

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO**

KAREN CORREIA PEREIRA

**MODELO DINÂMICO DE CRÉDITO UTILIZANDO ANÁLISE DE
SOBREVIVÊNCIA**

**SÃO PAULO
2014**

KAREN CORREIA PEREIRA

**MODELO DINÂMICO DE CRÉDITO UTILIZANDO ANÁLISE DE
SOBREVIVÊNCIA**

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como parte dos requisitos para a obtenção do título de Mestre em Economia.

Área de concentração: Macroeconomia

Orientador: Prof. Dr. Afonso de Campos Pinto

SÃO PAULO
2014

Pereira, Karen Correia.

Modelo Dinâmico de Crédito Utilizando Análise de Sobrevivência / Karen Correia Pereira - 2014.

52 f.

Orientador: Afonso de Campos Pinto

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Risco (Economia). 2. Créditos. 3. Análise de sobrevivência (Biometria). 4. Inadimplência (Finanças). I. Campos Pinto, Afonso. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 336.77

KAREN CORREIA PEREIRA

MODELO DINÂMICO DE CRÉDITO UTILIZANDO ANÁLISE DE SOBREVIVÊNCIA

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para obtenção do título de Mestre em Economia.

Área de Concentração:

Finanças

Data da aprovação:

___/___/___

Banca examinadora:

Prof. Dr. Afonso de Campos Pinto

Orientador

EESP - FGV

Prof. Dr. Juan Carlos Ruilova Terán

EESP - FGV

Prof. Dr. Roberto Anis Calfat

Inspere

AGRADECIMENTOS

Agradeço à minha família pelo apoio, incentivo e compreensão.

Ao meu marido Fernando por ter aceitado o desafio de vencer mais essa etapa juntos. Com certeza sua presença ao meu lado nesses dois anos de curso foi fundamental para que eu continuasse firme e persistente até o fim.

Ao meu orientador Prof. Dr. Afonso de Campos Pinto pela oportunidade de aprendizado e por toda atenção e apoio dispensado ao longo desses meses.

Agradeço ao Alessandro Zampieri pela oportunidade de crescimento que me foi proporcionada, pela força e incentivo desde o início e aos meus colegas de trabalho pela paciência nesses dois anos.

Agradeço o meu grande colega de turma Thiago por todo o apoio e troca de experiências no decorrer do curso.

“O que sabemos é uma gota; o que ignoramos é um oceano.”

Isaac Newton

RESUMO

Dado a importância da gestão de risco associada às operações de crédito, modelos estatísticos que avaliam o risco de inadimplência tornaram-se fundamentais na mensuração do risco associado a um cliente. Neste contexto, foi desenvolvido um modelo dinâmico de crédito utilizando variáveis características dos clientes, comportamentais e macroeconômicas através da Análise de Sobrevivência aplicada a dados discretos. Os resultados obtidos indicam que a inclusão de variáveis macroeconômicas provoca um efeito significativo, porém baixo, no ajuste do modelo. Entretanto, nota-se uma melhora expressiva no poder de previsão da taxa de inadimplência do portfólio quando aplicado a um conjunto de dados de validação.

Palavras-chave: Risco, Crédito, Modelos de Crédito, Análise de Sobrevivência, Inadimplência.

ABSTRACT

Statistical models became fundamental in risk measuring associated with a client, mainly when the risk management had grown up his importance associated to credit transactions.

In this context, a dynamic credit model was developed using client characteristics, behavioral and macroeconomic variables applying survival analysis to a discrete data.

The results indicated the macroeconomics variables inclusion lead to a significant, but low effect in model fit. However, there is a significant improvement in the default rate predictive power of the portfolio when applied to a validation dataset.

Keywords : Risk, Credit, Credit Models, Survival Analysis, Delinquency.

LISTA DE TABELAS

Tabela 1 - Exemplo da estrutura original dos dados	24
Tabela 2 - Exemplo da estrutura dos dados alterada	24
Tabela 3 - Exemplo de utilização das variáveis	29
Tabela 4 - Exemplo de estrutura da base de dados final	31
Tabela 5 - Exemplo de cálculo das probabilidades estimadas.....	33
Tabela 6 - Resumo dos modelos desenvolvidos.....	34
Tabela 7 - Variáveis significativas nos modelos desenvolvidos	37
Tabela 8 - Coeficientes estimados – Modelo 1	38
Tabela 9 - Coeficientes estimados – Modelo 2	40
Tabela 10 - Comparação dos ajustes dos modelos desenvolvidos	41
Tabela 11 - Comparação do poder de previsão no nível agregado.....	42

LISTA DE FIGURAS

Figura 1 - Censura à Direita	21
Figura 2 - Período da amostra de desenvolvimento e validação	28
Figura 3 - Comparação das taxas de default observada e estimada dos modelos desenvolvidos no período de validação	41
Figura 4 - Exemplo de curva de sobrevivência	43

Sumário

1. Introdução	12
2. Revisão Bibliográfica.....	14
3. Revisão Teórica.....	20
3.1 Análise de Sobrevida	20
3.1.1 Modelo de Sobrevida Discreto	22
3.1.1.1 Estrutura da Base de Dados	23
3.1.1.2 Estimativa dos Parâmetros	25
4. Metodologia.....	28
4.1 Descrição da Base de Dados.....	28
4.1.1 Variáveis Utilizadas	29
4.2 Especificação do Modelo.....	31
4.3 Modelos Desenvolvidos	33
4.4 Medidas de Desempenho.....	34
5. Resultados	37
5.1 Modelos Estimados.....	37
5.2 Medidas de Ajustes e de Previsão dos Modelos.....	41
5.3 Aplicação dos Modelos Estimados.....	42
6. Conclusão	44
REFERÊNCIAS	46
APÊNDICE A – Regressão Logística	49
APÊNDICE B – Lista de Variáveis Macroeconômicas	51
APÊNDICE C – Saída dos Modelos.....	52

1. Introdução

No decorrer dos últimos anos o mercado de crédito brasileiro vem apresentando significativas taxas de crescimento. De acordo com dados do Banco Central do Brasil (2014), a proporção do volume das operações de crédito com relação ao Produto Interno Bruto (PIB) apresentou um aumento de aproximadamente 25 pontos percentuais desde 2007, atingindo em fevereiro de 2014 o patamar de 55,8% do PIB.

Segundo Andrade (2012), entre os elementos responsáveis por esse crescimento verificado desde meados de 2003 destacam-se as mudanças institucionais deste mercado e os fatores econômicos.

Costa (2009) afirma que, entre as mudanças institucionais, destacam-se a facilidade de acesso popular aos bancos, devido, principalmente, à criação dos correspondentes não bancários e da conta simplificada e à expansão do crédito impulsionada pelo crédito consignado, aos consumidores e microcrédito. Segundo o Relatório Anual da Febraban (2012), houve um crescimento de 228,4% no número de correspondentes em comparação com 2008 e de 30% na quantidade de contas correntes no período de 2008 a 2012.

Segundo Andrade (2012), em relação aos fatores econômicos, a consolidação macroeconômica em termos de política fiscal e monetária foram as maiores contribuições para a expansão do crédito no Brasil. O autor destaca o crescimento da renda e do produto, o cenário externo positivo com elevada liquidez e a estabilidade dos preços que permitiu uma redução na taxa nominal de juros.

Diante deste cenário, a gestão do risco atrelada às operações de crédito é um fator de extrema importância, atribuindo elevada atenção aos modelos estatísticos utilizados na tomada de decisão. De acordo com Caouette *et al.* (1998), o principal objetivo no desenvolvimento de modelos é suprir as necessidades das instituições financeiras, fornecendo instrumentos de gestão e controle de risco de crédito que auxiliem na eficiência bancária, principalmente nos processos de concessão, de classificação de risco interno e na previsão de inadimplência. Nos últimos anos, dado a importância desses modelos, diversos autores trabalharam na construção de modelos de risco de crédito mais robustos e assertivos. Entre eles destacam-se Malik e Thomas (2009) e Bellotti e Crook (2013) que incluíram, em seus respectivos modelos, variáveis macroeconômicas utilizando a técnica estatística de Análise de Sobrevida. Este método

permite mensurar o tempo de sobrevivência até a ocorrência de um determinado evento de interesse, que no cenário de crédito está associado à inadimplência.

Neste contexto, o objetivo deste trabalho é o desenvolvimento de um modelo dinâmico de risco de crédito aplicado a dados brasileiros, a fim de calcular a probabilidade de inadimplência associada a um indivíduo. Serão utilizadas como variáveis explicativas do modelo as relacionadas ao cliente no momento da concessão do crédito, as comportamentais e variáveis macroeconômicas. A estimação dos parâmetros será realizada através da técnica estatística de Análise de Sobrevivência, porém aplicada a tempos discretos, dado que as informações utilizadas nesse estudo foram coletadas mensalmente.

A principal contribuição do estudo está relacionada ao aprimoramento dos modelos de risco de crédito, de tal forma que o modelo desenvolvido seja utilizado como uma ferramenta de medição do risco associado a um cliente e de previsão de inadimplência referente a uma determinada carteira de crédito.

A presente dissertação está organizada em 6 capítulos. Seguido à introdução apresentada neste capítulo inicial, o capítulo dois traz uma revisão bibliográfica sobre os principais estudos relacionados aos modelos de concessão de crédito. No capítulo três, a metodologia é apresentada, aprofundando-se nos conceitos da Análise de Sobrevivência, principal técnica a ser utilizada no desenvolvimento do trabalho. No capítulo quatro, a base de dados e as definições utilizadas para a construção do modelo são descritas. Os resultados são expostos no capítulo cinco. Por fim, no capítulo seis são apresentadas as conclusões e sugestões para estudos futuros.

2. Revisão Bibliográfica

Dada à importância do entendimento dos determinantes do risco de crédito, vários estudos foram realizados ao longo dos últimos anos com o objetivo de prever e mensurar o risco através de modelos estatísticos mais assertivos capazes de calcular a inadimplência associada a um empréstimo. Destacam-se neste âmbito os chamados modelos de *credit scoring*.

Segundo Sicsú (2010), o principal objetivo dos modelos de *credit scoring* é de mensurar, no instante da decisão do crédito, a probabilidade de perda para o credor associado ao empréstimo solicitado.

Entre as técnicas estatísticas empregadas na construção dos modelos de previsão de risco de crédito, destacam-se a análise discriminante, regressão logística e análise de sobrevivência.

Conforme Figini e Fantazzini (2009), o modelo desenvolvido por Altman (1968) utilizando a análise discriminante foi o primeiro a empregar técnicas estatísticas para estimar a probabilidade de inadimplência e foi considerado por muito tempo como o mais importante para o cálculo do risco. Neste modelo, Altman utilizou indicadores financeiros como precursores de falência de empresas.

Ao longo dos anos esta técnica foi sendo substituída pela regressão logística, de tal forma que atualmente esses modelos são os mais utilizados no mercado de crédito. Um dos pioneiros a empregar a regressão logística em um estudo de previsão de inadimplência foi Ohlson (1980), que desenvolveu um modelo baseado em um conjunto de dados de empresas americanas solventes e insolventes no período entre 1970 a 1976.

Uma característica destas técnicas é a utilização de informações do proponente coletadas no momento da decisão do crédito. Os modelos são estimados não levando em consideração que as variáveis mudam com o passar do tempo. Segundo Belotti e Crook (2013) tratam-se de modelos *cross-section*, que permitem a predição da probabilidade de *default* em um período pré-determinado de tempo, usualmente 18 meses. Por apresentarem essa característica, esses modelos são denominados modelos estáticos.

Conforme Almeida (2008), sob essa perspectiva, podemos adequar a técnica de modelagem levando em consideração uma resposta temporal do cliente à concessão. Neste contexto, surgem os chamados modelos dinâmicos de crédito. De acordo com Belotti e Crook (2013), entende-se por dinâmico um modelo de dados em painel cuja principal característica é a inclusão de variáveis que se modificam ao longo do período de estimação.

Segundo Almeida (2008) a técnica estatística que mais condiz com essa modelagem temporal é a Análise de Sobrevivência¹. A sua principal diferença em relação aos modelos estáticos é prever não somente se um cliente irá inadimplir, mas quando isso ocorrerá (BELOTTI e CROOK, 2009).

De acordo com Malik e Thomas (2009), a ideia de empregar a Análise de Sobrevivência na construção de modelos de *credit scoring* foi primeiramente proposta por Narain (1992). O autor aplicou o modelo de vida acelerada² para um conjunto de dados de empréstimos e mostrou que o modelo proposto apresentava bons resultados quanto às estimativas do número de falhas para cada período analisado.

Thomas, Ho e Scherer (2001) descreveram a utilização de um processo estocástico de cadeia de Markov como um modelo dinâmico de inadimplência. Entretanto na estrutura abordada não era permitido o uso de covariáveis. Alternativamente, apontaram a Análise de Sobrevivência como forma de construção de modelos dinâmicos, uma vez que esta técnica permite a inclusão de variáveis que se alteram no tempo, tais como as comportamentais e as macroeconômicas.

De acordo com Shumway (2001), os modelos de duração são mais apropriados que os estáticos para prever a falência. O autor demonstra que pelo fato dos modelos estáticos ignorarem que as informações mudam com o tempo, estes produzem estimativas viesadas e inconsistentes da probabilidade de *default*. Cita como vantagens dos modelos dinâmicos o uso de variáveis que se alteram ao longo da análise, além de permitirem o uso de informações macroeconômicas e de variáveis que relacionam o risco com o tempo de análise (dependência temporal). Em seu estudo, o autor desenvolveu um modelo de duração discreto e demonstrou que é possível estimá-lo através de um modelo logístico, dado que as funções de verossimilhança desses

¹ O termo Análise de Sobrevivência é muito utilizado na área médica. Outras terminologias são utilizadas para esta técnica, tais como Análise de Confiabilidade e Análise de Duração.

² Modelo de Vida Acelerada é uma classe de modelos da Análise de Sobrevivência. São conhecidos como *Accelerated Life Models* ou *Accelerated Failure Time Models – AFT Models*.

modelos são equivalentes sob determinadas condições na estrutura da base de dados que serão explicadas no próximo capítulo.

Leonardis e Rocci (2008) também construíram um modelo de sobrevivência discreto e apuraram que o poder preditivo deste modelo é superior ao modelo de regressão logística. Os autores utilizaram informações de balanço de 7711 empresas italianas no período de 1995 a 1998.

Francesca (2012) utilizou dados referentes a empréstimos do Banco Central da Itália e desenvolveu um modelo de duração discreto. O resultado obtido mostra que, considerando conjuntamente as variáveis explicativas e o tempo até a ocorrência do *default*, o ajuste do modelo é melhor quando comparado a técnicas tradicionais de Análise de Sobrevivência.

Um importante ponto a ser considerado no desenvolvimento dos modelos é que “[...]além das variáveis do perfil do tomador de crédito e da garantia, o cenário macroeconômico também pode influenciar a capacidade de pagamento dos tomadores de crédito” (ROCHA, 2012, p. 13). Segundo Crook e Belotti (2010) as condições da economia de um país afetam a chance de que o cliente será inadimplente no futuro.

Alguns estudos investigaram a importância das condições macroeconômicas para a estimação do risco de crédito (CARLING, 2007). Gross e Souleles (2002) utilizaram uma base de dados de contratos de cartão de crédito de uma grande companhia norte americana para estimar a inadimplência através de modelos dinâmicos equivalentes ao modelo de duração discreto proposto por Shumway (2001). Como resultado, a taxa de desemprego foi estatisticamente significativa com o sinal esperado, isto é, taxas maiores de desemprego estão associadas a níveis mais altos de inadimplência.

Antunes, Ribeiro e Antão (2005) consideraram empresas não financeiras do sistema de informação de crédito português (Central de Responsabilidades de Crédito) e estimaram um modelo para prever a inadimplência utilizando variáveis relacionadas ao negócio, às companhias e às condições macroeconômicas. As variáveis macroeconômicas taxa de crescimento do PIB e taxa de juros de curto prazo foram as mais significativas no processo de modelagem.

Carling *et al.* (2007) selecionaram empresas de uma carteira de crédito de um grande banco sueco do período de 1994 a 2000 e criaram um modelo de duração discreto para prever o tempo

até a inadimplência. Utilizaram tanto informações específicas das companhias quanto as relacionadas ao empréstimo, além de variáveis que refletiam as condições econômicas da época. Destacaram-se nesse estudo as seguintes variáveis macroeconômicas: curva de juros, hiato do produto e a expectativa sobre a economia sueca.

Nam *et al.* (2008) desenvolveram um modelo de sobrevivência discreto com variáveis variantes no tempo relacionadas às características financeiras das empresas e incorporaram a dependência macroeconômica à função de risco utilizando a volatilidade da taxa de câmbio. Foram selecionadas 367 empresas coreanas no período de 1991 a 2000. Os resultados mostraram que, com relação à previsão, este modelo é superior a um modelo estático.

Belotti e Crook (2009) desenvolveram um modelo de *credit scoring* com variáveis macroeconômicas utilizando Análise de Sobrevivência. Selecionaram de uma instituição financeira inglesa, 100.000 contratos de cartão de crédito no período de 1997 a 2005. Os autores analisaram variáveis do cliente, tais como idade, renda e situação no *bureau*, no momento da concessão de crédito. Entre as macroeconômicas, destacou-se a taxa de juros como sendo a mais significativa no modelo. Além disso, verificaram que, quando comparado à técnica usual de regressão logística, o modelo apresentou resultados superiores quanto ao poder preditivo.

Malik e Thomas (2009) incorporaram ao modelo de Análise de Sobrevivência informações específicas dos clientes através do escore de comportamento, além de fatores macroeconômicos. Na visão dos autores, o escore de comportamento captura o risco associado ao cliente, enquanto que as variáveis macroeconômicas explicam o risco sistêmico. Foram utilizadas informações de um banco inglês de janeiro de 2001 a dezembro de 2005 e foram considerados como inadimplentes clientes que apresentaram atrasos superiores a 90 dias no período de análise. Entre os modelos desenvolvidos, as variáveis macroeconômicas mais significativas foram taxa de juros, taxa de crescimento do PIB e inflação.

Männaso e Mayes (2009) consideraram no desenvolvimento do modelo de sobrevivência de tempo discreto informações de fatores macroeconômicos, estruturais e específicos de 600 bancos entre os anos de 1995 a 2004 para explicar a ocorrência de problemas bancários em dezenove países do Leste Europeu. Destacaram-se entre as variáveis macroeconômicas a razão crédito privado sobre PIB, taxa real de crescimento do PIB, variação cambial (moeda doméstica/dólar) e índice de mercado de ações.

Belotti e Crook (2013) estimaram um modelo dinâmico através de Análise de Sobrevivência discreta utilizando variáveis do cliente, de comportamento e macroeconômicas. Desenvolveram esse modelo com o objetivo de utilizá-lo como medida de risco para previsão e *stress testing*. A base de dados foi composta por 750.000 contratos de cartão de crédito do período entre 1999 e 2006 de uma instituição inglesa. Os autores concluíram que a inclusão de variáveis macroeconômicas, embora tenha proporcionado um pequeno aumento na performance do modelo, apresentou significativa melhora nas previsões da taxa de inadimplência no nível agregado. Entre as variáveis macroeconômicas testadas, taxa de juros e taxa de desemprego foram as mais importantes.

No Brasil, alguns autores se destacam pelo uso da Análise de Sobrevivência em modelos de risco de crédito, porém estes aplicam essa técnica considerando o tempo contínuo e poucos utilizam variáveis macroeconômicas.

Lima (2008) utilizou uma carteira de operações de crédito da Caixa Econômica Federal referente a 445.889 contratos destinados a pessoas físicas contratados entre janeiro de 2003 e dezembro de 2006 e desenvolveu um modelo de regressão de Cox³. Os resultados apontaram que, além das variáveis do perfil do tomador e das referentes ao contrato, a probabilidade de *default* dos clientes sofre influência de variáveis macroeconômicas, tais como a taxa de juros real e taxa de desemprego.

Batista, Divino e Orrillo (2011) estimaram um modelo de risco proporcional de Cox para uma carteira de crédito imobiliário de um grande banco brasileiro. A amostra correspondia a 637.704 contratos compreendidos entre janeiro de 1997 a agosto de 2009. O modelo final era composto por variáveis cadastrais e do negócio, além das variáveis macroeconômicas referentes a taxa de juros real da economia e crescimento real do PIB.

Divino e Rocha (2013) desenvolveram um modelo de regressão logística e um modelo de Cox com o objetivo de comparar diversas medidas estatísticas de performance. Utilizaram informações de contratos de crédito imobiliário de uma importante instituição financeira do período de janeiro de 2006 a dezembro de 2007. Neste estudo, o modelo de regressão logística apresentou desempenho superior ao modelo de análise de sobrevivência, principalmente quando a variável macroeconômica taxa básica da economia foi incluída no modelo.

³ Trata-se de um modelo semi-paramétrico de análise de sobrevivência desenvolvido por Cox em 1972 (Cox, 1972). Também é conhecido como modelo de riscos proporcionais.

Diferente do que foi adotado no mercado brasileiro até o momento, esta dissertação propõe o desenvolvimento de um modelo dinâmico de crédito utilizando variáveis disponíveis no momento da concessão relacionadas ao cliente, comportamentais, macroeconômicas e a técnica de Análise de Sobrevivência Discreta baseado no modelo proposto por Belotti e Crook (2013) e nas considerações a respeito do processo de estimação demonstradas por Shumway (2001).

Os principais conceitos teóricos e premissas utilizados na construção desse modelo são abordados no próximo capítulo.

3. Revisão Teórica

Neste capítulo será descrita a técnica utilizada no processo de modelagem dos dados referente à construção do modelo dinâmico de crédito, fornecendo suporte à metodologia proposta neste estudo. Os conceitos referentes à Análise de Sobrevivência são expostos, dando enfoque especial aos modelos de dados discretos.

3.1 Análise de Sobrevivência

A Análise de Sobrevivência é um conjunto de técnicas estatísticas que consiste na criação de um modelo que permite prever o tempo até a ocorrência de um evento de interesse. O tempo decorrido desde o início da análise até o evento de interesse é denominado **tempo de falha**. No caso aplicado aos modelos de concessão de crédito, considera-se como tempo de falha o período compreendido desde a contratação de um empréstimo até a ocorrência do evento de interesse. Geralmente este evento é definido como um problema de inadimplência levando em consideração atrasos superiores a uma quantidade definida de dias (60 ou 90 dias).

Entre as principais características de um modelo de sobrevivência, destaca-se a presença do conceito de **censura**. Dados censurados são aqueles em que há a observação parcial da variável resposta. No cenário de crédito, a censura pode ocorrer, principalmente, pelo fato do evento de interesse não ter sido observado no final do período de análise ou pelo contrato ter sido liquidado ao longo do estudo. Apesar dos dados incompletos, essas observações são importantes e devem ser levadas em consideração na estimação do modelo. Conforme Giolo e Colosimo (2006), as observações censuradas fornecem informações sobre o tempo de vida e a sua omissão acarretaria em conclusões viciadas considerando o cálculo das estatísticas de interesse.

Entre as formas de censura, podemos classificá-las como censura à direita, à esquerda ou intervalar. Neste estudo daremos destaque à censura à direita, caracterizada como a não observação do evento de interesse até o final do período de análise, conforme Figura 1.

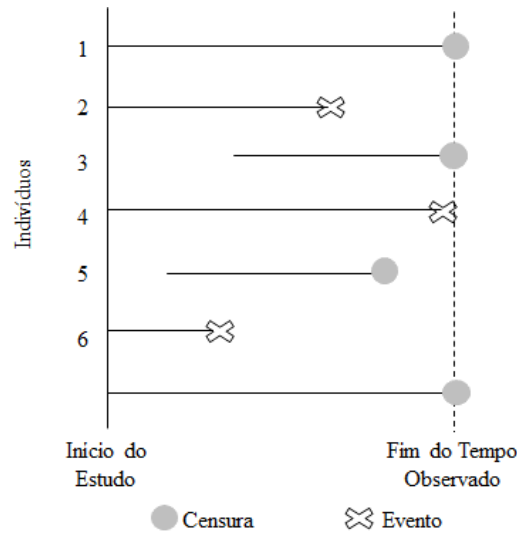


Figura 1 - Censura à Direita

O tempo de sobrevivência de uma observação pode ser definida de várias formas, dentre as quais se destacam a função de sobrevivência e a função de risco, descritas a seguir.

Considerando T como o tempo de falha do indivíduo (o tempo até a ocorrência do evento de interesse), T é uma variável aleatória contínua não-negativa que possui uma função de densidade de probabilidade $f(t)$ e função de densidade acumulada $F(t)$. A função de sobrevivência, denotada por $S(t)$, é a probabilidade do tempo de sobrevivência ser maior ou igual a um determinado tempo t , sendo definida da seguinte forma:

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(u) du \quad (1)$$

Notamos pela Equação 1 que a função de distribuição acumulada $F(t)$ é a probabilidade de um indivíduo não sobreviver até o instante t , ou seja, $F(t) = 1 - S(t)$.

A função de risco, denotada por $h(t)$, representa o risco de falha no tempo t , condicional à sobrevivência até o tempo t . É interpretada como o risco instantâneo de falha em t , conforme expressão abaixo:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T \leq t + \Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{S(t)} \quad (2)$$

Essa função descreve a forma como o risco muda ao longo do tempo, permitindo a especificação da sua distribuição. Essa característica torna a função de risco a preferida por muitos autores no processo de modelagem.

Consideramos até esse ponto que o tempo de falha apresenta a forma contínua, porém há estudos que utilizam as informações em dia, mês ou ano e para estes casos, a forma mais adequada de tratar os dados é adotar os modelos de sobrevivência discretos, conforme será abordado na próxima seção.

3.1.1 Modelo de Sobrevivência Discreto

Considerando T o tempo de falha de um determinado indivíduo, T é uma variável aleatória discreta com probabilidade

$$f_n = P(T = n) \quad (3)$$

em que $n \in (1, 2, 3, \dots)$, um conjunto de inteiros positivos. Para este caso, a função de sobrevivência é dada por:

$$S_n = P(T \geq n) = \sum_{k=n}^{\infty} f_k \quad (4)$$

A função de risco é a probabilidade condicional de que o evento irá ocorrer no tempo n , dado que não tenha ocorrido nos tempos anteriores a n . Essa função é definida como:

$$h_n = P(T = n | T \geq n) = \frac{f_n}{S_{n-1}} \quad (5)$$

A função de sobrevivência S_n pode ser reescrita em termos da função de risco da seguinte maneira:

$$S_n = (1 - h_1)(1 - h_2) \dots (1 - h_{n-1})(1 - h_n) = \prod_{k=1}^n (1 - h_k) \quad (6)$$

As Equações de 3 a 6 correspondem a um indivíduo genérico. Elas podem ser reescritas especificando explicitamente o indivíduo i a que se referem, como segue:

$$f_{i,n} = P(T_i = n) \quad (7)$$

$$S_{i,n} = P(T_i \geq n) = \sum_{k=n}^{\infty} f_{i,k} \quad (8)$$

$$h_{i,n} = P(T_i = n \mid T_i \geq n) = \frac{f_{i,n}}{S_{i,n-1}} \quad (9)$$

$$S_{i,n} = \prod_{k=1}^n (1 - h_{i,k}) \quad (10)$$

Podemos notar pelas Equações 6 e 10 que a probabilidade de sobrevivência até um determinado tempo n é dada pelo produto das probabilidades de não ter falhado em cada unidade de tempo observada. Dada essa característica, o banco de dados deve seguir uma estrutura pré-determinada, conforme será descrito na seção seguinte.

3.1.1.1 Estrutura da Base de Dados

O primeiro passo para utilização da Análise de Sobrevivência Discreta é a reestruturação do banco de dados de tal forma que, para cada unidade de tempo observado, cada indivíduo seja tratado como uma unidade de análise. É necessário a criação de uma base de dados com múltiplas observações para cada indivíduo da amostra até o primeiro evento de falha ou até a censura, definida anteriormente na seção 3.1.

Consideramos que para cada um dos N indivíduos, temos ao longo do período observado as informações referentes ao tempo de falha T_i e à indicadora de censura c_i , igual a 1 se não houve censura ou 0, caso contrário, com $i = 1, 2, \dots, N$. A partir desse conjunto de informações, uma nova variável binária $Y_{i,n}$ é definida para cada unidade do tempo, sendo codificada como:

$$Y_{i,n} = \begin{cases} 1, & n = T_i \text{ e } c_i = 1 \\ 0, & n = T_i \text{ e } c_i = 0 \\ 0, & n < T_i \end{cases}$$

Desta forma, o processo de falha de um indivíduo i pode ser considerado como uma sequência de respostas binárias que segue uma distribuição binomial:

$$Y_{i,n} = \begin{cases} 1, & \text{se } n = T_i \text{ e } c_i = 1 \\ 0, & \text{caso contrário} \end{cases}$$

A Tabela 1 sumariza a estrutura dos dados na sua forma original e a Tabela 2 exemplifica a forma que deve ser adotada para a estimação da Análise de Sobrevida Discreta.

Tabela 1 - Exemplo da estrutura original dos dados

Indivíduo i	T_i	c_i
1	5	1
2	2	0
3	6	0

Fonte: Elaboração Própria.

Tabela 2 - Exemplo da estrutura dos dados alterada

Indivíduo i	T_i	c_i	Tempo n	$Y_{i,n}$
1	5	1	1	0
1	5	1	2	0
1	5	1	3	0
1	5	1	4	0
1	5	1	5	1
2	2	0	1	0
2	2	0	2	0
3	6	0	1	0
3	6	0	2	0
3	6	0	3	0
3	6	0	4	0
3	6	0	5	0
3	6	0	6	0

Fonte: Elaboração Própria.

Neste exemplo, vamos considerar que o tempo de observação do estudo foi de 6 períodos. Para o indivíduo 1, que apresentou tempo de falha igual a 5 e que não foi censurado, a base de dados deverá conter 5 registros referentes a cada unidade de tempo observada e a variável $Y_{i,n}$ receberá o valor zero para os primeiros 4 registros e 1 para a última observação. No caso do indivíduo 2, em que o tempo de falha é 2 e houve censura, teremos duas linhas de observação com $Y_{i,n}$ igual a zero nesses dois registros. Por fim, para o indivíduo 3 com tempo de falha igual

a 6 e que foi censurado dado o final do período de observação, haverá 6 registros sendo que em cada uma dessas linhas a variável $Y_{i,n}$ será igual a zero.

Vale ressaltar que esta estrutura é montada “*pos facto*”, isto é, após o conhecimento do histórico de cada contrato. Além disso, o processo de modelagem é baseado nas informações dos clientes, porém na base de dados de estudo podemos ter indivíduos com mais de um contrato, de tal forma que cada um deles contribuirá com informações referentes aos clientes em diferentes períodos.

Observando o exemplo dado anteriormente, notamos que teremos distintas situações com relação ao acompanhamento dos indivíduos na estrutura da base de dados. São elas:

- Indivíduos cujos contratos serão acompanhados até o instante da falha, sendo este anterior ao final do período do estudo (indivíduo 1);
- Indivíduos que foram censurados e desapareceram da análise em um determinado período anterior ao final do estudo (indivíduo 2);
- Indivíduos que permanecerão até o final do período de acompanhamento e não apresentarão falha, porém serão censurados dado o fim do estudo (indivíduo 3).

No cenário do crédito, os indivíduos semelhantes ao 1 apresentaram problemas relacionados à inadimplência. Os que possuem perfil similar ao indivíduo 2 correspondem aos clientes que liquidaram o contrato ao longo do estudo, enquanto que o indivíduo 3 refere-se àqueles cujos contratos permaneceram vigentes e não apresentaram problemas de inadimplência até o encerramento do estudo.

3.1.1.2 Estimação dos Parâmetros

Dado que o objetivo da Análise de Sobrevivência é a criação de um modelo que permite prever o tempo até a ocorrência de um evento de interesse, devemos estimar os coeficientes relacionados a este modelo. Desta forma, consideramos um indivíduo i com tempo de observação $k = 1, \dots, n$ e que o processo de falha do indivíduo i é uma sequência de respostas binárias $y_{i,k}$, conforme definido anteriormente.

A contribuição de cada indivíduo para função de verossimilhança é dada por:

$$L_i = \prod_{k=1}^n h_{i,k}^{y_{i,k}} (1 - h_{i,k})^{(1 - y_{i,k})} \quad (11)$$

em que $h_{i,k}$ representa a função de risco do indivíduo i no tempo k .

Podemos observar que os casos censurados contribuem para a função de verossimilhança apenas com o fator $(1 - h_{i,k})$ em cada unidade de tempo observada. Já a contribuição das observações não censuradas é dada pelo elemento $(1 - h_{i,k})$ em todos os instantes de tempo anteriores a ocorrência do evento ($y_{i,k} = 0$) e por $h_{i,k}$ no momento em que o evento de interesse ocorre ($y_{i,k} = 1$).

Desta forma, para os N indivíduos da amostra, a função de verossimilhança será igual a:

$$L = \prod_{i=1}^N \prod_{k=1}^n h_{i,k}^{y_{i,k}} (1 - h_{i,k})^{(1 - y_{i,k})} \quad (12)$$

A Equação 12 corresponde à forma da função de verossimilhança de um modelo de regressão binária em que cada observação corresponde ao indivíduo i no instante k , cuja variável resposta é $y_{i,k}$. Essa condição implica que um modelo de sobrevivência discreto pode ser estimado através de programas de análise de dados dicotômicos, como por exemplo da regressão logística⁴.

Considerando o modelo logístico e os vetores de variáveis explicativas \mathbf{X} e de parâmetros $\boldsymbol{\beta}$, a função de risco é especificada da seguinte maneira:

$$h_n = \frac{1}{1 + \exp\{-\{\theta + \mathbf{X}'\boldsymbol{\beta}\}\}} \quad (13)$$

Com o objetivo de simplificar o modelo transformando-o em uma combinação linear das variáveis explicativas, a Equação 13 pode ser reescrita utilizando uma função *logito*, como segue abaixo.

⁴ Shumway (2001) denomina esse modelo de regressão como modelo logístico multiperíodo. Trata-se de um modelo logístico que é estimado utilizando as informações de cada um dos indivíduos para cada instante de tempo.

$$\text{logito } h_n = \log\left(\frac{h_n}{1-h_n}\right) = \theta + \mathbf{X}'\boldsymbol{\beta} \quad (14)$$

Informações adicionais sobre a técnica de Regressão Logística encontram-se no Apêndice A.

Vale ressaltar que o vetor de variáveis explicativas \mathbf{X} pode ser composto por mais de um tipo de variável, como por exemplo, variáveis relacionadas às características e ao comportamento dos indivíduos e às macroeconômicas.

O parâmetro θ deverá ser incluído na estimação do modelo como função do tempo com o objetivo de especificar a dependência do risco com relação ao tempo. Entre as formas mais usuais destacam-se as funções logarítmicas e polinomiais, conforme pode ser observado em Shumway (2001) e Belotti e Crook (2013).

Os conceitos descritos até aqui servirão como apoio à metodologia aplicada nesse estudo, detalhada no próximo capítulo.

4. Metodologia

Neste capítulo apresentamos os passos referentes ao desenvolvimento do modelo proposto nesse estudo baseado nos conceitos teóricos vistos no capítulo anterior. Iniciamos com o detalhamento da criação da base de dados, da lista de variáveis utilizadas no processo de modelagem e com a definição do evento de interesse. Em seguida, definimos a especificação do modelo desenvolvido e, por fim, explicamos as medidas de desempenho e de previsão utilizadas.

4.1 Descrição da Base de Dados

O banco de dados utilizado no desenvolvimento do estudo corresponde a uma determinada carteira de crédito de pessoas físicas de uma instituição financeira brasileira. Por motivos de confidencialidade, algumas informações não serão reveladas a fim de preservar as estratégias dessa instituição.

Desta carteira de crédito, foi selecionada uma amostra aleatória de 600.000 operações de crédito iniciadas no período de janeiro de 2009 a dezembro de 2012. A fim de mensurar o tempo até a ocorrência do evento de interesse, que será definido na seção 4.1.1 adiante, todos os contratos foram acompanhados mensalmente até o fim da sua vigência ou até o final do período de análise do estudo em dezembro de 2013.

Para o desenvolvimento do modelo foi considerado o período de janeiro de 2009 a dezembro de 2011 e o período entre janeiro a dezembro de 2012 foi utilizado no processo de validação, conforme a Figura 2 abaixo.

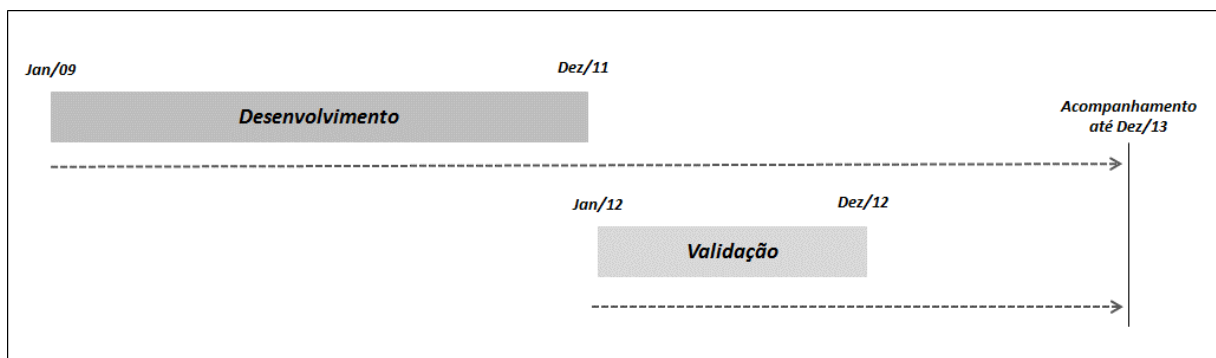


Figura 2 - Período da amostra de desenvolvimento e validação

Para a amostra de desenvolvimento, as operações foram acompanhadas, no mínimo, por 24 meses, enquanto que na validação o período mínimo de acompanhamento foi de 12 meses.

No processo de modelagem, a amostra de validação será utilizada no cálculo das medidas de previsão com o objetivo de verificar a eficiência do modelo desenvolvido em um período diferente do utilizado na sua construção.

4.1.1 Variáveis Utilizadas

O conjunto de variáveis utilizadas inclui as coletadas no momento da concessão do crédito, denominadas variáveis de *application*⁵, além das variáveis comportamentais e as macroeconômicas. As variáveis de *application* são estáticas, ou seja, seus valores não mudam ao longo dos meses observados. Já as comportamentais e as macroeconômicas são capturadas mensalmente. Maiores detalhes sobre as variáveis utilizadas serão dados adiante.

Seja n_i o número de meses desde que uma operação i foi iniciada (tempo de duração), w_i o vetor de variáveis de *application* da operação i , $x_{i,j}$ o vetor relacionado às variáveis comportamentais da operação i no período j , $j = 1, 2, \dots, n_i$ e z_j o vetor de variáveis macroeconômicas no período j , $j = 1, 2, \dots, n_i$, idêntico para todas as operações na mesma data. Podemos exemplificar a utilização das variáveis conforme a Tabela 3.

Tabela 3 - Exemplo de utilização das variáveis

Operação i	Tempo n_i	<i>Application</i> w_i	Comportamento $x_{i,j}$	Macroeconômica z_j
1	1	w_1	$x_{1,1}$	z_1
1	2	w_1	$x_{1,2}$	z_2
1	3	w_1	$x_{1,3}$	z_3
1	4	w_1	$x_{1,4}$	z_4
1	5	w_1	$x_{1,5}$	z_5
2	1	w_2	$x_{2,1}$	z_1
2	2	w_2	$x_{2,2}$	z_2
3	1	w_3	$x_{3,1}$	z_1
3	2	w_3	$x_{3,2}$	z_2
3	3	w_3	$x_{3,3}$	z_3
3	4	w_3	$x_{3,4}$	z_4
3	5	w_3	$x_{3,5}$	z_5
3	6	w_3	$x_{3,6}$	z_6

Fonte: Elaboração Própria.

⁵ Termo utilizado no ambiente de modelagem de concessão de crédito.

A Tabela 3 mostra 3 operações, sendo que a operação 1 possui tempo de vida $n_1 = 5$ períodos, a operação 2 tem tempo de vida $n_2 = 2$ períodos e a operação 3 apresenta $n_3 = 6$ períodos. As variáveis de *application* da operação $i = 1$ são dadas pelos componentes de \mathbf{w}_1 , idênticos para $j = 1, 2, \dots, n_1$. As comportamentais são dadas pelos componentes de $\mathbf{x}_{1,j}$ para $j = 1, 2, \dots, n_1$ e as macroeconômicas são especificadas por \mathbf{z}_j com $j = 1, 2, \dots, n_1$. De forma análoga, teremos as mesmas condições para a operação 2 com \mathbf{w}_2 , $\mathbf{x}_{2,j}$ e \mathbf{z}_j com $j = 1, 2, \dots, n_2$ e para a operação 3 com \mathbf{w}_3 , $\mathbf{x}_{3,j}$ e \mathbf{z}_j com $j = 1, 2, \dots, n_3$.

Entre as variáveis de *application*, destacam-se as relacionadas às características do proponente no momento do crédito. Com relação às comportamentais, é levado em consideração o histórico do relacionamento do cliente com a instituição financeira nos últimos 12 meses. Entre elas, destacam-se as variáveis relacionadas a contratos de crédito anteriores, a utilização de produtos específicos (por exemplo, cartão de crédito), a restrições financeiras e a sua reciprocidade financeira junto ao banco.

Em uma análise preliminar, as variáveis que possuíam em torno de 95% de valores ausentes foram desconsideradas do estudo. Por fim, foram pré-selecionadas 60 variáveis para serem utilizadas no processo de modelagem, sendo 15 variáveis de *application* e 45 comportamentais. Porém, por motivos de confidencialidade, somente as variáveis significativas no modelo serão descritas no próximo capítulo.

No conjunto de variáveis macroeconômicas, foram testadas 31 variáveis, conforme exposto no Apêndice B. Tratam-se de informações referentes à taxa de juros, ao produto interno bruto, à inflação e ao nível de confiança do consumidor, selecionadas de acordo com os estudos relacionados na revisão bibliográfica dessa dissertação.

Para estimação do modelo é necessário a definição da variável resposta. Neste caso, o evento de interesse foi definido de acordo com o comportamento de crédito, ou seja, uma operação foi considerada inadimplente (*default*) após 90 dias de atraso no pagamento das parcelas. A variável $d_{i,j}$ indica se, para uma determinada operação i , houve a ocorrência de *default* no tempo j , conforme a seguinte relação:

$$d_{i,j} = \begin{cases} 1, & \text{se a operação } i \text{ está inadimplente no tempo } j; \\ 0, & \text{caso contrário} \end{cases} \quad (15)$$

De acordo com o detalhado na seção 3.1.1.1, a partir do mês que foi observado o evento de ocorrência, a operação deixa de ser acompanhada, ou seja, não teremos mais informações desse contrato na base de dados a partir do instante posterior em que este se tornou inadimplente. De forma geral, considerando o conjunto de variáveis explicativas e a variável resposta, teremos a seguinte estrutura da base de dados no processo de estimação do modelo:

Tabela 4 - Exemplo de estrutura da base de dados final

Operação i	Tempo j	<i>Application</i>	Comportamento	Macroeconômica	<i>Default</i>
		w_i	$x_{i,j}$	z_j	$d_{i,j}$
1	1	w_1	$x_{1,1}$	z_1	0
1	2	w_1	$x_{1,2}$	z_2	0
1	3	w_1	$x_{1,3}$	z_3	0
1	4	w_1	$x_{1,4}$	z_4	0
1	5	w_1	$x_{1,5}$	z_5	1
2	1	w_2	$x_{2,1}$	z_1	0
2	2	w_2	$x_{2,2}$	z_2	0
3	1	w_3	$x_{3,1}$	z_1	0
3	2	w_3	$x_{3,2}$	z_2	0
3	3	w_3	$x_{3,3}$	z_3	0
3	4	w_3	$x_{3,4}$	z_4	0
3	5	w_3	$x_{3,5}$	z_5	0
3	6	w_3	$x_{3,6}$	z_6	0

Fonte: Elaboração Própria.

Neste exemplo, considerando que o tempo de observação do estudo foi de 6 meses, a operação 1, que se tornou inadimplente no mês 5 ($d_{1,5} = 1$), não foi mais acompanhada após esse instante, não existindo informação disponível para este contrato no mês 6. Já as outras operações permaneceram na base de dados até o instante em que elas foram censuradas, seja via encerramento do contrato ou término do período de análise.

4.2 Especificação do Modelo

Seja N o número total de operações no portfólio. Considerando os vetores de variáveis w_i ($i = 1, \dots, N$), $x_{i,j}$ ($i = 1, \dots, N$ e $j = 1, \dots, n_i$), z_j ($j = 1, \dots, \max(n_1, n_2, \dots, n_N)$) e a indicadora de *default* $d_{i,j}$ ($i = 1, \dots, N$ e $j = 1, \dots, n_i$) já definidos anteriormente, a probabilidade de *default* para cada operação i no tempo j , $P_{i,j}$, é dada por:

$$\begin{aligned}
P_{i,j} &= P(d_{i,j} = 1 \mid d_{i,k} = 0 \text{ para todo } k < j; \mathbf{w}_i, \mathbf{x}_{i,j}, \mathbf{z}_j) \\
&= \frac{1}{1 + \exp -(\alpha + \boldsymbol{\varphi}_j^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i,j}^T \boldsymbol{\beta}_3 + \mathbf{z}_j^T \boldsymbol{\beta}_4)}
\end{aligned} \tag{16}$$

$$i = 1, \dots, N \text{ e } j = 1, \dots, n_i$$

em que α é o intercepto e $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ e $\boldsymbol{\beta}_4$ são vetores dos coeficientes a serem estimados.

O parâmetro $\boldsymbol{\varphi}_j$ corresponde à função de transformação da função do tempo. De acordo com o exposto na seção 3.1.1.2, representações paramétricas do tempo devem ser incluídas no modelo a fim de especificar a dependência temporal das observações. Baseado em Shumway (2001), a função que permite capturar essa estrutura do risco sobre o tempo n é dada por $\log j$, sendo esta a utilizada no desenvolvimento dos modelos.

De acordo com a Equação 10, a probabilidade de sobrevivência estimada de um indivíduo i em um determinado tempo j , $\hat{S}_{i,j}$, é dada pela multiplicação das probabilidades de não falhar em cada período de tempo, condicional a não ter falhado anteriormente:

$$\hat{S}_{i,j} = \prod_{s=1}^j (1 - \hat{P}_{i,s}) \tag{17}$$

Desta forma, a probabilidade estimada de *default* do indivíduo i no tempo j , $\widehat{PD}_{i,j}$, será dada pela probabilidade de falha, conforme equação abaixo.

$$\widehat{PD}_{i,j} = 1 - \hat{S}_{i,j} \tag{18}$$

Para exemplificar o cálculo, consideremos a Tabela 5 com as probabilidades de sobrevivência e de falha para cada tempo observado, assumindo que os parâmetros dos modelos propostos no item 4.3 a seguir já foram determinados.

Tabela 5 - Exemplo de cálculo das probabilidades estimadas

Operação i	Tempo j	$\hat{P}_{i,j}$	$\hat{S}_{i,j}$	$\hat{PD}_{i,j}$
1	1	0,00	1,00	0,00
1	2	0,09	0,91	0,09
1	3	0,12	0,80	0,20
1	4	0,16	0,67	0,33
1	5	0,21	0,53	0,47
2	1	0,00	1,00	0,00
2	2	0,04	0,96	0,04
3	1	0,00	1,00	0,00
3	2	0,01	0,99	0,01
3	3	0,02	0,97	0,03
3	4	0,03	0,94	0,06
3	5	0,03	0,91	0,09
3	6	0,04	0,88	0,12

Fonte: Elaboração Própria.

A operação 1 apresenta ao final do quinto mês de observação uma probabilidade de *default* igual a 47%. Já a operação 2 possui uma probabilidade de *default* estimada em torno de 4% após 2 meses de análise, enquanto que o contrato 3 possui uma estimativa de *default* de 12% após 6 meses de observação.

Notamos que, conforme a Tabela 4 vista anteriormente, o contrato 1 tornou-se inadimplente no quinto mês de observação, evidenciado pela probabilidade alta de se tornar inadimplente ao longo do período analisado. Já os contratos 2 e 3 não são inadimplentes, apresentando probabilidades mais baixas de *default* no exemplo adotado acima.

4.3 Modelos Desenvolvidos

Para cumprir o objetivo do estudo, dois modelos foram desenvolvidos⁶ a fim de medir o impacto da inclusão das variáveis macroeconômicas no ajuste do modelo e na sua capacidade preditiva.

No primeiro modelo, denominado como Modelo 1, foi considerado o parâmetro relacionado ao tempo de duração, as variáveis de *application* e as comportamentais. Neste modelo, somente o vetor de parâmetros β_4 é considerado nulo, conforme expressão abaixo.

⁶ Para o desenvolvimento dos modelos foi utilizado o software estatístico SAS.

$$P_{i,j} = \frac{1}{1 + \exp -(\alpha + \boldsymbol{\varphi}_j^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i,j}^T \boldsymbol{\beta}_3)} \quad (19)$$

No Modelo 2, considerado o mais completo do estudo, foram adicionadas ao Modelo 1 as variáveis macroeconômicas. A equação deste modelo será dada por:

$$P_{i,j} = \frac{1}{1 + \exp -(\alpha + \boldsymbol{\varphi}_j^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i,j}^T \boldsymbol{\beta}_3 + \mathbf{z}_j^T \boldsymbol{\beta}_4)} \quad (20)$$

A Tabela 6 resume as características e expressão adotada para cada modelo descrito acima.

Tabela 6 - Resumo dos modelos desenvolvidos

Modelo	Descrição	Expressão
Modelo 1	Considera os parâmetros referentes ao tempo de duração, as variáveis de <i>application</i> e comportamentais	$P_{i,j} = \frac{1}{1 + \exp -(\alpha + \boldsymbol{\varphi}_j^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i,j}^T \boldsymbol{\beta}_3)}$
Modelo 2	Modelo completo: todos os coeficientes são estimados	$P_{i,j} = \frac{1}{1 + \exp -(\alpha + \boldsymbol{\varphi}_j^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i,j}^T \boldsymbol{\beta}_3 + \mathbf{z}_j^T \boldsymbol{\beta}_4)}$

Fonte: Elaboração Própria.

4.4 Medidas de Desempenho

Com o objetivo de avaliar os modelos desenvolvidos, foram adotadas algumas medidas de desempenho de acordo com duas dimensões de análise: (i) o ajuste do modelo, e (ii) a sua capacidade preditiva.

Como medidas para avaliar o ajuste do modelo utilizaremos as medidas *AIC* (*Akaike's Information Criteria*) e *SC* (*Schwarz's Criteria*). Esses critérios penalizam a verossimilhança de acordo com o número de parâmetros do modelo. São eles:

$$AIC = -2 * \text{Log } L + 2k \quad (21)$$

$$SC = -2 * \text{Log } L + k * \log(N) \quad (22)$$

em que $\log L$ corresponde ao logaritmo da verossimilhança, k é o número de coeficientes do modelo e N é a quantidade de observações. Valores menores dessas estatísticas indicam modelos com melhores ajustes.

Para avaliar o poder preditivo do modelo desenvolvido, os autores utilizam uma medida específica para o nível agregado (portfólio). Considerando que para um determinado tempo n temos N clientes, a taxa de *default* observada D_n e a taxa de *default* esperada $E(D_n)$ são dadas, respectivamente, por:

$$D_n = \frac{1}{N} \sum_{i=1}^N d_{i,n} \quad (23)$$

$$E(D_n) = \frac{1}{N} \sum_{i=1}^N PD_{i,n} \quad (24)$$

em que $d_{i,n}$ corresponde à indicadora de *default* do indivíduo i definida na Equação 15 e $PD_{i,n}$ se refere a probabilidade estimada de *default* do indivíduo i no tempo n , conforme Equação 18.

A medida de desempenho, DA , para previsão no nível agregado com tempo de observação variando de $n = 1$ até T será dada por:

$$DA = \sum_{n=1}^T |E(D_n) - D_n| \quad (25)$$

Quanto menor o valor de DA , menores são as diferenças entre o valor observado e esperado da taxa de *default* ao longo dos meses analisados, indicando modelos com melhor poder de previsão.

Com relação à metodologia do estudo, temos, resumidamente, as seguintes etapas:

- Seleção de uma amostra aleatória de operações iniciadas no período de estudo.
- Acompanhamento dos contratos selecionados no passo anterior até o encerramento da operação, até se tornarem inadimplentes ou até o fim do período de análise. Neste passo, as variáveis de *application*, comportamentais e macroeconômicas são coletadas mensalmente.
- Desenvolvimento dos modelos propostos: Modelo 1 e Modelo 2.

- Comparação do poder de ajuste e da capacidade preditiva dos modelos.

No próximo capítulo serão expostos os resultados obtidos, detalhando as variáveis significativas para os modelos desenvolvidos e os valores das medidas de performance.

5. Resultados

Neste capítulo, os resultados obtidos no desenvolvimento deste estudo são apresentados. Inicialmente, são expostas as variáveis significativas dos modelos construídos e são discutidos os efeitos dos coeficientes estimados. Por fim, apresentamos as medidas de ajuste e previsão e a aplicabilidade dos modelos.

5.1 Modelos Estimados

Entre as variáveis que foram testadas, conforme seção 4.1.1, a Tabela 7 apresenta a lista das que foram significativas nos modelos desenvolvidos⁷.

Tabela 7 - Variáveis significativas nos modelos desenvolvidos

Tipo	Parâmetros	Variável	Descrição
Parâmetro de Duração	β_1	LTEMPO	Logaritmo do tempo
<i>Application</i>	β_2	IN_RECEBE_CC	Indicadora de recebimento de salário por crédito em conta corrente
		QT_RELAC	Tempo de relacionamento (em mês)
		QT_RESTRITIVO	Quantidade de restritivos históricos
Comportamental	β_3	QT_SALDO_MAX	Quantidade de meses desde que o cliente apresentou o maior saldo devedor
		IN_FINANC_CARTAO	Indicadora se houve financiamento de cartão de crédito
Macroeconômica	β_4	VAR_21	Taxa de desemprego - Região metropolitana - Dessazonalizada
		VAR_1	Taxa de juros - Over / Selic

Fonte: Elaboração Própria.

Os coeficientes estimados para cada um dos parâmetros acima são expostos a seguir de acordo com a descrição dos modelos desenvolvidos e com a metodologia apresentada no capítulo anterior. Os valores referentes aos interceptos dos modelos não serão divulgados, dado a

⁷ Entende-se por variável significativa aquelas cujos coeficientes são estatisticamente diferentes de zero segundo o Teste de Wald. Neste teste a hipótese nula considera que o coeficiente é igual a zero contra o caso contrário (hipótese alternativa) e rejeita-se a hipótese nula quando p-valor inferior a 0,10.

confidencialidade dos dados. As saídas completas dos modelos estão disponíveis no Apêndice C.

A Tabela 8 expõe os parâmetros estimados referentes ao Modelo 1.

Tabela 8 - Coeficientes estimados – Modelo 1

Variável	Parâmetros	Descrição dos Parâmetros	Modelo 1
INTERCEPTO	α	Intercepto	-
LTEMPO	β_1	Logaritmo do tempo	0,9327
IN_RECEBE_CC	β_2	Cliente “Não recebe salário” por crédito em conta corrente	0,1240
QT_RELAC		Tempo de relacionamento (em mês)	-0,0045
QT_RESTRITIVO		Quantidade de restritivos históricos	0,0028
QT_SALDO_MAX	β_3	Quantidade de meses desde que o cliente apresentou o maior saldo devedor	-0,1361
IN_FINANC_CARTAO		“Não houve” financiamento de cartão de crédito	-1,1837

Fonte: Elaboração Própria.

Com relação ao comportamento das variáveis, analisando este modelo, o coeficiente relacionado ao tempo, $\log j$, foi significativo estatisticamente.

Para as outras variáveis independentes, o efeito de cada uma delas com relação à probabilidade de inadimplência pode ser avaliado por meio da interpretação dos seus respectivos coeficientes estimados. Coeficientes positivos estão relacionados a aumentos na probabilidade de inadimplência, enquanto que sinais negativos referem-se a uma diminuição da probabilidade de inadimplência.

Neste modelo, dado o sinal positivo da variável, constatamos que clientes que não recebem salário por crédito em conta corrente apresentam um comportamento pior com relação à inadimplência quando comparado aos que recebem. Esse é um comportamento esperado dado que clientes com recebimento por crédito em conta apresentam condições mais favoráveis de pagamento dos empréstimos adquiridos dado a existência de relacionamento com o banco.

O tempo de relacionamento possui um efeito negativo, isto é, quanto maior, melhor é o perfil do cliente. Para cada mês adicional de relacionamento, há uma diminuição na probabilidade de inadimplência, dado as demais variáveis fixas. Esse resultado justifica-se pelo fato de que clientes já conhecidos da instituição financeira e com histórico positivo de pagamento de

empréstimos conseguem condições mais favoráveis ao crédito como taxas de juros mais baixas, diminuindo a chance de se tornarem inadimplentes. Além disso, supõe-se que clientes novos que logo solicitam crédito estão em uma situação não muito favorável financeiramente, de tal forma que iniciam o relacionamento com a instituição financeira já atrelado a um empréstimo. Esse tipo de cliente apresenta um perfil desconhecido para o banco dado que o histórico de informações comportamentais é muito pequeno ou inexistente.

A quantidade de restritivos históricos no momento da concessão do crédito possui um efeito positivo sobre a inadimplência: quanto maior o número de restrições no passado, pior é o comportamento do cliente com relação à qualidade do crédito. Considera-se que estes clientes estão mais suscetíveis a possuírem problemas de inadimplência dado que já apresentaram restrições financeiras no passado recente.

Dado o sinal negativo do coeficiente, clientes que apresentaram o maior saldo devedor a mais tempo nos últimos 12 meses, possuem menor probabilidade de se tornarem inadimplentes. Esse efeito é esperado, pois quanto mais longe o cliente apresentou o maior saldo devedor na organização, menor é o nível de endividamento dele e, por consequência, menor é o seu comprometimento de renda. Desta forma, a chance de pagamento do empréstimo contraído é maior.

O coeficiente estimado para a variável que indica se houve financiamento de cartão de crédito nos últimos 12 meses apresentou sinal negativo, isto é, a probabilidade de inadimplência para os clientes sem essa característica é menor quando comparado aos indivíduos que financiaram o cartão de crédito. Podemos considerar que esse perfil se trata de clientes que enfrentam ou enfrentaram em um passado recente problemas com relação à sua capacidade de pagamento, estando mais suscetíveis a se tornarem inadimplentes.

A Tabela 9 abaixo refere-se aos parâmetros do Modelo 2.

Tabela 9 - Coeficientes estimados – Modelo 2

Variável	Parâmetros	Descrição dos Parâmetros	Modelo 2
INTERCEPTO	α	Intercepto	-
LTEMPO	β_1	Logaritmo do tempo	1,0228
IN_RECEBE_CC	β_2	Cliente “Não recebe salário” por crédito em conta corrente	0,1316
QT_RELAC		Tempo de relacionamento (em mês)	-0,0045
QT_RESTRICTIVO		Quantidade de restritivos históricos	0,0030
QT_SALDO_MAX	β_3	Quantidade de meses desde que o cliente apresentou o maior saldo devedor	-0,1353
IN_FINANC_CARTAO		“Não houve” financiamento de cartão de crédito	-1,1888
VAR_21	β_4	Taxa de desemprego - Região metropolitana - Dessazonalizada	0,1377
VAR_1		Taxa de juros - Over / Selic	0,7564

Fonte: Elaboração Própria.

No Modelo 2 notamos o mesmo comportamento das variáveis referentes às características e ao comportamento do cliente já vistas no Modelo 1, citado anteriormente. O diferencial deste modelo é a inclusão das variáveis macroeconômicas, no qual podemos observar que o sinal positivo da variável referente ao desemprego implica que elevações na taxa de desemprego provocam o aumento da inadimplência. Do ponto de vista econômico isso é esperado, dado que a taxa de desemprego é vista como um indicador de *stress* econômico. Clientes que perdem seu emprego, provavelmente terão maiores dificuldades em honrar seus compromissos de crédito.

A variável taxa de juros SELIC apresentou coeficiente positivo, indicando que aumentos implicam em elevações na taxa de inadimplência. A visão econômica para esse resultado é que aumentos na taxa de juros SELIC provocam um encarecimento do crédito, traduzido na forma de taxas de juros ao consumidor mais elevadas, tendo como consequência a diminuição da capacidade de pagamento dos clientes.

5.2 Medidas de Ajustes e de Previsão dos Modelos

As medidas de ajuste dos modelos desenvolvidos são mostradas na Tabela 10 abaixo.

Tabela 10 - Comparação dos ajustes dos modelos desenvolvidos

Modelo	AIC	SC
Modelo 1	143046	143126
Modelo 2	142823	142927

Fonte: Elaboração Própria.

Conforme os valores contidos na tabela acima, podemos observar que embora duas variáveis macroeconômicas sejam significativas estatisticamente (Modelo 2), a contribuição delas para o ajuste do modelo é fraca, sendo a diferença dos valores dos critérios *AIC* e *SC* entre os Modelo 1 e 2 muito pequena.

Considerando a medida de previsão para o nível agregado no período de validação, podemos notar no gráfico referente à Figura 3 as taxas mensais de *default* estimadas para os modelos desenvolvidos, comparando-as com as taxas de *default* observadas em cada mês da base de dados referente ao período de validação.

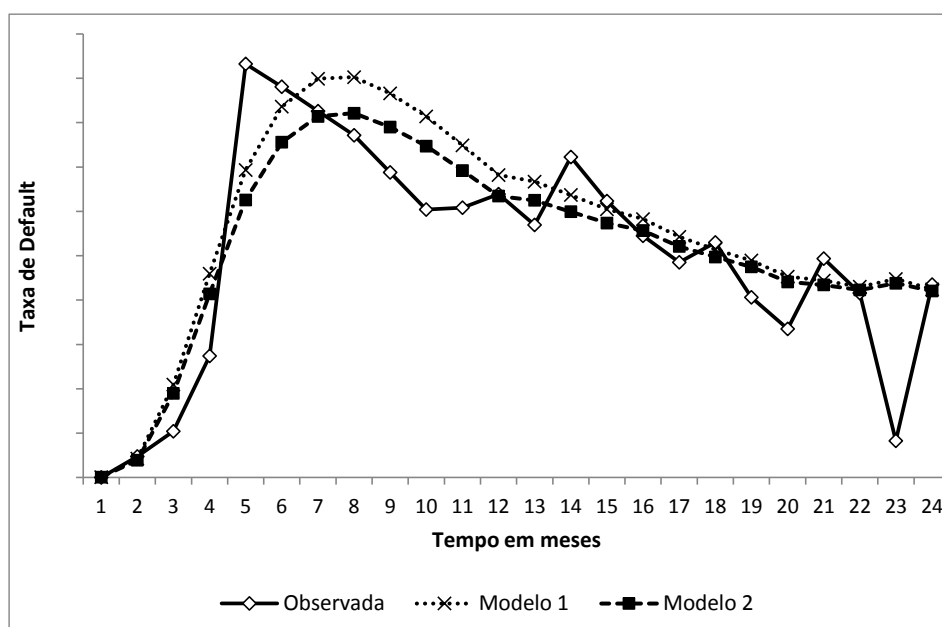


Figura 3 - Comparação das taxas de *default* observada e estimada dos modelos desenvolvidos no período de validação⁸

⁸ A escala do gráfico não foi mostrada por razões de confidencialidade.

Observa-se que aparentemente os modelos não possuem grandes diferenças, apresentando um comportamento estimado muito próximo do observado com relação à distribuição da taxa de *default*. Entretanto, quando analisamos a soma das diferenças mensais, a chamada *DA* explicada na seção 4.4, o Modelo 2 apresenta o menor valor, indicando um poder de previsão superior, conforme indicado na Tabela 11.

Tabela 11 - Comparação do poder de previsão no nível agregado

Modelo	DA
Modelo 1	0,1336
Modelo 2	0,0143

Fonte: Elaboração Própria.

Diante dos resultados apresentados, considerando o Modelo 2, podemos concluir que este modelo é composto por variáveis cujos efeitos são coerentes com as expectativas econômicas. Além disso, nota-se que a inclusão de variáveis macroeconômicas aumenta o ajuste do modelo de forma branda, porém há uma elevação do poder preditivo deste modelo com relação à taxa de inadimplência quando considera o portfólio como um todo.

5.3 Aplicação dos Modelos Estimados

Dado a estimação do modelo e a verificação da coerência dos efeitos das variáveis e das medidas de desempenho, o modelo desenvolvido poderá ser aplicado em diferentes situações, tais como na concessão e pré-aprovação de limites de crédito para clientes da carteira analisada.

Para efeito de exemplo da utilização do modelo, consideremos 2 clientes, denominados Cliente 1 e Cliente 2. De acordo com as variáveis e os parâmetros estimados do Modelo 2, podemos, no momento da concessão, construir a curva de sobrevivência de cada um deles, conforme Figura 4. Desta maneira, teremos as estimativas da sobrevivência no horizonte de tempo desejado, considerando que os valores das variáveis no tempo inicial permanecem inalterados.

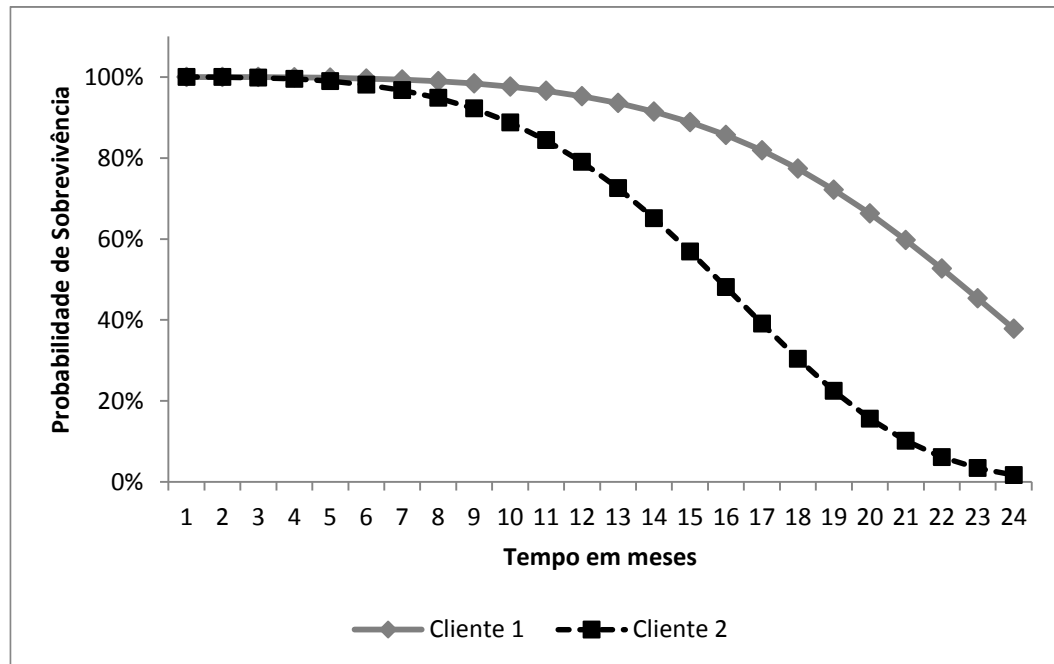


Figura 4 - Exemplo de curva de sobrevivência

Conforme podemos notar pela figura acima, o Cliente 1 apresenta perfil de comportamento creditício superior comparado ao Cliente 2. Neste exemplo, o Cliente 1 possui uma probabilidade de sobrevivência, isto é, de não se tornar inadimplente, no mês 12 de aproximadamente 95%, enquanto que neste mesmo instante o Cliente 2 teria uma probabilidade de sobrevivência em torno de 79%. Neste sentido, para o Cliente 1 o crédito seria concedido, enquanto que o Cliente 2 teria a operação recusada, considerando um específico ponto de corte adotado pela empresa.

De maneira análoga ao procedimento descrito acima, baseando-se nas curvas estimadas de sobrevivência, estratégias de crédito, tais como limites e operações pré-aprovadas, poderiam ser concedidas de acordo com o nível de risco aceito pela empresa. Desta forma, as características das operações em relação ao prazo e juros podem se adequar ao perfil do cliente.

6. Conclusão

O principal objetivo desse estudo foi o desenvolvimento de um modelo dinâmico de concessão de crédito através do uso da técnica estatística de Análise Sobrevivência na sua abordagem discreta. Os modelos dinâmicos apresentam uma abordagem mais flexível quando comparados aos usuais modelos de crédito estáticos, pois permitem a utilização de variáveis cujos valores se alteram ao longo do período do estudo.

Para a construção do modelo, foram utilizadas variáveis relacionadas ao cliente no momento da concessão do crédito, além das comportamentais e macroeconômicas, a fim de entender a relação de causa e efeito das variáveis significantes no modelo com a probabilidade de inadimplência.

Os resultados mostram que o tempo de relacionamento, quantidade de restritivos históricos e se o cliente recebe o salário por crédito em conta corrente são informações significantes estatisticamente. Entre as variáveis comportamentais, destacam-se a quantidade de meses desde que o cliente apresentou o maior saldo devedor na instituição financeira e se houve financiamento de cartão de crédito nos últimos 12 meses.

Entre as variáveis macroeconômicas estudadas, taxa de juros SELIC e de desemprego possuem uma relação significativa com o risco de inadimplência, entretanto nota-se que a contribuição dessas duas variáveis para o ajuste do modelo é fraca. Com relação à previsão da taxa de inadimplência, nota-se um aumento significativo do poder preditivo no nível agregado.

Esse desempenho é esperado dado que as condições da economia, representada nesse modelo pelas variáveis macroeconômicas, não atingem os clientes de forma pontual e, sim, afetam a probabilidade de *default* estimada de forma geral, tornando seu efeito notável no nível agregado. Os resultados alcançados estão de acordo com os obtidos em outros estudos que seguiram de base para a estrutura dessa dissertação, tais como Belotti e Crook (2013) e Gross e Souleles (2002).

A principal aplicação do modelo desenvolvido será de utilizá-lo como uma ferramenta de medição de risco associado a um proponente de crédito e de previsão de inadimplência, tornando a concessão de crédito mais eficiente. Esse aumento da qualidade de crédito permite a aplicação de taxas de juros diferenciadas de acordo com o risco do cliente, bem como a

utilização de estratégias de pré-aprovação de limites e crédito para indivíduos que apresentam riscos mais baixos de inadimplência.

Em estudos futuros, dados a utilização de variáveis macroeconômicas, sugere-se a utilização desse modelo para simulação de cenários de *stress testing*, criando diversas condições econômicas a fim de mensurar a perda inesperada associada a um portfólio de crédito. Além disso, há a possibilidade de ampliação do público estudado através do desenvolvimento de um modelo dinâmico aplicado a pessoas jurídicas e a outras carteiras de crédito.

REFERÊNCIAS

- ALLISON, P. D.. *Survival analysis using SAS: a practical guide*. 2.ed. Cary, NC: SAS Institute Inc., 2010
- ALMEIDA, M. P.. *Estimação bayesiana em modelos de sobrevivência: Uma aplicação em Credit Scoring*. 2008. 78f. Dissertação (Mestrado em Estatística) – Programa de Pós Graduação em Matemática e Estatística, Universidade Federal do Pará, Pará.
- ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporation bankruptcy. *Journal of Finance*, v. 23, n. 4, p. 589-609, 1968.
- ANDRADE, A. P. S.. *Avaliação da recente expansão do crédito no Brasil: boom ou crescimento sustentável?* 2012. 103f. Dissertação (Mestrado em Ciências) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.
- ANTUNES, A.; RIBEIRO, N.; ANTÃO, P. Estimating probabilities of default under macroeconomic scenarios. *Financial Stability Report*, p. 115, 2005.
- Banco Central do Brasil (BCB), Sistema Gerenciador de Séries Temporais – v2.1, [Online], Disponível em: <www3.bcb.gov.br/sgspub> [10 abril 2014]
- _____, Relatório Anual 2012, [Online], Disponível em: <<http://www.bcb.gov.br/pec/boletim/banual2012/rel2012p.pdf>> [15 janeiro 2014]
- BATISTA, S. R. F.; DIVINO, J. A. e ORRILLO, J.. Taxa de juros e default em mercados de empréstimos colateralizados. *Estudos Econômicos*, vol.41, n.4, p. 691-718, 2011.
- BELLOTTI, T.; CROOK, J. Credit Scoring with macroeconomic variables using survival analysis. *The Journal of the Operational Research Society*, v. 60, n. 12, p. 1669-1707, 2009.
- _____, Forecasting and Stress testing credit card default using dynamic models. *International Journal of Forecasting*, v. 29, n. 12, p. 563-574, 2013.
- CAOQUETTE J., ALTMAN E., NARAYANAN P. *Managing credit risk: the next great financial challenge*. New York: John Wiley & Sons, Inc. 1998.
- CARLING, K *et al.*. Corporate credit risk modeling and the macroeconomy. *Journal of Banking & Finance*, v. 31, n. 3, p. 845-868, 2007.
- COLOSIMO, E. A., GIOLO, S. R. *Análise de sobrevivência aplicada*. 1. ed. São Paulo: Edgard Blücher, 2006.
- COSTA, F. N., Bancos e crédito no Brasil: 1945-2007. *História e Economia: Revista Interdisciplinar*. v.4, n. 2, 2009.
- COX, D. R. Regression models and life tables. *Journal of the Royal Statistical Society*, v. 34, n. 2, p. 187-220, 1972.

CROOK, J. and BELOTTI, T., Time varying and dynamic models for default risk in consumer loans. *Journal of the Royal Statistical Society*, v. 173, n. 2, p. 283–305, 2010.

DIVINO, J. A.; ROCHA, L. C. S.. Probability of default in collateralized credit operations. *The North American Journal of Economics and Finance*, v. 25, p. 276-292, 2013.

FEBRABAN. Relatório Anual 2012 da Federação Brasileira dos Bancos. Febraban, 2012.

FRANCESCA, G., A discrete-time hazard model for loans: some evidence from Italian banking system. *American Journal of Applied Sciences*, v. 9, n.9, p. 1337-1346, 2012.

FIGINI S., FANTAZZINI D. Random survival forests models for SME credit risk measurement. *Methodology and Computing in Applied Probability*, v. 11, n. 1, p. 29-45 2009.

GROSS, D. B.; SOULELES, N. S. An empirical analysis of personal bankruptcy and delinquency. *Review of Financial Studies*, v. 15, n. 1, p. 319–347, 2002

HOSMER, D. W., Jr. and LEMESHOW, S.. *Applied logistic regression*, 2. ed., New York: John Wiley & Sons, 2000

LEONARDIS, D. D.; ROCCI, R. Assessing the default risk by means of a discrete-time survival analysis approach. *Applied Stochastic Models in Business and Industry*, v. 24, n. 4, p. 291-306, 2008.

LIMA, E. S.. *Taxa de default, mercados incompletos e análise de sobrevivência*. 2008. 43f. Dissertação (Mestrado em Finanças) – Universidade Católica de Brasília, Brasília.

MALIK M., THOMAS L.C., Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society*, v. 61, n. 3, p. 411-420. 2009

MÄNNASOO, K.; MAYES, D. G. Explaining bank distress in Eastern European transition economies. *Journal of Banking & Finance*, v. 33, n. 2, p. 244-253, 2009.

NAM, C. W, *et al.* Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, v. 27, n. 6, p. 493–506, 2008.

OHLSON, J. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, v. 18, n.1, p.109-131,1980.

ROCHA, Lineke C. S..*Avaliação de modelos usados para estimar a probabilidade de default*. 2012. 51f. Dissertação (Mestrado em Finanças) – Universidade Católica de Brasília, Brasília.

SICSÚ, A. L. *Credit Socring: desenvolvimento, implantação e acompanhamento*. 1 ed. Blucher: São Paulo. 2010.

SHUMWAY, T. Forecasting bankruptcy more accurately : A simple hazard model *. *The Journal of Business*, v. 74, n. 1, p. 101–124, 2001.

THOMAS, L. C.; HO, J.; SCHERER, W. T. Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, v. 12, n. 1, p. 89-103, 2001.

APÊNDICE A – Regressão Logística

A técnica de Regressão Logística é utilizada para descrever a relação entre um conjunto de variáveis independentes e uma variável resposta dicotômica que assume o valor 1 quando ocorre o evento de interesse e 0, caso contrário.

De forma geral, considerando um conjunto de p variáveis independentes, denotado pelo vetor $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$ e que Y é a variável resposta binária, temos que:

$$\pi(\mathbf{x}) = P(Y = 1 | \mathbf{x})$$

A forma específica do modelo de Regressão Logística é dada por:

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}} = \frac{1}{1 + e^{-g(\mathbf{x})}}$$

em que $g(\mathbf{x})$ corresponde a transformação *logito* dada por:

$$g(\mathbf{x}) = \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

A estimação dos parâmetros é dada pelo método da máxima verossimilhança, sendo que cada indivíduo i contribui para a função de verossimilhança da seguinte forma:

$$L_i = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{(1 - y_i)}$$

Considerando que as n observações da amostra são independentes, a função de verossimilhança é obtida por:

$$L = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{(1 - y_i)}$$

Os estimadores de máxima verossimilhança correspondem aos valores de β que maximizam a função L e são obtidos por métodos não iterativos, tais como Newton-Raphson.

Com relação à interpretação dos parâmetros, β corresponde à mudança na função *logito* para cada unidade de acréscimo em uma específica variável, considerando as demais constantes. Uma alternativa à interpretação dos coeficientes é a utilização da razão de *odds*, calculada da seguinte forma quando considera-se a população dividida em 2 grupos:

$$OR = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]}$$

A razão de *odds*, no caso de variável resposta dicotômica, corresponde à exponenciação do coeficiente calculado. Essa medida é amplamente usada, pois é tratada como um fator de associação entre as variáveis independentes e o evento de interesse.

APÊNDICE B – Lista de Variáveis Macroeconômicas

Segue a lista de variáveis macroeconômicas utilizadas no processo de modelagem dos dados.

Variável	Descrição	Fonte
var_1	Taxa de juros - Over / Selic - (% a.m.)	BCB
var_2	Vendas reais - indústria - índice (média 2006 = 100) - SP	FIESP
var_3	Usecheque - número de consultas - Unidade	ACSP/IEGV
var_4	SPC - número de consultas - Unidade	ACSP/IEGV
var_5	Inflação - IPCA - (% a.m.)	IBGE/SNIPC
var_6	Inflação - IGP-M - (% a.m.)	FGV
var_7	Índice de confiança do consumidor (ICC)	Fecomercio SP
var_8	Índice de condições econômicas atuais (ICEA)	Fecomercio SP
var_9	Índice de expectativas (IEC)	Fecomercio SP
var_10	Índice nacional de vendas - variação nominal - (% a.m.)	Abras
var_11	Vendas reais - varejo ampliado - índice dessaz. (média 2011 = 100)	IBGE/PMC
var_13	Inflação - INCC-M - (% a.m.)	FGV
var_14	Índice nacional de preços ao consumidor (INPC) - Var. % mensal	IBGE
var_15	Índice geral de preços do mercado (IGP-M) - Var. % mensal	FGV
var_16	Índice nacional de custo da construção (INCC) - Var. % mensal	FGV
var_17	Índice nacional de preços ao consumidor-amplio (IPCA) - Var. % mensal	IBGE
var_18	PIB mensal - Valores correntes (R\$ milhões) - R\$ (milhões)	BCB-Depec
var_19	Índice de Confiança do Consumidor - Índice	Fecomercio
var_20	Inadimplência no SPC - Taxa bruta (t-3) = novos registros/consultas - %	ACSP
var_21	Taxa de desemprego - Região metropolitana - Brasil (% a.m.)	IBGE
var_22	Indicadores da produção (2002=100) - Geral - Índice	IBGE
var_23	Taxa de câmbio - Livre - Dólar americano (venda) - Média de período - mensal - u.m.c./US\$	BCB-Depec
var_24	Taxa de juros - Selic acumulada no mês anualizada - % a.a.	BCB-Demab
var_25	Taxa de juros - Selic acumulada no mês - % a.m.	BCB-Depec
var_26	Índice de Atividade Econômica do Banco Central (IBC-Br)	BCB-Depec
var_27	Taxa de juros - CDI acumulada no mês - % a.m.	BCB-Depec
var_28	Indicador Serasa Experian de Atividade Econômica (PIB Mensal) (Média de 1995 = 100) - sem ajuste sazonal	Serasa
var_29	Indicador Serasa Experian de Atividade Econômica (PIB Mensal) (Média de 1995 = 100) - com ajuste sazonal	Serasa
var_30	Indicador Serasa Experian de Inadimplência do Consumidor - Sem Ajuste Sazonal (Média de 2009 = 100) - sem ajuste sazonal	Serasa
var_31	Indicador Serasa Experian de Atividade do Comércio - Sem Ajuste Sazonal (Média de 2000 = 100) - sem ajuste sazonal	Serasa
var_32	Indicador Serasa Experian de Atividade do Comércio - Com Ajuste Sazonal (Média de 2000 = 100) - com ajuste sazonal	Serasa

APÊNDICE C – Saída dos Modelos

As tabelas abaixo referem-se aos modelos desenvolvidos no estudo. As variáveis significantes e suas respectivas estimativas, erro padrão e nível de significância são apresentadas.

Modelo 1

Parâmetro	Estimativa	Erro Padrão	P-Valor
INTERCEPTO	-	0,0266	< 0,0001
LTEMPO	0,9327	0,0121	< 0,0001
IN_RECEBE_CC	0,1240	0,0164	< 0,0001
QT_RELAC	-0,0045	0,0001	< 0,0001
QT_RESTRICTIVO	0,0028	0,0002	< 0,0001
QT_SALDO_MAX	-0,1361	0,0022	< 0,0001
IN_FINANC_CARTAO	-1,1837	0,0175	< 0,0001

Modelo 2

Parâmetro	Estimativa	Erro Padrão	P-Valor
INTERCEPTO	-	0,1161	< 0,0001
LTEMPO	1,0228	0,0137	< 0,0001
IN_RECEBE_CC	0,1316	0,0164	< 0,0001
QT_RELAC	-0,0045	0,0001	< 0,0001
QT_RESTRICTIVO	0,0030	0,0002	< 0,0001
QT_SALDO_MAX	-0,1353	0,0022	< 0,0001
IN_FINANC_CARTAO	-1,1888	0,0176	< 0,0001
VAR_21	0,1377	0,0110	< 0,0001
VAR_1	0,7564	0,0675	< 0,0001