup>7</sup>.

<sup>7</sup> Em uma regressão existem pressupostos básicos para que o estimador de mínimos quadrados ordinários seja *BLUE* (*Best Linear Unbiased Estimator*), i.e., o melhor estimador linear não tendencioso. Entre os pressupostos destacam-se as premissas de que a variância dos erros é constante (homocedasticidade) e que não há correlação serial (correlação dos valores no tempo) e nem contemporânea entre os erros. Deste modo, verificar se tais pressupostos estão em conformidade é essencial para um estimador BLUE. Tais pressupostos são essenciais para a EFICIÊNCIA do estimador, não afetando a CONSISTÊNCIA. Estimadores não eficientes possuem a característica de alterar o intervalo de confiança dos parâmetros, tornando os testes t enganosos, assim, tornando imprecisa a decisão de "aceitar" a significância dos coeficientes estimados.

A avaliação de possíveis violações é realizada por meio dos seguintes testes de hipóteses: Breusch-Pagan LM (BPLM), Breusch-Godfrey/Wooldridge (BGW) e Breusch-Pagan test (BT test) para autocorrelação contemporânea, autocorrelação serial e heterocedasticidade, respectivamente. Resumidamente, os três testes, em sequência, possuem as seguintes hipóteses apresentadas nas 3.7, 3.8 e 3.9 (para maiores detalhes, verificar Breusch e Pagan, 1979; Breush e Godfrey, 1980; e Wooldridge, 2010).

$$BPLM = \begin{cases} H_0: \text{Ausência de autocorrelação contemporânea} \\ H_a: \text{Presença de autocorrelação contemporânea} \end{cases} \quad (3.7)$$

$$BGW = \begin{cases} H_0: \text{Ausência de autocorrelação serial} \\ H_a: \text{Presença de autocorrelação serial} \end{cases} \quad (3.8)$$

$$BPt = \begin{cases} H_0: \text{homocedasticidade} \\ H_a: \text{heterocedasticidade} \end{cases} \quad (3.9)$$

## 2.8 Considerações sobre o capítulo

Como discutido nesse capítulo, a seleção automática de modelos, objeto desta dissertação, é inspirada em uma rotina disponível em pacote *glmulti* para o *Software R*. No entanto, como este pacote não se utiliza das estatísticas para dados em painel, o modelo que será proposto no capítulo seguinte será feita de forma independente, sendo o aspecto contributo e inovador desta dissertação de mestrado.

Para os modelos em dados em painel, foi apresentado um breve resumo das vantagens da sua utilização. Apesar de ser uma ferramenta que consegue cruzar modelos *cross-section* e séries temporais, é necessário atentar-se para verificação dos testes de hipóteses adequados para que não se obtenha parâmetros estimados enviesados e não consistentes.

No capítulo 4, será apresentado as fases da seleção automática e a classificação dos modelos em dados em painel, objeto desta dissertação, e no

capítulo 5 será apresentada uma demonstração da aplicabilidade desses modelos para o caso brasileiro.

## 4

### Metodologia para seleção automática de modelos para dados em painel

Nesta seção são descritos os procedimentos utilizados na automação da busca exaustiva por modelos de regressão para dados em painel a serem utilizados na determinação do índice de Complexidade, um elemento crucial para definição das metas para as PNT.

#### 2.9

##### Visão geral da metodologia proposta

De forma resumida, a ANEEL, utilizando-se de variáveis socioeconômicas e de mercado para todas as concessionárias do Brasil, formulou diversas especificações de modelos econométricos para dados em painel com efeitos aleatórios, todos visando a determinação do Índice de Complexidade (IC), i.e., um índice que reflete a dificuldade do combate às perdas comerciais (PNT) de acordo com as características intrínsecas de cada área de concessão.

A ANEEL elencou uma longa lista de variáveis explicativas. Assim, a identificação dos melhores modelos de regressão para determinação do índice de complexidade requer uma busca exaustiva e automatizada das várias combinações possíveis de variáveis explicativas.

Conforme indicado na Figura 4.1, a seleção automática dos modelos para dados em painel é dividida em três fases. Na fase 1, ocorre a geração de possíveis modelos a serem testados. A partir dos conjuntos de variáveis explicativas candidatas são listadas todas as combinações de variáveis, i.e., os possíveis modelos. Já na fase 2, cada modelo é ajustado segundo três especificações: dados empilhados, painel com efeitos fixos e painel com efeitos aleatórios. Os ajustes são realizados por meio do pacote *plm* (Croissant & Millo, 2008). Por fim, na fase 3, analisam-se os requisitos necessários para classificar um modelo como viável, i.e.,

coeficientes de regressão estatisticamente significativos (ao nível alfa de 10%) e com sinais coerentes.

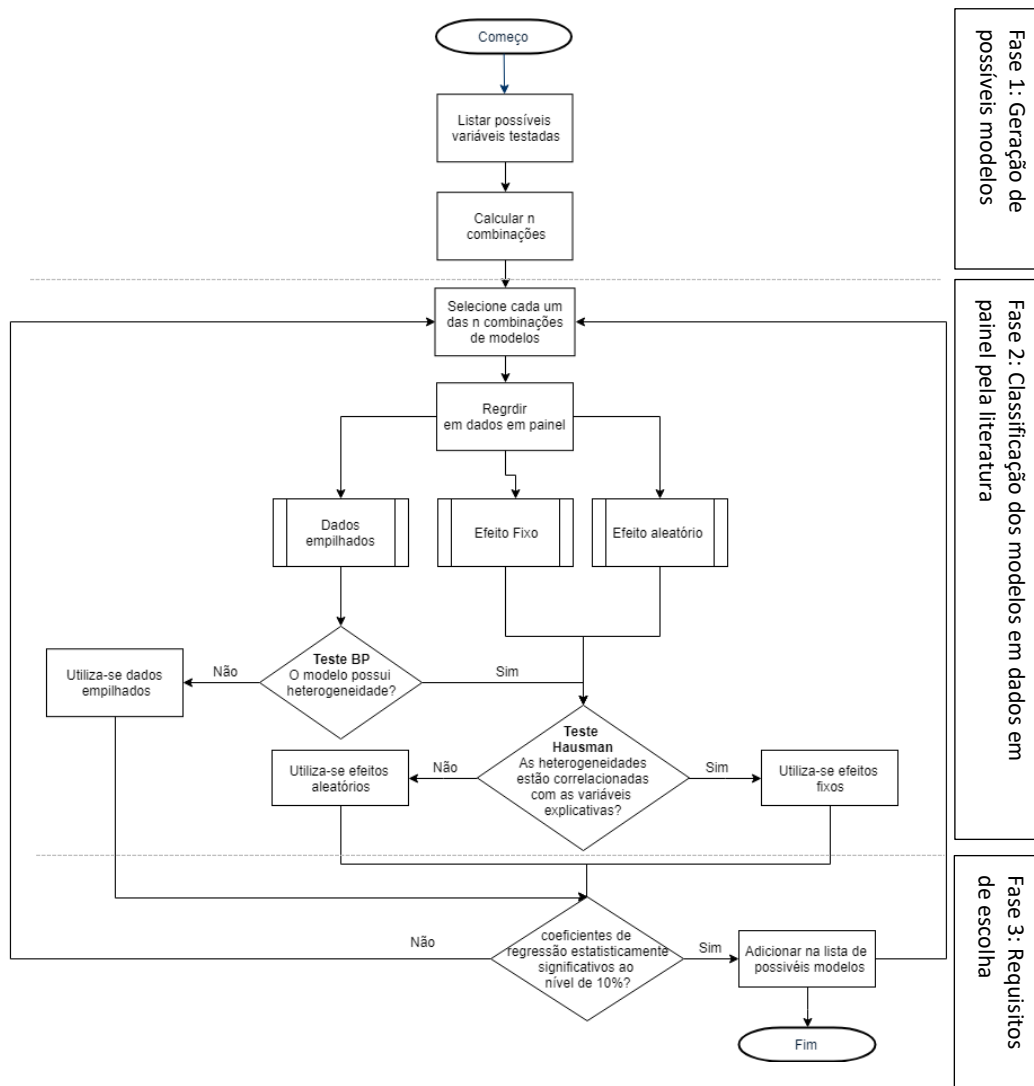


Figura 4.1 - Processo de decisão da escolha adequada do modelo de dados em painel



## 2.10

### Fase 1: Geração de possíveis modelos.

A partir da base de dados com  $m$  variáveis explicativas, são listadas todas as possíveis combinações de variáveis, totalizando  $2^m - 1$  combinações, cada um correspondente à um modelo de regressão.

## 2.11

### Fase 2: Classificação dos modelos em dados em painel

Na fase 2, cada combinação de variáveis identificada na fase 1 é avaliada segundo três tipos de modelos de regressão:

1. Dados empilhados;
2. Efeito fixo; e
3. Efeito aleatório.

Assim, utilizando todo arcabouço metodológico apresentado na seção 2.7, avalia-se qual das três opções acima fornece o melhor ajuste em cada combinação de variáveis. Adicionalmente, a avaliação envolve a análise de possíveis violações das premissas básicas assumidas nos modelos de regressão, mais especificamente a presença de heterocedasticidade, correlação contemporânea e autocorrelação serial dos erros. Embora tais violações não resultem em viés na estimação dos coeficientes de regressão, a presença delas implica em estimativas viesadas dos erros padrão e assim comprometem a inferência estatística dos coeficientes de regressão.

## 2.12

### Fase 3: Requisitos de escolha dos modelos

Nessa última fase, busca-se selecionar os melhores modelos. Em cada modelo avalia-se a significância estatística dos coeficientes de regressão (considera-se um nível de significância alfa de 10%). Assim, ao fim da busca exhaustiva, seleciona-se apenas os modelos que possuem todas as suas variáveis significativas. Após o modelo passar por todas as etapas, monta-se uma lista com os coeficientes estimados, p-valores, AIC e  $R^2$  ajustado de todos os modelos selecionados.

Apenas selecionar os modelos pelos critérios de testes hipóteses e avaliando a significância das variáveis não é suficiente para encontrar a melhor especificação.

Além disso, é preciso avaliar os sinais das estimativas dos coeficientes de regressão em cada modelo. Na Tabela 4.1 são indicados os sinais esperados dos coeficientes das variáveis do problema em tela. Além disso, são apresentadas as correlações das variáveis independentes candidatas e a variável dependente, i.e., o índice de PNT.

Tabela 4.1 – Resumo das variáveis e os sinais esperados dos coeficientes

Variável	Código	Sinal Esperado	Correlação
Densidade de moradores por dormitório	Admd	+	70%
Taxa de evasão escolar do ensino médio	Eem	+	-18%
Percentual de pessoas abaixo da linha da pobreza	Pob3	+	48%
Percentual de domicílios com rede geral de água encanada	Rga	-	-4%
Percentual de domicílios urbanos com coleta de lixo	Lixo.u	-	-13%
Índice de desigualdade social	Gini2	+	13%
Percentual de pessoas que vivem em domicílios subnormais	Sub2	+	67%
Inadimplência geral no setor de crédito do SFN	Inad	+	40%
Inadimplência de pessoas físicas no setor de crédito do SFN	Inad.pf	+	42%
Homicídio doloso	Homi_dolo	+	26%
Intervenções policiais e operações de guerra	Int_pol	+	20%
Furto de veículos	Furto_v	+	-18%
Roubo de veículos	Foubo_v	+	16%
Latrocínio	Latro	+	13%
Agressão por arma de fogo	Agraf	+	32%
Homicídio (óbitos por agressão)	Vio	+	40%
Homicídios (mortes violentas)	H.mvci	+	37%
Unidade consumidoras de baixa renda/unidades consumidoras de BT	Ucbr.Mb1UCbr	-	22%
Mercado de baixa renda/mercado B1 Total	Mbr.Mb1Mbr	-	11%
Mercado de baixa renda/mercado BT Total	Mbr.Mbt	-	17%
PIB per capita da área de concessão	PIB.PC	-	-19%
Tempo em anos	Time	*	-6%

Ressalta-se que alguns coeficientes de regressão apresentam sinais opostos ao esperado, conforme ilustrado no Quadro 4.1. Nesses casos, uma hipótese, é que as variáveis são pouco correlacionadas com o fenômeno de perdas não técnicas.

Quadro 4.1 – Correlação com o sinal esperado inconsistente

Correlação Negativa	Correlação positiva
Furto_v	Eem, Ucbr.Mb1UCbr, Mbr.Mb1Mbr e Mbr.Mbt

Ao analisar os modelos ao final do processo descrito nas seções anteriores (4.1 e 4.2), em geral, algumas variáveis irão ser mais recorrentes que outras, indicando maior relevância na modelagem de PNT. Desta forma, utiliza-se o critério de informação de Akaike (AIC) para auxiliar na seleção do modelo. O AIC é fundamentado na teoria da informação. Quando um modelo estatístico é usado para representar um determinado processo, a representação nunca será exata, ou seja, o modelo nunca será perfeito e certamente algumas informações serão perdidas. Quando um modelo estatístico é usado para representar um determinado processo (Taddy, 2019).

Uma solução direta para o problema seria selecionar o modelo que possui maior quantidade de variáveis explicativas, no entanto, os modelos devem ser parcimoniosos, i.e., não é desejável considerar sempre o modelo mais complexo como o melhor (Myung e Pitt, 1997, Myung *et al.*, 2000). Na realidade, o melhor modelo é aquele que fornece uma descrição adequada dos dados enquanto utiliza um número mínimo de parâmetros, conforme indicado por Myung e Pitt (1997) e Myung *et al.* (2000).

Para analisar os modelos deste artigo, serão considerados os pesos de Akaike, conforme Wagenmakers e Farrell (2004). Estes autores demonstram que os valores da AIC podem ser facilmente transformados nos chamados pesos de Akaike, que podem ser interpretados de maneira indireta como probabilidades de seleção dos modelos, podendo facilitar enormemente a interpretação dos resultados dos procedimentos de comparação do modelo AIC.

Na Equação 4.1, calcula-se, para cada modelo, as diferenças no AIC em relação ao AIC do melhor modelo candidato, ou seja, aquele que apresentou o menor AIC.

$$\Delta_i(AIC) = AIC_i - \min AIC \quad (4.1)$$

Pelas diferenças no AIC, pode-se obter uma estimativa da verossimilhança relativa  $L$  do modelo, conforme indicado na Equação 4.2, na qual  $\propto$  significa "é proporcional a".

$$L(M_i | data) \propto \exp \left\{ -\frac{1}{2} \Delta_i(AIC) \right\} \quad (4.2)$$

Na última etapa, as verossimilhanças relativas do modelo são normalizadas (ou seja, divididas pela soma das verossimilhanças de todos os modelos) para obter o Peso Akaike, conforme Equação 4.3.

$$w_i(AIC) = \frac{\exp \left\{ -\frac{1}{2} \Delta_i(AIC) \right\}}{\sum_{k=1}^k \exp \left\{ -\frac{1}{2} \Delta_k(AIC) \right\}} \quad (4.3)$$

em que  $\sum w_i(AIC) = 1$ .

O peso  $w_i(AIC)$  pode ser interpretado como a probabilidade do modelo  $i$  ser o melhor modelo (no sentido de AIC, que minimiza a Discrepância de Kullback-Leibler), conforme citado em Wagenmakers e Farrell (2004).

## 2.13 Considerações sobre o capítulo

Nesse capítulo foi proposto uma metodologia para definição de metas regulatórias de PNT partindo-se de uma árvore de decisão aplicando testes em dados em painel. Além disso, foi feita uma avaliação dos sinais esperados dos coeficientes de regressão. Foi proposto também o uso do Peso de Akaike para auxiliar a seleção dos modelos.

A vantagem da metodologia proposta reside na automação da seleção das variáveis explicativas a serem consideradas nos modelos para dados em painel utilizados na definição das metas regulatórias, evitando uma série de etapas *ad hoc* adotadas pelas ANEEL na abordagem atualmente em uso.

O capítulo 5 apresenta uma aplicação desse modelo às distribuidoras brasileiras, comparando os seus resultados com as perdas atuais das distribuidoras como forma de demonstrar não só a viabilidade da aplicação do modelo proposto, como também demonstrar as suas vantagens em relação ao modelo atualmente adotado pela ANEEL.

## 5

### Demonstração da aplicabilidade da metodologia proposta para o problema da perda não técnica no Brasil

Nesta seção, são apresentados os principais resultados obtidos pela metodologia proposta.

#### 2.14

#### Resultado da metodologia seleção automática

Na Figura 5.1, é apresentado o fluxograma das simulações. Observa-se que foram testados 1.097.789 modelos com diferentes combinações de variáveis. Como resultado das variáveis significativas, restaram 6.523 que podem ser divididos em 501 modelos com efeitos aleatórios e 6.022 modelos com efeitos fixos. Por fim, foram avaliados os sinais dos coeficientes para que estejam coerentes com a variável resposta de interesse. Desta forma, apenas 18 modelos de dados em painel com efeito aleatório e 42 modelos de dados em painel com efeito fixo mostraram coeficientes de regressão com sinais coerentes com as expectativas.

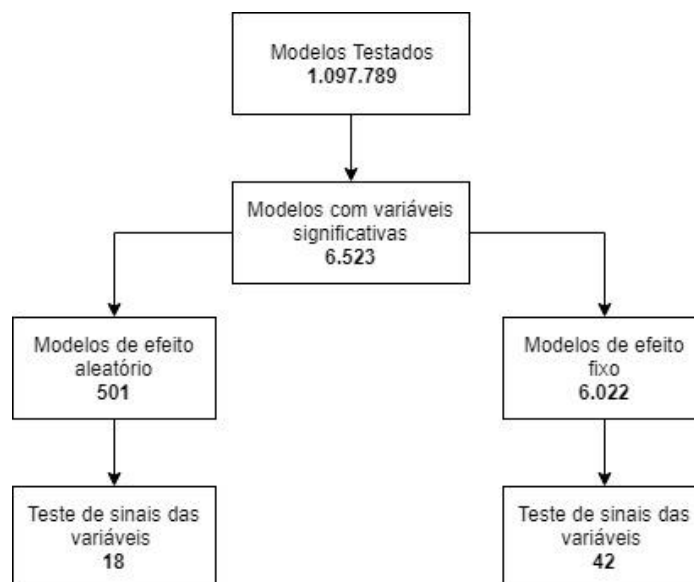


Figura 5.1 – Fluxograma da seleção dos modelos

Das 22 variáveis utilizadas, apenas 11 foram selecionadas em algum dos modelos de dados em painéis, sendo apresentadas na Tabela 5.1. Além disso, é possível verificar a recorrência em que uma mesma variável foi utilizada. Dessa forma, é possível verificar que existem alguns atributos que são melhores preditores das PNT do que outras, por exemplo, a inadimplência (Inad e Inad.pf) e o saneamento (Lixo.u), que são recorrentes nos modelos significativos.

Tabela 5.1 – Resultado do conjunto de variáveis significativas

#	Código	Nome da variável	Qtd. modelos
1	Inad	Inadimplência geral no setor de crédito do SFN	28
2	Inad.pf	Inadimplência de pessoa física no setor de crédito do SFN	24
3	Lixo.u	Percentual de domicílios urbanos com coleta de lixo	19
4	Pob3	Percentual de pessoas abaixo da linha da pobreza	17
5	Gini	Índice de desigualdade social	15
6	Time	Tempo em anos	14
7	Admd	Densidade de moradores por dormitório	11
8	Furto_V	Furto de veículos	7
9	PIB.PC	PIB per capita da área de concessão	4
10	Vio	Homicídio (óbitos por agressão)	2
11	Mbr.Mb1Mbr	Mercado de baixa renda/mercado B1 Total	1

Fonte: Elaboração Própria

A importância de uma variável explicativa pode ser avaliada pela soma dos pesos de Akaike de todos os modelos que incluem a variável (Silva & Pessanha, 2020). Na Figura 5.2, são apresentados os resultados da importância das variáveis explicativas aos 60 modelos selecionados. Observa-se que as inadimplências, o índice de Gini, o índice de pobreza, a coleta de lixo e densidade de moradores são as variáveis que concentram a maior importância.

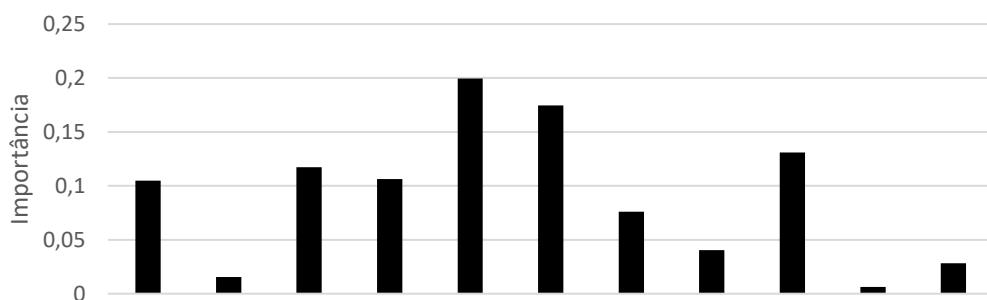


Figura 5.2 – Importância das variáveis explicativas

## 2.15 Utilização dos pesos de Akaike

Para a aplicação dos pesos de Akaike, optou-se por fazer de forma separada respeitando o grupo dos efeitos aleatórios e os de efeitos fixos. Na Tabela 5.2, são apresentados os resultados dos modelos de efeito aleatórios que tiveram maior impacto. Observa-se que os dois modelos concentram 100% das informações.

Tabela 5.2 – Resultado dos modelos com efeitos aleatórios

#	(Intercept)	pob3	inad	Inad.pf	AIC	R <sup>2</sup> -ajustado	DELTA AIC	Peso de AKAIKE	p-valor
1	0,01	0,38	1,33		- 2.749,87	0,17	0,57	43%	0,000
2	0,01	0,33	1,00	0,53	- 2.750,44	0,17	-	57%	0,000

Na Tabela 5.3, apresentam-se os resultados do modelo de efeito fixo. Pelo peso de Akaike, apenas um modelo foi suficiente para conter as informações de todos os outros.

Tabela 5.3 – Resultado dos modelos com efeitos fixos

Time	pob3	Inad.pf	AIC	R-ajustado	DELTA AIC	Peso de AKAIKE	p-valor
0,00	0,34	1,21	-2.817,83	0,10	-	100%	0,000



Conforme Tabela 5.2 e Tabela 5.3, a única variável que aparece em todos os modelos é a pob3. Na realidade, as variáveis de renda são as únicas que possuem o maior peso para descrever o fenômeno de perdas não técnicas.

## 2.16 Predição dos modelos

A seguir, na Tabela 5.4, apresentam-se as estimativas obtidas pelos modelos de regressão para os índices de perdas não técnicas das distribuidoras. Adicionalmente, na tabela são informados o porte e a região do Brasil atendida por cada distribuidora de energia elétrica. O conceito de porte encontra-se definido na NTº ANEEL 106/2015:

*“sendo consideradas de maior porte (Grupo 1) aquelas que possuam mercado maior que 1.000 GWh/ano e atendam mais que 500 mil unidades consumidoras ou que possuam mais do que 15.000 km de rede elétrica (...). As demais empresas são consideradas do Grupo 2. Isso porque empresas de menor porte tendem a apresentar maior facilidade de combate às perdas não técnicas que empresas grandes”.*  
**(Grifo nosso)**

Dessa forma, observa-se que os valores estimados dos modelos são similares. Para PNT real de 2017, chama atenção as perdas da distribuidora de grande, principalmente, a distribuidora AMAZONAS que possuem perdas maiores que o seu mercado de baixa tensão.

Tabela 5.4 – Resultado preditivo das distribuidoras de energia elétrica

Empresa	Região	Modelo efeito aleatório (#1)	Modelo efeito aleatório (#2)	Modelo efeito fixo	PNT 2017	Porte
AES-SUL	SU	8%	8%	8%	11%	Grande
AMAZONAS	NO	125%	124%	124%	124%	Grande
AMPLA	SE	32%	31%	31%	30%	Grande
BANDEIRANTE	SE	19%	19%	19%	13%	Grande
CAIUA	SE	2%	2%	2%	2%	Pequena
CEAL	NE	55%	54%	53%	36%	Grande
CEB	CO	8%	8%	8%	9%	Grande
CEEE	SU	28%	28%	28%	26%	Grande
CELESC	SU	4%	4%	4%	9%	Grande
CELG	CO	8%	8%	8%	10%	Grande

<b>Empresa</b>	<b>Região</b>	<b>Modelo efeito aleatório (#1)</b>	<b>Modelo efeito aleatório (#2)</b>	<b>Modelo efeito fixo</b>	<b>PNT 2017</b>	<b>Porte</b>
CELPA	NO	46%	46%	45%	39%	Grande
CELPE	NE	22%	21%	21%	18%	Grande
CELTINS	NO	7%	6%	7%	4%	Grande
CEMAR	NE	22%	21%	22%	11%	Grande
CEMAT	CO	12%	12%	12%	10%	Grande
CEMIG	SE	12%	12%	11%	14%	Grande
CEPISA	NE	43%	42%	43%	30%	Grande
CERON	NO	43%	42%	43%	50%	Grande
CFLO	SU	1%	0%	0%	0%	Pequena
CHESP	CO	3%	3%	3%	7%	Pequena
CJE	SE	3%	2%	2%	0%	Pequena
MOCOCA	SE	3%	2%	2%	7%	Pequena
SANTA CRUZ	SE	3%	2%	2%	5%	Pequena
NACIONAL	SE	1%	1%	1%	1%	Pequena
COCEL	SU	4%	3%	3%	5%	Pequena
COELBA	NE	12%	11%	11%	12%	Grande
COELCE	NE	8%	8%	7%	14%	Grande
COOPERALI ANCA	SU	6%	6%	6%	6%	Pequena
COPEL	SU	4%	4%	4%	4%	Grande
COSERN	NE	7%	6%	7%	3%	Grande
CPEE	SE	5%	5%	5%	6%	Pequena
PIRATININGA	SE	6%	6%	6%	9%	Grande
CPFL PAULISTA	SE	7%	6%	6%	10%	Grande
CSPE	SE	5%	4%	4%	11%	Pequena
DEMEI	SU	7%	6%	6%	3%	Pequena
DMED	SE	5%	4%	4%	3%	Pequena
EBO	NE	8%	8%	9%	3%	Pequena
EVP	SE	1%	0%	0%	1%	Pequena
BRAGANTIN A	SE	2%	2%	2%	2%	Pequena
JOAO CESA	SE	2%	1%	2%	2%	Pequena
EFLUL	SU	3%	3%	3%	3%	Pequena
ELEKTRO	SE	5%	4%	4%	8%	Grande
ELETROACR E	NO	26%	25%	26%	27%	Grande
ELETROCAR	SU	5%	5%	5%	6%	Pequena
ELETROPAU LO	SE	13%	12%	13%	10%	Grande
SANTA MARIA	SE	10%	9%	9%	3%	Pequena
EMG	SE	4%	4%	3%	4%	Grande
ENERSUL	CO	15%	15%	16%	7%	Grande
ENF	SE	4%	4%	3%	0%	Pequena
EPB	NE	16%	16%	17%	8%	Grande

Empresa	Região	Modelo efeito aleatório (#1)	Modelo efeito aleatório (#2)	Modelo efeito fixo	PNT 2017	Porte
ESCELSA	SE	23%	23%	22%	17%	Grande
Esse	NE	14%	14%	14%	8%	Grande
FORCEL	SU	1%	1%	1%	5%	Pequena
HIDROPAN	SU	1%	0%	0%	4%	Pequena
IENERGIA	SU	8%	8%	8%	9%	Pequena
LIGHT	SE	49%	48%	48%	51%	Grande
MUXFELDT	SU	2%	2%	2%	0%	Pequena
RGE	SU	6%	6%	6%	7%	Grande
SULGIPE	NE	12%	11%	11%	9%	Pequena
UHENPAL	SU	4%	4%	3%	0%	Pequena

Em geral, as PNT reais das empresas ficaram próximos com a meta definida pelos modelos. Na Quadro 5.1, apresenta-se uma síntese das empresas que tiveram os melhores e piores resultados dos modelos ao comparar suas perdas 2017. Observa-se que as empresas que estão performando abaixo dos resultados das regressões são empresas que possuem perdas reais altas, regionalmente concentrada nas regiões nordeste e sudeste. No entanto, para empresas que estão performando acima do esperado, atenta-se que são empresa que possuem perdas baixas, abaixo de 15%, exceto a CERON (NO) que possuem um patamar de perdas em 50%. São empresas que concentram mais nas regiões sul, sudeste e centro-oeste.

Quadro 5.1 – Sínteses das distribuidoras que tiveram os melhores e piores resultados dos modelos ao comparar suas perdas 2017

Empresas abaixo do esperado			Empresas acima do esperado		
CELPA	NO	Grande	CELESC	SU	Grande
CEMAT	CO	Grande	COELCE	NE	Grande
CELPE	NE	Grande	ELEKTRO	SE	Grande
ELETROPAULO	SE	Grande	PIRATININGA	SE	Grande
ESCELSA	SE	Grande	CPFL PAULISTA	SE	Grande
CEPISA	NE	Grande	AES-SUL	SU	Grande
BANDEIRANTE	SE	Grande	CELG	CO	Grande
CEAL	NE	Grande	CEMIG	SE	Grande
CELTINS	NO	Grande	RGE	SU	Grande
Esse	NE	Grande	CERON	NO	Grande
CEMAR	NE	Grande	CEB	CO	Grande
EPB	NE	Grande			
ENERSUL	CO	Grande			

COSERN	NE	Grande			
--------	----	--------	--	--	--

Estes resultados servem como balizadores para a definição das metas de perdas não técnicas regulatórias, sendo um *input* valioso para a ANEEL na definição de tais metas.

## 2.17 Discussão dos resultados

Resumidamente, foram feitas simulações exaustivas combinando todas as variáveis disponibilizada pela ANEEL. Consideraram-se apenas os modelos em que os coeficientes de regressão fossem estatisticamente significativos ao nível de significância de 10%, e com sinais coerentes com o esperado. Cabe ressaltar que as variáveis relacionadas a infraestrutura e inadimplência tiveram maior correlação com o fenômeno de perdas não técnicas.

Os resultados obtidos a partir da aplicação do modelo proposto para o caso brasileiro fomentam algumas discussões importantes, e mostram um pouco do potencial dos modelos.

Segundo Calili *et al.* (2017) a metodologia empregada pela ANEEL apresenta diversas limitações que acabam por resultar em distorções relevantes no processo de reconhecimento tarifário das perdas não técnicas. Além de que à metodologia atualmente empregada pela ANEEL é excessivamente complexa, frágil do ponto de vista do metodológico e, em certa medida incompleta, o que impõem ao órgão regulador a adoção de uma série de medidas *ad hoc* que acaba resultando em metas de perdas pouco condizentes com a realidade das distribuidoras. De modo geral, esse trabalho conseguiu simular modelos auxiliares, pautada pela literatura de dados em painel, com maior grau de parcimônia e robustez.

Outro ponto para ser discutido, é as escolhas das variáveis selecionadas. As simulações do modelo exaustivo resultaram no maior peso nas variáveis de renda e infraestrutura havendo indícios de uma maior correlação com fenômeno de PNT. Os resultados, sugerem por exemplo, que a introdução das variáveis de inadimplência de pessoa física e de coleta de lixo urbano tiveram maior recorrência

com as dos modelos simulados. Isso sugerem que esses fenômenos talvez estejam mais concentrados em algumas regiões específicas.

Os resultados apresentados também fomentam uma discussão importante sobre a exequibilidade das metas de PNT regulatórias atualmente estabelecidas pela ANEEL. O modelo proposto reforça um ponto que foi evidenciado no capítulo 2, que as PNT são modestamente reconhecidas nas tarifas, um problema particularmente relevante para algumas distribuidoras. Alguns desses casos, inclusive já reconhecidos, mesmo que de forma parcial, pela própria ANEEL, como no caso das Enel Rio e das distribuidoras recentemente privatizadas.

Há duas consequências deletérias desse fenômeno. Ao não reconhecer parte dos custos das distribuidoras com PNT, o regulador limita a capacidade de investimento da concessionária, inclusive no próprio processo de combate a PNT. Por outro lado, o próprio regulador passa a ter que adotar uma série de procedimento ad hoc para tentar lidar com essa situação, tornando o processo mais complexo e menos previsível.

## 6 Conclusões

O objetivo geral da dissertação foi selecionar as variáveis mais relevantes para o fenômeno de PNT e definir modelos que possam auxiliar no estabelecimento das metas regulatórias de redução de perdas não técnicas para as distribuidoras de energia elétrica brasileiras, equilibrando o princípio da regulação por incentivo e a complexidade da área de atuação de cada distribuidora

No capítulo 2, foi realizada uma breve explanação sobre o problema das PNT e dois principais conceitos que norteiam a Regulação do setor de distribuição de energia elétrica. Foram também identificadas as principais limitações no modelo atual adotado pela ANEEL para definição das metas regulatórias de PNT. Como demonstrado nesse capítulo a excessiva complexidade da metodologia atualmente empregada pela ANEEL na definição das metas de PNT regulatória, assim como a adoção de um conjunto grande critérios ad hoc, põem em risco um dos principais parâmetros que norteiam a regulação por incentivo: o estabelecimento de incentivos econômicos corretos para busca contínua pela melhoria da eficiência.

Os problemas apontados, apresentam indícios claros que há falhas no processo de comparação das PNT entre as distribuidoras que podem estar conduzindo tanto a um modesto reconhecimento da PNT de algumas distribuidoras quanto ao reconhecimento indevido da PNT em outras.

Com o objeto de formular um modelo alternativo, no capítulo 3 foram apresentados as ferramentas para seleção automática de modelos para dados em painel, um recurso resultante da evolução da tecnologia.

No capítulo 4 foi então formulado a metodologia e as premissas utilizadas para escolha das variáveis e dos modelos que auxiliam as definições de metas regulatórias de PNT a partir da metodologia em dados em painel. Para seleção dos modelos finais, foram feitas simulações exaustivas combinando todas as variáveis disponibilizada pela ANEEL. Como filtro, consideraram-se apenas os modelos em

que os coeficientes de regressão fossem estatisticamente significativos ao nível de significância de 10%, e com sinais coerentes com o esperado. Cabe ressaltar que as variáveis relacionadas a renda e inadimplência tiveram maior correlação com o fenômeno de perdas não técnicas.

O modelo proposto teria como vantagens não só tornar o processo de seleção das variáveis mais assertivas e direto, evitando uma série de etapas ad hoc adotadas pela ANEEL com o uso do seu modelo atual, como também tornaria o processo mais robusto em relação a seleção ao processo de seleção/omissão de variáveis. Um ponto bastante relevante é que o modelo proposto permitiria avaliar de uma forma mais criteriosa e objetiva o comportamento das distribuidoras no combate as PNT.

No capítulo 5 foi feita uma aplicação dos modelos propostos ao contexto do setor de distribuição de energia elétrica brasileiro para comparação das perdas reais das distribuidoras no ano de 2017. Os resultados oriundos dos efeitos aleatórios e de efeitos fixos foram similares. Além disso, ao comparar com as perdas reais de cada empresa, foi possível inferir quais são as distribuidoras ineficientes no combate às perdas.

A proposta apresentada nesse trabalho, no entanto, não é exaustiva. Uma comparação mais ampla, incluindo um conjunto de anos mais recentes, com os modelos da ANEEL poderia confirmar ou não alguns pontos levantados e poderia ser objeto de um trabalho futuro sobre o tema. Uma série de críticas levantadas no capítulo 2 em relação ao modelo da ANEEL, especialmente em relação as variáveis socioeconômicas, continua em aberto e também poderia ser objeto de futuros trabalhos.

Ademais a possibilidade de aplicação de outras variáveis para interagir com a análise de dados em painel utilizando a busca exaustiva dos modelos mais correlacionadas com os fenômenos de PNT. Contornando um problema citado por Calili *et al.* (2017) sobre a metodologia empregada pela ANEEL apresentar diversas limitações que resultam em resultados enviesados para o processo de reconhecimento tarifário das perdas não técnicas.

## 7

### Referências bibliográficas

Agrell, P.; Bogetoft, P. *Regulatory benchmarking: Models, analyses and applications*. **Data Envelopment Analysis Journal**, v.3, n. 1, p. 49-91, 2016.

Amin, S., Schwartz, G. A., Cardenas, A. A., & Sastry, S. S. (2015). Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure. **IEEE Control Systems**, v.35, p.66–81, 2015.

ANEEL. Submodulo 2.6 - **Procedimentos de Regulação Tarifária – PRORET**. 2015. Disponível em: < <http://www.aneel.gov.br/procedimentos-de-regulacao-tarifaria-proret>>. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 271/2010-SRE/ANEEL– Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica**. 2010. Disponível em: <http://www.aneel.gov.br/audiencias-publicas> >. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 031/2011-SRE/ANEEL– Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica**. 2011. Disponível em: <http://www.aneel.gov.br/audiencias-publicas> >. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 453/2013-SGT/SRM/ANEEL– Discussão conceitual sobre a metodologia de definição do nível regulatório de perdas na distribuição, a qualidade dos serviços e os incentivos regulatórios**. 2013a. Disponível em: <http://www.aneel.gov.br/consultas-publicas>>. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 494/2013-SRE/ANEEL– Disponibilização da base de dados**. 2013b. Disponível em: <http://www.aneel.gov.br/consultas-publicas>>. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 188/2014-SRE/ANEEL– Atualização do Índice de Complexidade Socioeconômica adotado para definição do nível regulatório de Perdas Não Técnicas na Distribuição**. 2014. Disponível em: <http://www.aneel.gov.br/audiencias-publicas> >. Acesso em: 07 nov.18.

\_\_\_\_\_. **Nota Técnica 406/2014-SRE/ANEEL– Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica**. 2014. Disponível em: <http://www.aneel.gov.br/audiencias-publicas> >. Acesso em: 07 nov.18.



\_\_\_\_\_. **Nota Técnica 106/2015-SGT/SRM/ANEEL– Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica.** 2015. Disponível em: <http://www.aneel.gov.br/audiencias-publicas> >. Acesso em: 07 nov.18.

\_\_\_\_\_. **Base de dados – Perdas.** 2019. Disponível em: < [https://www.aneel.gov.br/documents/15188137/0/Base\\_Perdas\\_Internet+-05-2019.xlsx/2e0051a3-93c9-a5b9-7f7b-2b0977adb00c](https://www.aneel.gov.br/documents/15188137/0/Base_Perdas_Internet+-05-2019.xlsx/2e0051a3-93c9-a5b9-7f7b-2b0977adb00c) >. Acesso em: 03 jun.19.

\_\_\_\_\_. **Nota Técnica Nº 46/2020 – SRM/SGT/ANEEL– Proposta de Consulta Pública para revisão da metodologia e atualização dos parâmetros dos Submódulos 2.2/2.2 A (Receitas Irrecuperáveis) e 2.6 (Perdas de Energia) dos Procedimentos de Revisão Tarifária - Proret.** 2020. Disponível em: < [https://www.aneel.gov.br/consultas-publicas?p\\_p\\_id=participacaopublica\\_WAR\\_participacaopublicaportlet&p\\_p\\_lifecycle=2&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_cacheability=cacheLevelPage&p\\_p\\_col\\_id=column-2&p\\_p\\_col\\_pos=1&p\\_p\\_col\\_count=2&participacaopublica\\_WAR\\_participacaopublicaportlet\\_idDocumento=39824&participacaopublica\\_WAR\\_participacaopublicaportlet\\_tipoFaseReuniao=fase&participacaopublica\\_WAR\\_participacaopublicaportlet\\_jspPage=%2Fhtml%2Fpp%2Fvisualizar.jsp](https://www.aneel.gov.br/consultas-publicas?p_p_id=participacaopublica_WAR_participacaopublicaportlet&p_p_lifecycle=2&p_p_state=normal&p_p_mode=view&p_p_cacheability=cacheLevelPage&p_p_col_id=column-2&p_p_col_pos=1&p_p_col_count=2&participacaopublica_WAR_participacaopublicaportlet_idDocumento=39824&participacaopublica_WAR_participacaopublicaportlet_tipoFaseReuniao=fase&participacaopublica_WAR_participacaopublicaportlet_jspPage=%2Fhtml%2Fpp%2Fvisualizar.jsp)>. Acesso em: 20 abr.20.

\_\_\_\_\_. **Base de Perdas disponibilizada no âmbito da Consulta Pública 029/2020 da Aneel.** 2020b. Disponível em: [https://www.aneel.gov.br/consultas-publicas?p\\_auth=nJuT0z50&p\\_p\\_id=participacaopublica\\_WAR\\_participacaopublicaportlet&p\\_p\\_lifecycle=1&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_col\\_id=column-2&p\\_p\\_col\\_pos=1&p\\_p\\_col\\_count=2&participacaopublica\\_WAR\\_participacaopublicaportlet\\_idParticipacaoPublica=3431&participacaopublica\\_WAR\\_participacaopublicaportlet\\_javax.portlet.action=visualizarParticipacaoPublica](https://www.aneel.gov.br/consultas-publicas?p_auth=nJuT0z50&p_p_id=participacaopublica_WAR_participacaopublicaportlet&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_pos=1&p_p_col_count=2&participacaopublica_WAR_participacaopublicaportlet_idParticipacaoPublica=3431&participacaopublica_WAR_participacaopublicaportlet_javax.portlet.action=visualizarParticipacaoPublica) >. Acesso em: 20 abr.20.

Antmann, P. “Strategies to achieve sustainable improvements in operational performance of electricity distribution companies.” **Work paper**, World Bank, 2009.

Arango, L. G.; Deccache, E.; Bonatto, B. D. ; Arango, H.; Pamplona, E. O. Study of Electricity Theft Impact on the Economy of a Regulated Electricity Company. **Journal Of Control Automation And Electrical Systems**, v.28, n.4, p. 567-575, 2017.

Baumol, W. J. O. Contestable Markets: An Uprising in the Theory of Industry Structure. **The American Economic Review**, Vol. 72, No. 1, p. 1-15, 1982.

Baumol, W. & Sidak, J. **Transmission Pricing and Stranded Costs in the Electric Power Industry**. The AEI Press, 1995.

Breusch, T. S., e A. R. Pagan. "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." **Review of Economic Studies**: 47, 1980: 239-254.

Breusch, T. S., e A. R. Pagan. "A Simple Test for Heteroskedasticity and Random Coefficient Variation." **Econometrica**, 47, 1979: 1287-1294.

Breusch, T. S., e L. G. Godfrey. "A Review of Recent Work on Testing for Autocorrelation in Dynamic Economic Models." **Discussion Paper \*8017**, University of Southampton., 1980.

Cai, Airong, Ruey S. Tsay, and Rong Chen. "Variable Selection in Linear Regression With Many Predictors." **Journal of Computational and Graphical Statistics** 18, no. 3 (2009): 573-91. Accessed April 10, 2021.

Calcagno, V., e C. Mazancourt. "glmulti: An R Package for Easy Automated Model Selection with ( Generalized ) Linear Models." **Journal Of Statistical Software**, 2010.

Calili, R. F.; Souza, R. C.; Simões, P.F.M. **Audiência 095/2016 -Termo aditivo ao Contrato de Concessão de Distribuição de Energia Elétrica da Enel Distribuição Rio Energia e Serviços S.A.: contribuições e críticas à metodologia de perdas não técnicas**. Contribuição a Audiência Pública ANEEL095 de 2016. 2017. Disponível em: < <http://www.ANEEL.gov.br/audiencias-publicas> >. Acesso em: 08 nov.18.

Croissant, Y., e G. Millo. "Panel Data Econometrics in R: The plm Package." **Journal of Statistical Software**, 2008: 1–43.

Depuru S., Wang L, Devabhaktuni V. Electricity theft: overview, issues, prevention and smart meter based approach to control theft. **Energy Policy**, v.39, p.1007–15, 2011.

Draper, H., e Smith, H., **Applied Regression Analysis (3rd ed.)**, New York: Wiley, 1998

Fourie, J.W. **A strategy for the management of energy losses in a local electricity distribution network**. Dissertação de Mestrado, Faculty of Engineering, the Built Environment and Information Technology, University Of Pretoria, 2005.

Gluzmann, P. e Panigo, D. Global Search Regression: A New Automatic Model-selection Technique for Cross-section, Time-series, and Panel-data Regressions. **Stata Journal**. 15. 2015.

Gopi, C, e Lalu, V. "Sensor network infrastructure for {AMI} in smart grid." **International Conference on Emerging Trends in Engineering, Science and Technology**, 2016.

Greene, W. H. **Econometric Analysis**. ADDISON WESLEY PUB CO INC, 2012.

Haney, A. B.; Pollitt, M. Efficiency analysis of energy networks: An international survey of regulators. **Energy Policy**, v.37, .n.12, p. 5814–5830, 2009.

Hausman, J.A. “Specification Tests in Econometrics.” **Econometrica**: 46, 1978: 1251–1271.

Hsiao, C. “Why Panel Data ? .” **IEPR WORKING PAPER** 05.33 , 2005.

Igarashi, Y., Takenaka, H., Nakanishi-Ohno, Y., Uemura, M., Ikeda, S., and Okada, M., “Exhaustive Search for Sparse Variable Selection in Linear Regression”, **Journal of the Physical Society of Japan**, vol. 87, no. 4, 2018.

Jamasb, T.; Pollitt, M.; Triebs, T. Productivity and efficiency of US gas transmission companies: A European regulatory perspective. **Energy Policy**, v.36, .n.9, p. 3398-3412,

Junior, G. A metropolização da criminalidade com a territorialização das UPPs: da migração a expansão. **Revista Tamoios**. 2016.

Kumbhakar, S.C. Production frontiers, panel data and time-varying technical inefficiency, **Journal of Econometrics**, v.46, pp. 201-211, 1990.

Kumbhakar, S. C.; Lovell, C.A.K. **Stochastic Frontier Analysis**. Cambridge University Press, 2003.

Kumbhakar, S.C.; Wang, H.J.; Horncastle, A.P. A Practitioner’s Guide to Stochastic Frontier Analysis Using Stata. **Cambridge University Press: Cambridge**, 2015.

Kumbhakar, S.C.; Lien, G. Yardstick Regulation of Electricity Distribution-Disentangling Short-run and Long-run Inefficiencies. **Energy Journal**, v.38, n.5, p. 17-38, 2017.

Kuosmanen, T; Kortelainen, M. Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. **Journal of Productivity Analysis**, v.38, p. 1-18, 2010.

Kuosmanen, T. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. **Economic Energy**, v.34, p. 2189-2199, 2012.

Kuosmanen, T.; Saastamoinen, A.; Sipilainen, T. What is the best practice for *benchmark* regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods. **Energy Policy**, v.61, p. 740-750, 2013.

Leite, D.; Pessanha, J.; Simões, P.; Calili, R. e Souza, R. A Stochastic Frontier Model for Definition of Non-Technical Loss Targets. **Energies**, 2020a.

Leite, D.; Calili, R. e Pessanha, J. Definição das metas regulatórias de perdas não técnicas para as distribuidoras de energia elétrica no Brasil: proposição de um modelo baseado na Análise de Fronteira Estocástica (SFA). **Dissertação de Mestrado, Programa de Pós-graduação em Metrologia. Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)**, 2020b.

MacKinnon, J. e White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, **Journal of Econometrics**, **29**, 1985.

Millard, R.; Emmerton, M. **Non-Technical Losses – How do other countries tackle the problem?** In: **Proceedings...** AMEU Convention. 2009.

Miller, A. **Subset Selection in Regression**, London: Chapman & Hall, 1990.

Mimmi, L. M., Ecer, S. An Econometric Study of illegal electricity connections in the urban favelas of Belo Horizonte, Brasil. **Energy Policy** **38**, pg. 5081 – 5097, 2010.

Mohassel, R. R., A. Fung, F. Mohammadi, e K. Raahemifar. “A survey on advanced metering infrastructure.” **Electrical Power and Energy Systems**, 2014: 63:473–484.

Myung, I. J., e M. A. Pitt. “Applying Occam’s razor in modeling cognition: A Bayesian approach.” **Psychonomic Bulletin & Review**, 1997: 79-95.

Myung, I. J., M. R. Forster, e M. W. Browne. “Special issue on model selection.” **Journal of Mathematical Psychology**, 2000.

northeast group, llc. “[http://www.northeast-group.com/.](http://www.northeast-group.com/)” **Northeast-group**. 2017. <http://www.northeast-group.com/reports/Brochure-Electricity%20Theft%20&%20Non-Technical%20Losses%20-%20Northeast%20Group.pdf> (acesso em 2020).

Penin, C. **Combate, prevenção e otimização das perdas comerciais de energia elétrica**. Tese de Doutorado, Escola Politecnica, Universidade São Paulo, 2008.

Ripley B. “**Selecting Amongst Large Classes of Models.**” Lecture, URL <http://www.stats.ox.ac.uk/~ripley/Nelder80.pdf>. 2003.

Shleifer, A. A Theory of yardstick competition. **Randon Jornal of Economics**, v.16, n.3, p. 319-327, 1985.

Silva, S. F. P, e Pessanha J. F. M. “Indicadores contábeis financeiros e a previsibilidade de insolvência das distribuidoras de energia elétrica” **LII Simpósio Brasileiro de Pesquisa Operacional**, 2020.

Simões, P.F.M.; Souza, R.C.; Calili, R.F.; Pessanha, J.F.M. Analysis and Short-term Predictions of non-Technical Loss of Electric Power Based on Mixed Effects Models. **Socio-Economic Planning Sciences**. v.71, 2020.

Smith, T.B. Electricity theft: a comparative analysis. **Energy Policy**, v.32, n.18, p. 2067-2076, 2004.

Sodeyama, K.; Igarashi, Y.; Nakayama. T.; Tateyama, Y. e Okada, M. Liquid electrolyte informatics using an exhaustive search with linear regression. **Phys. Chem. Chem. Phys.**, 2018, 20, 22585-22591

Szwarcwald, C. L.; Bastos, F. I.; Esteves, M. A. P.; Andrade, C. L. T; Paez, M. S.; Medici, E. V.; Derrico, M. Desigualdade de renda e situação de saúde: o caso do Rio de Janeiro. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 15, n. 1, p. 15-28, jan./mar. 1999.

Taddy, M. Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. McGraw Hill Professional, 2019

Venables W. N.; Ripley B. D. **Modern Applied Statistics with S-PLUS**. 3rd edition. Springer-Verlag, New York. 1997

Wang, K.Y., Tarr, G., Yang, J.Y. e Mueller, S. **Fast and approximate exhaustive variable selection for generalised linear models with APES**. Aust. N. Z. J. Stat., 61: 445-465. 2019.

Winther, T. Electricity theft as a relational issue: A comparative look at Zanzibar, Tanzania, and the Sunderban Islands, India. **Energy for Sustainable Development**, v.16, p. 111-119, 2012.

Wooldridge, J.M. (2010). **Econometric Analysis of Cross Section and Panel Data**, 2nd ed., MIT Press.

WORLD BANK. **Reducing Technical and Non-Technical Losses in the Power Sector. 2018**. Background Paper for the World Bank Group Energy Sector Strategy Disponível em: <  
<https://openknowledge.worldbank.org/bitstream/handle/10986/20786/926390WP0Box3800in0the0power0sector.pdf?sequence=1&isAllowed=y> >. Acesso em: 09 nov.18.

Yakubuab, O.; Babu, N. C.; Adjeib, O. Electricity theft: Analysis of the underlying contributory factors in Ghana. **Energy Policy**, v.123, n.9, p. 611-618, 2018.

Yurtseven, C. The causes of electricity theft: An econometric analysis of the case of Turkey. **Utilities Policy**, v.37, p. 70-78, 2015.