

Área: Métodos Quantitativos

Modelagem da proporção de inadimplentes entre tomadores de empréstimo no Banco do Povo de Goiás utilizando Regressão Beta

Sérgio de Holanda Rocha – Mestre em Economia (UFPE – PIMES)
Endereço: Rua Xavier Marques 275/1002 Graças, Recife-PE. Telefone: (81) 99652222.
E-mail: sergiodeholanda@hotmail.com

Felipe Resende Oliveira – Mestre em Economia (UFPE-PIMES)

Jailson Araújo de Lima Júnior – Mestre em Economia (UFPE-PIMES) e Economista da Companhia Hidroelétrica do São Francisco (CHESF).

Edivaldo Constantino das Neves Júnior – Mestre em Economia (UFPE-PIMES)

Ricardo Carvalho de Andrade Lima - Mestre em Economia (UFPE-PIMES)

Modelagem da proporção de inadimplentes entre tomadores de empréstimo no Banco do Povo de Goiás utilizando Regressão Beta

Resumo

O microcrédito é uma ferramenta importante para viabilizar o acesso de microempreendedores de baixa renda ao mercado de crédito. O presente trabalho teve por objetivo modelar a proporção mensal de inadimplentes entre os tomadores de empréstimos do Banco do Povo de Goiás. Para tal, utilizamos regressão beta, uma vez que a variável dependente assume valores dentro do intervalo (0,1). Dispusemos de dados de um período de 70 meses referentes à inadimplência, sexo, empregos gerados, valor médio dos empréstimos e variação no preço da cesta básica, obtidos junto ao Banco do Povo e ao Ipeadata. Ao final, avaliamos que o modelo que apresentou os melhores resultados foi que utilizou uma função de ligação log-log com parâmetro de dispersão variável. Os resultados sugerem que o único regressor estatisticamente significativo foi o valor médio dos empréstimos por mês. Considerando que maiores empréstimos estão associados a maiores expectativas de lucro, nossos resultados corroboram a lei do risco-retorno.

Palavras-chave: Regressão Beta; Microcrédito; Banco do Povo de Goiás.

Abstract

Microcredit is an important tool to guarantee low-income micro-entrepreneurs' access to the credit market. The present paper seeks to model the monthly proportion of default among borrowers from Banco do Povo de Goiás. As such, we use a beta regression model since our dependent variable is defined between the interval (0,1). We use aggregated data from a 70 months period regarding occurrence of default, gender, jobs created, mean value of loan and variation in prices index of basic goods in the city of Goiânia. The data used was taken from Ipeadata and Banco do Povo de Goiás. We conclude that the model that best fits our data is the one that uses a log-log link function with a variable parameter of dispersion. Our results suggest that the only significant regressor is the monthly mean value of loans. Considering that greater loans are associated to great profit expectations, our results seem to corroborate the relation between risk and return.

Keywords: Beta Regression; Microcredit; Banco do Povo de Goiás.

JEL classification: C13; C52; H81

1. Introdução

Segundo o BNDES, o microcrédito é a concessão de empréstimos de pequeno valor a microempreendedores que normalmente não teriam acesso ao mercado financeiro. A categoria também se caracteriza pelos juros mais baixos em comparação ao crédito tradicional. Neste contexto, Yunus (2000) argumenta que o microcrédito não é apenas uma política de financiamento, mas também uma ferramenta para promover o desenvolvimento econômico e a inclusão social, por oferecer oportunidades de investimento e crescimento a pequenos empreendedores. Devido a tais características, estes empréstimos são normalmente concedidos por bancos de fomento e os chamados bancos populares, ligados de alguma forma ao setor público.

Durante o período de alta inflação que perdurou no Brasil ao longo dos anos 80 e início dos anos 90, o mercado de crédito se manteve debilitado. De acordo com o Relatório do Banco Central (2010), o plano Real fez com que os índices de preços apresentassem trajetória declinante no período 1994-1998. Segundo Soares (2011), a estabilidade econômica se consolidou com a implementação do tripé definido por: regime flexível de taxa de câmbio em janeiro de 1999; adoção do regime de metas de inflação em junho de 1999 e da Lei de Responsabilidade Fiscal (LRF) em maio de 2000, que estabelece limites aos gastos públicos. Em conjunto, estes fatores introduziram o país em um período de relativa estabilidade macroeconômica.

Uma das consequências desta estabilidade foi o aumento da participação de pequenos e microempreendedores na economia como um todo. Contudo, estes investidores continuaram se deparando com problemas em conseguir crédito junto ao sistema financeiro tradicional por terem dificuldades em apresentar garantias ao emprestador, o que motivou o surgimento de novas linhas de financiamento.

Um dos bancos que oferecem este tipo de crédito é o Banco do Povo, criado em 1999 para atuar no estado de Goiás. O público alvo da instituição é a população de baixa renda exercendo atividade formal ou informal, em meio rural ou urbano. Os recursos emprestados são provenientes do Tesouro do estado. Segundo o Manual do Banco do Povo de Goiânia (Manual do agente de crédito ONG, 2010), a política de crédito é destinada ao financiamento de máquinas, equipamentos e/ou mercadorias para revenda, com um valor de até R\$ 4.000,00, com uma taxa de juros de 0,6% ao mês e dividido em até 18 meses. Ao todo, 246 municípios goianos já implantaram o Banco do Povo e, até janeiro de 2012, 91.938 contratos foram realizados, com os repasses totalizando R\$16.159.482 e 133.320 supostos empregos gerados. O valor médio liberado por contrato é de R\$1.724, que está de acordo com a proposta de pequenos valores.

É evidente que a viabilização de oferta de crédito para a população de baixa renda não é uma tarefa simples; logo, nota-se que o Banco do Povo tem um papel importante e, nesse sentido, vem fazendo um trabalho fundamental para o estado de Goiás. Neste contexto, o presente trabalho teve por objetivo modelar a proporção de inadimplentes entre os tomadores de empréstimos do Banco do Povo, utilizando regressão beta, uma vez que a variável dependente, y , assume valores dentro do intervalo $(0,1)$. Utilizamos dados de 70 meses referentes à inadimplência, sexo, empregos gerados, valor médio dos empréstimos e variação no preço da cesta básica, obtidos junto ao Banco do Povo e ao Ipeadata.

O trabalho está dividido em cinco seções. Após esta introdução, a seção 2 apresenta uma explanação sobre regressão beta. Em seguida, na seção 3, discutimos brevemente os dados disponíveis e justificamos a escolha das variáveis. Na seção 4, apresentamos a estratégia empírica, testando diversos modelos e tipos de estimação e seus respectivos diagnósticos para tomada de decisão quanto ao melhor modelo. Por

fim, na quinta seção, apresentamos os resultados para o modelo escolhido e discutimos suas interpretação e implicações.

2. Notas Metodológicas

A variável dependente que buscamos modelar é a proporção de inadimplentes entre os tomadores de empréstimo no Banco do Povo de Goiás. A metodologia utilizada foi a de regressão beta (Ferrari, Cribari-Neto, 2004), utilizada para modelagem de variáveis de resposta contínuas e restritas ao intervalo (0,1). Mais geralmente, qualquer variável de resposta x restrita a um intervalo (a, b) com a e b conhecidos pode ser modelada por este método se definirmos uma nova variável aleatória $y = (x - a)/(x - b)$, de forma que $y \in (0,1)$. Para proceder, primeiramente assumimos a hipótese de que a variável de resposta segue uma distribuição beta de parâmetros p e q tal que:

$$\pi(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1} \quad (1)$$

Em que $p, q > 0$ e $\Gamma(\cdot)$ é a função gama dada por:

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt \quad (2)$$

A restrição desta hipótese é atenuada pelo fato de a distribuição beta ser bastante flexível, assumindo diversas formas dependendo dos parâmetros p e q . Esta classe engloba a distribuição uniforme contínua, que é um caso especial da beta quando $p = q = 1$. De forma geral, temos que, se $y \sim Beta(p, q)$,

$$E(y) = \frac{p}{p+q} \text{ e } Var(y) = \frac{pq}{(p+q)^2(p+q+1)} \quad (3)$$

Os parâmetros em questão, no entanto, não são particularmente úteis de se modelar. A solução proposta pelos autores do método é uma reparametrização da distribuição beta para que esta fique em termos de sua própria média e de um novo parâmetro que pode ser interpretado como um índice de precisão. Mais especificamente, definimos $\mu = p/(p+q)$ e $\phi = p+q$ de forma que:

$$E(y) = \mu \text{ e } Var(y) = \frac{\mu(1-\mu)}{1+\phi} \quad (4)$$

Onde podemos ver que, para uma determinada média, a variância da distribuição é decrescente em ϕ , sendo este último, portanto, um parâmetro de precisão. Por outro lado, dado ϕ , a variância é maior à medida que a média se aproxima de 1/2, já que o termo $\mu(1-\mu)$ atinge seu máximo global em $\mu = 1/2$. Sob esta nova parametrização, a densidade dada por (1) toma a seguinte forma:

$$f(y; p, q) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu(\phi-1)} (1-y)^{(1-\mu)\phi-1} \quad (5)$$

Em que buscamos modelar μ . Para tal, sejam y_0, \dots, y_n variáveis aleatórias independentes e identicamente distribuídas seguindo a densidade dada por (5). Seja μ_t a média de cada uma das variáveis. O próximo passo é assumir que μ_t pode ser escrita em termos de uma função e um preditor linear, a saber,

$$g(\mu_t) = \eta_t = \sum_{i=1}^k x_{ti}\beta_i \quad (6)$$

Em que η_t é um preditor linear, $\beta = (\beta_1 \dots, \beta_k)^T$ é um vetor de parâmetros a serem estimados na regressão e $g(\cdot)$ é uma função de ligação $g: (0,1) \rightarrow \mathbb{R}$ estritamente crescente e duas vezes diferenciável. A presença desta função se faz necessária, já que $\mu_t \in (0,1)$ e $\eta_t \in \mathbb{R}$. A estimação é feita por máxima verossimilhança. Considerando uma amostra de tamanho n , a função verossimilhança será dada por:

$$L(\beta, \phi) = \prod_{t=1}^n L_t(\mu_t, \phi) \quad (7)$$

Na qual aplicamos uma transformação monotônica $\log(\cdot)$ - tal que os valores que maximizam a função não se alteram - e definimos a função log-verossimilhança $l(\beta, \phi) = \log L(\beta, \phi)$. Pelas propriedades dos logaritmos, sabemos que:

$$l(\beta, \phi) = \sum_{t=1}^n l_t(\mu_t, \phi), \quad (8)$$

Em que $l_t(\mu_t, \phi)$ é o logaritmo da densidade dada por (5) para cada t da amostra. Portanto,

$$l_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t + ((1 - \mu_t)\phi - 1) \log(1 - y_t) \quad (9)$$

Onde $\mu_t = g^{-1}(\eta_t)$. Os estimadores da regressão beta $\hat{\beta}$ e $\hat{\phi}$ são tais que $l(\beta, \phi)$ é maximizada em relação a β e ϕ . Tais estimadores não têm forma fechada e, portanto, devem ser obtidos através de maximização numérica com o auxílio de um algoritmo de otimização não linear. Ressalta-se, também, que a interpretação dos estimadores $\hat{\beta}_i$, $i = 1, \dots, k$ depende da função de ligação incorporada ao modelo.

3. Dados

Os dados utilizados na regressão são relativos a empréstimos realizados pelo Banco do Povo entre janeiro de 2004 e dezembro de 2010. Nossa variável de resposta é a proporção mensal de inadimplentes entre os tomadores de recursos. Ao todo, foram utilizadas 70 observações, pois não dispúnhamos de dados referentes a 14 meses do período em questão. Como agregamos nossa variável dependente de modo que ela apresentasse periodicidade mensal, tivemos de fazer o mesmo com as observações para os regressores obtidas na base de dados do Banco do Povo. Entre estes, usamos a proporção de pessoas do sexo feminino, o número médio de empregos gerados por empreendimento, o valor médio dos empréstimos e a variação no gasto com a cesta básica no município de Goiânia, todos referentes aos meses do período em questão. Estas variáveis foram retiradas do banco de dados referente às operações do Banco do Povo, à exceção da última, que foi obtida no IPEADATA.

O uso da porcentagem de mulheres busca detectar se o sexo do tomador de empréstimo afeta a probabilidade de este vir a ser inadimplente. O número de empregos gerados pode representar as despesas com que o empreendedor arca para dar continuidade ao seu projeto; por isto, esperamos que o número de empregados se relacione com a capacidade do investidor de honrar as parcelas. De forma semelhante, queremos testar se o valor médio dos empréstimos por mês acompanha a taxa de

inadimplência no mesmo período. Pode-se argumentar que empréstimos de maior valor estão associados a maiores expectativas de lucro e, portanto, maior risco. Por outro lado, variações no gasto com cesta básica podem significar realocações dos recursos das famílias com o objetivo de manter a qualidade de vida. Neste caso, esperamos que maiores variações no preço estejam associadas a maiores taxas de inadimplência.

Considerando os dados agregados ao longo de todo o período, temos que 11,9% dos empréstimos realizados incorreram em inadimplência e 68,8% dos mesmos foram realizados por mulheres. Em média, 1,8 empregos foram gerados por empreendimento. O valor médio dos empréstimos foi de R\$1.877 e a taxa média de crescimento no gasto da cesta básica foi de 0,56% ao mês.

4. Estratégia Empírica

4.1. Aplicação da modelagem de regressão beta

O nosso trabalho teve por objetivo modelar a proporção mensal de inadimplentes entre os tomadores de empréstimos do Banco do Povo, utilizando regressão beta, uma vez que a variável dependente, y , assume valores dentro do intervalo $(0,1)$. Note que poderíamos adotar a ideia de aplicar uma abordagem de Mínimos Quadrados Ordinários (MQO); contudo, incorreríamos em problemas devido ao fato de que y assume valores restritos ao intervalo $(0,1)$ enquanto os valores dos regressores variam em todo o conjunto dos reais. Alternativamente, poderíamos utilizar uma transformação logarítmica na variável explicada para tentar solucionar esse obstáculo, do tipo Box-Cox, por exemplo.

Porém, adotando essa estratégia, estaríamos sujeitos a limitações graves decorrentes da interpretação dos valores estimados da regressão, que seriam interpretados em termos da média de y transformada e não da média de y , que é o nosso interesse (valendo a desigualdade de Jensen). Além disso, variáveis na forma de proporções, taxas e porcentagens são tipicamente heteroscedásticas e possuem distribuição assimétrica, dificultando aproximações gaussianas para realização de testes de hipóteses e construção de intervalos de confiança (Cribari-Neto e Zeileis, 2010).

Esse cenário constitui motivação e justificativa para a modelagem do problema utilizando regressão beta. Com isto em mente, de início, ajustamos um modelo de Mínimos Quadrados Ordinários para mostrar sua inadequação em tratar os nossos dados. Em seguida, procedemos com a estimação de uma regressão beta em que usamos como função de ligação uma logit e consideramos o parâmetro de precisão fixo.

Além disso, verificamos a existência de pontos de influencia usando a distância de Cook (Cook, 1977) e analisamos o comportamento do parâmetro de precisão. A nossa investigação prossegue, buscando averiguar se o parâmetro de precisão é ou não fixo, usando, para tal, o teste da razão de verossimilhança, e, em seguida, o critério BIC para seleção de modelos. O próximo passo é tentar encontrar a melhor função de ligação, ajustando regressões com formas log e log-log. Utilizamos o pseudo- R^2 e o critério AIC de seleção de modelos para extrair nossas conclusões. Por fim, verificamos se o modelo está bem ajustado, usando como ferramenta o teste de diagnóstico inspirado no teste RESET de Ramsey sugerido por Cribari-Neto e Lima (2007).

4.2. Um ajuste com Mínimos Quadrados Ordinários (MQO)

Ajustamos uma regressão por mínimos quadrados ordinários para nossas 70 observações e obtivemos os seguintes resultados:

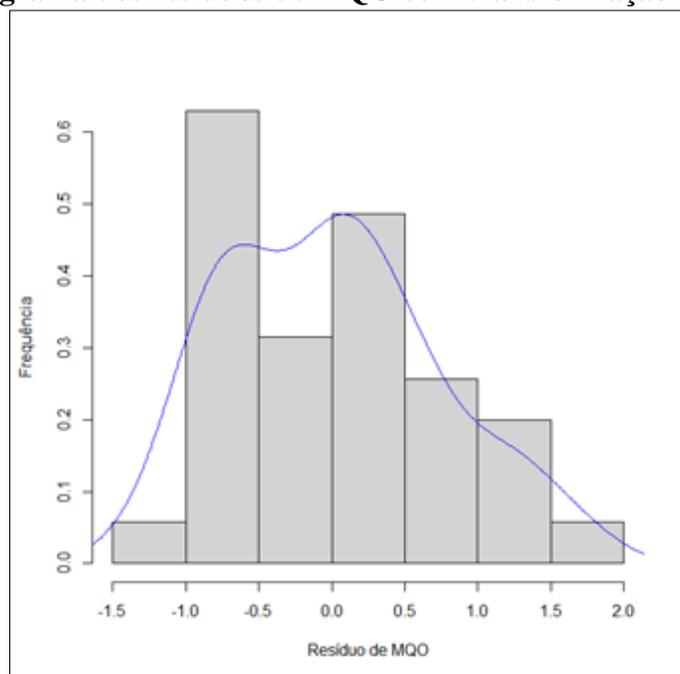
Tabela 1 – Resultados do MQO:

Variável	Estimativa	Erro-Padrão	Valor-t	Valor-p
Intercepto	-0,0392	0,107	-0,37	0,71
Mulher	0,00196	0,0954	0,02	0,98
Empregos	-0,0242	0,0262	-0,92	0,36
Valor	0,000115	0,000028	4,12	0
Cesta	0,000742	0,00365	0,21	0,83
R ²	0,32			
R ² Ajustado	0,28			

Fonte: Elaboração própria. Dados: Banco do Povo (GO) e Ipeadata.

Os resultados mostram que apenas a variável valor é estatisticamente significativa para explicar a proporção de inadimplentes, e que o poder de explicação do modelo é de 32,35%. Uma análise mais cuidadosa revela que esse modelo apresenta problemas de heteroscedasticidade. Para fundamentar nossa argumentação, utilizamos o teste de Koenker (1981) para identificar a presença da heteroscedasticidade. Assim, testamos a hipótese nula de que os dados são homoscedásticos *versus* a hipótese alternativa de que são heteroscedásticos. Os resultados mostram que, de fato, temos heteroscedasticidade em nossos dados, uma vez que obtivemos um *valor-p* de 0,013, rejeitando a hipótese nula a 5% de significância. Além disso, podemos aplicar uma transformação logarítmica na variável dependente no modelo padrão. Porém, a partir da figura 1 abaixo, percebemos que os dados são claramente assimétricos, constituindo mais um entrave para a validade dessa ferramenta. Desse modo, vamos seguir nosso estudo ajustando um modelo de regressão beta.

Figura 1 - Histograma dos resíduos de MQO com transformação logarítmica em y:



Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

4.3. Estimação via regressão beta

Uma vez que nossa variável dependente é contínua e restrita ao intervalo (0,1), procedemos com uma modelagem através de regressão beta. Como visto anteriormente, os autores sugeriram uma reparametrização envolvendo um parâmetro de dispersão. Para a modelagem de μ_t em termos do preditor linear, escolhemos, dentre algumas funções de ligação possíveis, as duas seguintes:

$$g(\mu) = \log\left\{\frac{\mu}{1-\mu}\right\} \text{ e } g(\mu) = -\log\{-\log(\mu)\} \quad (11)$$

Chamamos de logit e log-log, respectivamente.

Primeiramente, nosso objetivo é verificar se o parâmetro de dispersão é fixo ou variável. Para tanto, ajustamos uma regressão beta com função de ligação logit e parâmetro de precisão fixo. Os resultados (ver tabela 2) mostraram que a variável valor, bem como a variável intercepto, foram estatisticamente significantes a 1% de significância, e o modelo apresentou um poder de explicação de 34,8%. Note ainda que a estimação do modelo é realizada via máxima verossimilhança de forma numérica e que o pacote estatístico R utiliza como algoritmo de otimização o método *quasi-Newton* BFGS, que possui um melhor desempenho.

Tabela 2 - Resultados da regressão beta com dispersão fixa e ligação logit:

Variável	Estimativa	Erro-Padrão	Valor-t	Valor-p
Intercepto	-3,06	0,74	-4,12	0,00003
Mulher	-0,09	0,64	-0,14	0,88
Empregos	-0,13	0,18	-0,74	0,46
Valor	0,001	0,00422	4,22	0,00002
Cesta	0,005	0,02	0,24	0,81
ϕ	16,51	2,82	5,85	0
Log-Like	88,74 (6 gl)			
Pseudo R ²	0,348			
Nº Interações	18			

Fonte: Elaboração própria. Dados: Banco do Povo (GO) e Ipeadata.

Podemos analisar a qualidade do ajuste da regressão usando diferentes ferramentas de diagnóstico. O gráfico de probabilidade meio normal (ver figura 2, no apêndice) com envelope simulado mostra que temos pontos que se encontram fora da banda de confiança, o que denota que o modelo ainda não está bem ajustado. Além disso, a partir da figura 3, no Apêndice, que plota os resíduos ao longo da amostra, percebe-se que os pontos não estão distribuídos de forma aleatória. Ademais, visando a identificar pontos de alavanca, utilizamos como critério de análise a distância de Cook denotada por:

$$C_t = \frac{h_{tt}r_t^2}{k(1-h_{tt})^2} \quad (12)$$

A partir da figura 4 do Apêndice, constatamos a presença de pontos de alavanca. Dado esse cenário, optamos inicialmente por retirar essas observações da base de dados e verificar o comportamento do parâmetro de precisão. Os resultados mostraram que o $\hat{\phi}$ da regressão beta sem os pontos de alavanca (30.18113) é bem superior ao modelo

em que não retiramos essas observações (16.51487). A princípio, uma decisão plausível seria a exclusão dessas observações para prosseguir com nossa análise; contudo, adotando cautela, decidimos não retirá-las. Agora, devemos obter os resultados do ajuste da regressão com parâmetro de dispersão variável e, em seguida, aplicar o teste da razão de verossimilhança a fim de escolher o modelo mais adequado.

O modelo proposto por Ferrari e Cribari-Neto (2004) considera que o parâmetro de dispersão é fixo para as observações. Em muitas situações, essa hipótese pode não ser adequada, de modo que se torna necessário utilizar uma extensão do modelo de regressão beta introduzida por Simas, Barreto-Souza e Rocha (2010). Nesse caso, adiciona-se à modelagem da média uma estrutura de regressão para o parâmetro de dispersão. Desse modo, assume-se que:

$$g_1(\mu_i) = \eta_{1i} = x_i^T \beta \text{ e } g_2(\eta_i) = \eta_{2i} = z_i^T \gamma \quad (13)$$

Em que $\beta = (\beta_1 \dots, \beta_k)^T$ e $\gamma = (\gamma_1 \dots, \gamma_k)^T$ são os vetores de parâmetros a serem estimados no modelo. Com isto em mente, ajustamos um modelo de regressão beta com parâmetro de dispersão variável em que incorporamos o valor dos empréstimos como regressor no submodelo de dispersão. Como critério para escolher essa variável, testamos diversas combinações entre as variáveis explicativas e verificamos que a utilização da variável valor apresentava melhores resultados, sempre significantes. Os resultados podem ser vistos na tabela 3 abaixo:

Tabela 3 - Regressão beta com dispersão variável e ligação logit:

Variável	Estimativa	Erro-Padrão	Valor-t	Valor-p
Intercepto	-3,18	0,69	-4,603	0,0000001
Mulher	-0,3	0,64	-0,47	0,63
Empregos	-0,2	0,16	-1,264	0,2
Valor	0,001	0,00488	4,88	0,000002
Cesta	0,0001	0,02	-0,01	0,99
ϕ				
Intercepto	4,9	0,064	7,6	0
Valor	0	0	-3,17	0
Log-Like	92,93 (7 gl)			
Pseudo R ²	0,348			
Nº Interações	20			

Fonte: Elaboração própria. Dados: Banco do Povo (GO) e Ipeadata.

Perceba que a variável valor é estatisticamente significativa no submodelo do parâmetro de dispersão, sugerindo que, de fato, este último é variável. Após a estimação são realizados testes sobre os parâmetros do modelo com o intuito de confirmar ou não as hipóteses feitas sobre estes. Os testes mais comumente aplicados são o da razão de verossimilhança, score e Wald, pois suas estatísticas se baseiam na função de verossimilhança, que apresenta várias propriedades de otimalidade (Queiroz e Cribari-Neto, 2011).

Contudo, vamos utilizar o teste de razão de verossimilhança para escolher entre os dois modelos. Esse teste considera a hipótese nula de que a dispersão é fixa contra a hipótese alternativa de que a mesma é variável. Fazendo o teste, rejeitamos a hipótese nula a 1% de significância, o que indica que temos um modelo com dispersão variável.

Além disso, procedemos com um teste de seleção de modelos BIC para fundamentar com mais precisão nossa análise. Note que, no R, o BIC pode ser escrito

em termos de AIC quando $\log(n)$ é fornecido como o termo de penalidade k . Assim, encontramos que o modelo com dispersão variável apresentou um melhor ajuste, com AIC de -154,84, contra um AIC de -150,88 para o modelo com dispersão fixa, corroborando com nossos resultados anteriores. Até o presente momento do nosso estudo, concluímos que se deve utilizar um modelo com dispersão variável.

4.4. Seleção de diferentes funções de ligação

A seleção de uma função de ligação apropriada possui um grande impacto no ajuste do modelo, especialmente se verificarmos proporções perto de 0 ou de 1 nos dados. Assim, nosso objetivo nessa etapa do trabalho é estudar qual a melhor função de ligação para os nossos dados. Nesse sentido, ajustamos mais uma regressão beta em que alteramos a função de logit para log-log. Assim, temos duas regressões:

- (1) Ligação logit + Parâmetro de dispersão variável
- (2) Ligação log-log + Parâmetro de dispersão variável.

Perceba que, a partir das conclusões da subseção anterior, temos a informação de que devemos utilizar um modelo com dispersão variável. Como ferramenta de diagnóstico para avaliar a função de ligação, iremos utilizar o pseudo- R^2 e, em seguida, o critério de seleção de modelos AIC. O primeiro é dado por:

$$R_p^2 = Corr(\hat{\eta}, g(y)) \quad (14)$$

Os resultados mostram que o pseudo R^2 do modelo 1 (0.348875) é relativamente maior que o do modelo 2 (0.343392). Por outro lado, quando usamos o critério de seleção de modelos AIC constatamos que o modelo 2 apresenta um AIC -172,57, contra -171,86 para o modelo 1, tendo o modelo 2 uma melhor performance. Porém, atente que a diferença entre os valores obtidos é pequena. Esses resultados são contraditórios e não são capazes de conduzir a uma resposta específica sobre qual a função de ligação devemos utilizar.

Ferrari e Cribari-Neto (2004) argumentam que quando a variável de resposta não apresenta muitas observações localizadas no extremo da distribuição, a escolha da função de ligação não é tão relevante. Assim, o ajuste do modelo com várias funções de ligações não apresenta grandes disparidades no que diz respeito à log-verossimilhança. Realizando um teste com todas as funções de ligações disponíveis (ver tabela 4) é possível verificar que a função log-log apresenta melhores resultados, mas perceba que a diferença com outras funções é muito sutil. Corroborando com nossa argumentação, a partir da figura 7 do Apêndice, que mostra uma comparação dos resíduos absolutos $(y_i - \hat{\mu}_i)$ de ambos os modelos, verificamos que temos poucas observações acima e abaixo da diagonal, o que demonstra que os ajustes feito pelo logit e pelo log-log não apresentam grandes discrepâncias.

Tabela 4 - Seleção de função de ligação

Logit	Probit	Log-Log	Complementar	Cauchy	Log-log
92,9	93,1	92,76		91,14	93,28

Fonte: Elaboração própria

Assim, vamos considerar um modelo com função de ligação log-log e parâmetro de dispersão variável. Com o intuito de verificar se de fato temos um modelo bem

especificado, Cribari-Neto e Lima (2007) sugerem utilizar uma sequência de testes de diagnóstico baseados na ideia do teste RESET. Em outras palavras, incluiremos regressores auxiliares e testaremos suas significâncias. Caso esses regressores sejam estatisticamente significantes, então podemos concluir que nosso modelo está mal especificado. Os resultados mostram que não conseguimos rejeitar a hipótese nula de que o modelo está bem especificado, ou seja, temos um bom modelo.

5. Resultados e conclusões

A partir da seção anterior, concluímos que o modelo que apresenta o melhor desempenho é aquele com parâmetro de dispersão variável e com função de ligação log-log. Os resultados estão dispostos na tabela 5 abaixo:

Tabela 5 - Teste z para os coeficientes do modelo escolhido:

Variável	Estimativa	Erro-Padrão	Valor-t	Valor-p
Intercepto	-1,23	0,278	-4,4245	0,0000102
Mulher	-0,184	0,262	-0,7023	0,48
Empregos	-0,103	0,066	-1,5606	0,11
Valor	0,00044	0,0000906	4,85651	0,000001
Cesta	-0,00107	0,02	-0,0535	0,91
ϕ Intercepto	4,82	0,645	7,472868	0
ϕ Valor	-0,00101	0,000335	-3,01493	0,0025

Fonte: Elaboração própria. Dados: Banco do Povo (GO) e Ipeadata.

O ajuste do modelo acima pode ser verificado nas figuras 5 e 6 do Apêndice. Percebemos que, de fato, a única variável significativa obtida no modelo é o valor médio tomado a cada período. Apesar de seus valores não terem sido deflacionados, argumentamos que, se o efeito da variação dos preços fosse significativo, este deveria ser detectado pela variação no preço da cesta básica em Goiânia, pois tal índice deveria ser mais representativo em relação aos elementos da amostra.

Verificamos que as estimativas das variáveis de sexo e o número de empregos gerados não foram significantes para se associar à inadimplência. No entanto, reconhecemos que esta última variável não era promissora, pois pode ser superestimada, dado que os tomadores de empréstimos têm incentivos para reportar um número maior de empregos gerados a fim de ressaltar o benefício social de seu empreendimento.

De forma geral, podemos concluir que o valor médio dos empréstimos realizados ao longo do mês está relacionado positivamente ao índice de inadimplência. Como se espera que maiores investimentos estejam associados a maiores expectativas de lucro, verificamos a presença da lei de risco-retorno.

6. Referências bibliográficas

BANCO CENTRAL DO BRASIL. Disponível em: www.bcb.gov.br/?microfin/ 1. Sistema financeiro (Periódico). 2. Finanças (Periódico). Acesso em 16 de Dez. 2012.

BURTI, A. Sobrevivência e mortalidade das empresas paulistas de 1 a 5 anos. Serviços Brasileiro de Apoio às micro e pequenas empresas. SEBRAE. São Paulo, 2004.

CRIBARI-NETO, F., ZEILEIS, A. Beta Regression in R. Journal of Statistical Software, 2010.

FERRARI, S. L. P., CRIBARI-NETO, F. Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics, 2004.

PROGRAMA BANCO DO POVO. Manual do agente de crédito ONG. Goiás, 2010.(a)

PROGRAMA BANCO DO POVO. Manual do agente de crédito SLC/ SUIA. Goiás, 2010.(b)

SIMAS, A. B., BARRETO-SOUZA, W., ROCHA, A. V. Improved estimators for a general class of beta regression models. Computational Statistics & Data Analysis, 2010.

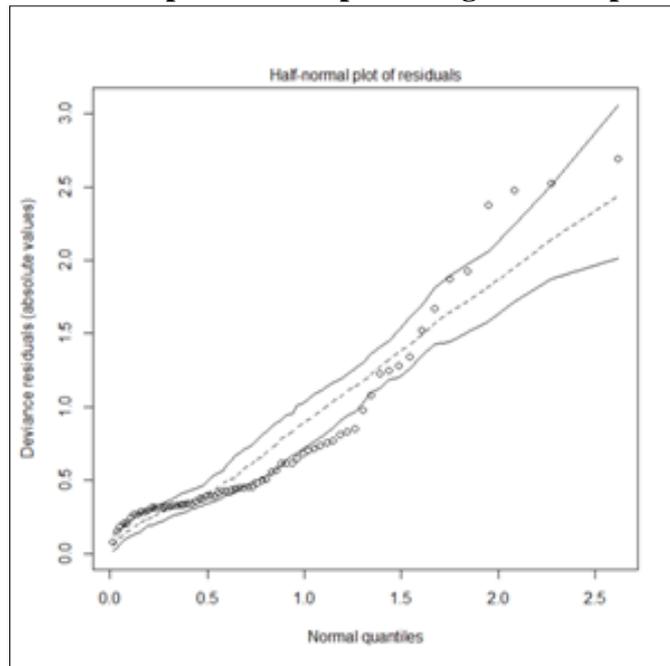
SMITHSON, M., VERKUILEN, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods, 11(1), 54-71, 2004.

SOARES, F.A.R. Economia Brasileira: Da primeira República ao Plano Real. Rio de Janeiro: Elsevier, 2011.

YUNUS, M. O banqueiro dos pobres. São Paulo: Ática, 2000.

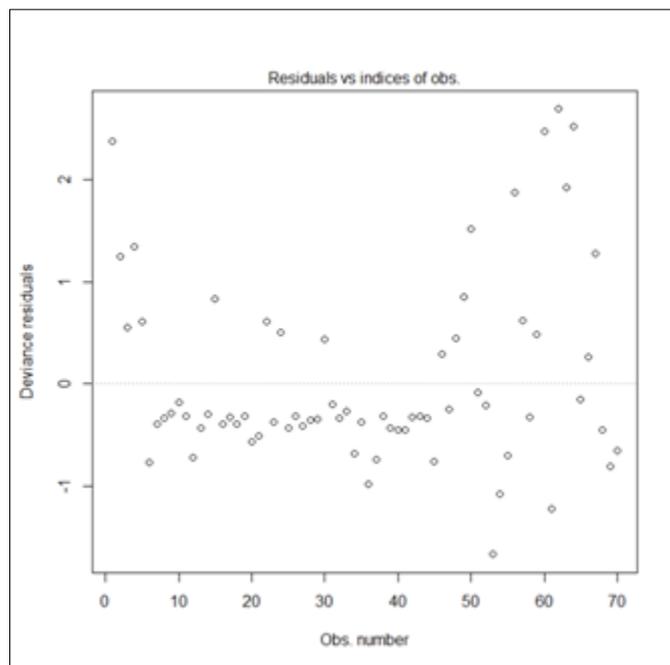
7. Apêndice - Figuras

Figura 2 - Envelope simulado para a logit com dispersão fixa.



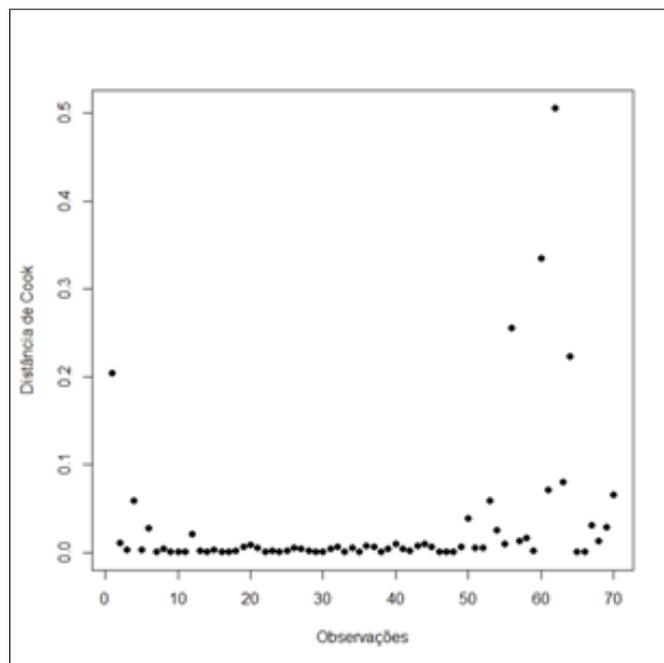
Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

Figura 3: Dispersão dos resíduos ao longo da amostra para a logit com dispersão fixa.



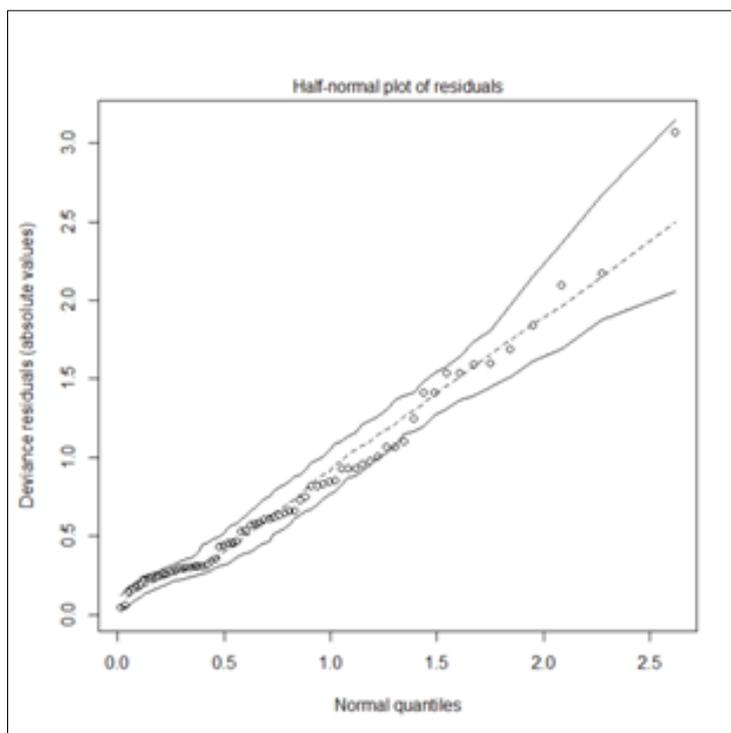
Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

Figura 4 - Distância de Cook para a logit com dispersão fixa.



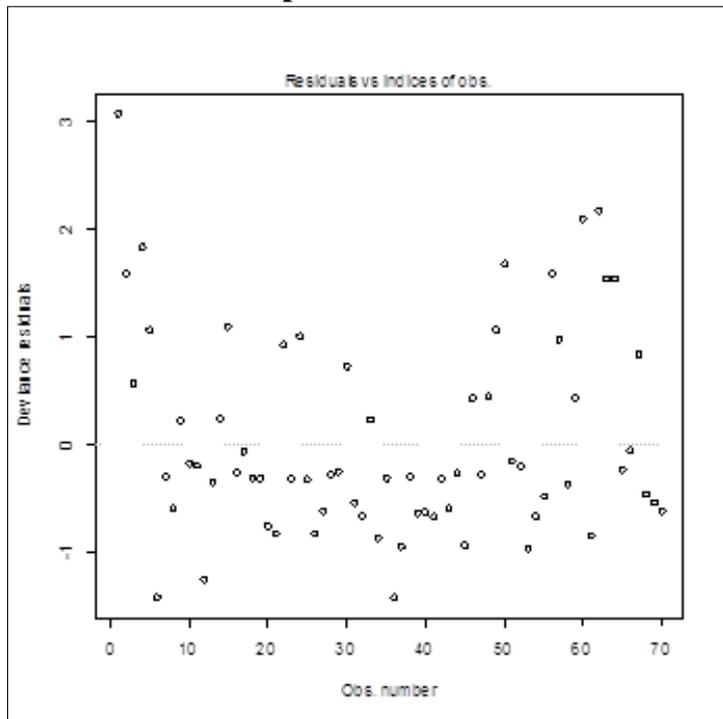
Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

Figura 5 - Envelope simulado para a log-log com dispersão variável.



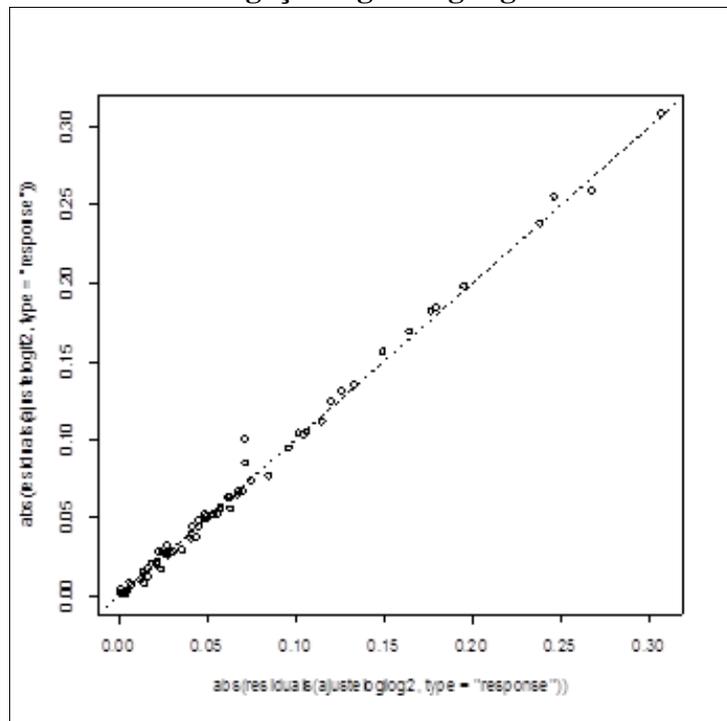
Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

Figura 6 - Dispersão dos resíduos ao longo da amostra para a log-log com dispersão variável.



Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.

Figura 7 - Comparação do valor dos resíduos da regressão beta com funções de ligação logit e log-log.



Fonte: Elaboração Própria com dados do Banco do Povo (GO) e Ipeadata.