

3

Análise de Dados

O objetivo desse capítulo é descrever o banco de dados utilizado, bem como detalhar porque a preferência por certas variáveis e não outras nessa fase inicial de busca por variáveis importantes na explicação das perdas não técnicas. Deve-se ressaltar que para realizar o estudo foram conseguidos dados de 47 das 56 empresas distribuidoras de energia elétrica, referentes ao ano civil de 2008, listadas pela ANEEL na nota técnica nº 271 de setembro de 2010.

3.1.

Variáveis Explicativas

A escolha das variáveis que compõem o grupo de “potenciais variáveis explicativas para um modelo final” decorreu, em grande medida, do levantamento da literatura existente sobre o assunto. Contudo, reserva-se espaço também para variáveis que ainda não figuram entre os principais trabalhos sobre perdas não técnicas, mas que poderiam muito bem ser utilizadas num modelo como o que se pretende estimar segundo especialistas do setor.

Dentre os fatores que podem impactar nas perdas não técnicas e que são considerados exógenos ao controle das empresas (poderiam ser usados no modelo de Yardstick Competition) podem ser relacionados: nível social dos consumidores, ausência do poder público na área de concessão, características da área ocupada, qualidade de vida, questões comportamentais, entre outros.

Para a estimação inicial do modelo (modelo completo) foram construídas 13 variáveis explicativas mais a variável de interesse, correspondentes ao ano base de 2008⁷. As variáveis explicativas testadas são listadas na Tabela 1.

⁷ Ver critérios de extrapolação para variáveis que não possuem coleta em 2008 no item 3.3 deste capítulo.

Tabela 1 – Lista de Variáveis

Variável	Fonte	Ano Base
Percentual de Domicílios com até um Morador por Cômodo	CENSO E PNAD/IBGE	2008
Percentual de Domicílios com até Dois Moradores por Cômodo	CENSO E PNAD/IBGE	2008
Percentual de Domicílios com Três ou Mais Moradores por Cômodo	CENSO E PNAD/IBGE	2008
Óbitos por Agressão	ANEEL/NT 271	2008
Proporção da População que Ganha até Meio Salário Mínimo	CENSO E PNAD/IBGE	2008
Taxa de Desemprego	CENSO E PNAD/IBGE	2008
Percentual de Domicílios com Cobertura de Abastecimento de Água	ANEEL/NT 271	2008
Percentual de Domicílios com Cobertura de Rede de Esgoto	CENSO E PNAD/IBGE	2008
Proporção da População que Ganha até Três Salários Mínimos	ANEEL/NT 271	2008
Taxa de Inadimplência no Setor de Crédito	ANEEL/NT 271	2008
Percentual de Domicílios Precários	ANEEL/NT 271	2007
Total de Consumidores em Baixa Tensão	ANEEL	2008
Região Geopolítica	IBGE e ANEEL	2008

Como citado anteriormente, as variáveis selecionadas tentam captar fatores que fogem à gerência das empresas e que influenciam no nível de perdas não técnicas, de acordo com a teoria de Yardstick Competition. Os fatores que as variáveis buscam reproduzir serão discutidos a seguir.

3.2.

Fatores

3.2.1.

Características gerais da ocupação

É indubitável que localização, material usado para construção, condições sanitárias, densidade de moradores e tipo de ocupação domiciliar sejam *proxies* de fatores socioeconômicos associados à ocupação domiciliar dos consumidores de energia elétrica.

Esses fatores estão empiricamente associados às características da ocupação e espera-se que eles capturem efeitos intrínsecos a áreas com problemas sociais, esquecidas pelo Estado, com baixa qualidade de vida, muitas vezes comuns a áreas favelizadas, características estas que podem alavancar os percentuais de perdas não técnicas. Segundo Rafael de Oliveira Marques, coordenador de planejamento da Light S/A, “pessoas que vivem em áreas que reúnem as características acima cultivam o pensamento que a energia elétrica de graça serviria como uma compensação a exclusão social vivida por eles”⁸. Ou seja, essas características se traduziriam também em um aspecto comportamental atrelado às irregularidades no consumo de energia.

Na Figura 2 observa-se um poste de luz em uma favela da área metropolitana do Rio de Janeiro, é válido notar que este local provavelmente é de difícil acesso para equipes de combate à perdas e regularização.

⁸ Entrevistado em setembro de 2010.



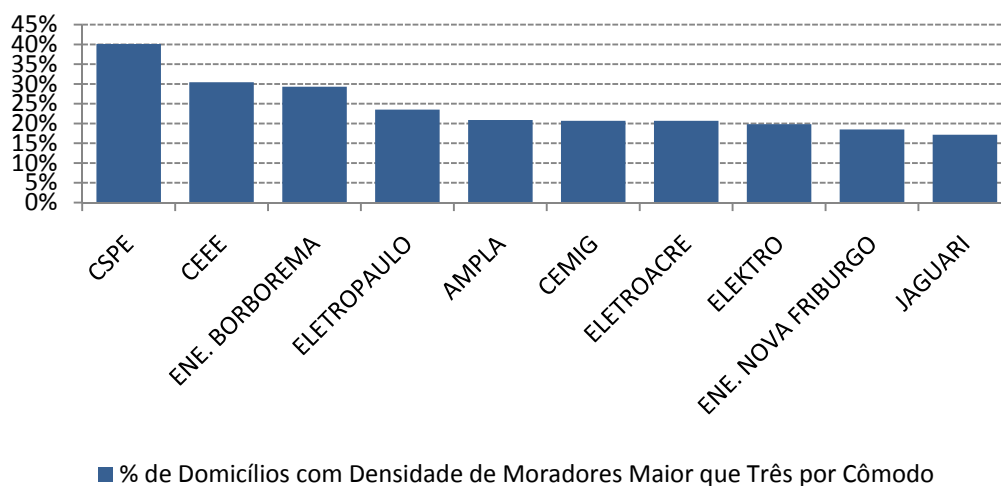
Fonte: Google Image

Figura 3 – Ligações Irregulares em uma Favela Carioca

Aparado na discussão acima, buscou-se traduzir o fator chamado de “características da ocupação” nas seguintes variáveis:

- i. Densidade de moradores por domicílio;
- ii. Percentual de domicílios classificados como precários;

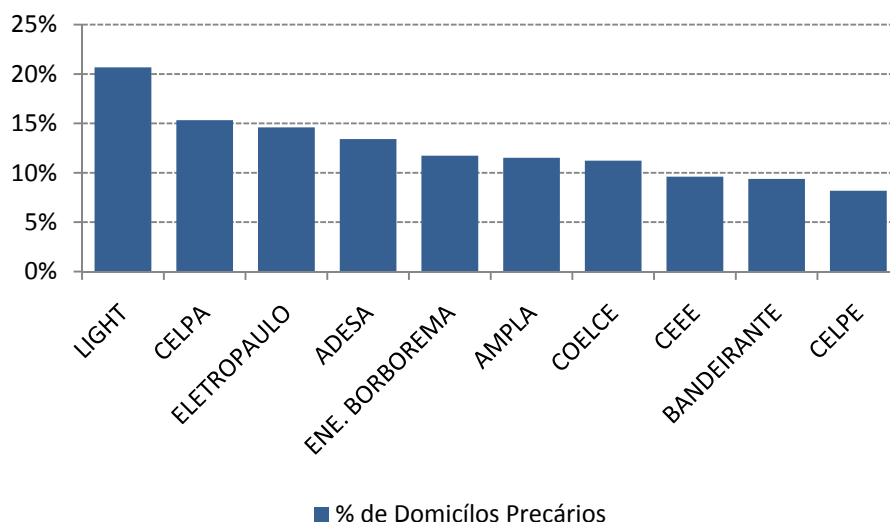
O gráfico da Figura 3 mostra que dentre as dez distribuidoras com maior densidade domiciliar – faixa de três ou mais moradores por cômodo estão as empresas: AMPLA, ELETROPAULO E CEMIG, que se localizam nos estados do Rio de Janeiro, São Paulo e Minas Gerais, respectivamente. Estes são os estados mais urbanizados do país e que notoriamente possuem muitas áreas com ocupação de favelas.



Fonte: IBGE

Figura 4 – Top 10 distribuidoras com maior percentual de domicílios com densidade de moradores maior do que três por cômodo.

O gráfico da Figura 4, a seguir, evidencia a grande quantidade de domicílios precários na área de concessão da Light, distribuidora carioca, além disso, encontram-se entre as distribuidoras com maior percentual de domicílios precários a AMPLA, também do Rio de Janeiro, e as paulistas Eletropaulo e Bandeirante. Novamente, percebe-se que Rio de Janeiro e São Paulo sofrem com problemas de ordem social. Não obstante, têm-se também distribuidoras do Norte e Nordeste como CELPA, ADESA, entre outras.



Fonte: ANEEL

Figura 5 – Top 10 distribuidoras com maior percentual de domicílios precários.

3.2.2.

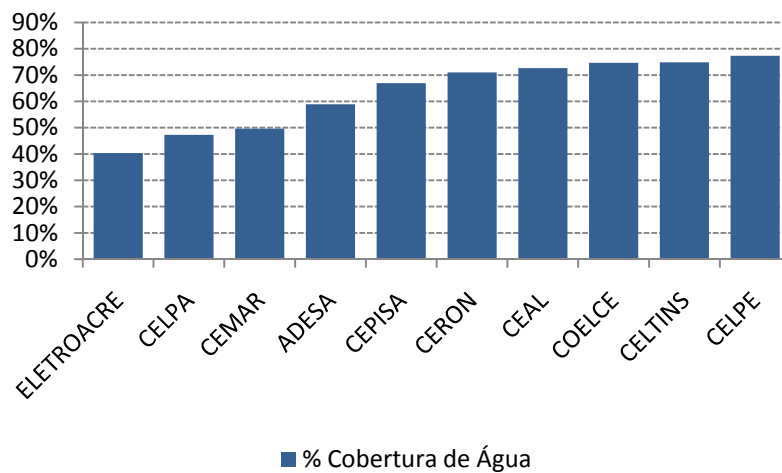
Acesso à infraestrutura

Outro fator relevante é o acesso aos serviços básicos de infraestrutura. Comunidades sem rede de abastecimento de água, rede canalizada de esgoto, coleta regular de lixo, e outros, tenderiam a apresentar maiores níveis de perdas não técnicas, de acordo com estudos anteriores. Estas variáveis refletem características das localidades ocupadas, ou seja, regiões com baixos percentuais de cobertura desses serviços podem ser entendidas como áreas marginalizadas, favelizadas e/ou de ocupação ilegal. Sendo assim, foram construídas e adicionadas ao estudo as seguintes variáveis:

- i. Percentual de domicílios com cobertura de abastecimento de água;
- ii. Percentual de domicílios com cobertura de rede de esgoto.

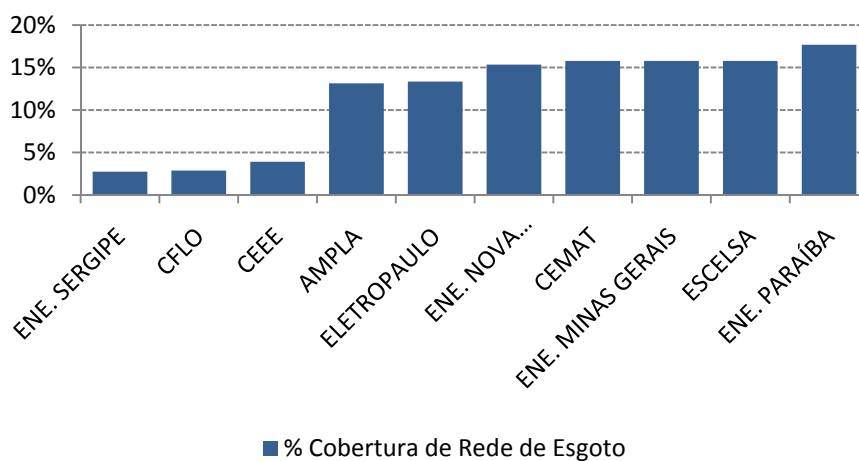
Na comparação das Figuras 5 e 6, percebe-se que existe uma variação mais abrupta com relação ao percentuais de cobertura de esgoto em cada distribuidora. Ainda comparando com os percentuais de abastecimento de água, conclui-se que este serviço parece ter maior abrangência do que a cobertura de esgoto. Isso pode

ser um indicio que a variável consiga traduzir melhor o efeito que se deseja captar, ou seja, problemas sociais associados à falta de infraestrutura dessas áreas.



Fonte: ANEEL

Figura 6 – Top 10 distribuidoras com menor percentual de cobertura de água encanada.



Fonte: ANEEL

Figura 7 – Top 10 distribuidoras com menor percentual de cobertura de rede de esgoto.

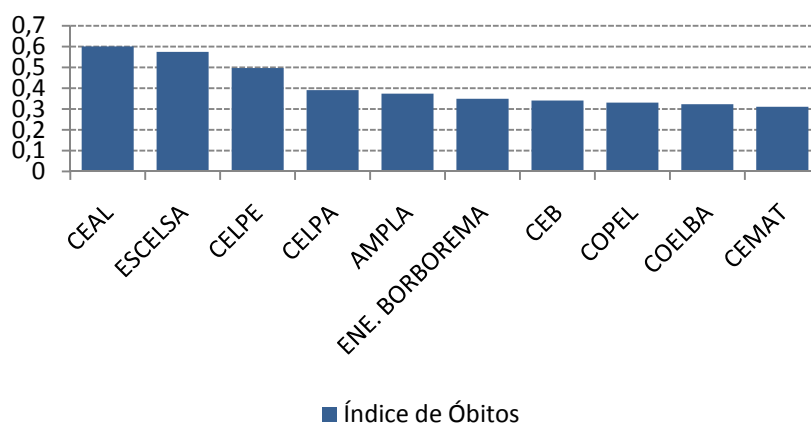
3.2.3.

Violência

Um dos fatores que se acredita determinante para a ocorrência de irregularidades no consumo é associado à violência. Áreas dominadas pelo chamado “poder paralelo” apresentam histórico de altíssimos níveis de perdas.

Segundo a linguagem popular, o “poder paralelo” é o poder exercido por criminosos dentro de comunidades pobres, o problema se dá na medida em que nesses lugares as equipes de regularização e manutenção não têm qualquer respaldo, ou segurança para trabalhar. O domínio dos criminosos dificulta a penetração de equipes que combatem as perdas, fazendo com que a quantidade de energia consumida de forma irregular continue crescendo. Estima-se que áreas de risco, onde não é possível fazer fiscalizações e regularizações, tipicamente apresentem perdas de mais de 90%. Sendo assim, a fim de captar efeitos relacionados a áreas de risco, optou-se pela variável “óbitos por agressão”.

Na análise do gráfico da Figura 7, destacam-se a CEAL, empresa de Alagoas, e a ESCELSA, do Espírito Santo, que ocupam os primeiros lugares no ranking de violência. Predominantemente, observam-se distribuidoras do Nordeste e Sudeste entre as empresas com maiores índices de óbitos por agressão em suas áreas de concessão.



Fonte: ANEEL

Figura 8 – Top 10 distribuidoras com maior índice de óbitos por agressão a cada cem mil habitantes.

Para se conseguir uma melhor idéia do problema vivido pelas distribuidoras, podem-se analisar dados divulgados pela Light S/A em 2009. Segundo a empresa, apesar de apenas 17% do total de seus consumidores habitarem regiões classificadas como áreas de risco, tais áreas seriam responsáveis por 39% do seu total de perdas não técnicas. Na Figura 8 pode-se observar uma matéria de jornal que exemplifica este problema vivido pela Light e por outras empresas.

EXPRESSO - RJ
Geral

Pág.: 4



03/09/2009

1/



► Policiais da Core ajudaram a resgatar os colegas e a equipe da Light. Carro da empresa levou quatro tiros

TÉCNICOS FORAM VERIFICAR DENÚNCIA E FICARAM NA MIRA DE BANDO

Tráfego encurrala equipe da Light

O que seria mais um dia comum para uma equipe de técnicos da Light acabou se transformando num pesadelo bem real. Com o apoio de policiais da Delegacia de Defesa dos Serviços Delegados, os técnicos foram verificar uma denúncia de que havia gatos de luz no Morro da Fé, na Vila da Penha, mas acaba-

ram atacados a tiros e encurralados por traficantes.

Equipes de outras delegacias, com o apoio do helicóptero da Polícia Civil, foram chamadas e conseguiram resgatar os funcionários da Light e os policiais.

O carro da Light foi atingido por quatro tiros, mas ninguém ficou ferido. Dois fun-

cionários da empresa ficaram em estado de choque.

Segundo a polícia, um bandido usando uma arma de mira telescópica estava apontando para os policiais. O tiroteio também assustou comerciantes e pedestres que passavam na Avenida Vicente de Carvalho, uma das mais movimentadas do bairro.

Fonte: Jornal Expresso

Figura 9 – Problemas em áreas de risco durante uma ação de combate as perdas não técnicas.

3.2.4.

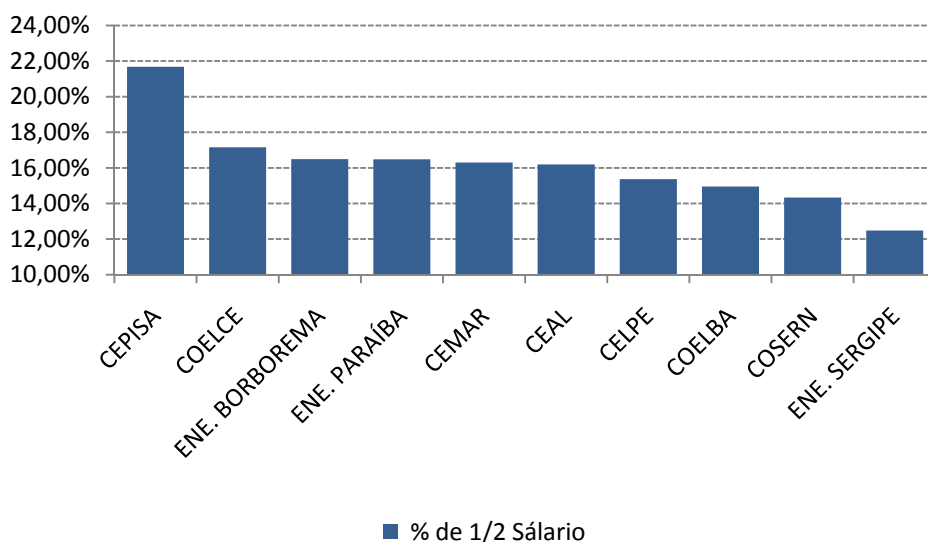
Renda

Outro fator que se busca captar é a disponibilidade de renda. Pode-se enxergar o problema sob duas óticas distintas.

A primeira seria o problema da falta de renda suficiente para manter a adimplência nas contas de luz, mesmo com baixo consumo. Nesse caso, o valor da conta de luz seria responsável por uma fatia muito representativa no orçamento familiar. Sendo assim, para continuar a ter acesso à eletricidade, opta-se por realizar uma ligação irregular, um “gato”. De acordo com essa ótica, a renda poderia explicar o nível de perdas não técnicas em situações associadas à pobreza quase que extrema. Assim, foram coletadas as variáveis abaixo:

- i. Taxa de desemprego;
- ii. Proporção de chefes do domicílio que ganham até meio salário mínimo;

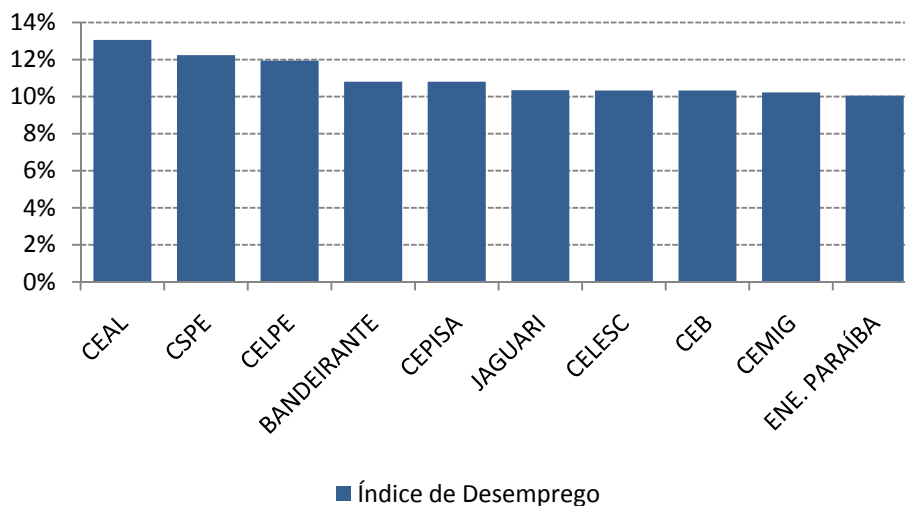
Diante do gráfico da Figura 9, observa-se uma variação bastante acentuada entre as informações da distribuidora Cepisa, do Piauí, para as demais. É interessante notar que as dez distribuidoras com maior percentual de chefes do domicílio ganhando até meio salário são da região Norte ou Nordeste, o que aponta o baixo nível de renda e provavelmente o alto peso da conta de luz no orçamento das famílias dessas regiões.



Fonte: IBGE

Figura 10 – Top 10 distribuidoras com maior percentual de chefe do domicílio com recebem até ½ salário.

O gráfico da figura abaixo aponta que, pelo menos no topo das distribuidoras com maior índice de desemprego, as taxas são quase constantes, ficando entre 12% e 10%.



Fonte: IPEA

Figura 11 – Top 10 distribuidoras com maior índice de desemprego

A segunda ótica pela qual se pode ter associação entre renda e perdas não técnicas se dá ao oposto da primeira. Agora seria exatamente a disponibilidade de renda que poderia aumentar o nível de perdas. Mas como isso faria sentido?

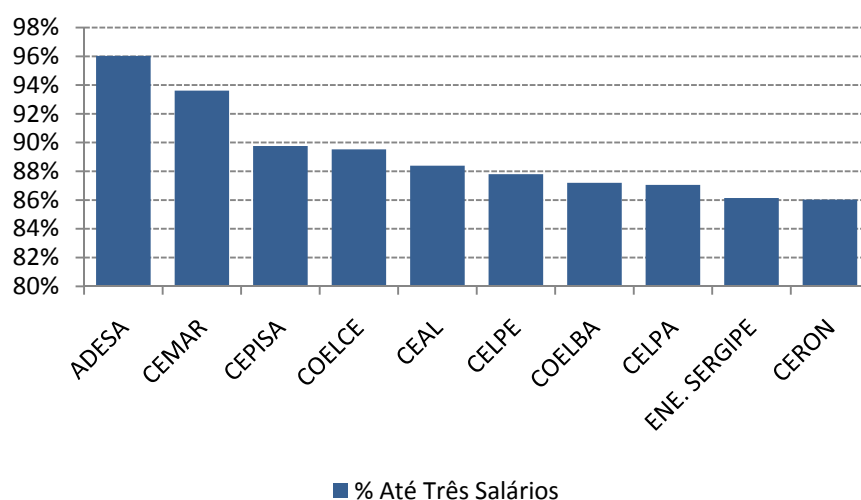
Segundo Yaccoub (2010), fatores comportamentais fazem com que os consumidores julguem ser mais importante saldar dívidas com crediários e outros, a pagar a conta de energia elétrica. Como uma consequência um tanto extrema, porém não incomum, segundo a autora, seria o consumidor que fica inadimplente junto à distribuidora fazer um “gato”, a fim de manter seu consumo de energia elétrica. Por isso, foram escolhidas duas variáveis para representar o fator renda.

- iii. Proporção de chefes de domicílios que ganham até três salários mínimos;
- iv. Taxa de inadimplência no setor de crédito.

A primeira, “proporção de chefes do domicílio que ganham até três salários mínimos”, é usada para mensurar a parcela de consumidores com renda suficiente para iniciar comprar a prazo, e que além do maior comprometimento do salário com o início do crédito teriam um aumento expressivo na conta de luz

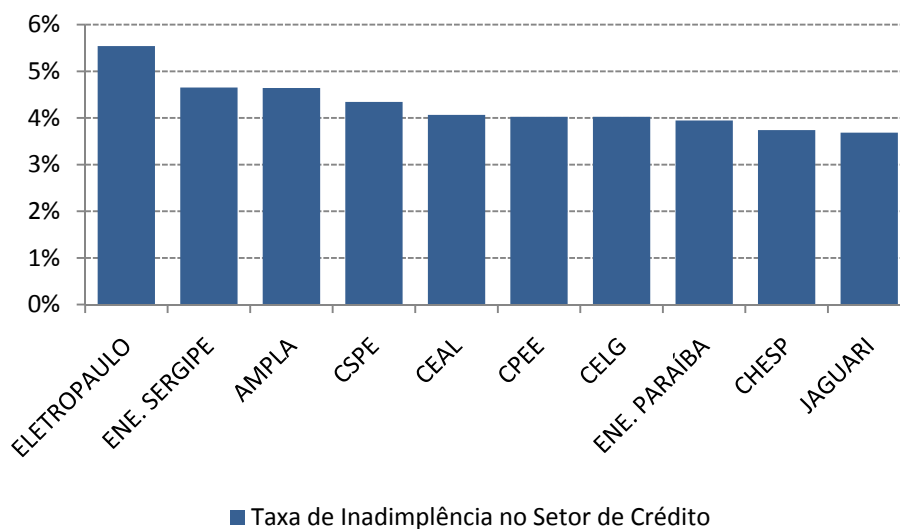
dependendo da compra que fizessem – um aparelho de ar condicionado, por exemplo. Outra variável selecionada é “taxa de inadimplência no setor de crédito”, uma vez que segundo a ordem de prioridades de Yaccoub (2010), espera-se que no caso de haver inadimplência no setor de crédito, a conta de luz já não foi paga e para continuar com o consumo normal essas famílias poderiam optar por fazer uma ligação irregular.

As Figuras 11 e 12 exemplificam as dez distribuidoras com maiores percentuais de pessoas que ganham até três salários mínimos e com maior taxa de inadimplência respectivamente.



Fonte: ANEEL

Figura 12 – Top 10 distribuidoras com maior percentual de chefes de domicílio que ganham até três salários.



Fonte: ANEEL

Figura 13 – Top 10 distribuidoras com maior taxa de inadimplência no setor de crédito.

A Figura 13, apresentada a seguir, trás uma matéria do jornal O GLOBO mostrando a grande demanda por aparelhos de ar condicionado. Como já citado anteriormente, além de pesar no orçamento, a aquisição de eletrodomésticos, invariavelmente, leva ao aumento do consumo de energia e da conta de luz.

PAÍS | RIO | CIDADES | ECONOMIA | MUNDO | CIÊNCIA | ESPORTES | CULTUR

Publicada em 09/11/2009 às 23h48m [Olá, francis](#)

[RIO 40 GRAUS](#)

Com calor, vendas de ar-condicionado sobem até 40% e redes varejistas ficam sem estoque

O Globo

☆☆☆☆ DÊ SEU VOTO | ☆☆☆☆ MÉDIA: 4,7 | [Comentários](#)

RIO e SÃO PAULO - O forte calor que se abateu sobre o Rio de Janeiro nos últimos dias fez os aparelhos de ar-condicionado e ventiladores sumirem das prateleiras das principais redes de varejo da cidade. É o que mostra a matéria de Bruno Rosa, Fabiana Ribeiro e Aguinaldo Novo, publicada nesta terça-feira no GLOBO. Em algumas lojas, as vendas desses produtos tiveram uma alta de até 40% em relação às semanas anteriores.

Ponto Frio, Ricardo Eletro, Casa&Video, Ambient Air, Tele Rio e Fast Shop já tiveram de pedir reposição de estoque para conseguir atender a crescente demanda. De acordo com estimativas, oito entre dez clientes foram à busca dos itens para enfrentar temperatura próxima à 40. Em lojas menores, a procura aumentou em até dez vezes.

Fonte: Jornal O GLOBO

Figura 14 – Demanda Acentuada por Eletrodomésticos

3.2.5.

Porte da Distribuidora

Entende-se que uma distribuidora de maior porte possivelmente teria que desprender maiores esforços para combater as perdas dado o tamanho de seu mercado e sua complexidade de gestão. A ANEEL reconhece que o universo de clientes que a distribuidora tem de atender pode impactar na velocidade de redução das perdas, mas não o testou no seu modelo de painéis. Sendo assim, propõem-se o teste da variável “número de consumidores na baixa tensão” a fim de capturar a dificuldade de gestão associada à dimensão do mercado que a empresa atende.

Região Geopolítica

Foi construída uma variável dummy, chamada também de binária, para diferenciar as regiões norte e nordeste das demais regiões do Brasil. O norte e nordeste são as regiões mais pobres do país e até sob o pretexto dos resultados de algumas variáveis socioeconômicas agrupadas por distribuidoras dessas duas regiões estarem entre os piores resultados observados, imagina-se que talvez, o fato das distribuidoras estarem alocadas em uma dessas duas regiões, por si só, ajude a explicar seus níveis de perdas.

3.3.

Construção das Variáveis

Para construir as variáveis explicativas em nível de cada distribuidora, é necessário agregar dados municipais de acordo com a área de concessão correspondente. O problema é que a fonte da maior parte dos dados com estratificação municipal é o CENSO e seus últimos resultados correspondem ao CENSO 2000 – realizado no ano 2000. Logo, há duas opções para trabalhar com os dados: estimar o modelo utilizando as variáveis no ano base 2000 ou adotar algum critério de extrapolação para os dados. O que se faz neste trabalho é reproduzir o fator de crescimento sugerido pela ANEEL nos seus estudos de perdas não técnicas.

Nas notas técnicas de perdas não técnicas, a agência sugere o seguinte procedimento: primeiro, são coletadas as variáveis municipais no CENSO 2000, em seguida elas são agregadas por área de concessão, de acordo com os municípios que cada concessão compreende. Em um segundo passo, no caso das mesmas variáveis também serem coletadas na PNAD⁹, calcula-se sua variação anual em nível estadual (até o ano de PNAD mais recente¹⁰), e a partir da variação estadual anual, extrapola-se o crescimento das variáveis construídas a partir do CENSO 2000, levando em consideração o estado em que a área de concessão se

⁹ Pesquisa Nacional de Amostragem Domiciliar. A coleta de informações é anual, mas suas estratificações são Federação, Unidades da Federação e Regiões Metropolitanas. Sendo assim, não trazem informações em nível municipal.

¹⁰ Até a data da elaboração deste trabalho, a PNAD mais recente é a do ano de 2008.

localiza predominantemente, ou seja, duas distribuidoras de um mesmo estado teriam a mesma taxa de crescimento empregada nas variáveis construídas sob essa premissa.

Para as variáveis que não são pesquisadas na PNAD, são usados os valores do CENSO 2000 sem atualização, e no caso da variável não ser coletada pelo CENSO e não estar disponível em nível municipal em qualquer outra pesquisa, a variável em questão é construída sob o menor nível de agregação geopolítico possível, por ordem estadual e depois federal.

3.4.

Variável de Interesse

Esta sessão tem por objetivo analisar a variável de interesse. No modelo, usa-se como variável de interesse a **“proporção de perdas não técnicas de acordo com a energia injetada no mercado de baixa tensão”**, que se denomina aqui apenas de **“perdas não técnicas”** como nomenclatura resumo. A variável é construída tomando-se como base somente os mercados de baixa tensão das distribuidoras, o que se deve a uma questão de padronização, já que, segundo a ANEEL, perdas não técnicas na alta tensão são naturalmente pequenas, logo para que uma distribuidora que atenda muitos clientes de alta não tenha seu verdadeiro percentual de perdas não técnicas subestimado, utiliza-se a energia injetada na baixa tensão para balizar a construção do percentual de perdas não técnicas. A fonte dos dados de perdas foi a nota técnica nº271 de 2010, e ainda segundo a ANEEL, o cálculo do montante de perdas não técnicas foi feito pelas próprias distribuidoras seguindo essa metodologia padronizada.

É importante notar que das 54 distribuidoras com dados de perdas não técnicas divulgadas pela ANEEL foram utilizadas no estudo apenas 47. Isso ocorreu por falta de dados disponíveis para a construção das variáveis explicativas e mudanças como fusões, além de outros acontecimentos do gênero que dificultaram a replicação dos dados para todas as distribuidoras.

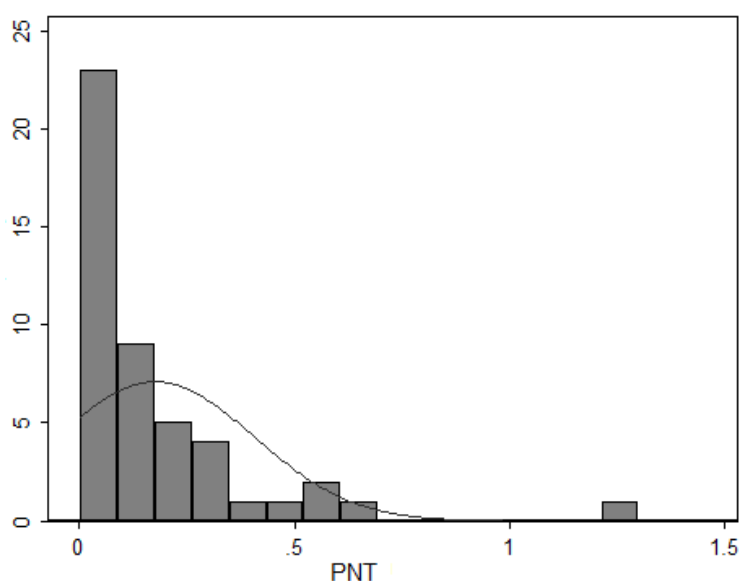
Análise Exploratória

A fim de conhecer melhor a variável de interesse, propõe-se uma análise exploratória dos seus dados.

Tabela 2 – Estatísticas Descritivas

Média	Mediana	Moda	Desvio Padrão	C.V. ¹¹
0,18	0,09	0,05	0,23	1,30

A distância entre as estatísticas de posição (média, moda e mediana), dá uma idéia que a distribuição não é simétrica. Para conhecer melhor a distribuição dos dados é feito um histograma - gráfico de frequências acumuladas.



.Figura 154 – Histograma da variável de interesse - % de Perdas Não Técnicas

Pelo formato do histograma, pode-se perceber que a variável de interesse não segue a distribuição normal, já que seu formato é assimétrico à esquerda. Para confirmar que a distribuição não é normal, foi feito um teste Jarque-Bera¹² e, segundo o p-valor do teste (resultado igual a 0,000), confirma-se a hipótese de não

¹¹ Coeficiente de variação.

¹² Teste que se baseia na curtose e assimetria da função de distribuição.

normalidade para um nível de confiança de 1%. Isso leva a um desdobramento bastante interessante: se os dados não seguem a distribuição normal, qual seria a sua verdadeira distribuição? Isso teria algum impacto na elaboração de um modelo para estimação de perdas não técnicas?

Para responder essas questões imagine um caso hipotético onde a distribuição da variável “perdas não técnicas” é normalmente distribuída. Nesse caso, a distribuição seria simétrica em formato de sino, a média seria bastante semelhante à moda e à mediana dos dados. Entretanto, fica difícil imaginar que a maioria das empresas do setor elétrico tem níveis de perdas não técnicas, concentrados numa certa média e com alguns valores extremos de baixíssimos percentuais de perdas, bem como altíssimos percentuais mais ou menos equidistantes.

Uma configuração hipotética seria um mercado com média de 25% de perdas e valores caudais (valores extremos), tendendo à 0%, no caso da cauda da esquerda e à 70% no caso da cauda da direita. Seria difícil pensar que as empresas sobreviveriam nesse ambiente, dada a magnitude do nível de perdas – média de 25%.

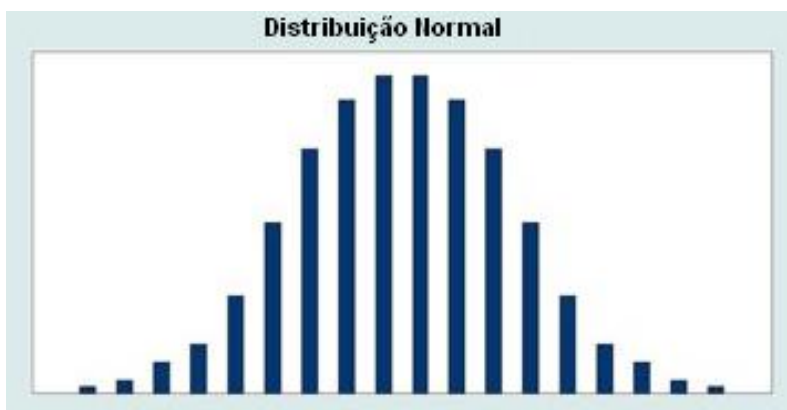


Figura 16 – Caso 1 Distribuição Simétrica

Em um hipotético caso 2, trabalhamos com um modelo de mercado onde a maioria das empresas tem perdas relativamente pequenas, e algumas outras empresas com níveis bastante altos de perdas não técnicas (estas possivelmente teriam problemas de gestão aliados a atuação numa área de concessão com alta complexidade socioeconômica). Esse modelo de mercado é mais semelhante à

realidade brasileira e a distribuição das perdas seria algo semelhante à Figura 14, o que pode ser representado hipoteticamente por uma curva como na Figura 16.



Figura 17 – Caso 2 Distribuição Assimétrica Positiva.

Na literatura econométrica, quando se trata de modelagem de dados que não seguem a distribuição normal, amplamente recorre-se à abordagem linear generalizada, também conhecidos como modelos lineares generalizados (MLG's). Os modelos lineares generalizados são uma classe de modelos, que incluem também o modelo linear gaussiano, e conseguem modelar, com maior acurácia, dados de diversas distribuições. O presente trabalho seguirá essa linha de pesquisa e o detalhamento da abordagem linear generalizada será feito a seguir.