

UTILIZAÇÃO DA FORMULAÇÃO LINEAR GENERALIZADA PARA ESTIMAR PERDAS NÃO TÉCNICAS EM UMA DISTRIBUIDORA DE ENERGIA ELÉTRICA

Francis de Asevedo

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil - 22451-900
francis.asevedo@light.com.br

Pedro Guilherme Ferreira

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil - 22451-900
ferreira@ele.puc-rio.br

RESUMO

Hoje, no Brasil, 8% de toda energia produzida é perdida em fraudes e consumos clandestinos (Perdas-Não-Técnicas). Neste trabalho procura-se desenvolver um modelo estatístico para a estimação dos níveis de perdas-não-técnicas de energia elétrica ao nível de alimentadores da área de concessão de uma distribuidora do estado da Paraíba. Além da previsão, tem-se por objetivo avaliar quais variáveis são importantes nessa análise (efeitos fixos, interações e termos quadráticos). Variáveis associadas a particularidades do estado da Paraíba serão testadas, como presença de irrigação, entre outras. Interações e termos quadráticos também serão usados no desenvolvimento do estudo. Dado a distribuição da variável de interesse, tal análise seguirá o arcabouço dos Modelos Lineares Generalizados. O estudo elaborado conseguiu estimar um bom modelo para a estimação de PNTs. O modelo trouxe informações de quais variáveis são importantes para explicar as PNTs, inclusive interações e termos quadráticos, ou seja, variáveis que não são encontradas em literaturas prévias.

PALAVRAS CHAVES: Perdas-Não-Técnicas, Alimentadores, Modelos Lineares Generalizados.

ABSTRACT:

Currently in Brazil, 8% of all electric energy produced is gone due to frauds and clandestine consume (Non-technical losses). This paper develops a statistical model to estimate the non-technical losses of electric electricity target according to distributor machines of certain area in the State of Paraíba. In addition to the predictive power, the objective of this work is to evaluate which variables are important in this context (fixed effects, interactions and squared terms). Variables associated to peculiarities of the State of Paraíba will be tested, like the existence of irrigation, among others. Interactions and squared terms are also going to be part of the analysis as well. Due to the distribution of the dependent variable, the modeling is going to follow the Generalized Linear Models techniques. The research conducted was able to estimate a good model to PNT's estimation. The model enables us to know what variables are important for the explanation of PNTs, including interactions and quadratic terms, that is, variables that are not found in previous literatures.

KEY WORDS: Non-technical losses, distributor machines, Generalized Linear Models.

1. INTRODUÇÃO

A reestruturação do setor elétrico brasileiro (SEB) iniciou-se em meados da década de noventa, concebida com base no modelo do setor elétrico do Reino Unido, um modelo orientado para o mercado e caracterizado pela privatização das empresas e pela desverticalização da indústria, onde a geração e a comercialização de eletricidade são organizadas em um ambiente competitivo, enquanto a transmissão e a distribuição permanecem como monopólios regulados (PESSANHA et. al., 2007).

Tal transformação fez com que as empresas ligadas ao setor começassem a se preocupar com toda a cadeia que envolve o mesmo, inclusive com as perdas de energia. Isto é, segundo a ABRADDEE (Associação Brasileira das Distribuidoras de Energia Elétrica), um aspecto importante na regulação do setor elétrico diz respeito à consideração de perdas de energia no cálculo tarifário. A fim de dar suporte aos processos de reajuste e revisão tarifária a ANEEL procura estabelecer limites regulatórios para as perdas de energia, em conformidade com os princípios de modicidade e de redução de custos e desperdício de energia elétrica que regem o regime de regulação.

De acordo com estimativas da ANEEL, os custos com Perdas-Não-Técnicas de energia elétrica são de aproximadamente de R\$ 5 bilhões anuais. Sendo assim foi lançada em novembro de 2008 a Nota Técnica nº 342, onde a Agência Reguladora elaborou formalmente uma proposta contendo metodologia e critérios a serem adotados no cálculo estimado da meta de Perdas-Não-Técnicas para cada concessionária no Brasil. A partir das estimativas de perdas máximas geradas por este modelo é que seria, então, determinado o patamar de PNTs média aceitável – o que figura entre os diferentes aspectos considerados no processo de reajuste tarifário.

Este fato, indubitavelmente, é uma motivação bastante forte para o interesse da sociedade como um todo no desenvolvimento de um bom modelo para a estimação da meta de PNTs.

O objetivo deste artigo será descobrir quais variáveis são determinantes na estimação de Perdas-Não-Técnicas (PNTs) ao nível de alimentadores na área de concessão de uma distribuidora de energia do estado da Paraíba e verificar com qual intensidade as mesmas atuam sobre a variável de interesse, PNTs.

Estudos empíricos apontam relações entre PNTs e o ambiente socioeconômico e institucional. Por isso, um passo fundamental do trabalho proposto será elucidar correlações entre variáveis socioeconômicas e os níveis de PNTs para então identificar as variáveis potencialmente relevantes para inclusão no modelo explicativo das perdas.

2. REFERENCIAL

Encontram-se diversos estudos nacionais e internacionais relatando que o nível de perdas comerciais (que corresponde, em geral, à maior parcela das Perdas-Não-Técnicas) pode estar associado a diversas variáveis socioeconômicas.

Até recentemente, havia pouca literatura sobre Perdas-Não-Técnicas de energia elétrica, havendo apenas estudos relativos a perdas técnicas e suas formas de mensuração. Um dos estudos mais abrangentes sobre o tema se encontra em Smith (2004). Neste artigo, o autor faz uma extensa discussão sobre o assunto, buscando definir as várias fontes de Perdas-Não-Técnicas existentes, realizando uma comparação internacional das formas de como o problema é tratado em diversos países e desenvolve uma interessante discussão a respeito das origens do problema.

Um trabalho recente sobre Perdas-Não-Técnicas de energia elétrica no Brasil é Araújo (2007). Nele o autor defende que as perdas podem ser explicadas por uma série de variáveis

socioeconômicas e de infra-estrutura. Entre as variáveis que o autor afirma estarem associadas às Perdas-Não-Técnicas de energia elétrica estão: educação, renda, desigualdade, localização geoeconômica da área de concessão e tarifa.

Com o objetivo de justificar seu nível de perdas atual, outras empresas têm elaborado estudos associando variáveis socioeconômicas ao problema. Por exemplo, a ESCELSA em trabalho encomendado à Aequus Consultoria, afirma que os indicadores sócio-econômicos do estado do Espírito Santo comprovam a tese de que existe uma relação entre a maior incidência de fraudes e furtos de energia elétrica e as regiões de elevada vulnerabilidade social. Os indicadores utilizados pela empresa são: emprego, massa salarial, PIB, violência (percentual de obtidos por agressões por 10 mil habitantes), e capacidade financeira dos governos locais.

3. BASE DE DADOS E METODOLOGIA

3.1. Base de Dados e Análise exploratória das variáveis

A fim de proporcionalizar as PNTs de acordo com o montante de energia que injetada no mercado de baixa tensão por cada alimentador, entende-se que a variável dependente a ser estimada será a razão entre Perdas-Não-Técnicas e Energia Injetada no Mercado de Baixa Tensão para cada Alimentador. Os dados foram coletados em 156 alimentadores distribuídos conforme a densidade de consumo da área de concessão. Estudos anteriores já mencionados no início do trabalho serviram de guia para a seleção de possíveis variáveis explicativas para este trabalho¹.

A representação de cada variável na base se dá de forma auto-explicativa, ou seja, a variável – Proporção de Domicílios Atendidos pela Rede de Distribuição de Água – será chamada de “*agua*”. Vale lembrar que os termos quadráticos serão representados com o mesmo código seguido pelo número 2, por exemplo, o termo quadrático da variável água é representado por “*agua2*”, e as interações serão representadas com os códigos das variáveis que interagem ligados por um “*underline*”.

É importante ressaltar que a busca de dados ao nível de alimentadores não é uma tarefa fácil, sendo que o CENSO 2000 do IBGE é a melhor fonte de coleta de variáveis ao nível municipal, entretanto como o modelo é desagregado ao nível de alimentadores e, uma vez que, no geral, um alimentador pode cobrir mais de um município, e/ou mais de um alimentador pode cobrir parte de um município torna-se necessário uma ponderação por número de unidades consumidoras atendidas pelo determinado alimentador.

Isto é, se um alimentador W passar por partes dos municípios A e B a variável PIB seria construída sendo a média do PIB em cada município ponderada pelo número de unidades consumidoras (UCs) atendidas pelo alimentador W, como mostra o exemplo:

$$PIB_{\text{alim.W}} = \frac{PIB_A * NC_A + PIB_B * NC_B}{NC_A + NC_B}$$

onde NC_i é o número de UCs da área i .

Ainda, é importante conhecer a distribuição de cada variável contemplada no estudo, principalmente a distribuição da variável resposta, afinal, empiricamente tem-se vestígios de que “PNTs/Energia Injetada em BT” não é normalmente distribuída e acrescenta um bom fator motivador deste trabalho. Para termos práticos a análise de um histograma já é uma boa pista da distribuição que cada variável segue.

Com a análise da figura 1, fica claro que a variável resposta não segue uma distribuição gaussiana. Pressupõe-se que essa variável pode obedecer, entre outras, a distribuição Gama ou Inversa Normal.

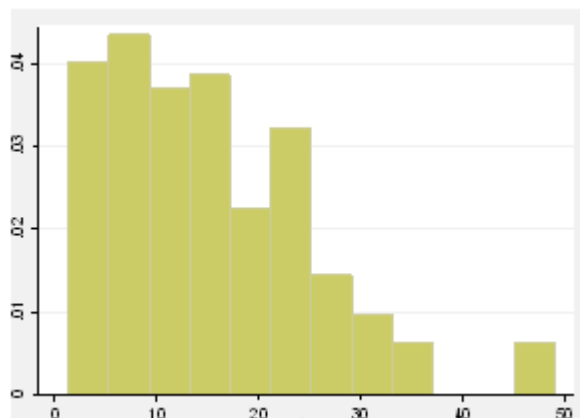


Figura 1 – Histograma de PNT/energia injetada em BT

A fim de procurar evidências de relações lineares entre as variáveis explicativas e a variável dependente plotou-se o diagrama de dispersão entre a variável dependente e cada uma das variáveis independentes candidatas. Para uma análise completa, deve-se obter a matriz de correlação simples entre as variáveis (y e x's e entre os x's).

Tabela 1– Matriz de correlação simples entre as variáveis

	pnt_ei	ar	geladeira	dom_adequados	agua	esgoto	analfabetismo	idh	violencia	renda	proprios	pea	inadimplencia	irrigante
pnt_ei	1,00													
ar	-0,10	1,00												
geladeira	-0,09	0,91	1,00											
dom_adequados	-0,19	0,84	0,83	1,00										
agua	-0,16	0,70	0,79	0,76	1,00									
esgoto	-0,21	0,46	0,47	0,61	0,49	1,00								
analfabetismo	0,16	-0,79	-0,80	-0,75	-0,66	-0,33	1,00							
idh	-0,16	0,94	0,92	0,81	0,68	0,43	-0,84	1,00						
violencia	-0,05	0,79	0,75	0,62	0,50	0,34	-0,63	0,82	1,00					
renda	-0,25	0,68	0,64	0,60	0,54	0,47	-0,41	0,66	0,53	1,00				
proprios	-0,33	-0,21	-0,19	-0,14	-0,09	-0,37	0,01	-0,17	-0,12	-0,24	1,00			
pea	-0,17	0,69	0,53	0,60	0,46	0,34	-0,56	0,64	0,58	0,47	0,00	1,00		
inadimplencia	0,31	-0,32	-0,30	-0,45	-0,39	-0,46	0,31	-0,30	-0,16	-0,40	0,03	-0,28	1,00	
irrigante	0,35	-0,39	-0,32	-0,46	-0,30	-0,26	0,36	-0,41	-0,39	-0,31	-0,17	-0,37	0,28	1,00

Fonte: elaboração própria dos autores

Analisando a tabela 1 observam-se variáveis explicativas correlacionadas, ou seja, possível presença de multicolinearidade. Um segundo efeito interessante é a captura do efeito de uma variável por outra altamente correlacionada com a primeira.

As correlações, em módulo, maiores do que 0,8 foram marcadas de vermelho. Percebe-se que o IDH está altamente correlacionado com as variáveis chamadas de variáveis de infraestrutura como cobertura da rede de água, esgoto e adequação da moradia. O IDH também é fortemente correlacionado com analfabetismo e violência, conforme esperado.

De acordo com as análises supracitadas é possível especular que, variáveis como analfabetismo, inadimplência e irrigantes teriam sinais positivos na equação estimada de regressão. Variáveis que indicam bons índices de desenvolvimento como cobertura de água, esgoto e domicílios adequados possivelmente entrariam com sinal negativo.

3.2. Metodologia

Nesta seção são brevemente expostos técnicas e arcabouços pertinentes aos modelos lineares generalizados (MLGs). Mesmo que a variável resposta seguisse distribuição normal (como demonstrado na seção anterior, não é normal) também se pode tratar a modelagem dentro dos procedimentos de MLG, uma vez que o modelo gaussiano é apenas um caso particular dos MLGs.

Deve-se verificar a significância estatística de interações e termos de ordem quadrática ao modelo de estudo, este é um assunto relevante no sentido que não há embasamento na literatura que aponte a influencia das interações no nível de PNT sendo assim interesse que se façam testes sobre a significância estatística de interações e termos quadráticos. São propostos testes e análises gráficas que serviram para corroborar a escolha ou não do modelo.

Distribuições que podem ser escritas segundo a equação 1, descrita abaixo pertencem a chamada família exponencial uniparamétrica, sendo assim, é possível sua modelagem seguindo os arcabouços dos MLGsⁱⁱ.

$$f(y; \theta, \phi) = \exp\{\phi^{-1}[y\theta - b(\theta)] + c(y, \phi)\} \quad (1)$$

A média e a variância da variável aleatória (v.a.) Y cuja distribuição pertence à família exponencial, na forma canônica são dadas pela equação 2.

$$E(Y) = \mu = b'(\theta) \quad e \quad \text{Var}(Y) = \phi b''(\theta) \quad (2)$$

Onde $b(\theta)$ é uma função conhecida da média, e ϕ é o parâmetro de dispersão da v.a.

O modelo linear generalizado é definido a partir das seguintes componentes: (a) **Componente Aleatória** - seja y um vetor de observações de tamanho n, denotando uma realização de um vetor de variáveis aleatórias Y, independentemente distribuídas com média μ . A distribuição de Y pode ser qualquer uma da família exponencial; (b) **Preditor Linear** - a parte sistemática do modelo continua sendo $\eta = X\beta$, onde $\eta_i = X_i\beta$ é a componente relativa à i-ésima observação; e, (c) **Função de Ligação (g)** - através desta função, faz-se a conexão entre a média das observações e a parte sistemática, da seguinte forma: $\mu_i = g^{-1}(\eta_i)$

A partir de testes e procedimentos propostos no trabalho, a escolha das componentes será feita a seguir, sem necessariamente seguir a ordem supracitada. No que diz respeito ao item a, deve-se dizer que a distribuição escolhida da família exponencial depende dos dados da variável resposta em questão, que podem ser discretos, contínuos, assimétricos ou proporções. Na tabela abaixo é apresentada a associação usual entre a distribuição e o tipo de dado estudado:

Tabela 1 – Algumas distribuições e tipos de dados

Distribuição	Tipos de Dados
Poisson	Contagens
Binomial	Proporções
Normal	Contínuos
Gama	Contínuos Positivos
Normal Inversa	Contínuos Positivos

A variável resposta é contínua e positiva. A análise do histograma da mesma mostra que as distribuições gama e normal inversa (inorm) são candidatas a distribuição do modelo.

Os coeficientes do modelo são estimados por um processo de otimização definido pela equação 3, Demétrio e Cordeiro (2008).

$$\beta^{(m+1)} = (X'W^{(m)}X)^{-1} X'W^{(m)}Z^{(m)} \quad (3)$$

Os coeficientes são obtidos pelo método dos quadrados mínimos ponderados, via um processo numérico que usa o algoritmo iterativo de Newton-Raphson. O método usual para iniciar o processo iterativo é especificar uma estimativa inicial e sucessivamente alterá-la até que a convergência seja obtida.

4. MODELAGEM E ANÁLISE DOS RESULTADOS

Foram elaborados dois modelos, sendo o primeiro modelo gama, utilizando função link log e o segundo modelo normal inversa usando também a função link log (o motivo da log será detalhado a seguir).

O critério de entrada e saída das variáveis no modelo de estudo seguiu a metodologia stepwise. O p-valor de entrada adotado foi de 0,12 e o de saída foi igual a 0,08. Neste processo foram testadas todas as variáveis, incluindo os termos quadráticos e as interações de até segunda ordem de todas as variáveis.

Como proposto no trabalho, ajustou-se um modelo com aproximadamente 90% das observações e seus resultados foram então estendido a todas as observações. Ou seja, uma análise “out of sample” foi desenvolvida. Assim o modelo faz a estimação em 141 observações (aproximadamente 90% dos dados) e a partir de então volta a fazer suas inferências na base toda.

A fim de escolher o melhor modelo, adotaram-se alguns procedimentosⁱⁱⁱ, sendo eles: (i) Análise da Deviance; (ii) Análise dos critérios de informação AIC e BIC; (iii) Análise dos resíduos; (iv) Interpretação dos parâmetros. Seguem a seguir os resultados das estimativas do modelo.

Tabela 3 - Modelo Gama link log

pnt_ei	coef.	z	$\rho > z $
Próprios	-0,3542	-5,51	0,000
Renda	-0,0009	-3,86	0,000
Esgoto	-0,0058	-2,83	0,005
Geladeira	-0,0111	3,02	0,003
Irrigante	0,2847	2,43	0,000
_cons	4,9393	9,09	0,000
Estatísticas			
AIC	7,3612	BIC	-612,5132
Log likelihood	-512,9659		

Fonte: elaboração própria dos autores

Tabela 4 - Modelo Normal Inversa link log

pnt_ei	coef.	z	$\rho > z $
Próprios	-0,3699	-5,23	0,000
Renda	-0,0008	-3,16	0,002
Esgoto	-0,0060	-2,59	0,009
Geladeira	-0,0127	3,35	0,001
Irrigante	0,3269	2,48	0,013
_cons	4,8976	8,46	0,000
Estatísticas			
AIC	9,2967	BIC	-660,9242
Log likelihood	-649,4213		

Fonte: elaboração própria dos autores

Vale ressaltar que na busca de melhores modelos (alinhados) ou outras funções de ligação a deviance é de extrema importância, pois é uma estatística usada na formulação de estatísticas de testes de hipótese para inclusão ou não de outras variáveis no modelo dentro dos arcabouços dos MLGs.

Como, neste momento, não são testados modelos alinhados, buscou-se escolher o melhor modelo via análise dos critérios de informação; análise dos resíduos e interpretação dos parâmetros.

4.1. Análise dos critérios de informação AIC e BIC

Os critérios AIC e BIC trazem informação do erro e também fazem uma penalização quanto ao nível de complexidade do modelo (número de parâmetros), quanto menor o AIC e BIC, melhor é o modelo estimado. Geralmente são muito próximos e se o AIC de um modelo A é menor que o AIC de um modelo B, o BIC também tende a seguir esta comparação. Porém, em certos casos, essas informações podem ser conflitantes, nos modelos em questão tem-se que o BIC é menor para o modelo ajustado sob distribuição normal inversa e o AIC é menor no modelo gama.

O BIC penaliza mais a complexidade do modelo talvez por isso o modelo gama tenha um BIC maior do que o modelo inorm, dado que o modelo gama tem oito parâmetros, contra sete do modelo inorm.

4.2. Análise dos Resíduos

Para saber mais sobre a qualidade dos ajustes, faz-se necessário o diagnóstico dos resíduos. É importante verificar a normalidade dos resíduos, assim como a não ocorrência de outliers.

No âmbito dos MLGs podem ser testados os resíduos de anscombe (que é uma transformação que normaliza os dados e estabiliza a variância), resíduos de deviance, entre outros^{iv}. Neste trabalho, as análises serão feitas a partir dos resíduos de anscombe, por apresentarem boas propriedades estatísticas, mas também poderiam ser feitas sobre os resíduos de deviance. Para tal, primeiro plotou-se um qq-plot dos resíduos de anscombe dos dois modelos candidatos.

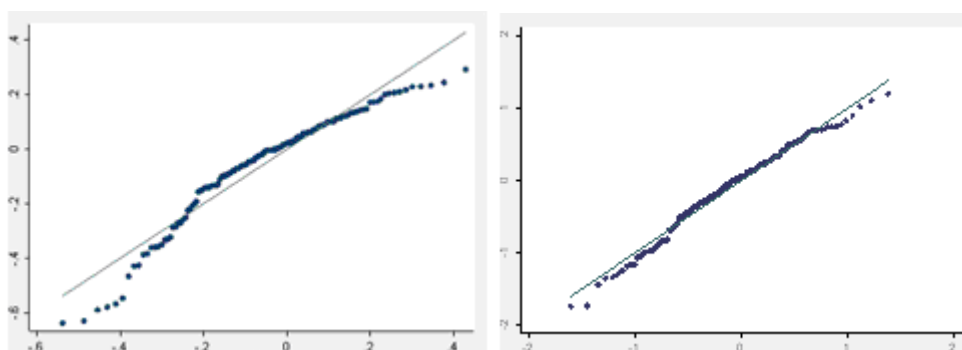


Figura 2 – Qq plot dos resíduos do modelo inorm (a) e qq plot dos resíduos do modelo Gama (b)

Observando a figura 2, percebe-se que o qq-plot dos resíduos do modelo normal inversa (a) apresenta um descolamento considerável da reta de 45°, enquanto o modelo gama parece ser bem comportado.

Para corroborar a análise visual, propõe um teste de normalidade dos resíduos, o teste Jarque-Bera (JB), Gujarati (2003). Este leva em consideração a assimetria e curtose dos resíduos e sua estatística de teste segue distribuição chi-quadrada com dois graus de liberdade.

O teste mostra que ao nível de 5% de confiança, não se rejeita a hipótese de normalidade (H_0) nos resíduos advindos do modelo ajustado com gama, pois este apresenta p-valor de 0,1189. No caso modelo inorm, rejeita-se H_0 , ao nível de 5% de confiança com o p-valor obtido igual a 0,0002.

Depois das análises realizadas a fim de definir o modelo de estudo, opta-se pelo modelo gama. Em que pese os dois modelos responderem bem aos testes e análises propostas, o modelo gama apresentou melhor resultados quanto a normalidade de seus resíduos e isso foi o fator preponderante à sua escolha.

4.3. Testando Interações e Termos Quadráticos

A primeira etapa da construção do modelo foi identificar os efeitos fixos. Não obstante, devem-se fazer testes que para verificar a significâncias de “não-linearidades”, ou seja, termos de ordem quadrática e interações das variáveis selecionadas na primeira etapa.

Para decidir pela inclusão ou não destas variáveis, usam-se testes do arcabouço dos MLGs. Com esse intuito Demétrio e Cordeiro (2008) propõem a aplicação do teste da razão de verossimilhança, teste de Wald e teste de score.

O primeiro usa as deviances dos modelos com e sem restrição, enquanto o segundo usa apenas a deviance do modelo sem restrição, já o terceiro usa o modelo reduzido.

Neste trabalho o teste de Wald foi aplicado com as seguintes hipóteses:

H_0 : Modelo com restrição ($\beta^*_1 = \beta^*_2 = \dots = \beta^*_k = 0$)

H_a : Modelo sem restrição (pelo menos $\beta^*_j \neq 0$, onde $j = 1, 2, \dots, k$)

Sua estatística de teste segue uma distribuição chi-quadrada com graus de liberdade igual ao número de restrições (parâmetros testados).

Foram testadas, ao nível de 10% de confiança, seis novas variáveis, entre termos quadráticos e interações. A estatística de teste segue distribuição chi-quadrada com 1 grau de liberdade e cada parâmetro foi testado isoladamente. Apenas a variável *Próprios_2* (correspondente ao termo “Próprios elevado ao quadrado”) foi considerada estatisticamente significativa segundo o teste de Wald apresentando um p-valor de 0,0402. Sendo assim o modelo de estudo passa a ter as variáveis “aprovadas” pelo stepwise + *Próprios_2*.

Tabela 5 – Modelo Final Normal Inversa link log

pnt_ei	coef.	z	$\rho > z $
Próprios	0,1279	1,61	0,108
Renda	-0,0010	-4,21	0,000
Próprios_2	-0,0011	-2,05	0,040
geladeira	0,0114	3,15	0,002
Irrigante	0,2825	2,46	0,014
_cons	-0,8893	-0,31	0,757
Estatísticas			
AIC	7,3667	BIC	-608,781
Log likelihood	-512,3576		

Fonte: elaboração própria dos autores

Com a opção pelo modelo com efeitos fixos mais o termo quadrático *Próprios_2*, uma nova análise de resíduos deve ser feita. Desta vez além dos resíduos de anscombe, foram calculados os resíduos da deviance.

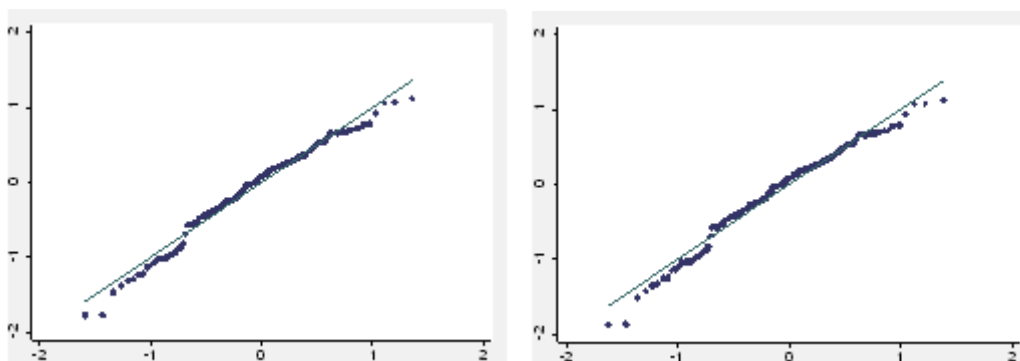


Figura 3 – Qq-plot dos resíduos de anscombe (a) e qq-plot dos resíduos da deviance (b)

O qq-plot mostra indícios de normalidade, mas para corroborar tal suspeita, fez-se o teste de normalidade JB. O teste foi feito para os resíduos de anscombe e para os resíduos da deviance.

Não se rejeita a hipótese de normalidade ao nível de 5% de confiança para os resíduos de anscombe, pois o p-valor do teste é 0,0743. Entretanto, os resíduos da deviance têm um p-valor um pouco mais baixo sendo de 0,0472, mesmo assim podemos considerá-los normais se tomarmos um nível de 1% de confiança.

4.4. Interpretação dos parâmetros

As variáveis que entraram no modelo segundo o *stepwise* são estatisticamente significantes ao nível de confiança de 10%. O termo constante não é significativo, em ambos os modelos, entretanto, deixa-se a constante na equação de estimação pela vantagem da não perda de certas propriedades estatísticas (Gujarati, 2000). No modelo gama, as variáveis do modelo estimado são as que seguem na tabela abaixo:

Tabela 2 – Coeficientes das variáveis do modelo de estudo

Variáveis	Coeficientes
<i>Próprios_2</i>	-0,001
<i>Próprios</i>	0,1279
<i>Renda</i>	-0,001
<i>Esgoto</i>	-0,005
<i>Irrigante</i>	0,2825
<i>Geladeira</i>	0,011

Fonte: elaboração própria dos autores

Variáveis de infra-estrutura têm sinal negativo, pois como discutido anteriormente, quanto mais acessos a serviços e menos precária for a condição de vida do consumidor, imagina-se que ele não se sente estimulado a fraudar.

Exceto pela variável *Próprios* (domicílios próprios) que no modelo aparece com sinal positivo, indicando quanto mais domicílios próprios mais PNT no alimentador i.

O que pode ser argumentado é que realmente domicílios próprios estariam mais propícios a cometer fraudes do que o seu complemento (alugados, cedidos e invasões). Domicílios alugados não cometeriam fraude, por fiscalização do próprio senhorio. Domicílios cedidos, segundo o IBGE, seriam domicílios cujo caseiro ou empregados seriam os moradores logo não teriam o porquê fraudar, dado que o dono da casa seria o responsável pelo pagamento da conta (geralmente baixa, pois os empregados podem não residir no local efetivamente). Domicílios invadidos realmente devem apresentar ligações clandestinas, estes podem ter pouco peso na amostra.

Com a entrada do termo quadrático *Próprios_2*, o efeito da variável *Próprios* é engolido, ou seja, deixa de ser significativo a 10% de confiança, como observado na Tabela 4. Esta permanece no modelo por recomendações de literaturas. As demais variáveis apresentam sinais pertinentes a uma análise empírica que segue o seguinte raciocínio: “quanto mais pobre (- renda), sem infra-estrutura (- esgoto) e com mais gastos com eletricidade (+ ter geladeira, + ser irrigante) maior a probabilidade de fraude”.

4.5. Adequação do Modelo

A fim de verificar se o modelo produziu resultados aderentes, foi elaborado um gráfico entre a variável resposta estimada e a observada. A relação de linearidade entre os pontos do gráfico mostra que o modelo proposto foi capaz de captar e estimar os efeitos entre as variáveis socioeconômicas e o nível de perda não técnica.

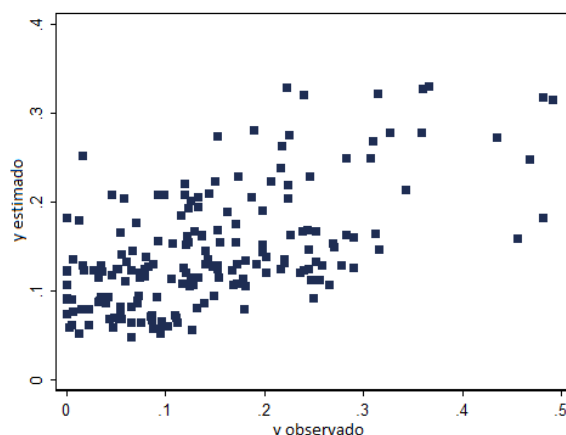


Figura 4 – Nível de PNT Estimado vs Obseado

5. CONSIDERAÇÕES FINAIS

O estudo elaborado conseguiu estimar um bom modelo para a estimação de PNTs. O modelo trouxe informações de quais variáveis são importantes para explicar as PNTs, inclusive interações e termos quadráticos, ou seja, variáveis que não são encontradas em literaturas prévias.

Propõe-se para trabalhos futuros testar variáveis mais difíceis de serem quantificadas, mas que certamente possuem influência nas PNTs como “existência de coronelismo, patronagem, sensação de impunidade, eficiência de inspeções e ações de combate”.

A abordagem da ANEEL se dá ao nível nacional, ou seja, há um modelo elaborado para todo o Brasil. Esse modelo não leva em consideração algumas particularidades bastante relevantes em cada uma das áreas de concessão como pode ser observado no modelo atual através da entrada da variável de irrigação. Empiricamente havia a suspeita que unidades consumidoras irrigantes seriam responsáveis pelo aumento do nível de PNT's.

Para finalizar, vale dizer que os arcabouços dos MLGs foram exaustivamente utilizados e enriqueceram a análise. O tema é atual e bastante relevância no âmbito econômico nacional.

6. REFERÊNCIAS

ANEEL (2007). “Metodologia de tratamento regulatório para Perdas-Não-Técnicas de energia elétrica para o segundo ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica. Nota Técnica 342”

ANEEL (2008). “Metodologia de tratamento regulatório para Perdas-Não-Técnicas de energia elétrica para o segundo ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica. Nota Técnica 348”.

ABRADEE (Associação Brasileira das Distribuidoras de Energia Elétrica) (2008). “Componente Social das Perdas-Não-Técnicas de Energia. Relatório Analítico”.

ARAÚJO, A. C. M (2007). “Perdas e Inadimplência na Atividade de Distribuição de Energia Elétrica. Tese (Doutorado em Ciências e Planejamento Energético)” - Coordenação dos Programas de Pós-Graduação de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

CONSULTORIA AEQUUS (2007). “Caracterização Socioeconômica dos Municípios da Área de Concessão da ESCELSA no Espírito Santo e sua Relação com os Índices de Perdas-Não-Técnicas de Energia Elétrica.”

CONSULTORIA TENDÊNCIAS (2007). “Determinantes Estruturais de Perdas Comerciais na Companhia Energética do Maranhão”.

CORDEIRO, G.M. e DEMÉTRIO, C.G.B. (2008). “Modelos Lineares Generalizados e Extensões (livro em preparação)”. Departamento de Estatística e Informática, UFRPE e ESALQ/USP.

CORDEIRO, Gauss M. (1986). “Modelos lineares generalizados”. Campinas, SP: 8º Simpósio Nacional de Probabilidade e Estatística, 286p.

DOBSON, Annette J. (1990). “An introduction to generalized linear models”. London: Chapman & Hall, 174 p.

GUJARATI, Damodar. (2003). “Econometria Básica”. Ed. campus.

LOYNES, R. M., “On Cox and Snell's General Definition of Residuals”. *Journal of the Royal Statistical Society*. Vol. 31, (1969), pp. 103-106.

McCULLAGH P; NELDER J.A. (1989). “Generalized Linear Models”. 2nd edition. London:Chapman and Hall

PESSANHA, J. M.; SOUZA, R. C., LAURENCEL, L. C. (2007). “Um modelo de análise envoltória de dados para o estabelecimento de metas de continuidade do fornecimento de energia elétrica”. *Pesqui. Oper.* [online]. 2007, vol.27, n.1, pp. 51-83. ISSN 0101-7438.

ⁱ As potenciais variáveis explicativas testadas foram: proporção de domicílios particulares permanentes com ar condicionado (**ar**); proporção de domicílios particulares permanentes com geladeira (**geladeira**); proporção de domicílios particulares permanentes com adequados (**dom_adequados**); proporção de domicílios atendidos pela rede geral de água (**água**); proporção de domicílios com ligação à rede urbana de esgoto (**esgoto**); analfabetismo (**analfabetismo**); Índice de desenvolvimento humano (**IDH**); razão entre óbitos por agressão e população residente (10.000) (**violência**); se alimentador cobre área irrigante ou não (variável dummy) (**irrigante**); rendimento médio nominal dos domicílios (**renda**); proporção de domicílios próprios (próprios); população economicamente ativa (**pea**); UC's com 3 ou mais faturas não pagas sobre o total de UC's (**inadimplência**). Vale lembrar que os nomes em negrito indicam os códigos das variáveis.

ⁱⁱ Para maiores detalhes ver Cordeiro e Demétrio (2008).

ⁱⁱⁱ Além das análises descritas foi elaborada após o ajuste do modelo, a Análise de Influência. Observou-se que não há observações que excedem o valor de referência da distância de cook.

^{iv} No MLG utiliza-se três principais tipos de resíduos: o de Pearson, o de Anscombe e aquele definido como componentes do desvio. Os três tipos de resíduos são casos especiais da definição de resíduos generalizados dada por Cox e Snell (1969).