

IBMEC SÃO PAULO

Mestrado Profissional em Economia

Jefferson Ferreira

**Modelos de previsão de perdas para
crédito massificado**

São Paulo

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Jefferson Ferreira

**Modelos de previsão de perdas para
crédito massificado**

Dissertação apresentada ao Programa
de Mestrado Profissional em Economia
da Faculdade IBMEC São Paulo para a
obtenção do título de Mestre em
Economia

Área de concentração: Finanças

Orientador: Prof. Dr. Rinaldo Artes – IBMEC SP

São Paulo

2008

Ferreira, Jefferson

Modelo de previsão de perdas para crédito massificado/Jefferson Ferreira; orientador Rinaldo Artes. São Paulo: Ibmec São Paulo, 2008. 53 f.

Dissertação (Mestrado – Programa de Mestrado Profissional em Economia. Área de concentração: Finanças – Faculdade Ibmec São Paulo).

1.Crédito 2.ZAIG 3.BEINF 4.Perdas 5.Inadimplência

FOLHA DE APROVAÇÃO

Jefferson Ferreira
Modelos de previsão de perdas para crédito massificado

Dissertação apresentada ao Programa de Mestrado Profissional em Economia da Faculdade IBMEC São Paulo para a obtenção do título de Mestre em Economia

Área de concentração: Finanças

Aprovado em: Julho/2008

Banca examinadora

Prof. Dr. Rinaldo Artes (orientador)
Instituição: IBMEC São Paulo

Profª. Dra. Andrea Maria Accioly Fonseca Minardi
Instituição: IBMEC São Paulo

Profª. Dra. Lúcia Pereira Barroso
Instituição: IME - USP

Agradecimentos

Agradeço a Deus pela inspiração e pela força dada à execução deste trabalho e por me ajudar a conciliar minha vida profissional, pessoal e o mestrado.

Aos meus pais por tudo que me ensinaram e pelo investimento e dedicação que sempre tiveram para minha formação.

À minha esposa Joana e aos meus filhos Mateus e o recém-chegado Tiago pela paciência e compreensão pelo pouco tempo que dediquei a eles neste período.

Ao professor Rinaldo Artes, grande orientador e excelente pessoa, pela dedicação e motivação, dando apoio na hora certa e debatendo os resultados e soluções propostas.

Ao Geraldo pelos muitos anos de convivência e pela grande colaboração em minha formação profissional.

Aos companheiros de mestrado, Augusto, Airton e Ronald pelas experiências trocadas ao longo do curso, principalmente em nossos trabalhos em grupo.

Aos meus colegas de trabalho e a todos que contribuíram de forma direta ou indireta, dando apoio ou idéias.

Resumo

Ferreira, Jefferson. **Modelo de previsão de perdas para crédito massificado**. São Paulo, 2008. 53 p. Dissertação (Mestrado) – IBMEC São Paulo, São Paulo, 2008.

Neste trabalho apresenta-se uma aplicação empírica para modelagem de perdas de uma carteira de crédito direto ao consumidor, cujas características são: grande quantidade de créditos com baixa exposição individual e pequena probabilidade individual de um tomador se tornar inadimplente, o que torna a ocorrência de perda um evento raro. A distribuição de perdas apresenta forte assimetria à direita, entre outras razões, devido à quantidade elevada de valores iguais a zero.

O objetivo da modelagem proposta é estimar a distribuição de perdas de crédito desta carteira, ajustando-a para grande massa de valores de perda iguais a zero, sendo algo muito pouco utilizado no mercado brasileiro de crédito e financiamento de varejo. Em termos práticos, há a possibilidade de aplicar os modelos estimados em novas carteiras de clientes para prever o montante de perda futura.

Propõe-se aqui a comparação de dois modelos: *ZAIG* e *BEINF*, apresentando-se a metodologia e os conceitos necessários a sua implementação e a avaliação dos resultados obtidos.

Palavras-chaves: Crédito; *ZAIG*; *BEINF*; perdas; inadimplência.

Abstract

Ferreira, Jefferson. **Models of expected loss for consumer credit**. São Paulo, 2008. 53 p. Dissertation (Mastership) – IBMEC São Paulo, São Paulo, 2008.

This essay shows an empirical application of modeling loss to a consumer credit portfolio, whose characteristics are: large amount of credits with low individual exposure and small individual default probability, which makes the occurrence of loss a rare event. The loss distribution shows extreme right skewness, among others reasons, due to a high quantity of a mass of zeros.

The goal of this modeling proposal is to estimate the loss distribution of this credit portfolio, adjusting it to great mass of zeros, which is unusual in the Brazilian consumer credit market. In practical terms, there is a possibility to apply estimated models in new portfolios to predict future loss of amount.

It is proposed in this study the comparison of two models: ZAIG and BEINF, showing the methodology and concepts necessary for its implementation and evaluation of results.

Keywords: Credit, ZAIG, BEINF, loss, default.

Sumário

1. Introdução	1
2. Conceitos de risco de crédito	3
2.1 Risco de crédito.....	3
2.2 Modelos de classificação de crédito.....	3
2.3 Modelos de risco de carteira	4
3. Metodologia	5
3.1 Modelo <i>ZAIG</i>	6
3.2 Modelo <i>BEINF</i>	8
3.3 Medidas de comparação dos modelos propostos	12
4. Descrição dos dados.....	14
4.1 A carteira.....	14
4.2 Variáveis utilizadas.....	14
4.3 Análise descritiva	19
5. Resultados	20
5.1 Modelo <i>ZAIG</i>	20
5.2 Modelo <i>BEINF</i>	26
5.3 Comparação dos modelos propostos.....	31
6. Conclusão	38
Apêndice A – Descrição das variáveis.....	40
Apêndice B – Análise descritiva.....	42
Apêndice C – Programas utilizados	47
Referências bibliográficas	51

1. Introdução

Entende-se por crédito direto ao consumidor, toda forma de empréstimos diretos a pessoas físicas para compras de mercadorias. O termo “crédito ao consumidor” é internacionalmente entendido como uma das muitas formas de comércio, a partir da qual um indivíduo obtém dinheiro, bens ou serviços com uma condição de promessa de pagar esse dinheiro, bem ou serviço, adicionada uma taxa, numa data futura específica.

Inerente à idéia do crédito está a idéia do risco, já que ele envolve a promessa de pagamento do valor emprestado em parcelas a serem pagas nos meses subseqüentes. Não se pode prever o futuro com perfeita acurácia, portanto, deve-se aceitar o fato de que nem todos os créditos concedidos serão pagos. Visa-se maximizar o valor emprestado minimizando-se o risco de inadimplência (não pagamento das parcelas do financiamento) por parte do devedor.

Dentro desta perspectiva é fundamental o correto dimensionamento das perdas esperadas na carteira dadas as diversas influências sofridas pelos consumidores, tais como desemprego, instabilidade econômica, entre outros fatores.

Muitos bancos e instituições financeiras baseiam suas decisões de concessão de empréstimos exclusivamente em análises subjetivas do perfil do indivíduo e de sua condição financeira no momento da solicitação, porém o uso de modelos de gestão de risco para crédito massificado pode melhorar o processo de tomada de decisão, reduzindo o risco de inadimplência.

A estabilidade econômica vivida nos últimos anos proporcionou um acentuado crescimento no volume de crédito disponível, incentivando a adoção de modelos estatísticos de avaliação individual do crédito, como os modelos de *Credit Scoring*. Tais modelos são denominados modelos de classificação de risco (ver Andrade, 2004).

Apesar da evolução no processo de concessão, o acompanhamento da inadimplência, das perdas e possíveis recuperações ainda são muito pouco usuais. Diversos modelos de gestão de risco podem ser implantados para controlar a inadimplência, porém mudanças de rumo ou do perfil dos clientes não são observadas rapidamente, o que pode levar a perdas expressivas para a instituição financeira.

Algumas poucas instituições brasileiras adotam modelos de riscos da carteira. Nesse caso, o objetivo é prever a perda total da carteira. Uma grande dificuldade na utilização destes modelos advém de uma das características de carteiras de crédito: A proporção de clientes inadimplentes é, em geral, muito pequena, o que pode gerar estimativas distorcidas das perdas, dadas as distribuições usualmente utilizadas.

O objetivo deste trabalho é estudar uma carteira de crédito direto ao consumidor e aplicar uma metodologia para quantificar a perda esperada. Para isto, apresentam-se dois modelos estatísticos que estimam, simultaneamente, a probabilidade de cada indivíduo tornar-se inadimplente e a perda esperada neste caso. Os modelos empregados têm aplicação recente em Economia, não se encontrando aplicação equivalente à proposta neste trabalho. Em uma situação prática, tais modelos podem ser aplicados na previsão de perdas futuras em novas carteiras de clientes.

Este trabalho se divide da seguinte forma:

No Capítulo 2, apresentam-se conceitos gerais sobre risco de crédito que serão utilizados.

No Capítulo 3, apresenta-se a metodologia dos modelos propostos, descrevendo sua conceituação.

No Capítulo 4 são descritos os dados e os métodos necessários à aplicação e no 5 os resultados encontrados em uma carteira de crédito direto ao consumidor. Por fim, as conclusões são apresentadas no Capítulo 6.

2. Conceitos de risco de crédito

Este capítulo destina-se a apresentação de conceitos e informações necessários ao desenvolvimento deste trabalho.

2.1 Risco de crédito

Conforme Siqueira (2000), o risco de crédito está associado à possibilidade de escolha do credor, ou seja, o risco é uma consequência da decisão de conceder um empréstimo com expectativa de ganho, sabendo-se, porém, que há a possibilidade de perda.

Segundo Andrade (2004) o risco está associado a três elementos:

- Ocorrência de evento negativo que pode levar a perdas.
- É uma opção de quem assume o risco.
- Quem assume este risco, visa obter lucro na operação, ou seja, que as receitas obtidas sejam maiores que as perdas ocorridas.

Desta forma, o conceito de risco de crédito está associado à probabilidade do credor não receber uma obrigação do tomador do empréstimo e, portanto, não obter o resultado esperado da operação, apresentando uma variabilidade sobre este valor. Siqueira (2000) diz que para se conhecer o risco devem-se estabelecer os possíveis resultados da operação e as probabilidades de ocorrência de cada um deles.

A fim de minimizar o risco advindo da atividade de concessão de créditos, as instituições utilizam-se de modelos de previsão do risco de inadimplência, que se refere ao atraso ou não pagamento de um compromisso, e modelos de quantificação das perdas ocorridas em caso de inadimplência. Para o primeiro caso, utilizam-se modelos de classificação de riscos e para o segundo, modelos de risco de carteiras.

2.2 Modelos de classificação de risco

Segundo Andrade (2004), estes modelos visam atribuir a um devedor ou a uma operação uma medida de risco associada à previsão da ocorrência da perda.

São utilizados no processo de avaliação de crédito, cumprindo dois objetivos básicos:

- Proporcionar uma decisão objetiva sobre a concessão dos empréstimos; se a classificação obtida apresentar maior risco que o desejável pela instituição, esta será rejeitada.
- Proporcionar eficiência operacional, dado que o processo se torna mais automatizado, reduzindo custos para a instituição.

O modelo mais utilizado para este objetivo é o *Credit Scoring*, que utiliza características cadastrais do proponente e da operação de crédito e, através de técnicas estatísticas, geram pontuação que representam a expectativa de inadimplência. Normalmente estas pontuações são utilizadas para gerar um *rating* de cada cliente/operação.

As técnicas mais utilizadas são regressão linear, regressão logística, análise discriminante e algumas técnicas mais sofisticadas como modelos de inteligência artificial e algoritmos genéticos. Comparações entre técnicas são discutidas em Rosa (2000) e Thomas, Edelman e Crook (2002) e o desenvolvimento de sistemas de *scoring* são amplamente discutidos por Sicsú (1998a,b) e Vasconcellos (2002).

2.3 Modelos de risco de carteira

Os modelos de risco de carteira, ao contrário dos modelos de classificação que observam as probabilidades individuais de inadimplência dos créditos, preocupam-se com dois objetivos básicos segundo Andrade (2004): o principal é determinação da distribuição de perdas de uma carteira (composta por todos os créditos individuais) em um determinado horizonte de tempo e, em segunda instância, a quantificação de riscos marginais. Prado, Bastos e Duarte (2000) apresentam uma aplicação desta segunda visão de utilização, cujo modelo de risco é segmentado por classificação de *rating* dos clientes/operações, controlando as possíveis concentrações de risco, estabelecendo limites de crédito, precificação adequada dos ativos e melhor avaliação de performance da carteira.

Diversos modelos foram desenvolvidos, porém os mais difundidos foram: CreditMetrics, KMV, CreditPortfolioView e o CreditRisk+. Uma análise comparativa entre estes modelos pode ser obtida em Crouhy, Galai e Mark (2000); a utilização e as limitações de cada um são apresentadas em Servigny e Renault (2004).

3. Metodologia

Os modelos propostos neste trabalho são baseados em uma classe de modelos estatísticos chamada GAMLSS¹ (Modelos Aditivos Generalizados para Posição, Escala e Forma).

Esta classe é uma evolução da classe GLM² (Modelos Lineares Generalizados) que foi introduzida por Nelder e Wedderburn (1972) e consiste basicamente em uma teoria de regressão para variáveis dependentes da família exponencial de distribuições (ver Spanos, 1999, por exemplo). Em GLM o modelo de regressão é dado por

$$E(y|x) = \mu = g(x^T\beta),$$

sendo y a variável dependente, x um vetor de variáveis independentes, β um vetor paramétrico e $g(\cdot)$ uma função de ligação³. Exemplos de distribuições e sugestões de funções de ligação podem ser encontrados em Lindsey (2005).

A classe de modelos de regressão GAMLSS foi proposta por Rigby e Stasinopoulos (2005). Nessa classe, a suposição da distribuição de probabilidade ser da família exponencial é relaxada e uma família mais geral é utilizada. Em princípio, exige-se que a função densidade de probabilidade (ou função de probabilidade) da variável dependente seja duplamente diferenciável nos parâmetros. A classe de modelos GAMLSS permite a construção de modelos de regressão para os parâmetros de muitas distribuições, tais como Binomial, Poisson, Binomial Negativa, Exponencial, Normal, Gamma, Gumbel, Weibull, ZIP, ZAIG, BEINF, por exemplo. Nesta abordagem, é possível modelar separadamente todos os parâmetros da distribuição como função de variáveis explicativas. Um componente aditivo não paramétrico pode ser incluído, porém o estudo dessa possibilidade não será objeto deste trabalho. Admita, por exemplo, que uma distribuição tenha um parâmetro de locação (μ), um parâmetro de escala (σ) e dois parâmetros de forma (ν e τ), a metodologia proposta permite estimar os seguintes modelos simultaneamente:

$$g_1(\mu) = x_1^T \beta_1 \quad g_2(\sigma) = x_2^T \beta_2 \quad g_3(\nu) = x_3^T \beta_3 \quad g_4(\tau) = x_4^T \beta_4,$$

¹ GAMLSS – Generalized Additive Models for Location, Scale and Shape

² GLM – Generalized Linear Models

³ Contínua, inversível e duplamente diferenciável.

sendo g_i funções de ligação, x_i vetores de variáveis independentes e β_i vetores paramétricos.

A estimação, nos casos tratados neste trabalho, é feita pelo método da máxima verossimilhança. Rigby e Stasinopoulos (2005) descrevem e sugerem o emprego de dois diferentes algoritmos de estimação: algoritmo RS, usado por Rigby e Stasinopoulos (1996) e algoritmo CG, dado por Cole e Green (1992).

Neste trabalho foram utilizadas duas distribuições específicas desta classe, ZAIG (Heller, Stasinopoulos e Rigby, 2006) e BEINF (ver Ospina e Ferrari, 2007, por exemplo), que possuem características importantes para o desenvolvimento do estudo. Essas distribuições podem ser obtidas a partir de misturas de variáveis aleatórias, uma contínua e outra discreta. Na aplicação, a distribuição contínua representa a perda em caso de inadimplência e a discreta relaciona-se à probabilidade de não haver perda (perda igual a zero) na classe ZAIG e as probabilidades de perda igual a zero e de perda total na BEINF. Nas próximas seções são apresentados alguns detalhes sobre essas distribuições.

3.1 Modelo ZAIG

O modelo ZAIG⁴ (Normal Inversa ajustada para o zero) caracteriza-se por permitir a existência de uma probabilidade positiva da variável resposta assumir o valor zero e, para valores maiores do que zero, que a variável dependente comporte-se como uma distribuição normal inversa, que é uma distribuição da família exponencial fortemente assimétrica à direita. Jong e Heller (2008) apresentam uma aplicação dessa distribuição na modelagem de sinistros de uma seguradora. O modelo ZAIG pode ser caracterizado por meio de uma mistura de distribuições.

Seja y_i o valor da perda ocorrida para indivíduo i . A função de probabilidade que representa o modelo é dada por:

$$f(y_i) = \begin{cases} v_i & , \text{ se } y_i = 0 \\ (1-v_i) h(y_i) & , \text{ se } y_i > 0 \end{cases}$$

⁴ ZAIG – Zero Adjusted Inverse Gaussian

na qual $v_i = 1 - \pi_i$ é a probabilidade de não ocorrência de perda e, portanto, π_i é a probabilidade de ocorrência da inadimplência e $0 \leq v_i, \pi_i \leq 1$; $h(y_i)$ é a função de densidade de probabilidade da Normal inversa, aplicada aos valores positivos de perda, sendo dada por:

$$h(y_i) = \frac{1}{\sqrt{2\pi y_i^3} \sigma_i} \exp \left[-\frac{1}{2y_i} \left(\frac{y_i - \mu_i}{\mu_i \sigma_i} \right)^2 \right],$$

sendo μ_i e $\sigma_i > 0$.

A média e a variância desta distribuição são dadas por: $E(y_i) = \pi_i \mu_i$ e $Var(y_i) = \pi_i \mu_i^2 (1 - \pi_i + \mu_i \sigma_i^2)$.

A estimação dos parâmetros da distribuição é realizada a partir do método da máxima verossimilhança.

A maximização da função logaritmo da verossimilhança pode ser feita pelos algoritmos RS ou CG, citados anteriormente.

No contexto desta dissertação, o modelo ZAIG será utilizado com o objetivo de representar dois fenômenos:

- 1) A probabilidade de ocorrência da inadimplência, ou seja, a chance de um determinado indivíduo gerar perdas pela falta de pagamento do financiamento;
- 2) A perda esperada em caso de inadimplência, ou seja, dada a ocorrência da inadimplência, qual o valor em reais da perda.

Estes dois fenômenos são modelados em função de variáveis explicativas atuando sobre os parâmetros μ (posição), σ (escala) e v (forma). As variáveis explicativas são incorporadas ao modelo através das funções de ligações. Utilizamos uma função logito para v e logarítmicas para μ e σ , conforme segue abaixo:

$$\begin{aligned} \log(\mu_i) &= \alpha_\mu + \mathbf{x}_{\mu_i}^T \beta_\mu \\ \log(\sigma_i) &= \alpha_\sigma + \mathbf{x}_{\sigma_i}^T \beta_\sigma \\ \log\left(\frac{v_i}{1-v_i}\right) &= \alpha_v + \mathbf{x}_{v_i}^T \beta_v \end{aligned} \quad (3.1)$$

em que x_{μ_i} , x_{σ_i} e x_{ν_i} são os vetores de variáveis explicativas para μ , σ e ν , respectivamente, sendo que estes podem ser distintos para cada parâmetro; β_{μ} , β_{σ} e β_{ν} são vetores paramétricos com os coeficientes a serem estimados para cada variável; α_{μ} , α_{σ} e α_{ν} são as constantes de cada modelo.

O modelo estimado consiste de três equações diferentes, uma para cada parâmetro da distribuição (μ , σ e ν), sendo que duas delas têm interpretação econômica: a equação de ν , que explica a ocorrência ou não da inadimplência e a equação de μ , que explica o valor de perda esperada dada a ocorrência da inadimplência. A equação de σ serve, dentre outras coisas, para controlar uma eventual heterocedasticidade.

3.2 Modelo BEINF

A distribuição Beta é usada para modelagem de dados em forma de frações, proporções ou porcentagens que estejam no intervalo contínuo e aberto entre zero e um. Contudo, esta distribuição não é adequada para modelar dados quando há grande concentração de zeros, de uns ou de ambos. Para corrigir este problema, utiliza-se uma distribuição misturada conforme descrita em Ospina e Ferrari (2007). Os autores sugerem o uso da distribuição Beta para modelar a variável resposta quando esta se encontra no intervalo contínuo e aberto (0,1) e distribuições de Bernoulli nos casos em que a variável assume os valores zero e um (se for o caso). Estes modelos misturados são conhecidos por *Inflated*, sugerindo a existência de observações com valores iguais a zero, um ou ambos, o que não é esperado numa distribuição Beta.

Nesta classe têm-se três modelos:

BEZI⁶ – Beta Inflada de Zeros, quando se tem uma massa de observações iguais a zero.

BEOI⁷ – Beta Inflada de Uns, quando se tem uma massa de observações iguais a um.

⁶ BEZI – Zero-Inflated Beta

⁷ BEOI – One-Inflated Beta

BEINF⁸ – Beta *Inflated*, quando há duas concentrações de observações: uma em valores iguais a zero e outra em valores iguais a um.

Para os dados utilizados neste trabalho utilizou-se a BEINF, para modelar o percentual do valor da perda em relação ao valor emprestado, pois a carteira apresenta uma grande concentração de observações iguais a zero, que representa os clientes que pagaram integralmente o empréstimo e outra bem menor, porém significativa, de valores iguais a um, que representa clientes que apresentaram perdas de 100% do valor emprestado, ou seja, em que ocorreu perda total.

A utilização desta distribuição tem por objetivo:

- 1) modelar a probabilidade de inadimplência;
- 2) modelar o percentual de perda observada nos casos em que ocorreu inadimplência;
- 3) modelar a probabilidade de indivíduos com perda total, ou seja, 100% de perda observada, representada pela massa de observações com valores iguais a um.

A função de probabilidade que representa o modelo é dada por:

$$f(y_i) = \begin{cases} \pi_{0i} & , y_i = 0 \\ (1 - \pi_{0i} - \pi_{1i}) h(y_i) & , y_i \in (0,1) \\ \pi_{1i} & , y_i = 1 \end{cases}$$

sendo: y_i o percentual de perda ocorrida para indivíduo i ; π_{0i} a probabilidade de não ocorrência de perdas para este indivíduo, dada por:

$$\pi_{0i} = \frac{v_i}{(1 + v_i + \tau_i)}, \quad v_i, \tau_i \text{ são parâmetros de forma, sendo } v_i, \tau_i > 0;$$

π_{1i} a probabilidade de ocorrência de perda total, dada por:

$$\pi_{1i} = \frac{\tau_i}{(1 + v_i + \tau_i)} e$$

$h(y_i)$ é a função de densidade da Beta para os valores positivos (menores que um) de percentual de perda, mas diferentes de 100%. Esta função pode ser dada por:

$$h(y_i) = \frac{\Gamma\left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right)}{\Gamma\left(\frac{\mu_i(1 - \sigma_i^2)}{\sigma_i^2}\right) \Gamma\left(\frac{(1 - \mu_i)(1 - \sigma_i^2)}{\sigma_i^2}\right)} y_i^{\frac{\mu_i(1 - \sigma_i^2)}{\sigma_i^2} - 1} (1 - y_i)^{\frac{(1 - \mu_i)(1 - \sigma_i^2)}{\sigma_i^2} - 1}$$

⁸ BEINF – Zero and one-Inflated Beta, Beta Inflated

na qual, $\Gamma(\cdot)$ é a função *gama* dada por $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$; μ é o parâmetro de posição (média), sendo $0 < \mu_i < 1$; σ é o parâmetro de escala, sendo $0 < \sigma_i < 1$. Por meio dessa parametrização da Beta, temos que, dado que $y_i \in (0,1)$, $E(y_i) = \mu_i$ e $Var(y_i) = \sigma_i^2 \mu_i (1 - \mu_i)$.

A estimação dos parâmetros é realizada pelo método de máxima verossimilhança. A função de verossimilhança pode ser fatorada em três termos (ver detalhes em Ospina e Ferrari, 2007), sendo dada por:

$$L(\theta) = \prod_{i=1}^n f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i) = L_1(\theta_1) L_2(\theta_2) L_3(\theta_3).$$

sendo θ o conjunto de todos os parâmetros a serem estimados; θ_1 o conjunto de parâmetros relativos à ν_i , θ_2 relativos à τ_i e θ_3 relativos à μ_i e σ_i .

Assim como mencionado no modelo ZAIG o estimador de máxima verossimilhança pode ser encontrado por meio dos algoritmos RS ou CG.

A média e a variância da distribuição BEINF são dadas por:

$$E(y_i) = (1 - \pi_{0i} - \pi_{1i}) \mu_i + \pi_{1i} \text{ e}$$

$$Var(y_i) = \frac{\pi_{0i} \pi_{1i}}{\pi_{0i} + \pi_{1i}} + (1 - \pi_{0i} - \pi_{1i}) \left[\sigma^2 (1 - \mu_i) \mu_i + (\pi_{0i} + \pi_{1i}) \left(\frac{\pi_{1i}}{\pi_{0i} + \pi_{1i}} - \mu_i \right)^2 \right].$$

Uma diferença importante entre os modelos propostos, ZAIG e BEINF, está na variável dependente, que no primeiro caso representa valores de perdas em Reais e no segundo o percentual de perda ocorrida em relação ao valor total emprestado. Além disso, o modelo BEINF, utiliza um parâmetro a mais, τ , que tem uma influência direta na probabilidade de um indivíduo apresentar perda total (π_1).

Estes parâmetros são modelados em função de variáveis explicativas que são incorporadas através das funções de ligações, sendo uma função logito para μ e σ e logarítmica para ν e τ , conforme segue:

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha_{\mu} + \mathbf{x}_{\mu_i}^T \beta_{\mu}$$

$$\log\left(\frac{\sigma_i}{1-\sigma_i}\right) = \alpha_\sigma + \mathbf{x}_{\sigma_i}^T \beta_\sigma \quad (3.2)$$

$$\log(v_i) = \alpha_\nu + \mathbf{x}_{v_i}^T \beta_\nu$$

$$\log(\tau_i) = \alpha_\tau + \mathbf{x}_{\tau_i}^T \beta_\tau$$

em que \mathbf{x}_{μ_i} , \mathbf{x}_{σ_i} , \mathbf{x}_{v_i} e \mathbf{x}_{τ_i} são os conjuntos de variáveis explicativas para μ , σ , ν e τ , respectivamente, sendo que estes podem ser distintos para cada parâmetro; β_μ , β_σ , β_ν e β_τ são os coeficientes a serem estimados para cada variável; α_μ , α_σ , α_ν e α_τ são as constantes de cada modelo. Três das equações acima têm interpretação econômica: a equação de ν , está ligada à probabilidade de não ocorrência de inadimplência, a equação de μ , explica o percentual de perda esperada dada a ocorrência da inadimplência e a equação de τ , ligada à ocorrência de indivíduos com perda total do valor emprestado. A equação de σ serve para modelar a variância da distribuição.

3.3 Medidas de comparação dos modelos propostos

Para escolha do melhor modelo dentro de cada uma das distribuições (ZAIG e BEINF), utilizou-se o Critério de Informação de Akaike Generalizado (GAIC⁹) (ver Akaike, 1983 e Mittelhammer, Judge e Miller, 2000) e o melhor modelo será o que apresentar o menor valor desta estatística, que é dada por:

$$GAIC(t) = -2 \sum_{i=1}^n l(\hat{\mu}_i, \hat{\sigma}_i, \hat{\nu}_i, \hat{\tau}_i) + t \, df$$

sendo t a penalização que permite a generalização do critério, n o número de parâmetros estimados e df o número de graus de liberdade residual do modelo.

Pela fórmula acima, vê-se que quanto maior a penalização, maior é a sensibilidade da estatística a inclusão de novas variáveis no modelo. A escolha de qual deve ser a penalização está ligada aos objetivos do modelo estimado.

Neste trabalho, deu-se preferência a $t = 2$ para a comparação, que é o padrão utilizado pelo pacote estatístico adotado (ver Capítulo 5) e que representa o AIC original

⁹ GAIC – Generalized Akaike Information Criterion

estabelecido por Akaike (1974), pois não há qualquer restrição quanto a inclusão de novas variáveis nos modelos propostos.

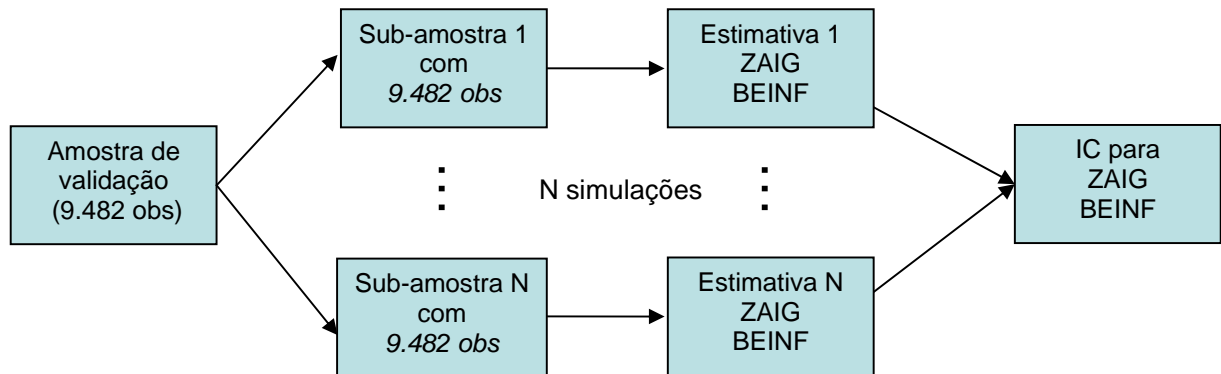
Paralelamente, utilizou-se também este mesmo critério com diferentes penalizações e o Critério de Informação Bayesiano de Schwarz (SBC¹⁰). O SBC também é um caso especial do GAIC, onde $t = \log(n)$.

Escolhido o melhor modelo de cada distribuição, partiu-se para a comparação entre as duas distribuições a fim de se definir a que melhor se ajusta ao problema proposto no trabalho. Para tanto, utilizou-se o método *Bootstrap* (ver Spanos, 1999), processo de reamostragem em que se retiram sucessivas sub-amostras com reposição de uma dada amostra, conforme apresentado no Quadro 3.1.

Seguiram-se os passos abaixo, aplicados na amostra de validação separada para a comparação:

- Aplicação dos modelos selecionados, que representam as equações (3.1) e (3.2), em cada indivíduo da amostra de validação, calculando-se a probabilidade de inadimplência e a perda esperada em caso de inadimplência;
- Geração de 5.000 sub-amostras com reposição, obtidas a partir das observações da amostra definida. Cada sub-amostra possui 9.482 observações, que é o tamanho da amostra de validação;
- Soma do valor total das perdas estimadas pelos modelos em cada uma das sub-amostras selecionadas;
- A partir das 5.000 observações de perda estimada em cada modelo, determina-se o intervalo de confiança das estimativas;
- Comparar os resultados obtidos para os modelos ZAIG e BEINF.

¹⁰ SBC – Schwarz Bayesian Information Criterion

Quadro 3.1 – Diagrama ilustrativo do método *Bootstrap*

4. Descrição dos dados

Neste capítulo são apresentados os dados referentes à carteira de empréstimos analisada nesta dissertação. Foram consideradas variáveis sobre as características dos clientes e da operação de empréstimo, além de informações sobre os pagamentos efetuados por estes, que definem sua condição de bom cliente ou inadimplente.

Inadimplência é definida como a ausência de pagamento de qualquer uma das parcelas ao longo do contrato sem retomada dos pagamentos.

4.1 A carteira

A carteira selecionada¹¹ para a aplicação refere-se a 47.412 financiamentos para aquisição de bens de consumo no varejo, em que todas as parcelas já estão vencidas, isto é, a carteira está encerrada e, portanto, bem definida em relação à inadimplência. Esta carteira foi dividida aleatoriamente em 37.930 observações para desenvolvimento dos modelos e 9.482 para validação dos mesmos, o que representam 80% e 20% do total da carteira respectivamente.

4.2 Variáveis utilizadas

4.2.1 Variáveis de comportamento

Utilizaram-se as seguintes variáveis que refletem o comportamento do cliente em relação ao pagamento do empréstimo:

PCI – Perda em Caso de Inadimplência: refere-se ao valor principal, em R\$, não recebido em decorrência da inadimplência do cliente, apresentando valores entre zero (quando não há ocorrência de inadimplência) e o valor total do empréstimo concedido (quando nenhum pagamento é efetuado pelo cliente).

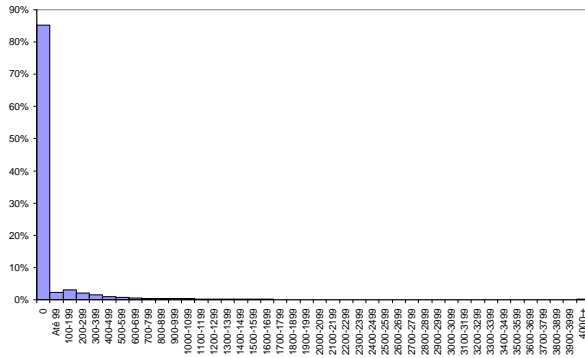
PPCI – Percentual de Perda em Caso de Inadimplência: é a razão entre a PCI e o valor emprestado, apresentando valores entre 0 e 100%, que representam a não ocorrência da inadimplência e a falta total de pagamento das parcelas, respectivamente.

¹¹ Por questões de sigilo, algumas informações da carteira foram omitidas.

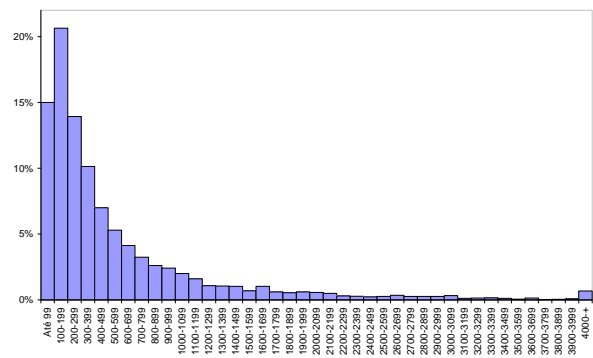
Os Gráficos 4.1 e 4.2 trazem, respectivamente, as distribuições de frequência das variáveis PCI e PPCI.

Gráfico 4.1 – Distribuição da variável PCI

(A) Todas as observações



(B) Valores maiores que zero

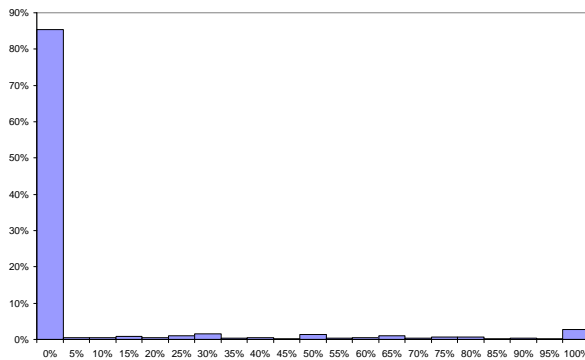


Vê-se claramente, no Gráfico 4.1A, uma forte assimetria à direita da distribuição de PCI e a grande concentração de observações com valor igual a zero, representando cerca de 85% do total.

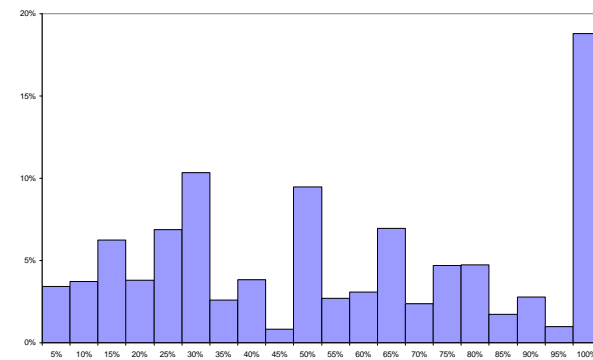
No Gráfico 4.1B observa-se a mesma distribuição sem a grande massa de zeros. Nota-se grande concentração entre os valores R\$ 100 e R\$ 199 e queda rápida até R\$ 1.200. O valor médio de financiamento da carteira é R\$ 665 e seu desvio-padrão é de R\$ 833.

Gráfico 4.2 – Distribuição da variável PPCI

(A) Todas as observações



(B) Valores maiores que zero



A distribuição de PPCI, representada no Gráfico 4.2A, apresenta características semelhantes à distribuição de PCI: forte assimetria à direita e a grande massa de zeros, representando 85% do total.

No Gráfico 4.2B tem-se a mesma distribuição sem a grande massa de zeros, na qual se observa uma concentração de valores iguais a um e o valor médio de PPCI é 56,6%. Verifica-se comportamento distinto ao observado para PCI: há uma alta ocorrência de valores iguais a 100%, ou seja, a distribuição de PPCI apresenta duas concentrações de valores – uma, maior, no 0% e uma, menor, no 100%.

A Tabela 4.1 apresenta estatísticas gerais da carteira segmentadas pelas amostras de modelagem e validação.

Tabela 4.1 – Estatísticas gerais da carteira

Estatísticas	Amostra		Total geral
	Modelagem	Validação	
Clientes	37.930	9.482	47.412
Inadimplentes	5.586	1.419	7.005
% Inadimplentes	14,70%	14,97%	14,77%
Valor Financiado	25.259.446	6.352.336	31.611.782
PCI Total	3.181.148	814.265	3.995.413
PCI médio*	569	574	570
DP PCI*	765	853	783
PPCI médio*	56,6%	58,4%	57,0%
DP PPCI*	30,6%	30,6%	30,6%

Considerando-se apenas os clientes inadimplentes

A inadimplência total da carteira avaliada é de 14,77%, o que equivale a 7.005 clientes que apresentaram faltas de pagamentos e conseqüentemente gerando perdas que totalizaram R\$ 3,995 milhões para R\$ 31,611 milhões de valor emprestado. Os valores de PCI e PPCI médios e seus respectivos desvios-padrão são calculados a partir dos clientes inadimplentes, não contemplando, portanto, valores iguais a zero.

4.2.2 Variáveis do cliente

As seguintes variáveis do cliente foram utilizadas:

CS - Credit Score¹² – classificação de risco de inadimplência, dado em classes - foram considerados 10 níveis de risco, sendo o nível 1 o de maior proporção de clientes inadimplentes que decresce até o nível 10 com a menor proporção.

ID - Idade – idade do cliente no momento da concessão, em anos. Espera-se uma relação decrescente com a inadimplência, ou seja, quanto maior a idade do cliente menor seria a inadimplência observada.

REN - Renda – renda mensal total do cliente, em R\$. A relação com a inadimplência também pode ser decrescente.

CR - Comprometimento de renda – proporção da renda que é comprometida com a prestação do financiamento, dado em percentual. Quanto maior o comprometimento da renda maior pode ser a chance do cliente tornar-se inadimplente.

CEP - região do domicílio do cliente. A inadimplência varia conforme a região de residência do cliente.

Todas as variáveis acima foram categorizadas conforme apresentado no Apêndice A.

4.2.3 Variáveis da operação

As seguintes variáveis da operação foram utilizadas:

VF - Valor Financiado – valor emprestado para a compra do produto negociado dado em R\$. Quanto maior o valor financiado, maior tende a ser a perda em caso de inadimplência.

logVF - Log (Valor Financiado)¹³ – Logaritmo do valor financiado. Esta variável foi selecionada a fim de ajustar a estimativa do valor de perda.

TX - Taxa de juros – taxa ao mês aplicada ao financiamento. Apresenta comportamento crescente com a inadimplência, ou seja, quanto maior a taxa de juros, maior tende a ser a inadimplência.

PZ - Prazo – número de parcelas com vencimento mensal. Também tem relação crescente, ou seja, quanto mais longo o prazo de pagamento do financiamento, maior pode ser a chance de inadimplência.

¹² Modelo obtido através de regressão logística, conforme descrito no item 2.2.

¹³ Foi utilizada apenas no modelo ZAIG e não foi categorizada.

PRD - Produto¹⁴ – tipo de produto adquirido no financiamento. A inadimplência varia muito conforme o objetivo da compra, ou seja, com o produto adquirido.

Todas as variáveis acima, exceto logVF, foram categorizadas conforme apresentado no Apêndice A.

4.2.4 Variáveis de interação

Foram utilizadas cinco variáveis de interação. Essa escolha baseou-se na expectativa de que o efeito resultante da combinação dessas variáveis seja maior do que o efeito da presença dessas variáveis individualmente. Nas variáveis escolhidas, a associação entre elas pode potencializar (ou reduzir) a inadimplência, ou seja, duas características que aumentam a chance de inadimplência separadamente, quando associadas podem aumentá-la (ou diminuí-la) a níveis muito maiores (menores).

As variáveis definidas são:

VF x PZ – Interação entre Valor Financiado (VF) e Prazo (PZ). Quanto mais longo o prazo para pagamento e maior o valor financiado, maior a chance de inadimplência.

VF x CS – Interação entre Valor Financiado (VF) e *Credit Score* (CS). O efeito positivo em relação à chance de inadimplência dos níveis mais altos do Credit Score pode ser anulado pelo efeito negativo do maior valor financiado, sendo que o último deve prevalecer, ou seja, aumenta a chance de inadimplência.

VF x REN – Interação entre Valor Financiado (VF) e Renda (REN). O efeito positivo dos valores mais altos da renda pode ser anulado pelo efeito negativo do maior valor financiado, sendo que o último deve prevalecer.

PZ x CS – Interação entre Prazo (PZ) e *Credit Score* (CS). O efeito positivo dos níveis mais altos do Credit Score pode ser anulado pelo efeito negativo do prazo mais longo, sendo que o último prevalece.

CR x REN – Interação entre Comprometimento de Renda (CR) e Renda (REN). O efeito positivo dos valores mais altos da renda pode ser anulado pelo efeito negativo do comprometimento maior, apresentando resultados variáveis.

¹⁴ Por questões de sigilo, não foi possível descrever os referidos produtos.

4.3 Análise descritiva

A fim de conhecer as relações existentes entre as variáveis explicativas, inadimplência, PCI e PPCI, realizaram-se análises descritivas com cada uma delas, conforme apresentadas no Apêndice B.

Para aplicação dos modelos propostos neste estudo, cada categoria das variáveis apresentadas foi transformada em uma variável indicadora (*dummy*), que assume valor 1 (um) na ocorrência da categoria em questão e 0 (zero) caso contrário. Este artifício é utilizado a fim de incorporar o efeito não linear das variáveis estudadas nos modelos. É necessário excluir uma das categorias de cada variável para evitar o efeito de multicolinearidade, dado que uma categoria é combinação linear das demais. Como critério de exclusão, optou-se pelas categorias com maior percentual de clientes inadimplentes, porém outros critérios poderiam ser usados sem comprometimento do estudo proposto. As categorias excluídas são apresentadas no Apêndice A.

5. Resultados

Utilizou-se, nos procedimentos computacionais de estimação, o pacote *GAMLSS*, desenvolvido por Rigby e Stasinopoulos (manual disponível em Stasinopoulos, Rigby e Akantziliotou, 2008), escrito para o *software* livre R¹⁵ (*software* destinado à computação estatística). As macros elaboradas para a estimação dos modelos e análise dos resultados estão disponíveis no Apêndice C.

5.1 Modelo ZAIG

Nesta seção foram ajustados os modelos descritos pelas equações dadas em (3.1).

Primeiramente gerou-se um modelo básico (Modelo 1) sem utilização de variáveis explicativas, obtendo-se estimativas diretas para os parâmetros da distribuição aplicada à modelagem dos valores de perda. Esse modelo será considerado como referência na comparação com outros modelos. As estimativas são apresentadas na Tabela 5.1 do trabalho.

Tabela 5.1 – Modelo sem variáveis explicativas

Parâmetros	Estimativas	Erro Padrão	Estat t	P-Valor
μ	6,34500	0,0190	334,70	0,0%
σ	(2,82400)	0,0095	(298,50)	0,0%
π	1,75600	0,0145	121,20	0,0%

Estabeleceram-se três modelos iniciais que utilizaram todas as variáveis explicativas apresentadas no Capítulo 4:

Modelo 2) Sem interações: utiliza todas as variáveis explicativas, exceto as variáveis de interação;

Modelo 3) Completo: utiliza todas as variáveis explicativas, acrescentando as de interação, com o objetivo de verificar o ganho proporcionado pela associação existente entre as variáveis explicativas do trabalho;

¹⁵ www.r-project.org

Modelo 4) Completo com *offset*: acrescenta ao modelo anterior a variável $\log VF$ com coeficiente fixo e igual a 1, conforme descrito abaixo

$$\log(\mu_i) = \alpha_\mu + \log VF + x_{\mu_i}^T \beta_\mu \quad (5.1)$$

A inclusão do $\log VF$ se justifica na medida em que o valor da perda está limitado ao valor do empréstimo. Desse modo, a inclusão dessa variável pode trazer um efeito positivo à qualidade do modelo. A idéia do uso do *offset* vem da análise da variável $PPCI$, empregada no modelo BEINF. Essa variável é definida por $PPCI = \frac{PCI}{VF}$ e aplicando-se o logaritmo em ambos os lados da equação temos:

$$\log(PPCI) = \log\left(\frac{PCI}{VF}\right) = \log(PCI) - \log(VF).$$

A seleção dos modelos é feita por meio do menor valor da estatística calculada pelo critério GAIC com penalização igual a 2, conforme descrito no item 3.3. Segundo a Tabela 5.2, verifica-se que:

- a utilização de variáveis explicativas melhora significativamente a qualidade do modelo, dada a queda expressiva da estatística GAIC2 ao se comparar o Modelo 2 com o Modelo 1;
- a introdução de interações melhora o modelo, porém de forma menos intensa, quando se compara o Modelo 3 com o 2;
- a introdução do termo *offset* do $\log VF$ (Modelo 4), apresenta o menor valor de GAIC2 entre as quatro opções definidas.

A partir do modelo completo com *offset*, utilizou-se o método *stepwise* disponível no pacote apresentado, que retira sucessivamente as variáveis menos significativas de modo a se chegar a um modelo com menor valor da estatística GAIC (com penalidade 2) entre todos os modelos possíveis. Este método foi repetido para os modelos de cada um dos parâmetros, isto é, μ , σ e ν , nesta ordem. Como se vê na Tabela 5.2, o modelo que apresenta o menor GAIC2, apesar de não ser o objetivo da seleção, também apresentou valores menores para as estatísticas de GAIC com penalidades iguais a 3 e a 5, além da estatística SBC.

Tabela 5.2 – Estatísticas de seleção dos modelos

Modelo	GAIC2	GAIC3	GAIC5	SBC
Modelo 1 - Sem var explicativas	112.517	112.520	112.526	112.543
Modelo 2 - Completo sem interações	101.228	101.435	101.849	102.996
Modelo 3 - Completo com interações	101.006	101.318	101.945	103.671
Modelo 4 - Completo com <i>offset</i>	100.837	101.149	101.776	103.511
Modelo Selecionado	100.739	100.828	101.006	101.499

Os modelos selecionados para μ , σ e ν são apresentados a seguir:

Tabela 5.3 – Modelo selecionado para μ

Variáveis	Coeficientes	Erro Padrão	Estat t	P-Valor
Intercepto	(0,58644)	0,0134	(43,76)	0,0%
logVF	1,00000	-	-	-
PZ_CS1	0,41755	0,0271	15,41	0,0%
PZ_CS3	0,09772	0,0263	3,71	0,0%
PZ_CS4	0,06096	0,0342	1,78	7,5%
PZ_CS5	0,04823	0,0281	1,72	8,6%
CR_REN7	0,05137	0,0290	1,77	7,7%
CR_REN8	0,10045	0,0348	2,89	0,4%

Notação disponível no Apêndice A

Verifica-se que poucas variáveis são significantes para representar a perda média em caso de inadimplência: apenas alguns níveis da variável de interação Prazo x *Credit Score* e da variável Comprometimento de renda x Renda.

Pelas estatísticas descritivas apresentadas no Apêndice B, grande parte das variáveis aparenta ter relação com o valor de perda esperada. Contudo, com a utilização da variável logVF, criou-se uma referência entre o valor financiado e o valor da perda, pois clientes com valores financiados maiores tendem a apresentar maiores valores de perdas. Esta associação também ocorre entre logVF e muitas das variáveis independentes como, por exemplo, Renda, em que clientes com maiores rendas tendem a ter maiores valores financiados e, como mencionado anteriormente, tendem a ter maiores valores de perda. Desta forma, muito da explicação de uma série de variáveis independentes já está embutida em logVF, o que permite a exclusão dessas variáveis do modelo final.

Tem-se, da Tabela 5.3, que a relação existente entre a perda esperada e a variável PZ x CS é positiva e decrescente, ou seja, para níveis mais baixos de PZ_CS (PZ_CS1, etc.) maior é a perda esperada.

A variável CR_REN é significativa em apenas dois níveis: CR_REN7 e CR_REN8, sendo os coeficientes positivos e crescentes, o que está em concordância com a Tabela B.15.

Para o modelo de σ também se utilizou a variável logVF, porém como variável do modelo (sem utilização do recurso *offset*) por apresentar melhor resultado.

Tabela 5.4 – Modelo selecionado para σ

Variáveis	Coefficientes	Erro Padrão	Estat t	P-Valor
Intercepto	1,56675	0,4151	3,77	0,0%
logVF	(0,67044)	0,0527	(12,73)	0,0%
CS9	(0,46985)	0,0759	(6,19)	0,0%
CS10	(0,79942)	0,1065	(7,51)	0,0%
VFm200	(0,58711)	0,1551	(3,79)	0,0%
VF200	(0,66203)	0,1235	(5,36)	0,0%
VF300	(0,58851)	0,1074	(5,48)	0,0%
VF400	(0,53385)	0,0965	(5,53)	0,0%
VF500	(0,47670)	0,0885	(5,39)	0,0%
VF600	(0,36387)	0,0778	(4,68)	0,0%
VF750	(0,26572)	0,0661	(4,02)	0,0%
VF1000	(0,18547)	0,0496	(3,74)	0,0%
REN4500	0,14398	0,0590	2,44	1,5%
VF_PZ1	(0,63261)	0,1216	(5,20)	0,0%
VF_PZ2	(0,40933)	0,0837	(4,89)	0,0%
VF_PZ3	(0,22188)	0,0518	(4,29)	0,0%
VF_REN1	0,27487	0,0718	3,83	0,0%
VF_REN2	0,29314	0,0638	4,60	0,0%
VF_REN3	0,22945	0,0604	3,80	0,0%
VF_REN4	0,15267	0,0565	2,70	0,7%
VF_REN5	0,15107	0,0524	2,88	0,4%
VF_REN6	0,08307	0,0487	1,71	8,8%
VF_REN7	0,09183	0,0418	2,20	2,8%
PZ_CS1	(0,82024)	0,0809	(10,14)	0,0%
PZ_CS3	(0,53625)	0,0475	(11,29)	0,0%
PZ_CS4	(0,37121)	0,0460	(8,07)	0,0%
PZ_CS5	(0,43499)	0,0401	(10,84)	0,0%
PZ_CS6	(0,24696)	0,0345	(7,15)	0,0%
PZ_CS7	(0,17729)	0,0303	(5,86)	0,0%

As variáveis de interação são as que aparecem em maior número neste modelo: VF_PZ, VF_REN e PZ_CS nas quais a variabilidade é menor nos níveis inferiores de

cada variável, ou seja, a variabilidade aumenta com o incremento nos níveis destas variáveis.

A variabilidade é menor nos níveis 9 e 10 do *Credit Score* em relação aos demais níveis e a renda acima de R\$ 4.500 apresenta variabilidade significativamente maior em relação aos demais níveis desta variável.

Em relação ao valor financiado, conforme já visto, clientes com valores financiados maiores tendem a apresentar maiores valores de perdas, o que pode gerar maior variabilidade nesta situação. Mantendo-se as variáveis Renda e Prazo fixas, verifica-se que há maior variabilidade da perda para valores financiados maiores.

Tabela 5.5 – Modelo selecionado para ν

Variáveis	Coeficientes	Erro Padrão	Estat t	P-Valor
Intercepto	(2,51520)	0,1890	(13,31)	0,0%
VF400	0,11140	0,0576	1,93	5,3%
VF1000	0,15570	0,0592	2,63	0,9%
VF1500	0,12010	0,0670	1,79	7,3%
PZ1	4,24520	0,1499	28,32	0,0%
PZ3	3,16090	0,1213	26,05	0,0%
PZ6	2,38710	0,1165	20,49	0,0%
PZ9	1,99190	0,1154	17,26	0,0%
PZ12	1,34580	0,1154	11,66	0,0%
PZ18	0,77610	0,1167	6,65	0,0%
ID20	0,55450	0,1444	3,84	0,0%
ID25	0,87630	0,1425	6,15	0,0%
ID30	1,02190	0,1427	7,16	0,0%
ID35	1,22650	0,1444	8,49	0,0%
ID40	1,35220	0,1459	9,27	0,0%
ID45	1,45690	0,1489	9,79	0,0%
ID50	1,50640	0,1548	9,73	0,0%
ID55	1,59800	0,1503	10,63	0,0%
PRD2	(0,89170)	0,0511	(17,45)	0,0%
PRD4	0,47670	0,0755	6,32	0,0%
TX1	0,82600	0,0850	9,72	0,0%
TX2	0,69580	0,0849	8,20	0,0%
TX3	0,68520	0,0834	8,21	0,0%
TX4	0,64330	0,0792	8,13	0,0%
TX5	0,45370	0,0739	6,14	0,0%
TX6	0,21550	0,0681	3,17	0,2%
TX7	0,17480	0,0648	2,70	0,7%
REN1000	(0,14310)	0,0507	(2,82)	0,5%
REN1500	(0,24370)	0,0664	(3,67)	0,0%
REN2000	(0,41400)	0,0871	(4,75)	0,0%
REN2500	(0,55140)	0,1064	(5,18)	0,0%
REN3000	(0,37970)	0,1156	(3,28)	0,1%

(Continuação)

REN3500	(0,48590)	0,1211	(4,01)	0,0%
REN4500	(0,36120)	0,1140	(3,17)	0,2%
Cep1	0,41320	0,1702	2,43	1,5%
Cep3	0,18360	0,0682	2,69	0,7%
Cep5	0,25780	0,0626	4,12	0,0%
Cep6	0,39170	0,0754	5,19	0,0%
Cep7	0,27480	0,0583	4,72	0,0%
Cep8	0,28570	0,0545	5,24	0,0%
Cep10	0,45210	0,0736	6,14	0,0%
CRm5	0,39210	0,0819	4,79	0,0%
CR5	0,21410	0,0532	4,03	0,0%
CR10	0,11540	0,0459	2,51	1,2%
VF_PZ1	0,45250	0,1357	3,33	0,1%
VF_PZ2	0,44700	0,0946	4,73	0,0%
VF_PZ3	0,16450	0,0711	2,31	2,1%
PZ_CS3	0,64210	0,0895	7,17	0,0%
PZ_CS4	0,65880	0,0833	7,91	0,0%
PZ_CS5	0,49740	0,0716	6,95	0,0%
PZ_CS6	0,25880	0,0599	4,32	0,0%
CR_REN1	(0,10140)	0,0602	(1,69)	9,2%
CR_REN6	0,12530	0,0519	2,42	1,6%

Da Tabela 5.5 tem-se:

- 1) Este é o modelo com o maior número de variáveis significantes, porém com pequena participação das variáveis de interação: apenas alguns níveis de VF_PZ, PZ_CS e CR_REN, onde as duas primeiras apresentam maior proporção de clientes sem perdas nos níveis inferiores e a última no nível 1 apresenta resultado contrário ao esperado, porém esta é a variável com o maior valor-p do modelo (9,2%) sendo a de menor significância.
- 2) Das variáveis do cliente e da operação apenas a variável *Credit Score* não apresentou qualquer um de seus níveis significantes para ν , exceto interagindo com Prazo.
- 3) A probabilidade de adimplência aumenta com a idade do cliente.
- 4) As variáveis Taxa e Comprometimento de renda têm relação decrescente com a probabilidade adimplência, ou seja, a probabilidade tende a diminuir com o incremento nos níveis destas variáveis, porém Comprometimento de renda apresenta apenas três níveis significativos.

5) O produto 2 apresenta coeficiente negativo indicando que este nível apresenta probabilidade de adimplência abaixo da média, enquanto o produto 4 tem probabilidade acima da média.

6) A variável CEP apresentou resultados variáveis em relação aos coeficientes, porém não é uma variável ordinal.

7) As variáveis Valor Financiado, Renda também apresentaram resultados variáveis em relação aos coeficientes, sendo que Valor Financiado apresentou apenas três níveis significativos, dos quais dois (VF400 e VF1500) têm valor-p superior a 5%, mostrando menor significância deste resultado. Em relação à Renda os níveis de menor renda não são significativos e todos os coeficientes são negativos, indicando que maiores rendas apresentam menor probabilidade de clientes sem perdas.

Todos os resultados encontrados são coerentes e estão em concordância com as tabelas no Apêndice B, exceto a variável Valor Financiado, que é pouco significativa, e Renda que apresentou comportamento oposto ao esperado, possivelmente devido a multicolinearidade com outra variável.

5.2 Modelo BEINF

Nesta seção foram ajustados os modelos definidos em (3.2).

Como no modelo ZAIG, gerou-se um modelo básico (Modelo A) sem utilização de variáveis explicativas, a fim de estimar os parâmetros da distribuição BEINF. As estimativas são apresentadas na Tabela 5.6.

Tabela 5.6 – Modelo sem variáveis explicativas

Parâmetros	Estimativas	Erro Padrão	Estat t	P-Valor
μ	(0,14110)	0,0139	(10,16)	0,0%
σ	(0,10370)	0,0138	(7,52)	0,0%
π	1,97400	0,0145	136,20	0,0%
τ	(1,41400)	0,0307	(46,06)	0,0%

Estabeleceram-se dois modelos iniciais que utilizaram todas as variáveis explicativas apresentadas no Capítulo 4:

Modelo B) Sem interações: utiliza todas as variáveis explicativas, exceto as variáveis de interação;

Modelo C) Completo: utiliza todas as variáveis explicativas, acrescentando as de interação.

A variável logVF é desnecessária neste modelo, pois a variável dependente já se encontra na forma adequada aos objetivos do trabalho.

Assim como no modelo ZAIG, entre os modelos A, B e C, o melhor é o que apresenta as variáveis de interação (Tabela 5.7). A partir da aplicação do método *stepwise*, chegou-se a um modelo com menor GAIC2. O modelo escolhido apresentou também bons valores para as estatísticas GAIC3, GAIC5 e SBC.

Tabela 5.7 – Estatísticas de seleção dos modelos

Modelo	GAIC2	GAIC3	GAIC5	SBC
Modelo A - Sem var explicativas	36.449	36.453	36.461	36.483
Modelo B - Completo sem interações	29.283	29.514	29.976	31.256
Modelo C - Completo com interações	29.015	29.331	29.963	31.715
Modelo Selecionado	28.799	28.900	29.102	29.662

Os modelos selecionados para μ , σ , ν e τ são apresentados a seguir:

Tabela 5.8 – Modelo selecionado para μ

Variáveis	Coefficientes	Erro Padrão	Estat t	P-Valor
Intercepto	(0,15050)	0,0283	(5,33)	0,0%
PZ3	(0,40630)	0,0489	(8,30)	0,0%
PZ6	(0,24220)	0,0468	(5,17)	0,0%
PZ9	(0,13390)	0,0464	(2,89)	0,4%
PZ_CS1	0,15050	0,0418	3,60	0,0%
PZ_CS2	0,15070	0,0290	5,20	0,0%
PZ_CS3	0,48620	0,0619	7,85	0,0%
PZ_CS4	0,33250	0,0634	5,25	0,0%
PZ_CS5	0,38290	0,0592	6,47	0,0%
PZ_CS6	0,18210	0,0517	3,53	0,0%
PZ_CS7	0,19630	0,0437	4,49	0,0%

Da análise da Tabela 5.8, temos:

1) Poucas variáveis foram significantes: apenas Prazo e a variável de interação Prazo x *Credit Score*.

2) Apenas os prazos menores são significativos e seus coeficientes são negativos e crescentes, o que está em concordância com a análise descritiva (Apêndice B), ou seja, quanto menor o prazo menor é o percentual de perda esperada.

3) Todos os níveis da variável PZ_CS foram significativos, sendo os coeficientes positivos e decrescentes, ou seja, os níveis superiores da variável apresentam menores percentuais de perda.

O modelo para a variabilidade do percentual de perda é dado por:

Tabela 5.9 – Modelo selecionado para σ

Variáveis	Coefficientes	Erro Padrão	Estat t	P-Valor
Intercepto	0,06958	0,0326	2,14	3,3%
CS2	(0,06664)	0,0404	(1,65)	9,9%
CS3	(0,14823)	0,0477	(3,11)	0,2%
CS4	(0,15881)	0,0639	(2,49)	1,3%
CS5	(0,16765)	0,0708	(2,37)	1,8%
CS6	(0,14961)	0,0835	(1,79)	7,3%
CS7	(0,55376)	0,1008	(5,49)	0,0%
CS8	(0,52522)	0,1137	(4,62)	0,0%
CS9	(1,13719)	0,1410	(8,07)	0,0%
CS10	(2,89176)	0,1681	(17,21)	0,0%
VFm200	0,56382	0,1011	5,58	0,0%
ID50	0,13245	0,0616	2,15	3,2%
ID55	0,10480	0,0554	1,89	5,9%
PRD1	0,11817	0,0586	2,02	4,4%
PRD4	0,22157	0,0691	3,21	0,1%
REN2000	(0,13083)	0,0632	(2,07)	3,9%
Cep10	0,15345	0,0602	2,55	1,1%
CR30	(0,15860)	0,0718	(2,21)	2,7%
VF_PZ1	(2,72217)	0,1668	(16,32)	0,0%
VF_PZ2	(1,66877)	0,1198	(13,93)	0,0%
VF_PZ3	(0,88610)	0,0841	(10,54)	0,0%
VF_PZ4	(0,55777)	0,0770	(7,24)	0,0%
VF_PZ5	(0,36883)	0,0635	(5,81)	0,0%
VF_PZ6	(0,30412)	0,0558	(5,45)	0,0%
VF_PZ7	(0,19552)	0,0438	(4,47)	0,0%
CR_REN1	0,28130	0,0614	4,58	0,0%
CR_REN2	0,24967	0,0526	4,75	0,0%
CR_REN3	0,13619	0,0502	2,71	0,7%
CR_REN4	0,12727	0,0425	3,00	0,3%

Temos que para σ :

1) As variáveis *Credit Score* e Valor Financiado x Prazo foram as que apresentaram o maior número de categorias significantes, todas com coeficientes negativos, sendo que

para o Credit Score esta relação foi decrescente, ou seja, quanto maior o *Credit Score* menor é a variabilidade do percentual de perda e para VF_PZ a relação é crescente, ou seja, os níveis superiores da variável apresentam maior variabilidade. Os resultados estão em conformidade com a análise descritiva.

2) A variável CR_REN apresentou apenas quatro níveis com coeficientes positivos e decrescentes.

3) As variáveis Valor Financiado, Idade, Produto, Renda, CEP e Comprometimento de renda apresentaram apenas um ou dois níveis significantes.

O modelo para ν é o que apresenta o maior número de variáveis significativas, assim como ocorreu com o modelo ZAIG, conforme tabela a seguir:

Tabela 5.10 – Modelo selecionado para ν

Variáveis	Coefficientes	Erro Padrão	Estat t	P-Valor
Intercepto	(2,16300)	0,1882	(11,49)	0,0%
VF400	0,11240	0,0574	1,96	5,0%
VF1000	0,18080	0,0574	3,15	0,2%
PZ1	4,14900	0,1544	26,88	0,0%
PZ3	3,03870	0,1201	25,31	0,0%
PZ6	2,22170	0,1152	19,28	0,0%
PZ9	1,83710	0,1144	16,06	0,0%
PZ12	1,21240	0,1146	10,58	0,0%
PZ18	0,64980	0,1159	5,61	0,0%
ID20	0,57280	0,1447	3,96	0,0%
ID25	0,89940	0,1428	6,30	0,0%
ID30	1,05400	0,1430	7,37	0,0%
ID35	1,25530	0,1448	8,67	0,0%
ID40	1,38070	0,1463	9,44	0,0%
ID45	1,49360	0,1494	10,00	0,0%
ID50	1,53020	0,1552	9,86	0,0%
ID55	1,62130	0,1508	10,75	0,0%
PRD2	(0,91570)	0,0513	(17,87)	0,0%
PRD4	0,41360	0,0753	5,50	0,0%
TX1	0,80700	0,0850	9,50	0,0%
TX2	0,68590	0,0848	8,09	0,0%
TX3	0,67970	0,0835	8,14	0,0%
TX4	0,64800	0,0792	8,18	0,0%
TX5	0,45350	0,0740	6,13	0,0%
TX6	0,21580	0,0682	3,17	0,2%
TX7	0,17360	0,0649	2,68	0,7%
REN1000	(0,11510)	0,0500	(2,30)	2,1%
REN1500	(0,20900)	0,0653	(3,20)	0,1%
REN2000	(0,37710)	0,0860	(4,38)	0,0%
REN2500	(0,49760)	0,1053	(4,73)	0,0%

(Continuação)

REN3000	(0,32050)	0,1145	(2,80)	0,5%
REN3500	(0,42580)	0,1195	(3,57)	0,0%
REN4500	(0,29250)	0,1118	(2,62)	0,9%
Cep1	0,38960	0,1701	2,29	2,2%
Cep3	0,16850	0,0683	2,47	1,4%
Cep5	0,23400	0,0625	3,74	0,0%
Cep6	0,37350	0,0754	4,95	0,0%
Cep7	0,25670	0,0583	4,40	0,0%
Cep8	0,26630	0,0545	4,89	0,0%
Cep10	0,43450	0,0737	5,90	0,0%
CRm5	0,31640	0,0777	4,07	0,0%
CR5	0,17060	0,0517	3,30	0,1%
CR10	0,09940	0,0454	2,19	2,9%
VF_PZ1	0,32750	0,1404	2,33	2,0%
VF_PZ2	0,38010	0,0929	4,09	0,0%
VF_PZ3	0,12970	0,0708	1,83	6,7%
PZ_CS1	0,89710	0,1308	6,86	0,0%
PZ_CS3	0,75440	0,0897	8,41	0,0%
PZ_CS4	0,67720	0,0833	8,13	0,0%
PZ_CS5	0,51240	0,0717	7,14	0,0%
PZ_CS6	0,26090	0,0601	4,34	0,0%

Em relação à ν :

- 1) Das variáveis de interação, somente VF_PZ e PZ_CS foram significantes, com coeficientes positivos e decrescentes, ou seja, os níveis superiores destas variáveis apresentam menor probabilidade de clientes sem perda.
- 2) Como no modelo ZAIG apenas a variável *Credit Score* não apresentou qualquer um de seus níveis significantes para ν , entre as demais variáveis.
- 3) As demais variáveis apresentaram comportamento semelhante aos já descritos na análise dos resultados do modelo ZAIG.
- 4) Exceto para a variável Renda, todos os resultados encontrados são coerentes e estão em concordância com a análise descritiva.

No modelo BEINF há um parâmetro a mais, τ , associado à proporção de clientes com perda total, cujo modelo encontra-se na Tabela 5.11.

Poucas variáveis foram significantes neste modelo e apenas *Credit Score* e Prazo x *Credit Score* apresentaram dois níveis. As demais apresentaram apenas um único nível.

Tabela 5.11 – Modelo selecionado para τ

Variáveis	Coefficientes	Erro Padrão	Estat t	P-Valor
Intercepto	(1,39680)	0,0492	(28,37)	0,0%
CS2	(0,36420)	0,0866	(4,21)	0,0%
CS3	(0,37200)	0,0984	(3,78)	0,0%
VF1500	(0,72930)	0,1555	(4,69)	0,0%
PRD4	(0,44150)	0,1735	(2,54)	1,1%
Cep2	0,50130	0,1762	2,84	0,4%
Cep9	0,14810	0,0742	2,00	4,6%
PZ_CS1	1,88210	0,1112	16,93	0,0%
PZ_CS3	0,32460	0,1042	3,12	0,2%
CR_REN8	0,43070	0,1164	3,70	0,0%

5.3 Comparação dos modelos propostos

Nos tópicos anteriores, escolheu-se o melhor modelo de cada uma das distribuições estudadas, segundo o menor valor da estatística dada pelo critério de *Akaike*.

Neste tópico, realizou-se a comparação entre estes dois modelos selecionados a fim de escolher a distribuição que melhor representa os objetivos deste trabalho, que é, em primeira instância, calcular a probabilidade de inadimplência da carteira de crédito escolhida e como objetivo final calcular o percentual de perda esperada para a mesma. Para isto, os modelos foram aplicados nas amostras da seguinte forma:

- ZAIG: Para cada indivíduo da amostra, calcula-se a probabilidade de inadimplência, dada por π (ou seja, $1-\nu$), e o valor de perda esperada em caso de inadimplência (PCI), dado por μ . A perda total esperada da carteira é dada por:

$$P_{Total} = \sum_{i=1}^n \hat{\pi}_i \hat{\mu}_i ,$$

em que n é o número de clientes da amostra;
 π_i é a probabilidade de inadimplência e
 μ_i é a perda em caso de inadimplência.

- BEINF: Para cada indivíduo da amostra, calcula-se a probabilidade de inadimplência, dada por $1-\pi_0$, e o percentual de perda esperada em caso de

inadimplência (PPCI), dado em função de μ e τ . A perda total esperada da carteira é dada por:

$$P_{Total} = \sum_{i=1}^n VF_i * PPCI_i,$$

sendo:

$$PPCI_i = (1 - \hat{\pi}_{0i} - \hat{\pi}_{1i}) \hat{\mu}_i + \hat{\pi}_{1i}$$

com $\hat{\pi}_{0i} = \frac{\hat{V}_i}{(1 + \hat{V}_i + \hat{\tau}_i)}$ e $\hat{\pi}_{1i} = \frac{\hat{\tau}_i}{(1 + \hat{V}_i + \hat{\tau}_i)}$. Logo, tem-se:

$$PPCI_i = \frac{\hat{\mu}_i + \hat{\tau}_i}{(1 + \hat{V}_i + \hat{\tau}_i)}$$

O percentual de perda total esperada é obtido pelo quociente da perda total e valor financiado total.

5.3.1 Resultados gerais

Na Tabela 5.12, tem-se os resultados encontrados a partir dos modelos selecionados, na amostra utilizada para desenvolvimento do modelo:

Tabela 5.12 – Resultados gerais para amostra de modelagem

Modelos	VF	Perda	%PI	%Perda	Δ PI	Δ Perda
Observado	25.259.446	3.181.203	14,70%	12,59%	-	-
ZAIG	25.259.446	3.271.986	14,73%	12,95%	0,20%	2,85%
BEINF	25.259.446	3.145.500	14,69%	12,45%	-0,07%	-1,12%

VF – soma do valor financiado da amostra;

Perda – soma de perda da amostra;

%PI – proporção de clientes inadimplentes;

%Perda – Percentual da perda da amostra, ou seja, Perda em relação a VF;

Δ PI – Variação do PI estimado de cada modelo em relação ao valor observado, isto é:

$$\Delta PI = \frac{\% PI_{Estimado} - \% PI_{Obs}}{\% PI_{Obs}};$$

Δ Perda – Variação do valor da perda estimada por cada modelo em relação ao valor observado, isto é: $\Delta Perda = \frac{Perda_{Estimada} - Perda_{Obs}}{Perda_{Obs}}$.

Vê-se que ambos os modelos apresentaram percentual de inadimplência (dado por %PI) extremamente próximos ao observado na amostra, em que a variação do valor estimado pelo modelo e o valor real observado, dado por ΔPI , é próxima a zero.

Em relação ao valor da perda esperado, o erro das estimativas foi muito baixo, sendo que o modelo ZAIG estimou valor 2,85% maior do que o observado, enquanto o modelo BEINF estimou valor 1,12% menor do que o esperado.

Na Tabela 5.13 apresentam-se os resultados da aplicação dos modelos selecionados na amostra de validação.

Tabela 5.13 – Resultados gerais para amostra de validação

Modelo	VF	Perda	%PI	%Perda	ΔPI	$\Delta Perda$
Observado	6.352.336	814.265	14,97%	12,82%	-	-
ZAIG	6.352.336	854.077	15,02%	13,45%	0,33%	4,89%
BEINF	6.352.336	822.785	14,98%	12,95%	0,07%	1,05%

VF – soma do valor financiado da amostra;

Perda – soma de perda da amostra;

%PI – proporção de clientes inadimplentes;

%Perda – Percentual da perda da amostra, ou seja, Perda em relação a VF;

ΔPI – Variação do PI estimado de cada modelo em relação ao valor observado, isto é:

$$\Delta PI = \frac{\%PI_{Estimado} - \%PI_{Obs}}{\%PI_{Obs}};$$

$\Delta Perda$ – Variação do valor da perda estimada por cada modelo em relação ao valor

observado, isto é: $\Delta Perda = \frac{Perda_{Estimada} - Perda_{Obs}}{Perda_{Obs}}$.

Nesta amostra encontraram-se resultados semelhantes aos descritos anteriormente, em que a variação entre a estimativa do percentual de inadimplência e a inadimplência efetivamente observada é muito baixa, indicando que os modelos são muito eficientes para estimar a probabilidade de inadimplência, e a variação do valor de perda é maior no modelo ZAIG do que no modelo BEINF, indicando que este último é forte candidato a escolha de melhor modelo, pois apresentou variação do valor de perda muito baixa. Apesar da variação do valor de perda maior, a estimativa encontrada para o modelo ZAIG é bastante satisfatória.

Na Tabela 5.14 tem-se uma comparação dos modelos dos parâmetros de cada uma das distribuições.

Tabela 5.14 – Variáveis presentes em cada modelo

Parâmetros	Nos dois modelos	Somente ZAIG	Somente BEINF
μ	PZ_CS1 PZ_CS3 PZ_CS4 PZ_CS5	logVF CR_REN7 CR_REN8	PZ3 PZ_CS2 PZ6 PZ_CS6 PZ9 PZ_CS7
σ	CS9 CS10 VFm200 VF_PZ1 VF_PZ2 VF_PZ3	logVF VF_REN3 VF200 VF_REN4 VF300 VF_REN5 VF400 VF_REN6 VF500 VF_REN7 VF600 PZ_CS1 VF750 PZ_CS3 VF1000 PZ_CS4 REN4500 PZ_CS5 VF_REN1 PZ_CS6 VF_REN2 PZ_CS7	CS2 REN2000 CS3 Cep10 CS4 CR30 CS5 VF_PZ4 CS6 VF_PZ5 CS7 VF_PZ6 CS8 VF_PZ7 ID50 CR_REN1 ID55 CR_REN2 PRD1 CR_REN3 PRD4 CR_REN4
ν	VF400 REN1000 VF1000 REN1500 PZ1 REN2000 PZ3 REN2500 PZ6 REN3000 PZ9 REN3500 PZ12 REN4500 PZ18 Cep1 ID20 Cep3 ID25 Cep5 ID30 Cep6 ID35 Cep7 ID40 Cep8 ID45 Cep10 ID50 CRm5 ID55 CR5 PRD2 CR10 PRD4 VF_PZ1 TX1 VF_PZ2 TX2 VF_PZ3 TX3 PZ_CS3 TX4 PZ_CS4 TX5 PZ_CS5 TX6 PZ_CS6 TX7	VF1500 CR_REN1 CR_REN6	PZ_CS1
τ			CS2 Cep2 CS3 Cep9 VF1500 PZ_CS1 PRD4 PZ_CS3 CR_REN8

Observa-se que:

μ - Ambos os modelos apresentaram poucas variáveis significativas, sendo a variável de interação Prazo x *Credit Score* a única presente nos dois modelos, porém com coeficientes bastante diferentes. Vale ressaltar que no modelo ZAIG a perda esperada é obtida diretamente por μ enquanto que no modelo BEINF a perda esperada é obtida em função de μ e de τ , o que não permite uma comparação direta entre os coeficientes encontrados.

σ - Os modelos apresentaram variáveis significantes muito diferentes.

ν - Os modelos que determinam a probabilidade de clientes sem perdas apresentam comportamento muito parecido: têm praticamente as mesmas variáveis significantes e seus coeficientes têm valores muito próximos para as mesmas. Apenas a variável PZ_CS1 está no modelo BEINF e não está presente no modelo ZAIG e as variáveis VF1500, CR_REN1 e CR_REN6 aparecem no ZAIG, porém com baixa significância e não estão presentes no BEINF.

Os modelos de ν são os que apresentam maior número de variáveis significantes, o que permite estabelecer com maior precisão os fatores que definem a inadimplência em uma carteira de crédito. Os resultados encontrados nas estimativas de ambos os modelos foram muito próximos do valor real observado, tanto na amostra de desenvolvimento do modelo quanto na amostra de validação.

τ - Este modelo é exclusivo para a distribuição BEINF e apresenta poucas variáveis significantes.

5.3.2 Estimativas intervalares

Nas tabelas a seguir, encontram-se os resultados da aplicação da metodologia de comparação descrita no item 3.3, utilizando-se o método de *Bootstrap* para este fim.

A partir da amostra de validação, com as estimativas de valores de perdas de cada modelo calculadas para todas as observações, faz-se uma amostragem aleatória com reposição, cujo tamanho amostral é exatamente o número total de observações selecionadas para validação do modelo, que no caso deste estudo 9.482 observações.

Chama-se este processo de reamostragem e para cada sub-amostra calcula-se a perda estimada pelos modelos ZAIG e BEINF.

Repetiu-se este processo para 5.000 sub-amostras e desta forma, obtém-se distribuições de valores de perda estimada para cada modelo.

A partir estas distribuições determinaram-se intervalos de confiança de 95% e 99% para as perdas estimadas.

Na Tabela 5.15 encontram-se os resultados da aplicação do método de comparação.

Tabela 5.15 – Intervalos de confiança para perda estimada de cada modelo

Modelo	IC	Lim Inferior	Lim Superior	Distância	Média
ZAIG	95%	13,03%	13,85%	0,82%	13,44%
	99%	12,96%	13,95%	0,99%	13,44%
BEINF	95%	12,55%	13,34%	0,79%	12,95%
	99%	12,48%	13,46%	0,98%	12,95%

A perda observada (real) na amostra de validação foi 12,82% (Tabela 5.13), tem-se:

1) As diferenças médias encontradas pelas perdas estimadas em relação à perda real foi de 0,62% para o modelo ZAIG e 0,13% para o BEINF, sendo a perda estimada superior a perda real observada em ambos os casos. Estas diferenças representam variações relativas de 4,87% e 1,04% sobre o valor de perda real esperado, respectivamente para ZAIG e BEINF.

2) As amplitudes dos intervalos de confiança de 95% foram de 0,82% e 0,79%, para ZAIG e BEINF respectivamente, sugerindo uma maior variabilidade para os resultados obtidos a partir do modelo ZAIG.

3) Analisando-se os intervalos de confiança de 95% para as perdas estimadas pelo ZAIG ([13,03%;13,85%]) e pelo BEINF ([12,55%;13,34%]), apenas o intervalo para o modelo BEINF contém o verdadeiro valor da perda observada (12,82%). Isto também se verifica para os intervalos de confiança de 99%.

Comparando-se os resultados apresentados acima, verifica-se que:

a) Os dois modelos apresentaram amplitudes dos intervalos de confiança para as perdas estimadas muito próximas, com uma pequena vantagem para o modelo BEINF.

b) A distribuição BEINF apresentou menor diferença entre a perda estimada pelo modelo e perda real observada (0,13%), enquanto que a distribuição ZAIG apresentou diferença em torno de 0,62%.

c) Apenas o modelo BEINF apresentou o valor de perda real observada dentro do intervalo de confiança de 95% e de 99% para as perdas estimadas, isto é, o valor de 12,82% está contido nos intervalos.

Esses fatos sugerem, apesar do bom desempenho do modelo ZAIG, na análise desta carteira, uma pequena vantagem para o modelo BEINF. Isso talvez se deva ao fato de nos modelos BEINF ser possível modelar separadamente a probabilidade de perda total, o que pode trazer benefícios ao modelo como um todo. Vale ressaltar que o modelo ZAIG, apesar de não conter o valor da perda esperada em seu intervalo de confiança, também é uma alternativa interessante de análise, haja vista a aderência e o pequeno erro encontrado em suas estimativas.

6. Conclusão

As carteiras de crédito massificado são muito influenciadas pelo perfil de seus clientes e de suas condições financeiras no momento do empréstimo.

Muitas instituições utilizam-se de ferramentas como modelos de *Credit Scoring*, a fim de proporcionar decisões objetivas em relação ao risco dos clientes, porém poucas se utilizam de modelos que quantificam adequadamente a inadimplência e a perda esperada em suas carteiras.

O objetivo deste trabalho foi apresentar modelos para prever a perda financeira esperada em carteiras novas, isto é, quantificar adequadamente a expectativa de perda no momento da concessão do crédito.

Uma das dificuldades dos modelos tradicionais está no fato da inadimplência na carteira ser um evento relativamente raro, o que pode gerar estimativas distorcidas das perdas, dada a grande quantidade de clientes sem perda. Os dois modelos estatísticos da classe GAMLSS apresentados, buscam corrigir esta limitação através de ajustes nas estimativas, modelando explicitamente a probabilidade de ocorrência da inadimplência. Estes modelos são obtidos a partir de misturas de distribuições: uma Bernoulli para indicar a ocorrência de inadimplência¹⁶ e uma contínua para indicar a perda em caso de inadimplência (PCI).

O modelo ZAIG utiliza a distribuição Normal inversa para a parte contínua, representada por PCI; esta distribuição acomoda bem uma forte assimetria à direita, esperada para esta variável. Para a parte discreta, utilizou-se um modelo logito para a ocorrência de inadimplência.

O modelo BEINF utiliza a distribuição Beta para a parte contínua, representada pelo percentual de perda em caso de inadimplência (PPCI), sendo indicada para variáveis que se encontram no intervalo aberto de 0 a 100%. Para a parte discreta, utilizaram-se modelos log-lineares para os parâmetros que definem a não ocorrência de inadimplência (quando perda igual a zero) e para a ocorrência de perda total do valor emprestado, isto é, quando PPCI é igual a 100%.

¹⁶ No caso do modelo Beinf, há uma segunda variável Bernoulli que indica a possibilidade de perda total do valor financiado.

Neste trabalho foram incorporadas variáveis explicativas relacionadas ao perfil do cliente e à operação de empréstimo na modelagem dos parâmetros das distribuições, o que trouxe ganhos de estimação ao modelo.

Através dos resultados obtidos, verifica-se que o modelo BEINF foi ligeiramente superior ao ZAIG para a carteira analisada, mas ambos são úteis para a determinação da perda esperada e da probabilidade de inadimplência em carteiras de crédito massificado, tornando-se uma alternativa interessante aos modelos disponíveis atualmente.

Apêndice A – Descrição das variáveis

Variável	categoria	Descrição
Credit Score	CS1*	classificação de risco nível 1 - maior risco de inadimplência
	CS2	classificação de risco nível 2
	CS3	classificação de risco nível 3
	CS4	classificação de risco nível 4
	CS5	classificação de risco nível 5
	CS6	classificação de risco nível 6
	CS7	classificação de risco nível 7
	CS8	classificação de risco nível 8
	CS9	classificação de risco nível 9
	CS10	classificação de risco nível 10 - menor risco de inadimplência
Idade	ID<20*	clientes com menos de 20 anos
	ID20	clientes de 20 a 25 anos
	ID25	clientes de 25 a 30 anos
	ID30	clientes de 30 a 35 anos
	ID35	clientes de 35 a 40 anos
	ID40	clientes de 40 a 45 anos
	ID45	clientes de 45 a 50 anos
	ID50	clientes de 50 a 55 anos
Renda	REN<500*	Renda pessoal inferior a R\$ 500
	REN500	Renda pessoal de R\$ 500 a R\$750
	REN750	Renda pessoal de R\$ 750 a R\$1000
	REN1000	Renda pessoal de R\$ 1000 a R\$1500
	REN1500	Renda pessoal de R\$ 1500 a R\$2000
	REN2000	Renda pessoal de R\$ 2000 a R\$2500
	REN2500	Renda pessoal de R\$ 2500 a R\$3000
	REN3000	Renda pessoal de R\$ 3000 a R\$3500
	REN3500	Renda pessoal de R\$ 3500 a R\$4500
REN4500	Renda pessoal superior a R\$ 4500	
Comprometimento de Renda	CRm5	comprometimento de renda inferior a 5%
	CR5	comprometimento de renda de 5% a 10%
	CR10	comprometimento de renda de 10% a 15%
	CR15*	comprometimento de renda de 15% a 20%
	CR20	comprometimento de renda de 20% a 25%
	CR25	comprometimento de renda de 25% a 30%
	CR30	comprometimento de renda de 30% a 35%
CR35	comprometimento de renda superior a 35%	
CEP	Cep1	CEP residencial com início 01
	Cep2	CEP residencial com início 02
	Cep3	CEP residencial com início 03
	Cep4	CEP residencial com início 04
	Cep5	CEP residencial com início 05
	Cep6	CEP residencial com início 06
	Cep7	CEP residencial com início 07
	Cep8	CEP residencial com início 08
	Cep9	CEP residencial com início 09

(Continuação)

	Cep10 Cep11*	CEP residencial com início 10 Demais regiões
Valor Financiado	VFm200	valor financiado inferior a R\$ 200
	VF200	valor financiado de R\$ 200 a R\$ 300
	VF300	valor financiado de R\$ 300 a R\$ 400
	VF400	valor financiado de R\$ 400 a R\$ 500
	VF500	valor financiado de R\$ 500 a R\$ 600
	VF600	valor financiado de R\$ 600 a R\$ 750
	VF750	valor financiado de R\$ 750 a R\$ 1000
	VF1000	valor financiado de R\$ 1000 a R\$ 1500
	VF1500	valor financiado de R\$ 1500 a R\$ 2000
	VF2000*	valor financiado superior a R\$ 2000
Log(Val Fin)	LogVF	Logaritmo do valor financiado
Taxa de juros	TX1	taxa de juros do financiamento inferior a 4% ao mês
	TX2	taxa de juros do financiamento de 4% a 4,5% ao mês
	TX3	taxa de juros do financiamento de 4,5% a 5% ao mês
	TX4	taxa de juros do financiamento de 6% a 6% ao mês
	TX5	taxa de juros do financiamento de 6% a 7% ao mês
	TX6	taxa de juros do financiamento de 7% a 8% ao mês
	TX7	taxa de juros do financiamento de 8% a 9% ao mês
	TX8*	taxa de juros do financiamento superior a 9% ao mês
Prazo	PZ1	número de parcelas inferior a 3
	PZ3	número de parcelas de 3 a 5
	PZ6	número de parcelas de 6 a 8
	PZ9	número de parcelas de 9 a 11
	PZ12	número de parcelas de 12 a 17
	PZ18	número de parcelas de 18 a 23
PZ24*	número de parcelas igual a 24	
Produto	PRD1	produto do tipo 1
	PRD2	produto do tipo 2
	PRD3*	produto do tipo 3
	PRD4	produto do tipo 4
Interação	VF_PZ	Produto de valor financiado e Prazo. Com 8 níveis.
	VF_CS	Produto de valor financiado e credit score. Com 8 níveis.
	VF_REN	Produto de valor financiado e renda. Com 8 níveis.
	PZ_CS	Produto de prazo e credit score. Com 8 níveis.
	CR_REN	Produto de comprometimento de renda e renda. Com 8 níveis.

* Para aplicação dos modelos excluíram-se as categorias com maior percentual de clientes inadimplentes a fim de evitar o efeito de multicolinearidade.

Apêndice B – Análise descritiva

Tabela B.1 – Correlação entre as variáveis

<i>Variáveis</i>	Taxa Juros	Val Financ	Prazo	Idade	Renda	<i>Credit Score</i>
Taxa Juros	1					
Val Financ	-0,26	1				
Prazo	-0,13	0,58*	1			
Idade	-0,03	0,00	-0,04	1		
Renda	-0,25	0,30*	0,04	0,11	1	
<i>Credit Score</i>	-0,22	-0,36*	-0,82*	0,20	0,09	1
Comp Renda	0,09	0,21	-0,05	-0,08	-0,30*	0,00

* Escolheram-se correlações superiores a $|0,30|$ para definição das interações utilizadas no modelo.

Tabela B.2 – *Credit score*

<i>Credit Score</i>	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
CS1	2.743	1.515	55,2%	817	843	59,3%	31,9%
CS2	3.608	1.323	36,7%	646	800	52,7%	30,6%
CS3	3.782	922	24,4%	510	641	51,4%	29,7%
CS4	3.681	552	15,0%	384	693	57,1%	29,2%
CS5	4.073	454	11,1%	315	811	56,8%	29,2%
CS6	3.883	295	7,6%	248	300	60,7%	28,5%
CS7	3.984	224	5,6%	264	355	64,4%	29,0%
CS8	4.549	155	3,4%	423	663	61,3%	29,5%
CS9	4.091	98	2,4%	421	676	70,0%	27,5%
CS10	3.536	48	1,4%	519	713	70,1%	25,5%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.3 – Idade do cliente

IDADE	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
ID<20	325	99	30,5%	477	548	57,0%	28,4%
ID20	3.628	791	21,8%	526	649	58,4%	30,8%
ID25	6.167	1.081	17,5%	530	678	57,7%	30,0%
ID30	6.819	1.061	15,6%	593	844	56,2%	30,5%
ID35	5.773	793	13,7%	572	711	57,1%	31,3%
ID40	5.116	644	12,6%	663	957	56,7%	30,7%
ID45	3.829	451	11,8%	576	787	54,2%	30,5%
ID50	2.547	281	11,0%	603	775	52,5%	29,4%
ID55+	3.726	385	10,3%	536	724	55,6%	31,2%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.4 – Renda do cliente

Renda	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
REN<500	6.870	1.222	17,8%	250	266	56,3%	30,2%
REN500	7.429	1.156	15,6%	374	356	56,0%	29,9%
REN750	4.911	760	15,5%	507	541	57,0%	30,8%
REN1000	7.158	985	13,8%	651	749	57,2%	30,8%
REN1500	3.768	549	14,6%	839	944	57,7%	31,9%
REN2000	2.160	278	12,9%	817	790	56,5%	30,1%
REN2500	1.280	179	14,0%	887	860	58,3%	29,6%
REN3000	1.300	141	10,8%	941	1.059	56,4%	30,8%
REN3500	1.167	137	11,7%	1.233	1.226	59,2%	32,2%
REN4500	1.887	179	9,5%	1.499	1.848	51,6%	31,5%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.5 – Comprometimento de renda

Compr renda	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
CRm5	4.834	480	9,9%	403	497	56,5%	31,2%
CR5	8.426	1.073	12,7%	550	689	57,3%	30,4%
CR10	7.394	1.131	15,3%	581	739	56,2%	30,2%
CR15	6.179	1.133	18,3%	541	754	56,8%	31,0%
CR20	4.529	834	18,4%	580	739	57,9%	31,1%
CR25	2.877	478	16,6%	598	759	55,8%	30,4%
CR30	1.495	209	14,0%	660	823	53,7%	29,7%
CR35	2.196	248	11,3%	889	1.366	55,0%	29,4%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.6 – CEP de residência do cliente

CEP	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
Cep1	377	51	13,5%	419	434	57,8%	32,5%
Cep2	849	149	17,6%	951	1.006	58,6%	32,1%
Cep3	2.241	391	17,4%	568	649	56,4%	30,4%
Cep4	9.033	1.069	11,8%	479	644	57,8%	30,4%
Cep5	3.437	443	12,9%	646	797	52,1%	29,8%
Cep6	2.096	290	13,8%	752	1.308	53,4%	31,3%
Cep7	3.398	559	16,5%	423	495	54,9%	30,7%
Cep8	4.533	630	13,9%	456	604	56,4%	29,1%
Cep9	7.574	1.135	15,0%	392	480	60,0%	30,6%
Cep10	2.172	313	14,4%	577	682	57,3%	31,3%
Cep11	2.220	556	25,0%	1.135	1.098	53,8%	30,8%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.7 – Valor financiado

VF	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
VFm200	7.727	406	5,3%	95	45	66,6%	28,4%
VF200	6.993	637	9,1%	149	75	59,7%	28,8%
VF300	5.231	641	12,3%	195	105	56,0%	29,8%
VF400	3.540	512	14,5%	248	132	55,2%	29,1%
VF500	2.409	427	17,7%	314	167	57,5%	30,3%
VF600	2.372	518	21,8%	383	209	57,1%	30,9%
VF750	2.452	579	23,6%	485	278	55,8%	31,2%
VF1000	2.803	598	21,3%	659	386	54,2%	31,3%
VF1500	1.974	479	24,3%	855	531	49,4%	30,4%
VF2000	2.429	789	32,5%	1.748	1.335	56,4%	32,1%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.8 – Prazo da operação

Prazo	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
PZ1	8.998	292	3,2%	233	362	76,0%	25,0%
PZ3	17.763	1.458	8,2%	261	352	57,7%	28,7%
PZ6	4.375	1.130	25,8%	388	687	54,7%	30,0%
PZ9	2.475	731	29,5%	558	593	52,4%	30,8%
PZ12	2.615	1.170	44,7%	706	704	56,8%	31,8%
PZ18	1.261	533	42,3%	1.219	1.089	51,4%	31,2%
PZ24	443	272	61,4%	1.510	1.133	59,0%	32,7%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.9 – Taxa de juros

Taxa Fin	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
TX1	3.932	660	16,8%	991	1.081	53,1%	31,4%
TX2	4.510	722	16,0%	1.048	1.022	55,4%	31,4%
TX3	4.987	445	8,9%	603	787	54,5%	30,4%
TX4	6.194	474	7,7%	574	765	56,6%	29,0%
TX5	5.449	605	11,1%	424	589	57,7%	29,7%
TX6	5.467	920	16,8%	414	635	59,0%	30,3%
TX7	5.193	1.154	22,2%	342	355	56,9%	30,9%
TX8	2.198	606	27,6%	327	290	58,4%	30,3%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.10 – Produto adquirido

Produto	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
PRD1	11.407	536	4,7%	444	831	57,4%	29,2%
PRD2	18.195	3.453	19,0%	335	326	58,4%	30,3%
PRD3	4.051	1.325	32,7%	1.140	1.074	52,4%	31,5%
PRD4	4.277	272	6,4%	1.012	1.082	53,0%	30,0%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.11 – Interação entre Valor financiado e Prazo

VF x PZ	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
VF_PZ1	3.651	135	3,7%	92	20	78,6%	15,6%
VF_PZ2	5.641	255	4,5%	129	31	67,8%	15,2%
VF_PZ3	5.909	400	6,8%	155	46	60,8%	16,9%
VF_PZ4	4.621	395	8,5%	176	59	55,3%	17,5%
VF_PZ5	4.909	646	13,2%	234	106	55,2%	21,5%
VF_PZ6	3.368	639	19,0%	306	153	56,1%	25,6%
VF_PZ7	4.160	1.111	26,7%	396	229	55,9%	29,5%
VF_PZ8	5.671	2.005	35,4%	1.106	814	54,1%	32,0%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.12 – Interação entre Valor financiado e Renda

VF x REN	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
VF_REN1	5.420	498	9,2%	118	41	61,6%	19,9%
VF_REN2	5.639	657	11,7%	180	70	57,0%	20,9%
VF_REN3	5.587	712	12,7%	233	95	57,1%	21,8%
VF_REN4	5.430	785	14,5%	315	139	56,7%	22,9%
VF_REN5	5.085	786	15,5%	425	194	56,5%	23,8%
VF_REN6	3.300	614	18,6%	589	296	56,1%	25,6%
VF_REN7	3.482	676	19,4%	853	447	53,9%	25,4%
VF_REN8	3.987	858	21,5%	1.537	881	55,6%	27,2%
Total peral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.13 – Interação entre Valor financiado e *Credit score*

VF x CS	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
VF_CS1	4.667	304	6,5%	114	36	68,3%	18,4%
VF_CS2	5.215	606	11,6%	164	67	59,6%	21,6%
VF_CS3	5.794	769	13,3%	221	94	59,1%	22,9%
VF_CS4	5.071	721	14,2%	286	126	56,5%	22,7%
VF_CS5	5.685	879	15,5%	375	175	55,0%	23,1%
VF_CS6	4.665	934	20,0%	588	307	55,2%	26,2%
VF_CS7	3.325	698	21,0%	973	519	53,3%	25,9%
VF_CS8	3.508	675	19,2%	1.650	896	53,4%	25,2%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.14 – Interação entre Prazo e Credit score

PZ x CS	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
PZ_CS1	3.572	158	4,4%	213	97	82,0%	17,6%
PZ_CS2	5.478	150	2,7%	250	62	68,2%	11,9%
PZ_CS3	4.490	438	9,8%	196	80	61,0%	20,0%
PZ_CS4	5.350	345	6,4%	353	156	59,4%	16,4%
PZ_CS5	4.642	570	12,3%	299	146	58,7%	21,7%
PZ_CS6	4.339	926	21,3%	348	219	56,0%	27,0%
PZ_CS7	4.890	1.293	26,4%	503	424	56,7%	29,7%
PZ_CS8	5.169	1.706	33,0%	1.031	757	51,1%	30,0%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Tabela B.15 – Interação entre Comprometimento de renda e Renda

CR x REN	N Clientes	N Inad	% Inad	PCI médio	DP PCI	PPCI médio	DP PPCI
CR_REN1	4.663	581	12,5%	142	61	57,8%	21,9%
CR_REN2	5.311	748	14,1%	195	86	58,6%	23,1%
CR_REN3	4.548	700	15,4%	264	126	56,1%	23,5%
CR_REN4	5.932	924	15,6%	375	181	56,6%	23,8%
CR_REN5	4.741	821	17,3%	534	272	54,0%	24,1%
CR_REN6	4.157	690	16,6%	745	375	55,7%	24,2%
CR_REN7	4.726	758	16,0%	1.083	543	56,8%	24,3%
CR_REN8	3.852	364	9,4%	1.778	728	59,0%	19,8%
Total geral	37.930	5.586	14,7%	569	765	56,6%	30,6%

Apêndice C – Programas utilizados

```

#####
##### Carregando dados e bibliotecas #####
#####

dadosmod=read.delim2(file="C:\\dm.txt", header=TRUE, sep="\t", quote="", dec=",", fill=TRUE
, comment.char="")
dadosval=read.delim2(file="C:\\dv.txt", header=TRUE, sep="\t", quote="", dec=",", fill=TRUE
, comment.char="")

library(gamlss)
library(gamlss.dist)
library(gamlss.nl)
library(gamlss.mx)
library(gamlss.cens)
library(gamlss.boot)
library(gamlss.tr)
library(gamlss.rsm)

detach(dadosval)
attach(dadosval)
attach(dadosmod)

#####
##### ZAI G #####
#####

#----- Modelo genérico-----
modgen=gamlss(Perda~1, famliy=ZAI G)
GAI C(modgen, k=3)
GAI C(modgen, k=5)

plot(modgen)
qqnorm(modgen$residuals, main = "Normal Q-Q Plot",
        xlab = "Quantis Teóricos", ylab = "Quantis amostrais",
        plot.it = TRUE, datax = FALSE);qqline(modgen$residuals)

hist(modgen$residuals, freq=FALSE, col="light blue", main="Histograma",
      xlab="Resíduos do modelo", ylab="Densidade", nclass=30)

summary(modgen)

pdf.plot(obj=modgen, min=0.001, max=3000, step=5)
rm(modgen)

#----- Modelo com var explicativas
modZ<-gamlss(Perda~offset(logVF)+PZ_CS1+...+CR_REN8
            , si.gma.fo=Perda~logVF+CS9+...+PZ_CS7
            , nu.fo=Perda~VF400+...+CR_REN6
            , famliy=ZAI G(), method=RS())

modZmu=stepGAI C(modZ) # Aplicação do método stepwise para mu
modZsig=stepGAI C(modZmu, what="si.gma") # Aplicação do método stepwise para sigma
modZnu=stepGAI C(modZsig, what="nu") # Aplicação do método stepwise para nu
modZ=modZnu # Modelo selecionado

#----- Estatísticas do modelo
plot(modZ)
qqnorm(modZ$residuals, main = "Normal Q-Q Plot",
        xlab = "Quantis Teóricos", ylab = "Quantis amostrais",
        plot.it = TRUE, datax = FALSE);qqline(modZ$residuals)

hist(modZ$residuals, freq=FALSE, col="light blue", main="Histograma",
      xlab="Resíduos do modelo", ylab="Densidade", nclass=30)

summary(modZ)
GAI C(modZ, k=3)
GAI C(modZ, k=5)

```



```

#----- Aplicando o modelo na amostra de validação

#---- dados do modelo
MUZ=predict(modZ, what="mu", newdata=NULL, type="response")
NUZ=predict(modZ, what="nu", newdata=NULL, type="response")
PIZ=1-NUZ
PerdaZ=PIZ*MUZ

#---- dados de validação
MUZval=predict(modZ, what="mu", newdata=dadosval, type="response", data=dadosmod)
NUZval=predict(modZ, what="nu", newdata=dadosval, type="response", data=dadosmod)
PIZval=1-NUZval
PerdavalZ=PIZval*MUZval

rm(modZ) # Finalizando o modelo

#*****
#***** BEINF *****
#*****

#----- Modelo genérico
modgen=gamlss(PercPerda~1, famliy=BEINF)
plot(modgen)
qqnorm(modgen$residuals, main = "Normal Q-Q Plot",
        xlab = "Quantis Teóricos", ylab = "Quantis amostrais",
        plot.it = TRUE, datax = FALSE);qqline(modgen$residuals)

hist(modgen$residuals, freq=FALSE, col="light blue", main="Histograma",
      xlab="Resíduos do modelo", ylab="Densidade", nclass=30)

summary(modgen)
GAI C(modgen, k=3)
GAI C(modgen, k=5)

pdf.plot(obj=modgen, min=0, max=1, step=0.05)
rm(modgen)

#----- Modelo com var explicativas
modB<-gamlss(PercPerda~PZ3+...+PZ_CS7
            , sigma.fo=PercPerda-CS2+...+CR_REN4
            , nu.fo=PercPerda-VF400+...+PZ_CS6
            , tau.fo=PercPerda-CS2+...+CR_REN8
            , famliy=BEINF(), method=RS())

modBmu=stepGAI C(modZ) # Aplicação do método stepwise para mu
modBsig=stepGAI C(modBmu, what="sigma") # Aplicação do método stepwise para sigma
modBnu=stepGAI C(modBsig, what="nu") # Aplicação do método stepwise para nu
modBtau=stepGAI C(modBnu, what="tau") # Aplicação do método stepwise para tau
modB=modBtau # Modelo selecionado

#----- Estatísticas do modelo
#plot(modB)
qqnorm(modB$residuals, main = "Normal Q-Q Plot",
        xlab = "Quantis Teóricos", ylab = "Quantis amostrais",
        plot.it = TRUE, datax = FALSE);qqline(modB$residuals)

hist(modB$residuals, freq=FALSE, col="light blue", main="Histograma",
      xlab="Resíduos do modelo", ylab="Densidade", nclass=30)

summary(modB)
GAI C(modB, k=3)
GAI C(modB, k=5)

#----- Aplicando o modelo na amostra de validação
#---- dados do modelo
MUB=predict(modB, what="mu", newdata=NULL, type="response")
NUB=predict(modB, what="nu", newdata=NULL, type="response")
TAUB=predict(modB, what="tau", newdata=NULL, type="response")
PI0B=MUB/(1+NUB+TAUB)
PI1B=TAUB/(1+NUB+TAUB)

```

```

PPCI B=(1-PI 0B-PI 1B)*MUB+PI 1B
PerdaB=PPCI B*dadosmod$val _fi n

#---- dados de val i dação
MUBval =predi ct(modB, what="mu", newdata=dadosval , type="response", data=dadosmod)
NUBval =predi ct(modB, what="nu", newdata=dadosval , type="response", data=dadosmod)
TAUBval =predi ct(modB, what="tau", newdata=dadosval , type="response", data=dadosmod)
PI 0Bval =NUBval / (1+NUBval +TAUBval )
PI 1Bval =TAUBval / (1+NUBval +TAUBval )
PPCI Bval =(1-PI 0Bval -PI 1Bval )*MUBval +PI 1Bval
Perdaval B=PPCI Bval *dadosval $val _fi n

rm(modB)          # Fi nal izando o modelo

#####
#***** Simulação Bootstrap *****
#####

#----carregando dados na matrix de simulação
dsi mul =matri x(0, nrow(dadosval ), 4)
dsi mul [, 1]=dadosval $val _fi n
dsi mul [, 2]=dadosval $Perda
dsi mul [, 3]=Perdaval Z
dsi mul [, 4]=Perdaval B

col names(dsi mul )=c("VF", "Perda", "PZAI G", "PBEI NF")

#----função que simula as amostras
boosi mul =functi on(nsi mul , nobs)
{
  amostra=matri x(0, nobs)
  resul tado=matri x(0, nsi mul , 5)
  Vobs=0
  Vzai g=0
  Vbei nf=0
  Dzai g=0
  Dbei nf=0
  for (i in 1:nsi mul )
  {
    Res_amostra=matri x(0, 4)
    amostra=sampl e(1: nrow(dsi mul ), nobs, repl ace=TRUE)
    for (j in 1:nobs)
      for (l in 1:4)
        Res_amostra[l]=Res_amostra[l]+dsi mul [amostra[j], l]
    Vobs=sum(Res_amostra[2])/sum(Res_amostra[1]) #PPCI real
    Vzai g=sum(Res_amostra[3])/sum(Res_amostra[1]) #PPCI equi val ente ZAI G
    Vbei nf=sum(Res_amostra[4])/sum(Res_amostra[1]) #PPCI BEI NF
    Dzai g=Vobs-Vzai g
    Dbei nf=Vobs-Vbei nf
    resul tado[i, 1]= Vobs          # valores reais observados
    resul tado[i, 2]= Vzai g        # valores estimados pela ZAI G
    resul tado[i, 3]= Vbei nf       # valores estimados pela BEI NF
    resul tado[i, 4]= Dzai g        # Di ferença Obs x ZAI G
    resul tado[i, 5]= Dbei nf       # Di ferença Obs x BEI NF
  }
  Quanti s=matri x(0, 5, 9)
  col names(Quanti s)=c("Q1%", "Q2. 5%", "Q5%", "Q25%", "Q50%", "Q75%", "Q95%", "Q97. 5%", "Q99%")
  Medi a=matri x(0, 5)
  for(i in 1:5)
  {
    Quanti s[i, 1]=quanti le(resul tado[, i], 0.01)
    Quanti s[i, 2]=quanti le(resul tado[, i], 0.025)
    Quanti s[i, 3]=quanti le(resul tado[, i], 0.05)
    Quanti s[i, 4]=quanti le(resul tado[, i], 0.25)
    Quanti s[i, 5]=quanti le(resul tado[, i], 0.50)
    Quanti s[i, 6]=quanti le(resul tado[, i], 0.75)
    Quanti s[i, 7]=quanti le(resul tado[, i], 0.95)
    Quanti s[i, 8]=quanti le(resul tado[, i], 0.975)
    Quanti s[i, 9]=quanti le(resul tado[, i], 0.99)
    Medi a[i]=mean(resul tado[, i])
  }
}

```

```
    }  
    return(Quantis, Media)  
}
```

#----Simulação

```
boosi mul (5000, 9482)
```

Referências bibliográficas

- Andrade, F.W.M. (2004). *Desenvolvimento de modelo de risco de portfólio para carteiras de crédito a pessoas físicas*. São Paulo: Tese de Doutorado, EAESP/FGV.
- Akaike, H. (1974). *A new look at the statistical model identification*. Trans. Autom. Control, 19, 716-723.
- Akaike, H. (1983). *Information measures and model selection*. Bull. Int. Statistics Institution, 50, 277-290.
- Cole, T.J.; Green, P.J. (1992). *Smoothing reference centile curves: the LMS method and penalized likelihood*. Statistical Med., 11, 1305-1319.
- Crouhy, M.; Galai, D.; Mark, R.A. (2000). *Comparative analysis of current credit risk models*. Journal of Banking & Finance, 24, p.59-117.
- Crouhy, M.; Galai, D.; Mark, R.A. (2001). *Risk Management*. New York, McGraw-Hill.
- Heller, G.; Stasinopoulos, M.; Rigby, B. (2006). *The zero-adjusted inverse Gaussian distribution as a model for insurance claims*. Proceedings of the 21th International Workshop on Statistical Modelling, J. Hinde, J. Einbeck, and J. Newell (Eds.), 226-233.
- Jong, P. e Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.
- Lindsey, J.K. (2005). *Applying Generalized Linear Models*. New York, Springer

- Liviano, D.; Arauzo, J.M.C. (2007). *Agglomeration and Location: a Nonparametric approach*. Catalúnia: Relatório técnico, Universitat Rovira I Virgili. http://gandalf.fcee.urv.es/departaments/economia/web/compartit/documents_treball/apartats/2007/documents/200705. Acesso em 13/05/2008.
- Mittelhammer, R.C.; Judge, G.G; Miller, D. (2000). *Econometric Foundation*. New York: Cambridge University Press.
- Nelder, J.A.; Wedderburn, R.W.M. (1972). *Generalized linear models*, Journal of the Royal Statistical Society, 135, 370-384.
- Ospina, R.; Ferrari, S.L.P. (2007). *Inflated Beta distributions*. São Paulo: Relatório técnico, USP. http://arxiv.org/PS_cache/arxiv/pdf/0705/0705.0700v3.pdf Acesso em 13/05/2008.
- Prado, R.G.A; Bastos, N.T. Duarte Junior, A.M. (2000). *Gerenciamento de riscos de crédito em bancos de varejo no Brasil*. Tecnologia de Crédito, 19, 7-30.
- Rigby, R.A.; Stasinopoulos, D.M. (1996). *A semi-parametric additive model for variance heterogeneity*. Statistical Comput. 6, 57-65.
- Rigby, R.A.; Stasinopoulos, D.M. (2005). *Generalized additive models for location, scale and shape (with discussion)*. Applied Statistics, 54, 507-554.
- Rigby, R.A.; Stasinopoulos, D.M. (2006). *Statistical Modelling using GAMLSS in R*. Disponível em <http://www.londonmet.ac.uk/gamlss/>. Acesso em 13/05/2008.
- Rosa, P.T.M. (2000). *Modelos de "credit scoring": Regressão Logística, CHAID, Real*. São Paulo: Dissertação de Mestrado, USP.

- Servigny, A.; Renault, O. (2004). *Measuring and Managing Credit Risk*. New York, McGraw-Hill.
- Sicsú, A.L. (1998a). *Desenvolvimento de um sistema de credit scoring – Parte I*. Tecnologia de Crédito, 4, 63-76.
- Sicsú, A.L. (1998b). *Desenvolvimento de um sistema de credit scoring – Parte II*. Tecnologia de Crédito, 5, 57-68.
- Siqueira, J.O. (2000). *Risco: da filosofia a administração*. Convent Selecta, 3, Editora Mandruvá.
- Spanos, A. (1999). *Probability Theory and Statistical Inference*. Cambridge: Cambridge University Press.
- Stasinopoulos D. M.; Rigby R.A. (2007). *Generalized additive models for location scale and shape (GAMLSS) in R*. Journal of Statistical Software, V. 23
- Stasinopoulos, D.M.; Rigby, R.A.; Akantziliotou, C. (2008). *Instructions on how to use the GAMLSS package in R. Second edition*. Disponível em <http://www.londonmet.ac.uk/gamlss/>. Acesso em 13/05/2008.
- Thomas, L.C.; Edelman, D.B.; Crook, J.N. (2002). *Credit scoring and its applications*. Philadelphia: SIAM.
- Vasconcellos, M.B. (2002). *Proposta de método para análise de concessão de crédito a pessoas físicas*. São Paulo: Dissertação de Mestrado, USP.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)